Correlations between stochastic epidemics in two interacting populations

Sophie R. Meakin[a],[*], Matt J. Keeling[b]

[a] EPSRC & MRC Centre for Doctoral Training in Mathematics for Real-World Systems, University of Warwick, United Kingdom
[b] Zeeman Institute: SBIDER, Mathematics Institute and School of Life Sciences, University of Warwick, United Kingdom

ABSTRACT

It is increasingly apparent that heterogeneity in the interaction between individuals plays an important role in the dynamics, persistence, evolution and control of infectious diseases. In epidemic modelling two main forms of heterogeneity are commonly considered: spatial heterogeneity due to the segregation of populations and heterogeneity in risk at the same location. The transition from random-mixing to heterogeneous-mixing models is made by incorporating the interaction, or coupling, within and between subpopulations. However, such couplings are difficult to measure explicitly; instead, their action through the correlations between subpopulations is often all that can be observed. Here, using moment-closure methodology supported by stochastic simulation, we investigate how the coupling and resulting correlation are related. We focus on the simplest case of interactions, two identical coupled populations, and show that for a wide range of parameters the correlation between the prevalence of infection takes a relatively simple form. In particular, the correlation can be approximated by a logistic function of the between population coupling, with the free parameter determined analytically from the epidemiological parameters. These results suggest that detailed case-reporting data alone may be sufficient to infer the strength of between population interaction and hence lead to more accurate mathematical descriptions of infectious disease behaviour.

1. Introduction

The incorporation of heterogeneity is an increasingly important feature of epidemiological models, with spatial-structure (Grenfell and Bolker, 1998; Xia et al., 2004; Viboud et al., 2006) and risk-structure (Schenzle, 1984; Keeling and Grenfell, 1997; Keeling and White, 2010) most prominently considered. This more realistic heterogeneous structure has marked influences on many properties including: invasion dynamics, leading to travelling waves in spatial systems (Viboud et al., 2006; Diekmann, 1978; Grenfell et al., 2002) and aggregation in certain groups for risk-structured populations (Schenzle, 1984); endemic behaviour, breaking the simple relationships between proportion susceptible and the basic reproductive ratio that hold for simpler models (Keeling and Rohani, 2008); persistence, generally acting to increase the persistence within stochastic populations (Keeling, 2000c; Hagenaars et al., 2004); and control, leading to targeted interventions (Keeling and White, 2010; Christley et al., 2005; Wallinga et al., 2010). Therefore, structuring the population has profound and wide-reaching implications.

A common modelling paradigm that captures multiple forms of heterogeneity in epidemic models is the metapopulation-type model (Gilpin and Hanski, 1991; Hanski, 1998; Hanski and Gaggiotti, 2004;

Keeling et al., 2010). In such models, the population is divided into multiple interacting, or 'coupled', subpopulations, where within-population interactions, and hence transmission, typically occur at a much higher rate than between-population interactions. Although metapopulation models are often considered as a form of spatial model, their compartmental structure can equally well apply to age or risk structured mixing, or to multiple host species. In the case of two coupled subpopulations, which we focus on in this work, these could refer to two distinct communities, different risk groups (e.g. high and low risk) or different age groups (e.g. adults and children).

In general, the difference between the metapopulation framework and the standard homogeneous mixing models is the way that transmission, or interaction, between the subpopulations is incorporated. The interaction between subpopulations is often represented as a matrix of transmission rates within and between populations, which has clear links to the number of cases generated by each group and hence to the basic reproductive ratio, R_0, through the dominant eigenvalue (Diekmann et al., 1990; Heesterbeek, 2002). When dealing with M populations, this transmission matrix has M^2 terms, which creates unidentifiability problems when attempting to estimate parameters from endemic equilibria, as we only have M pieces of information (Grenfell and Anderson, 1985).

* Corresponding author.
E-mail address: s.meakin@warwick.ac.uk (S.R. Meakin).

For spatial models, there have been some efforts to construct mechanistic models of between-population interaction (Belik et al., 2011; Sattenspiel and Dietz, 1995; Keeling and Rohani, 2002). Other approaches estimate the coupling using a generalised gravity transmission model in combination with data, for example: commuter mobility data (Viboud et al., 2006; Balcan et al., 2009), mobile phone data (used as a proxy for human mobility) (Tizzoni et al., 2014) or spatiotemporal time series of disease incidence (Xia et al., 2004), where coupling parameters are estimated so that simulated epidemic sizes, fade-out lengths and spatial synchrony patterns replicate observed ones. However, the rules governing between-population interactions are complex and good data on relevant movements between populations is rare, especially in developing countries where epidemiological models are likely to be applied. Moreover, even with access to good data it is far from clear how a given pattern of movements should translate into a single phenomenological transmission parameter for the subpopulations concerned. Radiation models have been proposed more recently as an alternative to gravity models and only require spatial distribution of population to estimate coupling (Simini et al., 2012; Masucci et al., 2013), although this model sometimes fail to describe human mobility at smaller scales (Masucci et al., 2013; Yang et al., 2014). For age-structured models, the transmission matrix is often based upon diary-based records of interactions (Mossong et al., 2008; Danon et al., 2013; Read et al., 2014), but the same issues regarding translation of social interactions into a transmission rate between age-groups applies.

However, for a stochastic system, the $\frac{1}{2}M(M-1)$ correlations between the levels of infection in the subpopulations may help to mitigate this unidentifiability, especially if the transmission matrix can be assumed to have some form of symmetry. Such long-term data on disease incidence is more widely available (Olsen and Schaffer, 1990; Grenfell and Harwood, 1997). From this data we can estimate the correlation between epidemics in distinct subpopulations and, given a relationship between the coupling strength and the correlation, we can then infer possible movement parameters. This extra information that is accrued from the correlations between subpopulations is a substantive motivating element for this study.

Due to the nonlinearity of the processes invoked by transmission, exact analysis of stochastic epidemics is often mathematically intractable. Computer simulation of the epidemic process, whilst commonly used and clearly useful, provides only a limited insight into the dynamics and can be computationally expensive. Approximation methods circumvent these issues and allow us to derive analytic results and develop intuition about the expected behaviour and variability of the epidemic process. One widely used method is the moment closure approximation, whereby differential equations for the means, variances and covariances can be derived either from first principles or from the Kolmogorov forward equations (Keeling, 2000a,b; Lloyd, 2004; Keeling and Ross, 2008). The most commonly used moment closure approximation, and the one used throughout this paper, assumes that third-order cumulants and higher are equal to zero; this is equivalent to assuming the distribution of states follows a multivariate normal distribution (Whittle, 1957). Alternative approximations have been proposed based on different distributional assumptions: the multiplicative moment closure approximation of Keeling (2000a) assumes a multivariate log-normal distribution; (Nåsell, 2003) assumes a binomial distribution; while (Krishnarajah et al., 2005) assumes a beta-binomial distribution.

In this paper we derive an approximation for the correlation between the level of infection in two distinct subpopulations as a function of the relative transmission rates, or the coupling, between them; this improves upon the results of Keeling and Rohani (2002) and corrects an error in the parametrisation of the original approximation. Using a multivariate normal moment closure approximation we derive this approximation for the simple case of two identical subpopulations and provide conditions under which we expect this result to hold. We also numerically evaluate our model and compare our analytic

approximation to stochastic simulations of the epidemic process.

2. Model formulation

2.1. A simple stochastic epidemic model

We begin by introducing the notation for a simple stochastic *SIR* model, with births, deaths, transmission and recovery. At any time $t \in [0, \infty)$, individuals are in one of three states: susceptible, infected or recovered. A given susceptible individual meets other individuals at rate $k > 0$. We assume that these encounters are sufficiently close that if the other individual is infected, then transmission of infection occurs with probability τ and the susceptible individual immediately becomes infected and infectious to others. Reverting to standard epidemiological modelling notation, we let the transmission rate be $\beta = k\tau$. Susceptible individuals can also succumb to infection independent of contact with infected individuals in the populations; this occurs at rate $\epsilon > 0$, the external import rate. Infected individuals recover from infection at rate $\gamma > 0$, after which they become immune to further infection. Susceptible, infected and recovered individuals all die at rate $\mu > 0$, independent of infection status; we assume that a death is immediately followed by the birth of a susceptible individual, and hence the total population size remains constant. The basic reproductive ratio, R_0, for this process is $R_0 = \beta/(\gamma + \mu)$. Let $S(t), I(t), R(t) \in \{0, 1, 2, \dots\}$ denote the number of susceptible, infected and recovered individuals, respectively, at time $t \geq 0$. If we take the (constant) population size to be N then we can reduce the dimensionality of the system by setting $R(t) = N - S(t) - I(t)$.

Equivalently, we can write down the Kolmogorov forward equation, also known as the ensemble or master equation, for this process. Let $p_t(s, i)$ denote the probability that there are s susceptible individuals and i infectious individuals in the population at time t; then the Kolmogorov forward equations are given by

$$\frac{\mathrm{d}p_t(s,i)}{\mathrm{d}t} = \left(\frac{\beta}{N}(s+1)(i-1) + \epsilon(s+1) \right) p_t(s+1, i-1) + \gamma(i+1)p_t(s, i+1)$$
$$+ (\mu(i+1)p_t(s-1, i+1) + \mu(N-(s-1)-i)p_t(s-1, i)$$
$$- \left(\frac{\beta}{N}si + \epsilon s + \gamma i + \mu i + \mu(N-s-i) \right) p_t(s, i).$$

$$(1)$$

This formulation is extremely useful for describing the complete nature of a given stochastic system, but for large population sizes the method is impractical as the number of ordinary differential equations (ODEs) grows quadratically with the population size. However, such equations have the advantage that they can be mechanistically used, together with the closure assumption, to derive equations for means, variances and other moments of interest.

2.2. A stochastic epidemic model for coupled populations

Now consider a pair of identical populations of size N. We assume the populations are the same size for analytic tractability; we discuss the relaxation of this assumption in Sections 3 and 4. Furthermore, we assume that both populations exhibit the same population dynamics as the simple stochastic epidemic model described in Section 2.1; however, we now assume that a proportion $\sigma \in [0, 1]$ of an individual's contacts are with individuals in the other population. In this way, σ describes the interaction, or 'coupling', between the two populations, and the force of infection in each population depends on the number of infected individuals in both populations. Changing σ does not change the basic reproductive ratio in this model, but simply determines the distribution of secondary cases between the two subpopulations.

We now let $S_j(t), I_j(t), R_j(t) \in \{0, 1, 2, \dots\}$ denote the number of susceptible, infected and recovered individuals, respectively, in population $j = 1, 2$ at time $t \geq 0$; and again insist that population sizes

Table 1

A summary of the transition rates of the four-dimensional Markov chain epidemic model $\{(S_1(t), I_1(t), S_2(t), I_2(t)) : t \geq 0\}$ from state (s_1, i_1, s_2, i_2) with birth/death rate $\mu > 0$, contact rate $\beta > 0$, external import rate $\epsilon > 0$, recovery rate $\gamma > 0$ and coupling $\sigma \in [0, 1]$.

Population	Event	Transition	Rate
$j, k \in \{1, 2\}$, $k \neq j$	Infection	$s_j \to s_j - 1$, $i_j \to i_j + 1$	$\beta s_j[(1 - \sigma)i_j + \sigma i_k]/N + \epsilon s_j$
	Recovery	$i_j \to i_j - 1$, $r_j \to r_j + 1$	γi_j
	Death of infected	$s_j \to s_j + 1$, $i_1 \to i_j - 1$	μi_j
	Death of recovered	$s_j \to s_j + 1$, $r_j \to r_j - 1$	$\mu(N - s_j - i_j)$

remain constant: $N = S_j(t) + I_j(t) + R_j(t)$, $\forall t \geq 0$, $j = 1, 2$. The transition rates for the resulting four-dimensional Markov chain from state (s_1, i_1, s_2, i_2) at time t are summarised in Table 1. Again the Kolmogorov forward equations for this process can be formulated; these are shown in the Supplementary Information.

2.3. Moment closure approximations in the coupled stochastic epidemic model

An exact analysis of the coupled stochastic epidemic model is mathematically intractable. Instead we consider the approximate behaviour of the first- and second-order central moments of the process. For the coupled stochastic epidemic model there are eight distinct first- and second-order central moments, five of which are 'within-population' and three of which are 'between-population'. Since the two populations are identical, there are symmetries within the system that can be exploited: for example, $\mathbb{E}[S_1] = \mathbb{E}[S_2]$ and $Var(S_1) = Var(S_2)$, and similarly for other central moments. We denote the within-population central moments by

$$
\begin{aligned}
\overline{S} &= \mathbb{E}[S_1] = \mathbb{E}[S_2] \\
\overline{I} &= \mathbb{E}[I_1] = \mathbb{E}[I_2] \\
C_{SS} &= Var(S_1) = Var(S_2) \\
C_{II} &= Var(I_1) = Var(I_2) \\
C_{SI} &= Cov(S_1, I_1) = Cov(S_2, I_2)
\end{aligned}
$$

and denote the between-population central moments by

$$
\begin{aligned}
\hat{C}_{SS} &= Cov(S_1, S_2) \\
\hat{C}_{II} &= Cov(I_1, I_2) \\
\hat{C}_{SI} &= Cov(S_1, I_2) = Cov(S_2, I_1).
\end{aligned}
$$

Using the Kolmogorov forward equation, we can write down an ODE for each of the eight first- and second-order central moments, details of this method can be found in existing literature on moment closure approximations in epidemiological modelling (Keeling and Rohani, 2002; Keeling, 2000b; Lloyd, 2004; Keeling et al., 2000); or the differential equation for $\mathbb{E}[X]$ can be calculated from first principles using:

$$
\frac{d\mathbb{E}[X]}{dt} = \sum_{\text{events}} \text{rate of event} \times \text{change in } X \text{ due to event.} \tag{2}
$$

Due to the non-linearity of the infection term in the model, the ODE for an nth-order moment will depend on one or more $(n + 1)$th-order moments. To fully define the system of ODEs we would therefore have to write down an infinite set of equations. To circumvent this problem we use a moment closure approximation, which truncates this set of equations at some order. Here, we make a second-order moment closure approximation, which assumes that third- and higher-order cumulants are equal to zero. In this way, third-order moments can be written in terms of the mean and covariance. This is equivalent to assuming that

the random variable has a multivariate normal (MVN) distribution (Whittle, 1957) and so we refer to this approximation as a second-order MVN moment closure approximation. The resulting set of eight ODEs and their derivation can be found in the Supplementary Information. Note that a first-order moment closure approximation assumes that second- and higher-order cumulants are equal to zero; this approximation returns the standard set of ODEs for the SIR-model, which describe the stochastic process in the large-population limit.

3. Results

We derive a theoretical approximation for the correlation at endemic equilibrium between the number of infected individuals in population 1 and the number of infected individuals in population 2 as a function of the coupling, σ. We define the correlation between the number of infected individuals in each population at endemic equilibrium as:

$$
\rho = \frac{\text{Cov}(I_1, I_2)}{\sqrt{\text{Var}(I_1)\text{Var}(I_2)}},
$$

which, in the case of two identical populations where the variances are equal and using our earlier notation, simplifies to:

$$
\rho = \frac{\hat{C}_{II}^*}{C_{II}^*},
$$

where X^* denotes the quantity X at endemic equilibrium.

3.1. Theoretical result

For two identical populations we find that we can write the correlation as a sigmoidal function plus a correction term that is often relatively small:

$$
\rho = \frac{\sigma}{\xi + \sigma} - \Delta, \tag{3}
$$

where

$$
\xi = \frac{N(\gamma + \mu) - \beta\overline{S}^*}{\beta\overline{S}^*} \tag{4}
$$

and

$$
\Delta = \frac{(\beta\overline{I}^* + N\epsilon)\frac{\hat{C}_{SI}^*}{C_{II}^*}}{\beta(1 - \sigma)\overline{S}^* - N(\gamma + \mu)}. \tag{5}
$$

To derive this result, we use the moment equation for $\hat{C}_{II}$ derived in the Supplementary Information:

$$
\frac{d\hat{C}_{II}}{dt} = 2\frac{\beta}{N}\sigma\overline{S}C_{II} + 2\left(\frac{\beta}{N}(1 - \sigma)\overline{S} - \gamma - \mu\right)\hat{C}_{II} + 2\left(\frac{\beta}{N}\overline{I} + \epsilon\right)\hat{C}_{SI}. \tag{6}
$$

At equilibrium $d\hat{C}_{II}/dt = 0$ and if we divide by $2C_{II}^*/N$, then

$$
0 = \beta\sigma\overline{S}^* + (\beta(1 - \sigma)\overline{S}^* - N(\gamma + \mu))\rho + (\beta\overline{I}^* + N\epsilon)\frac{\hat{C}_{SI}^*}{C_{II}^*}, \tag{7}
$$

and hence we have the following approximation for the correlation that we will henceforth refer to as the MVN correlation:

$$
\begin{aligned}
\rho &= \frac{-\beta\sigma\overline{S}^*}{\beta(1 - \sigma)\overline{S}^* - N(\gamma + \mu)} - \frac{(\beta\overline{I}^* + N\epsilon)\frac{\hat{C}_{SI}^*}{C_{II}^*}}{\beta(1 - \sigma)\overline{S}^* - N(\gamma + \mu)} \\
&= \frac{\sigma}{\left(\frac{N(\gamma + \mu) - \beta\overline{S}^*}{\beta\overline{S}^*}\right) + \sigma} - \Delta \\
&= \frac{\sigma}{\xi + \sigma} - \Delta. \tag{8}
\end{aligned}
$$

Moreover, if we can show that $\Delta \ll 1$ then we have the following simplified approximation for the correlation

$$\rho \approx \frac{\sigma}{\xi + \sigma}. \tag{9}$$

If we relax the assumption that the two populations are of equal size, then we do not obtain this simple relationship between the coupling and correlation. The model is less amenable to analytic methods and the resulting approximation now depends on $r = N_1/N_2$ and on the equilibrium values $\overline{S}_j^*$ and $Var(I_j)$, $j = 1, 2$. To fully define the approximation we would need to estimate these values from data or through simulation. This result and full derivation are given in the Supplementary Information.

We can also derive an alternative approximate expression for ξ that is independent of $\overline{S}^*$, hence eliminating the need to find the equilibrium of the 8-dimensional ODE model. By ignoring the effects of both imports and correlations and taking the large population limit, we can find an approximation to S^*, which leads to the following expression:

$$\xi \approx \xi' = \frac{\epsilon(\gamma + \mu)}{\mu(\beta - \gamma - \mu)} = \frac{\epsilon}{\mu(R_0 - 1)}. \tag{10}$$

Full details of this derivation can be found in the Supplementary Information. This parametrisation of ξ is preferable to the original (Eq. (4)) as it removes the need to estimate the number of susceptible individuals in the population at endemic equilibrium, either from data or through simulation. In addition, this alternative parametrisation provides intuition into how the epidemic parameters directly impact the correlation. We can see that as R_0 increases then the correlation also increases. Conversely, as the external import rate ϵ increases, then the correlation decreases: as ϵ increases then external infections mask the effect of the between-population infections. Given the appeal of the simpler from of Eq. (10), in the work that follows we evaluate the approximation of the correlation ρ by the sigmoidal function $\sigma/(\xi' + \sigma)$.

The MVN moment closure approximation holds in the large-population limit (i.e $N \to \infty$) and assumes that the distribution of states is a multivariate normal distribution; this follows from the results of (Kurtz, 1970, 1971), which show that a stochastic process can be approximated by a deterministic processes in the large population limit. Further error in approximation comes from assuming that $\Delta \ll 1$ and that ξ is constant and equal to ξ'. In the following section, our aim is to understand whether our approximation (Eq. (9)) and expression for ξ' (Eq. (10)) are generic to a wider range of assumptions and parameters.

Given parameter values for our coupled stochastic epidemic model, we are able solve the underlying ODEs and hence check numerically that Δ is small and calculate ξ. The absolute error introduced into our approximation by assuming that $\Delta \ll 1$ is given by Δ; the error relative to the correlation ρ is given by Δ/ρ. The absolute error introduced into our approximation by assuming that ξ is constant (Eq. (10)) is $[\rho - \sigma/(\xi' + \sigma) + \Delta]$, or equivalently $[\sigma/(\xi + \sigma) - \sigma/(\xi' + \sigma)]$; the error relative to the correlation ρ is given by $[\sigma/(\xi + \sigma) - \sigma/(\xi' + \sigma)]/\rho$. In this paper we take 0.1 as a threshold for the absolute error and 0.25 as a threshold for the error relative to the correlation ρ. If the absolute or relative error exceed 0.1 or 0.25 respectively, then we say that the approximation fails.

3.2. Numerical results

To compare our analytic predictions to simulation results, we need to define a set of base parameters. In the majority of the numerical analysis we will utilise parameters for a highly-transmissible measles-like endemic disease in the UK (Anderson and May, 1992), although we note that a full model of measles requires both seasonality (Earn et al., 2000; Rohani et al., 2002; Grenfell and Bolker, 1995) and age-structure (Schenzle, 1984; Keeling and Grenfell, 1997; Bolker, 1993). We will also use parameters representing mumps (Anderson and May, 1992), rubella (Anderson and May, 1992), chickenpox (Anderson and May, 1992), whooping cough (Anderson and May, 1992), smallpox (Keeling and Rohani, 2008) and influenza (Cauchemez et al., 2004; Biggerstaff

Table 2
Epidemiological parameters for seven infectious diseases in the UK; across all diseases we take $N = 10^5$, $\mu = 5.5 \times 10^{-5}$ days^{-1} and $\epsilon = 5.5 \times 10^{-5}$ days^{-1}. We also give the value of the parameter $\xi' = \epsilon/(\mu(R_0 - 1))$ taken in our approximation for the correlation.

Disease	Basic reproductive ratio R_0	Average infectious period γ^{-1} (days)	ξ'
Measles	17	13	0.0625
Mumps	12	21	0.0909
Rubella	7	17	0.1667
Chickenpox	11	20	0.1
Whooping cough	17	22	0.0625
Smallpox	5	7	0.25
Influenza	2	4	1

et al., 2014). For all diseases, we consider two identical populations of size $N = 10^5$ where $\mu = 5.5 \times 10^{-5}$ days^{-1} and $\epsilon = 5.5 \times 10^{-5}$ days^{-1}. Disease-specific parameters are given in Table 2. The numerical integration of ODEs is performed using the MATLAB ode45 solver with a relative error tolerance of 10^{-5}.

Fig. 1 shows the equilibrium values of the first-order central moments $\overline{S}^*$ and $\overline{I}^*$ and second-order central moments C_{II}^* and $\hat{C}_{II}^*$ for a measles-like endemic disease in the UK as the coupling parameter σ is varied between 0 and 1. These results are obtained by numerical integration of the 8-dimensional ODEs given in the Supplementary Information, and therefore only depend on the MVN moment closure approximation. We note that all curves broadly show a sigmoidal pattern (although $\overline{S}^*$ has a minimum and $\overline{I}^*$ a maximum at $\sigma = 0.5$), with $\overline{S}^*$ and C_{II}^* decreasing with the coupling and $\overline{I}^*$ and $\hat{C}_{II}^*$ increasing with the coupling.

For these measles-like epidemic parameters, we compare the MVN correlation ρ (Eq. (8)) and our approximation $\sigma/(\xi' + \sigma)$, $\xi' = 0.0625$, to the results of full stochastic simulations (Fig. 2a). We simulate the stochastic process over a 200 year period using the Gillespie algorithm, with a burn-in period of 50 years, and generate 1000 realisations of the process for each value of σ. The correlation is calculated as a time-weighted Pearson correlation coefficient for $50 < t \le 200$ years. From this comparison we draw three conclusions. Firstly, all three correlations follow a sigmoidal relationship increasing from zero for low coupling to a value close to one when the coupling is largest- although we note that values of $\sigma > 0.5$ do not match with our idealised view of a metapopulation in which within-population transmission is larger than between-population transmission. Secondly, the remarkably close agreement between ρ and the simulation results, suggest our use of the MVN moment closure approximation is justified. Finally, $\sigma/(\xi' + \sigma)$ is a reasonable approximation for the MVN correlation ρ as the difference between the two curves is small.

We also compare the MVN correlation ρ and our approximation $\sigma/(\xi' + \sigma)$ for six other infectious diseases in the UK: mumps, rubella, chickenpox, whooping cough, smallpox and influenza (Fig. 2b). Interestingly we observe that our approximation underestimates the correlation for diseases with a high R_0 (e.g. whooping cough) and overestimates the correlation for diseases with a low R_0 (e.g. smallpox and influenza). We attribute this to the differential action of Δ and the approximation to ξ across epidemiological parameters. However, across all diseases the difference between ρ and our approximation is small, hence we can relate the phenomenological coupling parameter, σ, to the correlation between the number of infected individuals in two populations by $\rho = \sigma/(\xi' + \sigma)$.

In Fig. 3, we evaluate the two main sources of error in our approximation (Eq. (9)), introduced by assuming that $\Delta \ll 1$ and that ξ is constant and equal to ξ'. For measles-like parameters, Δ is small in both absolute and relative terms (green lines in Fig. 3); importantly, Δ never exceeds our chosen thresholds of 0.1 and 0.25 for the absolute and relative error respectively, and has a diminishing impact on the
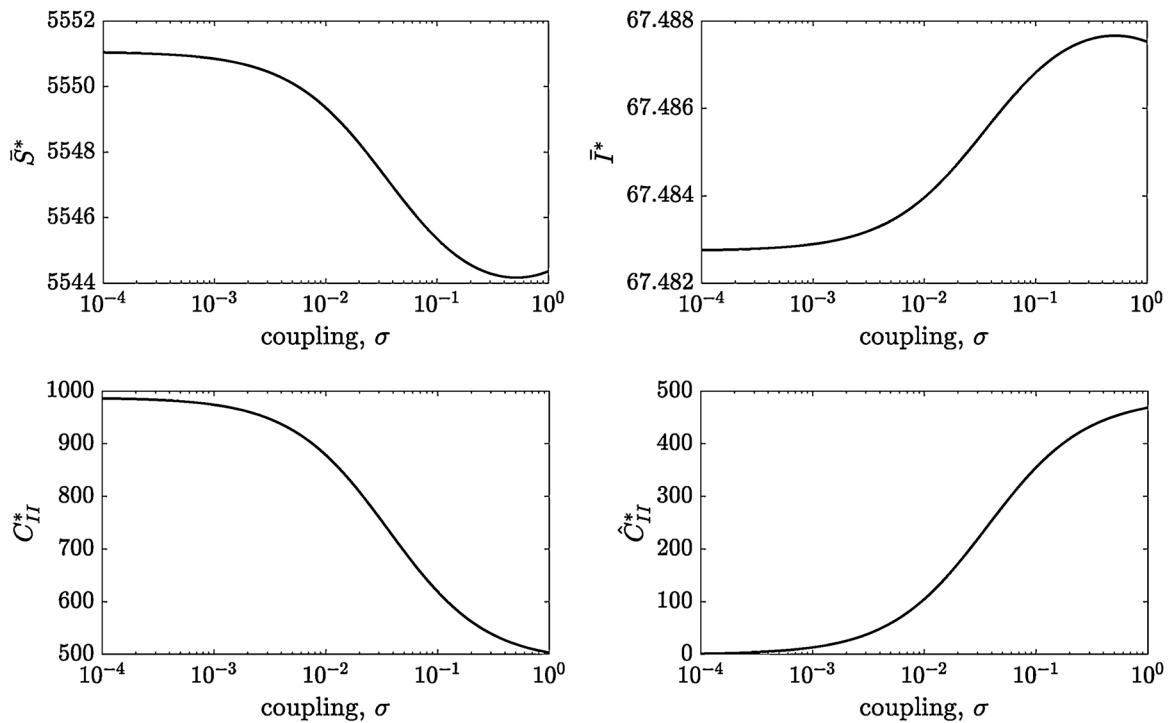
Fig. 1. The effect of the coupling, σ, on key mean variables $\bar{S}^*$, $\bar{I}^*$, C_{II}^* and $\hat{C}_{II}^*$ for a measles-like endemic disease in the UK ($N = 10^5$, $\mu = 5.5 \times 10^{-5}$, $R_0 = 17$, $\gamma^{-1} = 13$ and $\epsilon = 5.5 \times 10^{-5}$), calculated from the ODEs given in the Supplementary Information.

correlation when the coupling σ exceeds 0.065. The error introduced into the approximation by assuming ξ is constant is given by $[\rho - \sigma/(0.0625 + \sigma) + \Delta] = [\sigma/(\xi + \sigma) - \sigma/(0.0625 + \sigma)]$ (yellow lines in Fig. 3). Again, we observe that this error is well within our chosen thresholds of 0.1 and 0.25 for the absolute and relative error respectively. Moreover, this error is approximately one order of magnitude smaller than Δ, which tells us that for measles-like parameters the main source of error is our approximation is due to assuming that $\Delta \ll 1$. Overall, these findings suggest that our simple approximation (Eq. (9)) should hold for these parameters across the entire range of coupling values.

3.2.1. Smaller population sizes

Whilst the MVN moment closure approximation holds in the large population limit (Kurtz, 1970, 1971), we may often be interested in much smaller populations where the impact of stochasticity is more pronounced. For $N = 10^2, 10^3, 10^4, 10^5$ we compare our approximation for the correlation to stochastic simulations. We generate 1000 realisations of each (N, σ) pair using the same method as described in Section 3.2 and calculate the mean of all first- and second-order central moments and the mean and variance of the correlation. Since $\xi' = \epsilon/(\mu(R_0 - 1))$ is independent of N then we take $\xi' = 0.0625$ for all N.

In Fig. 4 we compare the stochastic simulations for each of $N = 10^2$, $10^3, 10^4, 10^5$ to our approximation $\sigma/(0.0625 + \sigma)$. We find that, for a given σ, decreasing the population size leads to weaker correlations; equivalently, this means that in smaller populations stronger coupling is required to achieve the same level of correlation. This is because in smaller populations the correlation between the two populations is reduced by independent stochastic effects acting on the two populations. Despite this, we find even for $N = 10^3$ the correlation between the two populations is well approximated by $\sigma/(\xi' + \sigma)$; only at very small population sizes, $N = 100$, is our approximation a poor estimate of the correlation.

Although it is simpler to take $\xi = \xi'$, this value can also be calculated as $\xi = (N(\gamma + \mu) - \beta\bar{S}^*)/\beta\bar{S}^*$, for some specific value of $\bar{S}^*$. This method requires the numerical integration of the ODEs given in the

Supplementary Information; however, we find that for $N \lesssim 10^{4.2} \approx 16$, 000 the numerical solution to the system of ODEs "blows up". Therefore, we cannot use this method for calculating ξ to parametrise our approximation in smaller populations. This phenomenon occurs since we assume that the distribution of states follows a multivariate normal distribution and at low levels of infection this leads to a significant proportion of the distribution being negative. For example, for $N = 10^5$ then $\bar{I}^* \approx 67$, but for $N \lesssim 10^{4.2}$ then $\bar{I} \lesssim 10.6$. Zero infectious cases should act as a boundary for the distribution, and hence as $\bar{I}$ reduces the multivariate normal assumption breaks down.

3.2.2. Parameter sensitivity analysis

We perform a brief parameter sensitivity analysis to understand how the correlation between the number of infected individuals in the two populations is affected by the value of the epidemiological parameters. In the first half of the analysis, we use the parameters for a measles-like disease in the UK ($N = 10^5$, $\mu = 5.5 \times 10^{-5}$, $R_0 = 17$, $\gamma^{-1} = 13$ and $\epsilon = 5.5 \times 10^{-5}$) and independently change the value of each of the four parameters μ, β, ϵ and γ. We show the impact of these epidemiological parameters on $\xi' = \epsilon/(\mu(R_0 - 1))$. For each set of epidemiological parameter values we also calculate Δ and compare $\xi \sim \xi'$ across all values of coupling σ to determine the range of parameter values for which our approximation holds, that is, for which the absolute and relative errors are within our chosen thresholds of 0.1 and 0.25 respectively for all coupling values.

We find that the values of each of the four key parameters have a profound impact on the correlation between the number of infected individuals in the two populations, but that our approximation holds for a wide range of realistic values (Fig. 5a). The correlation increases with the birth rate, μ, the basic reproductive ratio, $R_0 = \beta/(\gamma + \mu)$, varied by changing β, and the mean infectious period, γ^{-1}; increases in the external import rate, ϵ, lead to a decrease in the correlation. The exact region in which our approximation fails is a complex trade-off between all parameters; however, for all four parameters the approximation fails (that is, either the absolute or relative error exceeds our chosen thresholds) as ξ'^+ becomes smaller; failures occur for
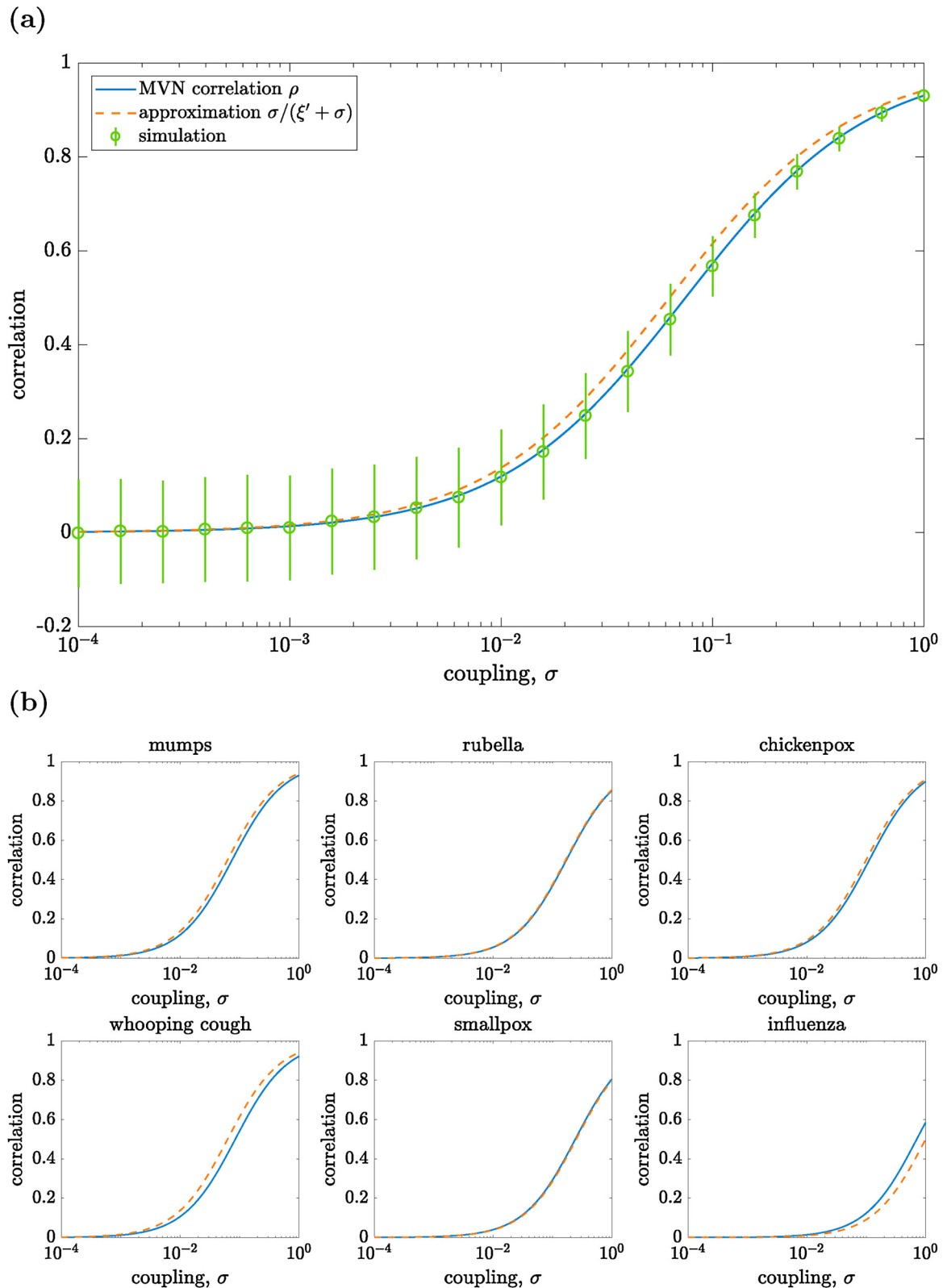
Fig. 2. Comparing the MVN correlation ρ and our approximation $\sigma/(\xi' + \sigma)$ to stochastic simulations, for (a) a measles-like endemic disease in the UK ($N = 10^5$, $\mu = 5.5 \times 10^{-5}$, $R_0 = 17$, $\gamma^{-1} = 13$ and $\epsilon = 5.5 \times 10^{-5}$; $\xi' = 0.0625$) and (b) for parameters representing mumps, rubella, chickenpox, whooping cough, smallpox and influenza (parameter values are given in Table 2). We generate 1000 realisations of the process for each value of σ and calculate the correlation as a time-weighted Pearson correlation coefficient for $50 < t \leq 200$ years; error bars represent $\pm$ 2 standard deviations.

$\mu \gtrsim 5.62 \times 10^{-5}$, $\beta \gtrsim 1.35$ ($R_0 \gtrsim 17.5$), $\epsilon \lesssim 4.61 \times 10^{-5}$ and $\gamma^{-1} \gtrsim 13$. This failure mode is due to the growing importance of the correction term Δ relative to our approximation $\sigma/(\xi' + \sigma)$.

In the second half of the analysis, we focus on the external import rate ϵ, as this is generally the most difficult parameter to estimate. For each of the epidemiological parameter sets representing mumps,
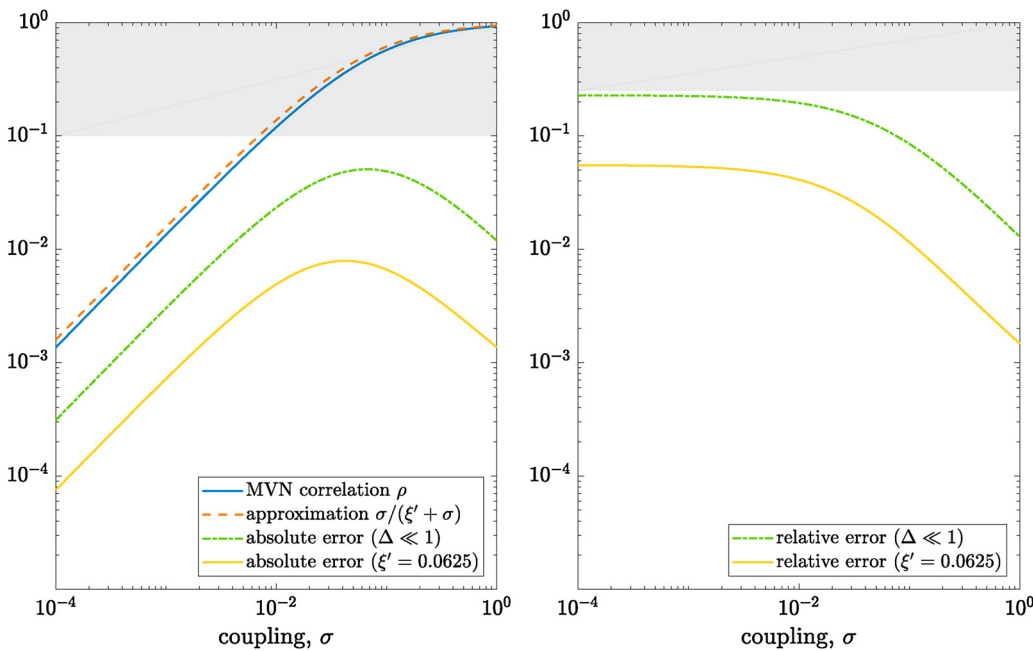
Fig. 3. Evaluating the sources of error in our approximation for a measles-like endemic disease in the UK ($N = 10^5$, $\mu = 5.5 \times 10^{-5}$, $R_0 = 17$, $\gamma^{-1} = 13$ and $\epsilon = 5.5 \times 10^{-5}$), taking 0.1 and 0.25 as thresholds for the absolute (left) and relative (right) error, respectively. We compare the MVN correlation ρ and our approximation $\sigma/(\xi' + \sigma)$ to the two sources of error in our approximation: assuming $\Delta \ll 1$, and assuming that ξ is constant and equal to $\xi' = \epsilon/(\mu(R_0 - 1)) = 0.0625$.

rubella, chickenpox, whooping cough, smallpox and influenza, and for each value of ϵ, we show $\xi' = \epsilon/(\mu(R_0 - 1))$. For each diseases parameter set we also determine a range of values for ϵ for which the approximation holds, that is, for which the absolute and relative error are within our chosen thresholds of 0.1 and 0.25 respectively.

We find that the external import rate ϵ has a significant impact on the correlation (Fig. 5b). For all diseases we consider, increasing the external import rate leads to a higher value of ξ' and thus predicts a lower correlation for a given coupling strength: as the external import rate is increased, external infections mask the effect of the between-population infections. For measles, mumps, rubella, chickenpox, whooping cough and smallpox, the approximation fails (that is, either the absolute or relative error exceeds our chosen thresholds) as ϵ become smaller. We also observe that our approximation fails at lower values of ϵ in diseases with a lower R_0: for example, our approximation

for whooping cough ($R_0 = 17$) fails for $\epsilon \lesssim 8.71 \times 10^{-5}$, whereas our approximation for rubella ($R_0 = 7$) fails for $\epsilon \lesssim 9.78 \times 10^{-6}$. However, for influenza, our approximation fails as ϵ becomes larger. Further analysis (given in the Supplementary Information) shows that for influenza, $\xi' = \epsilon/\mu(R_0 - 1)$) significantly overestimates ξ for large values of ϵ, so assuming that $\xi = \xi'$ leads to a large error in our approximation. However, for smallpox, rubella, chickenpox, mumps and whooping cough the difference between ξ' and ξ is small and so we do not observe failure for large values of ϵ in the range of values that we consider; the value of ϵ would have to be unrealistically high to observe such an effect.

4. Discussion

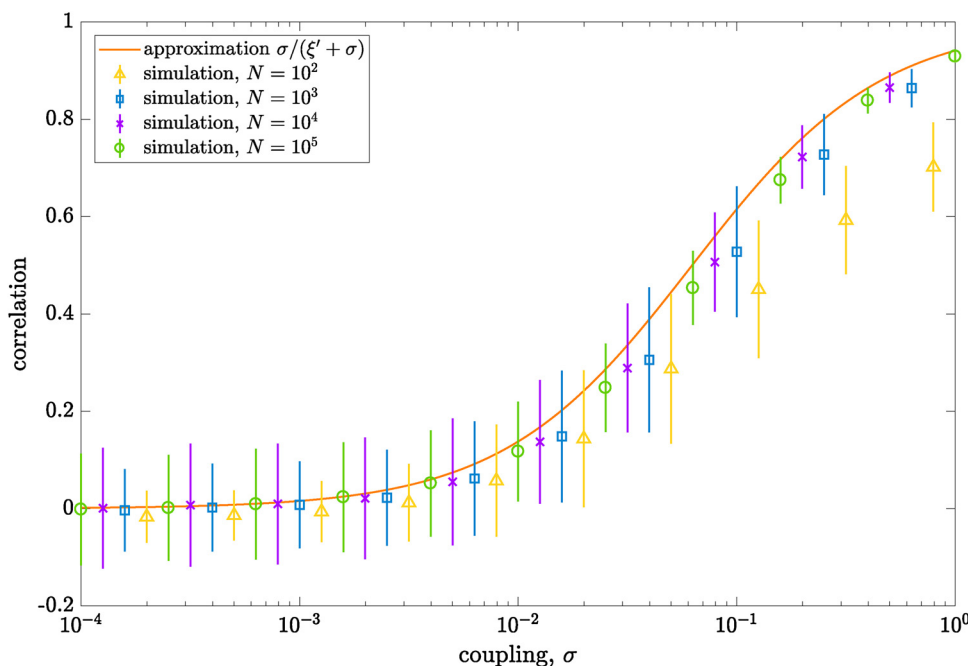A limitation of metapopulation-type models within epidemiological

Fig. 4. Comparing our approximation $\sigma/(\xi' + \sigma)$, $\xi' = 0.0625$, to stochastic simulations for a measles-like endemic disease in the UK ($\mu = 5.5 \times 10^{-5}$, $R_0 = 17$, $\gamma^{-1} = 13$ and $\epsilon = 5.5 \times 10^{-5}$) and $N = 10^2, 10^3, 10^4, 10^5$. We simulate the stochastic process over a 200 year period using the Gillespie algorithm, with a burn-in period of 50 years, and generate 1000 realisations of the process for each of (N, σ) pair. The correlation is calculated as a time-weighted Pearson correlation coefficient for $50 < t \leq 200$; error bars represent ± 2 standard deviations.
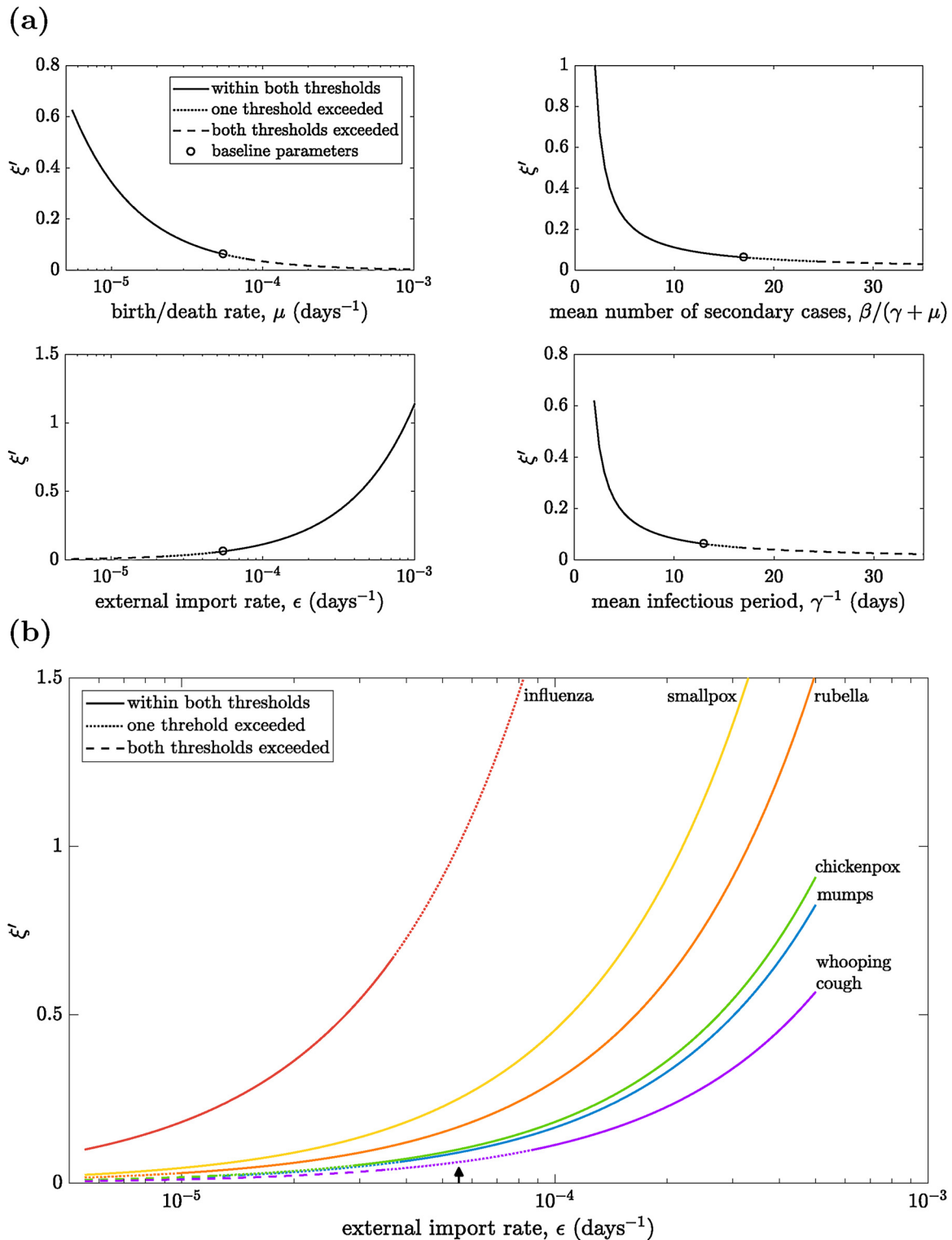
(a)

(b)

Fig. 5. Sensitivity analysis for our approximation to the epidemic parameters. The parameter ξ in our approximation is calculated as $\xi' = \epsilon/(\mu(R_0 - 1))$; our approximation holds if both the absolute and relative error is within our chosen thresholds of 0.1 and 0.25 respectively (represented by a solid line). The approximation fails due to one or both of the absolute and relative errors exceeding our chosen thresholds (dotted and dashed lines respectively). This analysis is performed for (a) each of the four epidemiological parameters μ, β, ϵ and γ with baseline parameters $\mu = 5.5 \times 10^{-5}$, $R_0 = 17$, $\gamma^{-1} = 13$ and $\epsilon = 5.5 \times 10^{-5}$ (shown by a circle) and (b) for the external import rate ϵ for the given diseases with baseline parameters given in Table 2 (shown by an arrow on the *x*-axis). $N = 10^5$ throughout.

modelling is how to infer the coupling between subpopulations. Sufficiently rich data on relevant interactions is often lacking, especially in developing countries, and it is unclear how such data should translate into a single phenomenological coupling parameter. In light of

data on disease incidence being more widely available, we derive an approximation for the correlation, ρ, between the number of infected individuals in two identical populations as a function of the coupling parameter σ, providing a one-to-one mapping between the correlation

and the coupling.

The results presented here refine the analysis of (Keeling and Rohani, 2002) and correct an error in the original derivation of ξ. Our numerical results for a measles-like infection show substantial correlation for all but the weakest coupling. These findings are consistent with similar studies focussing on persistence and spatial synchronisation of measles outbreaks (Lloyd, 2004; Bolker and Grenfell, 1996), despite differences in the characterisation of the basic model. An analytic relationship between the coupling and correlation has been previously derived (Rozhnova et al., 2012) in a more general setting and yields similar numerical results: their relationship is derived through the van Kampen system-size expansion and analysis of the power spectrum. However, we believe that our results provide a significantly simpler relationship between correlations and epidemiological parameters, providing greater intuition and analytical traction. In addition, throughout we compare our analytically tractable results to solution of the moment-based ODEs (given in the Supplementary Information) and to numerical simulation, providing a deeper understanding of the parameter ranges over which the simple results hold and hence the range of applications where the methods are of use. We also differentiate between different modes of failure in our approximation between diseases with low and high basic reproductive ratios.

Our work also offers an alternative parametrisation of ξ (Eq. (10)) that depends only on the epidemiological parameters and holds in the large-population limit; however, our numerical analysis shows that this parametrisation also leads to a good qualitative approximation in populations of size $N = 10^3$. This parametrisation is preferable to the original as it provides intuition and insight into how the epidemic parameters affect the correlation. In addition, it removes the need to estimate the number of susceptible individuals in the population at endemic equilibrium, either from data or through simulation. This is particularly useful in smaller populations where we find that the MVN moment closure approximation fails numerically and the solution to the ODEs 'blows up'. This type of failure is well-documented in the literature and typically attributed to large negative covariances, frequent global extinctions or when the distribution of states is bimodal (Keeling, 2000a,b; Lloyd, 2004; Krishnarajah et al., 2005; Nåsell, 1999). In the limit as $N \to \infty$, (Nåsell, 1999) shows that the distribution of states conditioned on non-extinction is approximately normal when R_0 is greater than 1; however, this does not explain why the MVN moment closure approximation sometimes appears to hold for smaller populations, such as in our own analysis and the wider literature (Isham, 1995).

Our model is sufficiently general that it can describe multiple forms of heterogeneity in the population including spatial, age and risk heterogeneity; however, a limitation of the model it that the underlying SIR model is too simple describe the full dynamics of many diseases. For example, as noted previously a full model of measles dynamics should include both seasonality and age structure. These limitations should be addressed before using our results to infer the between-population coupling parameters. We have also shown that adding complexity reduces the analytical tractability of the model, such as with populations of unequal size; in the most general case, analysis of the model may require a computational, rather than analytical, approach. Finally, whilst data on disease incidence in each of the subpopulations is more widely available than mobility data, our results are still limited by the availability and quality of such data. In particular, our results will be affected by under-reporting of infections.

Our results provide a method by which the coupling can be estimated from the correlation between the number of infected individuals in two populations using data on disease incidence. Crucially, this allows us to estimate the coupling between subpopulations even in the absence of data on human mobility, thus circumventing one of the main limitations of metapopulation models. At present our model considers the mathematically tractable case of two identical populations at endemic equilibrium. Future research should aim to address the

limitations outlined above by improving the underlying epidemic model, for example by incorporating seasonality or age structure. The current model can easily be extended to multiple identical interacting populations when the underlying graph is a symmetric graph, such as the complete graph or k-regular infinite tree graph. This holds since any adjacent populations will have identical neighbourhoods. These extensions will significantly improve the realism of the model and validate the use of the results in the inference of between-population coupling parameters.

5. Conclusion

A limitation of metapopulation models is how to infer the coupling between subpopulations. This paper relates the correlation between the number of infected individuals in two identical populations as a function of the coupling, providing a one-to-one mapping between the correlation and the coupling. Combined with data on disease incidence in each of the subpopulations, this result provides a method by which the between-population coupling can be estimated, even in the absence of information on the population mobility.

Authors' contributions

M.J.K. developed the initial concepts. S.R.M. performed the detailed mathematical analysis. Both authors played a role in writing the manuscript.

Competing interests

We have no competing interests.

Funding

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.epidem.2018.08.005.

References

Anderson, R.M., May, R., 1992. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford, UK.

Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A., 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc. Natl. Acad. Sci. 106 (51), 21484–21489.

Belik, V., Geisel, T., Brockmann, D., 2011. Natural human mobility patterns and spatial spread of infectious diseases. Phys. Rev. X 1, 1–5. https://doi.org/10.1103/PhysRevX.1.011001.

Biggerstaff, M., Cauchemez, S., Reed, C., Gambhir, M., Finelli, L., 2014. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect. Dis. 14, 1–20.

Bolker, B.M., 1993. Chaos and complexity in measles models: a comparative numerical study. IMA J. Math. Appl. Med. Biol. 10 (2), 83–95. https://doi.org/10.1093/imammb/10.2.83. ISSN 14778599.

Bolker, B.M., Grenfell, B.T., 1996. Impact of vaccination on the spatial correlation and persistence of measles dynamics. Proc. Natl. Acad. Sci. USA 93 (22), 12648–12653. https://doi.org/10.1073/pnas.93.22.12648. ISSN 0027-8424.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A.J., Boelle, P.Y., 2004. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. Stat. Med. 23, 3469–3487. https://doi.org/10.1002/sim.1912.

Christley, R.M., Pinchbeck, G.L., Bowers, R.G., Clancy, D., French, N.P., Bennett, R., 2005. Infection in social networks: using network analysis to identify high-risk individuals. Am. J. Epidemiol. 162 (10), 1024–1031. https://doi.org/10.1093/aje/kwi308.

Danon, L., Read, J.M., House, T.A., Vernon, M.C., Keeling, M.J., 2013. Social encounter networks: characterizing Great Britain. Proc. R. Soc. B.

Diekmann, O., 1978. Thresholds and travelling waves for the geographical spread of infection. J. Math. Biol. 6 (2), 109–130.

Diekmann, O., Heesterbeek, J.A.P., Metz, J.A.J., 1990. On the definition and the

computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. 28, 365–382.

Earn, D.J.D., Rohani, P., Grenfell, B.T., Bolker, B.M., 2000. A simple model for complex dynamical transitions in epidemics. Science 287 (5453), 667–670.

Gilpin, M.E., Hanski, I.A. (Eds.), 1991. Metapopulation Dynamics: Empirical and Theoretical Investigations. Academic Press.

Grenfell, B.T., Anderson, R.M., 1985. The estimation of age related rates of infection from case notifications and serological data. Epidemiol. Infect. 95 (2), 419–436.

Grenfell, B.T., Bolker, B.M., 1995. Spatial heterogeneity, nonlinear dynamics and chaos in infectious diseases. Stat. Methods Med. Res. 4 (2), 160–183.

Grenfell, B.T., Bolker, B.M., 1998. Cities and villages: infection hierarchies in a measles metapopulation. Ecol. Lett. (1), 63–70.

Grenfell, B.T., Harwood, J., 1997. (Meta)population dynamics of infectious diseases. Trends Ecol. Evol. 12 (10), 395–399. https://doi.org/10.1016/S0169-5347(97)01174-9.. ISSN 01695347, http://www.sciencedirect.com/science/article/pii/S0169534797011749.

Grenfell, B.T., Bjørnstad, O.N., Kappey, J., 2002. Travelling waves and spatial hierarchies in measles epidemics. Nature 414 (6865), 716–723. https://doi.org/10.1038/414716a.

Hagenaars, T.J., Donnelly, C.A., Ferguson, N.M., 2004. Spatial heterogeneity and the persistence of infectious diseases. J. Theor. Biol. 229, 349–359. https://doi.org/10.1016/j.jtbi.2004.04.002.

Hanski, I.A., 1998. Metapopulation dynamics. Nature 396, 41–49.

Hanski, I.A., Gaggiotti, O.E. (Eds.), 2004. Ecology, Genetics, and Evolution of Metapopulations. Academic Press.

Heesterbeek, J.A.P., 2002. A brief history of R0 and a recipe for its calculation. Acta Biotheor. 50 (3), 189–204.

Isham, V., 1995. Stochastic models of host-macroparasite interaction. Ann. Appl. Probab. 5 (3), 720–740.

Keeling, M.J., 2000a. Multiplicative moments and measures of persistence in ecology. J. Theor. Biol. 205 (2), 269–281. https://doi.org/10.1006/jtbi.2000.2066. ISSN 0022-5193.

Keeling, M.J., 2000b. Metapopulation moments: coupling, stochasticity and persistence. J. Anim. Ecol. 69 (5), 725–736.

Keeling, M.J., 2000c. Evolutionary trade-offs at two time-scales: competition versus persistence. Proc. R. Soc. Lond. B: Biol. Sci. 267 (1441), 385–391.

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

Keeling, M.J., Rohani, P., 2002. Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecol. Lett. 5 (1), 20–29. https://doi.org/10.1046/j.1461-0248.2002.00268.x. ISSN 1461023X.

Keeling, M.J., Rohani, P., 2008. Modelling Infectious Diseases in Humans and Animals. Princeton University Press.

Keeling, M.J., Ross, J.V., 2008. On methods for studying stochastic disease dynamics. J. R. Soc. Interface 5 (19), 171–181. https://doi.org/10.1098/rsif.2007.1106. ISSN 1742-5689.

Keeling, M.J., White, P.J., 2010. Targeting vaccination against novel infections: risk, age and spatial structure for pandemic influenza in Great Britain. J. R. Soc. Interface 8 (58), 661–670. https://doi.org/10.1098/rsif.2010.0474. ISSN 1742-5689.

Keeling, M.J., Wilson, H.B., Pacala, S.W., 2000. Reinterpreting space, time lags, and functional responses in ecological models. Science 290 (5497), 1758–1761.

Keeling, M.J., Danon, L., Vernon, M.C., House, T.A., 2010. Individual identity and movement networks for disease metapopulations. Proc. Natl. Acad. Sci. 107 (May (19)), 8866–8870.

Krishnarajah, I., Cook, A., Marion, G., Gibson, G., 2005. Novel moment closure approximations in stochastic epidemics. Bull. Math. Biol. 67, 855–873. https://doi.org/10.

1016/j.bulm.2004.11.002.

Kurtz, T.G., 1970. Solutions of ordinary differential equations as limits of pure jump Markov processes. J. Appl. Probab. 7 (1), 49–58.

Kurtz, T.G., 1971. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. J. Appl. Probab. 8 (2), 344–356.

Lloyd, A.L., 2004. Estimating variability in models for recurrent epidemics: assessing the use of moment closure techniques. Theor. Popul. Biol. 65 (1), 49–65. https://doi.org/10.1016/j.tpb.2003.07.002. ISSN 00405809.

Masucci, A.P., Serras, J., Johansson, A., Batty, M., 2013. Gravity versus radiation models: on the importance of scale and heterogeneity in commuting flows. Phys. Rev. E 88 (2), 022812. https://doi.org/10.1103/PhysRevE.88.022812.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5 (3), 381–391. https://doi.org/10.1371/journal.pmed.0050074. ISSN 15491277.

Nåsell, I., 1999. On the time to extinction in recurrent epidemics. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 61 (2), 309–330.

Nåsell, I., 2003. Moment closure and the stochastic logistic model. Theor. Popul. Biol. 63, 159–168. https://doi.org/10.1016/S0040-5809(02)00060-6.

Olsen, L.F., Schaffer, W.M., 1990. Chaos versus noisy periodicity: alternative hypotheses for childhood epidemics. Science 249 (4968), 499–504. https://doi.org/10.1126/science.2382131.

Read, J.M., Lessler, J., Riley, S., Wang, S., Tan, L.J., Kwok, K.O., Guan, Y., Jiang, C.Q., Cummings, D.A.T., 2014. Social mixing patterns in rural and urban areas of southern China. Proc. R. Soc. B 281 (June (1785)).

Rohani, P., Keeling, M.J., Grenfell, B.T., 2002. The interplay between determinism and stochasticity in childhood diseases. Am. Nat. 159 (5), 469–481. https://doi.org/10.1086/339467. ISSN 0003-0147.

Rozhnova, G., Nunes, A., Mckane, A.J., 2012. Phase lag in epidemics on a network of cities. Phys. Rev. E 85 (5), 051912. https://doi.org/10.1103/PhysRevE.85.051912.

Sattenspiel, L., Dietz, K., 1995. A structured epidemic model incorporating geographic mobility among regions. Math. Biosci. 128 (1), 71–91.

Schenzle, D., 1984. An age-structured model of pre- and post-vaccination measles transmission. Math. Med. Biol.: J. IMA 1 (2), 169–191.

Simini, F., González, M.C., Maritan, A., Barabási, A.L., 2012. A universal model for mobility and migration patterns. Nature 484 (7392), 96–100. https://doi.org/10.1038/nature10856. ISSN 00280836.

Tizzoni, M., Bajardi, P., Decuyper, A., Kon, G., King, K., Schneider, C.M., 2014. On the use of human mobility proxies for modeling epidemics. PLoS Comput. Biol. 10 (7). https://doi.org/10.1371/journal.pcbi.1003716.

Viboud, C., Bjørnstad, O.N., Smith, D.L., Simonsen, L., Miller, M.A., Grenfell, B.T., 2006. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science 312 (5772), 447–451. https://doi.org/10.1126/science.1125237. ISSN 1095-9203.

Wallinga, J., van Boven, M., Lipsitch, M., 2010. Optimizing infectious disease interventions during an emerging epidemic. Proc. Natl. Acad. Sci. 107 (January (2)), 923–928.

Whittle, P., 1957. On the use of the normal approximation in the treatment of stochastic processes. J. R. Stat. Soc. Ser. B (Methodol.) 19 (2), 268–281.

Xia, Y., Bjørnstad, O.N., Grenfell, B.T., 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. Am. Nat. 164 (2), 267–281. https://doi.org/10.1086/422341. ISSN 1537-5323.

Yang, Y., Herrera, C., Eagle, N., Gonza, M.C., 2014. Limits of predictability in commuting flows in the absence of data for calibration. Sci. Rep. 4, 5662. https://doi.org/10.1038/srep05662.