

## Approximate disease dynamics in household-structured populations

P.J Dodd and N.M Ferguson

*J. R. Soc. Interface* 2007 **4**, 1103-1106  
doi: 10.1098/rsif.2007.0231

### References

[This article cites 7 articles](#)

<http://rsif.royalsocietypublishing.org/content/4/17/1103.full.html#ref-list-1>

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *J. R. Soc. Interface* go to: <http://rsif.royalsocietypublishing.org/subscriptions>

# Approximate disease dynamics in household-structured populations

P. J. Dodd\* and N. M. Ferguson

*Department of Infectious Disease Epidemiology, Imperial College London,  
St Mary's Campus, Norfolk Place, London W2 1PG, UK*

We argue that the large-dimensional dynamical systems which frequently occur in biological models can sometimes be effectively reduced to much smaller ones. We illustrate this by applying projection operator techniques to a mean-field model of an infectious disease spreading through a population of households. In this way, we are able to accurately approximate the dynamics of the system in terms of a few key quantities greatly reducing the number of equations required. We investigate linear stability in this framework and find a new way of calculating the familiar threshold criterion for household systems.

**Keywords:** model reduction; projection operators; household epidemics

## 1. INTRODUCTION

In mathematical models of biological dynamics, the state space can easily become large, causing a proliferation in the number of equations needed. This can make models difficult to deal with and insight unlikely. Usually, one is interested only in the behaviour of a few special quantities describing the system. It would be useful if one could find an approximate description of such systems only in terms of these quantities of interest.

Projection operator techniques (e.g. Grabert 1982; Rau and Müller 1996) were developed in the context of statistical mechanics for extracting the dynamics of coarse-grained thermodynamic variables, given the underlying laws on a much larger state space. From a dynamical system governing the vectors  $X$  and  $Y$ ,

$$\frac{dX}{dt} = F(X, Y),$$

$$\frac{dY}{dt} = G(X, Y),$$

one can write down a dynamical law for the vector  $X$  which refers only to  $X$  itself, at the price of acquiring history dependence (e.g. Givon *et al.* 2004),

$$\frac{dX}{dt} = \bar{F}(X) + N(X_0, Y_0, t) + \int_0^t ds M[t, s, X]. \quad (1.1)$$

The rewrite equation (1.1) is no more tractable than the original problem, but in situations where this approach works, the function  $\bar{F}$  is relatively simple and is thought of as dominating the dynamics. The function  $N$ , which depends only on time and the initial conditions, is called the noise term. This is typically rapidly decaying and is often neglected altogether. The integral part of

equation (1.1) is known as the memory term and is often well approximated by a kernel function integrated against some function of  $X$ . Typically, the ‘memory’ decays rapidly, and the first port of call is a Markovian approximation, with the kernel approximated as proportional to a  $\delta$  function.

We apply this approach to a susceptible–infected–recovered (SIR) model which describes the spread of an infectious disease through a population of smaller subgroups. Any individual can infect any other, but those in the same group have an enhanced hazard of doing so. Such multilevel mixing models have been much studied recently with household structure in mind (e.g. Ball *et al.* 1997; Becker *et al.* 2003; Ball & Becker 2006). Results have centred around final epidemic sizes, threshold behaviour and static vaccination policies. We show that the dynamics can be rewritten in the form

$$\frac{dS}{dt} = -\beta SI - N - \int_0^t dx k[t, x; I] \beta S_x I_x,$$

$$\frac{dI}{dt} = \beta SI - \nu I + N + \int_0^t dx k[t, x; I] \beta S_x I_x, \quad (1.2)$$

where  $S_x$  and  $I_x$  are the total number of susceptibles and infecteds, respectively, at time  $x$ ;  $N$  is the noise function as above; and  $k$  is the kernel which we will go on to approximate. The advantage is that while the number of equations needed to describe this system increases like  $n^2$  (where  $n$  is the subgroup size), this formulation only ever needs two equations. Indeed, if the population consists of subgroups of different sizes, then the number of equations increases like the maximum group size cubed. These two equations are those of the usual freely mixing SIR model, but with corrections which account for the underlying inhomogeneity of mixing.

We go on to analyse linear stability about the disease-free equilibrium and obtain a stability discriminant which can be identified with the household reproduction

\*Author for correspondence (peter.dodd@imperial.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2007.0231> or via <http://www.journals.royalsoc.ac.uk>.

number,  $R_*$ . This  $R_*$  is loosely defined as the expected number of houses infected by one singly infected household and determines whether a disease can spread or not, according to whether it is greater or less than unity.

## 2. BASIC MODEL

Let  $y_{ij}$  be the number of households which contain  $i$  susceptible individuals,  $j$  infecteds and, by implication,  $n-i-j$  recovered individuals. Our dynamical law is specified by

$$\begin{aligned}\dot{y}_{i,j} = & -\beta I y_{i,j} + \beta I(i+1)y_{i+1,j-1} - \delta ij y_{i,j} \\ & + \delta(i+1)(j-1)y_{i+1,j-1} - \nu j y_{i,j} + \nu(j+1)y_{i,j+1},\end{aligned}\quad (2.1)$$

where  $I = \sum_{ij} j y_{ij}$  is the total number of infecteds in the population;  $\beta$  is the global infection strength;  $\nu$  is the recovery rate; and  $\delta$  is the difference between the local (within-group) infection strength and  $\beta$ . Where  $i$  and  $j$  imply a negative number of susceptible, infected or recovered, the corresponding  $y$  is taken to be zero. This can be thought of as the mean-field version of the master equation, which is exact in the infinite population limit, or as a deterministic starting point in its own right (though a deterministic model could be valid only for large total populations).

Taking a linear combination of these equations, we obtain the rate laws for the total number of infecteds and the total number of susceptibles,  $S = \sum_{ij} i y_{ij}$ ,

$$\begin{aligned}\dot{S} &= -\beta SI - \delta \mathbf{v}^T \mathbf{y}, \\ \dot{I} &= \beta SI - \nu I + \delta \mathbf{v}^T \mathbf{y},\end{aligned}\quad (2.2)$$

where  $\mathbf{v}$  is the vector whose elements are given by  $v_{ij} = ij$ . When the difference between local and global infection strengths,  $\delta$ , vanishes, the group structure dissolves and we are left with the usual mass action SIR equations. We will choose to eliminate  $y_{n,0}$  and  $y_{0,n}$  in favour of  $S$  and  $I$ , using their definitions. Henceforth, the vector  $\mathbf{y}$  is understood with these two elements deleted. If  $\mathbf{C}$  is the vector with all elements zero except  $C_{n-1,1} = 1$ , the dynamics of  $\mathbf{y}$  take the form

$$\dot{\mathbf{y}} = -\mathbf{A}(t)\mathbf{y} + \beta SI \cdot \mathbf{C},$$

which can be formally solved with a matrix integrating factor  $U(t, s)$ , such that  $U(t, t) = 1$  and  $dU/dt = -\mathbf{A}U$ , to give

$$\mathbf{v}^T \mathbf{y}(t) = \mathbf{v}^T U(t, 0) \mathbf{y}(0) + \int_0^t d\tau \mathbf{v}^T U(t, \tau) \mathbf{C} \beta S(\tau) I(\tau).\quad (2.3)$$

This brings equations (2.2) into the advertised form of equation (1.2), with the kernel given by  $k = \mathbf{v}^T U \mathbf{C}$ . We will restrict to initial conditions only with singly infected households, for which  $N(t) \propto k(t, 0)$ . The equations (2.2) are exactly equivalent to the full set of equations in equation (2.1), but are still difficult to compute because the time dependence of  $\mathbf{A}$  leads  $U$  to depend on repeated time-integrals of  $I$ . We introduce an approximation (A1), equivalent to assuming that the part of  $\mathbf{A}$  proportional to  $\nu$  commutes with the rest. This simplifies things significantly because  $\mathbf{v}^T$  turns out to be

a right eigenvector of both the  $\nu$ -part of  $\mathbf{A}$  and the part proportional to  $I$ . The upshot is that one may approximate the kernel of equation (1.2) by

$$k(t', t) = \exp[-\nu(t' - t) - 2\beta(J(t') - J(t))] f_n(t' - t),$$

where  $J = \int_0^t d\tau I(\tau)$  is the cumulative number of infections (equal to  $R/\nu$  in the case without vaccination). More details can be found in the electronic supplementary material. The functions,  $f_n$ , are determined in a standard way by solving a linear set of constant coefficient differential equations. These differential equations are equivalent to the master equation of the SI model, and a general solution is given in Bailey (1975). For  $n=1-5$ , we have

$$\begin{aligned}f_1(t) &= 0, \\ f_2(t) &= e^{-\delta \cdot t}, \\ f_3(t) &= 2e^{-2\delta \cdot t} \cdot [1 + 2\delta \cdot t], \\ f_4(t) &= 3e^{-3\delta \cdot t} \cdot [1 - 8(1 - e^{-\delta \cdot t}) + 12\delta \cdot t], \\ f_5(t) &= 4e^{-4\delta \cdot t} \cdot [1 - 24(1 - e^{-2\delta \cdot t}) + 18\delta \cdot t e^{-2\delta \cdot t} \\ &\quad + 36\delta \cdot t].\end{aligned}$$

Representative performance of approximation (A1) is depicted in figure 1. More details on the error behaviour can be found in the supplementary information. In Bailey (1975), an asymptotic solution for large  $n$  is also given, which could be adapted to provide a convenient large- $n$  asymptotic approximation to  $f_n$ .

## 3. LINEAR STABILITY AND THRESHOLDS

To study the linear stability of the disease-free equilibrium, one considers the linearized version of the differential equation for  $I$  given by equations (2.2) and (2.3). Not all elements of  $\mathbf{y}$  are dynamical at linear order in  $I$ . Let  $L$  denote  $\mathbf{A}$  with  $I=0$ , and the rows and columns referring to the non-dynamical elements of  $\mathbf{y}$  excised. The matrix  $L$  is invertible and a linear function of  $\nu$  and  $\delta$ . To determine stability, one Laplace transforms the linearized dynamics and asks whether the solution's transform has any poles with positive real part. In the electronic supplementary material, we show that such poles are absent (implying stability), if

$$\tilde{R} = R_g(1 + \delta \mathbf{v}^T L^{-1} \mathbf{C}) < 1$$

(where  $R_g = \beta S_0/\nu$  is the reproductive ratio for only global infections and  $S_0$  is the initial number of susceptibles) and that if this inequality is reversed, the system is unstable. The quantity  $\tilde{R}$  can be identified with the household reproduction ratio  $R_*$  and provides a single compact formula which is easy to implement. Given this, one can compare the threshold behaviour of the full system and the system under approximation (A1) for different  $\nu$  and  $\delta$ , and see that their thresholds are very similar. The case  $n=2$  is degenerate, with the full and approximate systems sharing the same threshold. For concreteness, the  $n=2$  and  $n=3$  cases yield (for more details, see the electronic

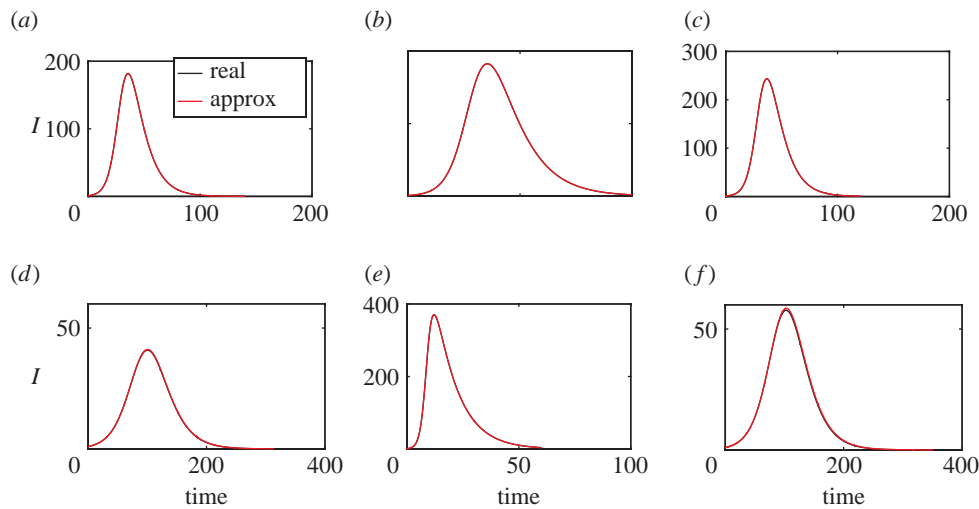


Figure 1. Comparison of approximate (red) and real (black) epidemic curves for populations of  $m$  houses of size  $n$ . Cases have  $(n, \delta/\beta, R_g = \beta S_0/\nu, m)$  given by: (a) (2,2,3,300); (b) (3,2,3,200); (c) (4,2,3,200); (d) (2,8,1.5,300); (e) (3,8,1.5,200) and (f) (4,8,1.5,200).  $m$  is chosen so as to accentuate errors.

supplementary material)

$$\tilde{R}_2 = R_g \left[ 1 + \frac{\delta}{(\delta + \nu)} \right],$$

$$\tilde{R}_3 = R_g \left[ \frac{(\nu^3 + 6\delta^3 + 11\nu\delta^2 + 6\nu^2\delta)}{(\delta + \nu)^2(2\delta + \nu)} \right].$$

#### 4. DISTRIBUTIONS OF HOUSEHOLD SIZES

More realistically, one can consider a population of households of different sizes. Let  $y_{i,j;n}$  be the number of houses of size  $n$  with  $i$  susceptibles and  $j$  infecteds. If there is no death, there is no mechanism to change the total number of individuals in a household, and the governing equations form parallel sets with the same structure as equation (2.1), but with global infections at each  $n$  still forced by the total number of infecteds,  $I$ . In addition to approximation (A1), we will make the approximation (A2) that  $S_n$ , the number of susceptibles residing in houses of size  $n$ , is a constant fraction of the total number of susceptibles, i.e.

$$S_n(t) \approx \frac{S_n(0)}{S(0)} S(t) \equiv p_n S(t). \quad (3.1)$$

While not particularly accurate in its own right, this approximation does allow us to obtain a good approximation to the dynamics, once again in the form equation (1.2), where now the noise and kernel are given by

$$N(t) = \delta e^{-\nu t - 2\beta J} F_0(t),$$

$$k(t, t') = \delta e^{-\nu(t-t') - 2\beta J(t) + 2\beta J(t')} F(t-t'),$$

$$F_0(t) = \sum_n i_n f_n(t),$$

$$F(t) = \sum_n p_n f_n(t),$$

with  $i_n$  being the number of houses of size  $n$  singly infected at  $t=0$ . We note that one could easily extend this procedure to the case where different household sizes have a different within-house infection strength,  $\delta_n$ . Such variability has been shown to give a better fit to the real data for influenza (Cauchemez 2004).

#### 5. CONTROL MEASURES

Many globally applied control measures can be represented as a change in transmission parameters  $\beta$  and  $\delta$ , brought about, for example, by social distancing measures or mass prophylaxis. Most simply,  $\beta$  and/or  $\delta$  become functions of time. More realistically, they could be modelled as being functions of  $I$  or  $J$ , if a response is initiated by a threshold level of infections. None of these possibilities spoils the structure which has allowed us develop our approximate dynamics. Instead of  $\beta J$ , one would have  $\int dt \beta I$ ; and the functions  $f_n$  would be computed from differential equations with time-varying coefficients.

One might also be interested in interventions which are household-based responses to infection: interventions, such as vaccination, applied to a household on finding an infection there. Let us assume that an intervention transfers the remaining susceptible individuals of a house directly to the removed class and is applied after detection of the index case (or not at all). In terms of the  $y$  variables, one can represent this scenario by the addition of  $-\kappa y_{n-1,1}$  to the  $y_{n-1,1}$  rate equation, and  $\kappa y_{n-1,1}$  to the  $y_{0,n}$  rate equation. The new parameter  $\kappa$  contains information about the coverage of the measure, and the rate with which it is applied. The cumulative number of cases  $J$  and  $R/\nu$  now need to be distinguished. One can proceed very much as before, but with two kernels—one proportional to  $\delta$  and another proportional to  $\kappa$ —which appear asymmetrically in the  $\dot{S}$  and  $\dot{I}$  equations (both in the  $S$  equation; but only the  $\delta$  kernel in the  $I$  equation; more details are provided in the supplementary information). One can again approximate the dynamics, so it depends only on  $S$ ,  $I$  and  $J$ , though the accuracy is worse than for the simple case.

#### 6. DISCUSSION

In this paper, we have often used the word ‘household’. This should not be taken too literally—our approach is particularly convenient for large subgroup sizes  $n$ , such as might be appropriate to the description of animal populations. Equations of the form (1.2) can then be

efficiently solved by the analogues of Runge–Kutta techniques (Lubich 1982), and one avoids dealing with the  $n^2$  or  $n^3$  equations of the conventional approach.

Projection operator techniques should be applicable to a wide range of biological problems beyond the example of this paper. In their full generality, they can produce equations of motion of the form (1.1) for mean variables of stochastic systems. Their use requires an insightful choice of relevant variables, and some effort in setting up the formalism for the problem of interest; but they can provide a completely fresh starting point for approximations and a substantial gain in simplicity.

This work was supported by the UK BBSRC. The authors would like to thank Christophe Fraser and Lorenzo Pellis for their useful conversations.

## REFERENCES

- Bailey, N. T. J. 1975 *The mathematical theory of infectious diseases and its applications*. London, UK: Charles Griffin and Co.
- Ball, F. G. & Becker, N. G. 2006 Control of transmission with two types of infection. *Math. Biosci.* **200**, 170–187. (doi:10.1016/j.mbs.2005.12.024)
- Ball, F. G., Mollison, D. & Scalia-Tomba, G. 1997 Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7**, 46–89. (doi:10.1214/aoap/1034625252)
- Becker, N. G., Britton, T. & O'Neill, P. D. 2003 Estimating vaccine effects on transmission of infection from household outbreak data. *Biometrics* **59**, 467–475. (doi:10.1111/1541-0420.00056)
- Cauchemez, S. *et al.* 2004 A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23**, 3469–3487. (doi:10.1002/sim.1912)
- Givon, D., Kupferman, R. & Stuart, A. M. 2004 Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity* **17**, R55–R127. (doi:10.1088/0951-7715/17/6/R01)
- Grabert, H. 1982 *Projection operator techniques in non-equilibrium statistical mechanics*. Berlin, Germany: Springer.
- Lubich, Ch. 1982 Runge–Kutta theory of volterra integro-differential equations. *Numerische Mathematik* **40**, 119–135. (doi:10.1007/BF01459081)
- Rau, J. & Müller, B. 1996 From reversible quantum microdynamics to irreversible quantum transport. *Phys. Rept.* **272**, 1–59. (doi:10.1016/0370-1573(95)00077-1)