From Individuals to Populations: A Symbolic Process Algebra Approach to Epidemiology

Chris McCaig, Rachel Norman and Carron Shankland

Abstract. Is it possible to *symbolically* express and analyse an individual-based model of disease spread, including realistic population dynamics? This problem is addressed through the use of process algebra and a novel method for transforming process algebra into Mean Field Equations. A number of stochastic models of population growth are presented, exploring different representations based on alternative views of individual behaviour. The overall population dynamics in terms of mean field equations are derived using a formal and rigorous rewriting based method. These equations are easily compared with the traditionally used deterministic Ordinary Differential Equation models and allow evaluation of those ODE models, challenging their assumptions about system dynamics. The utility of our approach for epidemiology is confirmed by constructing a model combining population growth with disease spread and fitting it to data on HIV in the UK population.

1. Introduction

Epidemiology has benefited for many years from a symbolic approach: Ordinary Differential Equations (ODEs) have been used to capture the spread of disease since the Kermack and McKendrick models of the 1920s [15]. More sophisticated models have been built since, capturing features of population dynamics such as growth, seasonality, metapopulations, and networks of contact. A disadvantage of this approach is that the equations express population features which are difficult

to measure in the field. One such feature is the transmission rate of a disease (from an infected individual to a susceptible individual in the population) which incorporates aspects of likelihood of appropriate contact, and likelihood of contracting the disease following contact. Mathematical biologists have for many years written down the ODE description believing that behaviour at the population level translates simply and intuitively from assumptions about individual interactions. Turner et al [29] showed that this is not necessarily the case. There is a need for a rigorous relation between the actions of individuals and the outcome at a population level.

An alternative approach is to base models on individual behaviour, for example, probabilistic Cellular Automata models [1]. The most common way to interrogate the model is simulation, but full exploration of the model requires instantiation over a range of parameter values. Ensuring that all important areas of parameter space have been covered incurs heavy computational expense, and may even be impossible. Limited algebraic analysis is available through methods such as pair approximation [14].

A third approach combines the advantages of symbolic modelling with those of individual-based modelling: process algebra. Process algebra has increasingly been used to model a wide range of biological systems [6, 20, 24, 26, 27]. The benefits of using process algebra to study such systems are twofold. First, process algebra allows symbolic, precise and unambiguous expression of a model. Second, process algebra has a formal mathematical semantics, allowing rigorous investigation of the model via a range of techniques.

Our work uses the discrete time process algebra *Weighted Synchronous Calculus of Communicating Systems* (WSCCS)[28]. The semantics of WSCCS can be viewed as a Discrete Time Markov Chain (DTMC). Simulation can be used to explore the model. Steady state analysis can be carried out, and properties of the Markov Chain computed, e.g. probability of being in a particular state, or average number of occurrences of an action before a specific event occurs. As with cellular automata, such investigation can be computationally expensive for multiple parameter values. Our previous work [17, 18] has been to facilitate *symbolic* analyses of the model by developing a rewriting-based method to derive Mean Field Equations (MFEs) from a description of a system in WSCCS. The MFEs describe the average behaviour of the system at the population level and are analogous to traditional ODE models. Although our focus is on biological systems, the technique can be applied to any system characterized by large numbers of similar components cooperating dynamically. The MFEs provide an approximation of the system dynamics (the DTMC corresponding to the WSCCS description). The MFEs are amenable to analysis using established algebraic techniques developed by mathematical biologists for ODEs. The key advantage of our approach is that biological observations of individuals can be exploited in making the (individual based) WSCCS model, and the MFEs are derived automatically and efficiently.

In this paper we consider the problem of using process algebra to accurately represent population growth and thereby construct useful models of disease spread.

Population dynamics have been of interest to modellers for centuries. In 1798 Malthus [16] proposed a simple exponential growth model based on compound interest but noted that this was unrealistic, since when a population becomes very large, access to resources will become restricted, limiting further growth in the population. Verhulst proposed the logistic growth model [30] to overcome this deficiency and this is still widely used to describe density dependent growth. Clearly the question of how exactly to model growth remains contentious since other candidates have been proposed [3, 10, 11, 25]. The question of population growth is of importance because given a model of disease spread, the addition of a fluctuating population can alter the dynamics of the epidemic. Therefore, developing a suitable model of population growth is an important step in producing realistic models of disease spread which can be analysed to provide predictive information about potential impact of epidemics, and to evaluate control strategies.

Related Work. Others have investigated individual based models of population dynamics and related their behaviour to population level equations. Sumpter [26] developed a simple WSCCS model of population growth and derived MFEs for the model. Brännström and Sumpter [4] presented individual based (not process algebra) models of competition which could be used to derive several existing population models but notably not Verhulst's logistic equation. The work presented here improves on previous work by applying a rigorous method across a range of different models of population growth.

Outline of the Paper. Sect. 2 gives a brief description of the syntax and semantics of the language used in our models (WSCCS), and an outline of the method for deriving MFEs. In Sect. 3 WSCCS models of population dynamics are presented, which include density dependent growth in a variety of formulations (in either births or deaths, and introduced implicitly by enriching the WSCCS language or explicitly by including agents representing resources for which the population competes). The resultant changes in overall population dynamics are explored, comparing the derived MFEs to traditional population level equations for population dynamics. In Sect. 4 we add disease spread to our model of population dynamics and fit the resulting model to data from the literature on HIV prevalence in the UK. Our results are summarised and conclusions drawn in Sect. 5.

2. Background

2.1. WSCCS Syntax and Semantics

In WSCCS the basic components are *actions* and the *processes* (or *agents*) that carry out those actions. The actions are chosen by the modeller to represent activities in the system. For example, *infect*, *send*, *receive*, *throw dice*, and so on. Essentially, processes order actions in time, providing sequences of actions, choices between actions, and actions in parallel. When actions occur in parallel we denote the multiplication by $a\#b$. Actions form an Abelian group, such that $a\#\bar{a} = \sqrt{}$ where $\bar{a}$ is the inverse of the action a and $\sqrt{}$ is the identity action. Actions occur

instantaneously and have no duration. WSCCS is a probabilistic process algebra, meaning that the decision to move from one state to another can be a probabilistic one. The formal syntax and semantics of WSCCS is presented in Tofts [28]. The main details are repeated here for the convenience of the reader.

The possible WSCCS expressions are given by the following BNF grammar:

$$A ::= X \mid a\!:\!A \mid \Sigma\{w_i.A_i|i \in I\} \mid A \times B \mid A\lceil L \mid \Theta(A) \mid A[S] \mid X \stackrel{\text{def}}{=} A \ .$$

Here $X \in Var$, a set of process variables; $a \in Act$, an action group; $w_i \in \mathscr{W}$, a set of weights; S a set of renaming functions, $S : Act \rightarrow Act$ such that $S(\sqrt{}) = \sqrt{}$ and $\overline{S(a)} = S(\overline{a})$; action subsets $L \subseteq Act$ with $\sqrt{} \in L$; and arbitrary indexing sets I. The informal interpretation of the operators is as follows:

- 0 a process which cannot proceed, representing deadlock ;
- X the process bound to the variable X ;
- $a\!:\!A$ a process which can perform the action a becoming the process A ;
- $\Sigma\{w_i.A_i|i \in I\}$ the weighted choice between processes A_i, the weight of A_i being w_i. Considering a large number of repeated experiments of this process, we expect to see A_i chosen with relative frequency $w_i/\Sigma_{i \in I}w_i$. This is therefore also referred to as probabilistic choice. The binary plus operator can be used in place of the indexed sum, i.e. writing $\Sigma\{1_1.a\!:\!0, 2_2.b\!:\!0|i \in \{1,2\}\}$ as $1.a\!:\!0 + 2.b\!:\!0$. Weights are generally positive natural numbers or reals, but may also incorporate the special weight ω which is greater than all natural numbers. In combination with the operator $\Theta(A)$ this yields a form of *priority*: in a choice between an ω weight and a natural number weight, the process with the ω weight *must* be taken if possible. Prioritised weights are written $m\omega^n$ where $m, n \geq 0$;
- $A \times B$ the synchronous parallel composition of A and B. At each stage each process must perform an action with the composed process performing the composition of the individual actions, e.g. $a\!:\!A \times b\!:\!B$ yields $a\#b\!:\!(A \times B)$. This is a powerful operator: models are constructed by describing simple individuals and composing a number of those in parallel. McCaig [17] introduces an extended notation $A\{n\}$ which is syntactic sugar for n instances of process A in parallel, where $n \in \mathbb{N}$;
- $A\lceil L$ a process which can only perform actions in the group L. This operator is used to enforce communication on actions $b \notin L$. Two processes in parallel may communicate when one carries out an action and the other carries out the matching co-action, e.g. *infect* and $\overline{infect}$. Communication can be used to model passing of information from one process to another, or to coordinate activity. Such communication is strictly two-way; that is, only two processes may interact on this action ;
- $\Theta(A)$ discards all the non-ω weighted parts of the process A, i.e. ensures prioritised actions are always executed in preference to non-prioritised actions ;
- $A[S]$ represents A relabelled by the function S (we do not use relabelling in this paper, but it is included for completeness) ;

$$\overline{a:A \xrightarrow{a} A} \qquad\qquad \overline{\sum\{w_i.A_i | i \in I\} \xmapsto{w_i} A_i}$$

$$\frac{A \xrightarrow{a} A' \quad B \xrightarrow{b} B'}{A \times B \xrightarrow{a\#b} A' \times B'} \qquad\qquad \frac{A \xmapsto{w} A' \quad B \xmapsto{v} B'}{A \times B \xmapsto{wv} A' \times B'}$$

$$\frac{A \xrightarrow{a} A' \quad B \xmapsto{w} B'}{A \times B \xmapsto{w} A \times B'} \qquad\qquad \frac{A \xmapsto{w} A' \quad B \xrightarrow{a} B'}{A \times B \xmapsto{w} A' \times B}$$

$$\frac{A \xrightarrow{a} A' \quad a \in L}{does_L(A)} \qquad\qquad \frac{A \xmapsto{w} A' \quad does_L(A')}{does_L(A)}$$

$$\frac{A \xrightarrow{a} A' \quad a \in L}{A\lceil L \xrightarrow{a} A'\lceil L} \qquad\qquad \frac{A \xmapsto{w} A' \quad does_L(A')}{A\lceil L \xmapsto{w} A'\lceil L}$$

$$\frac{A \xrightarrow{a} A'}{A[S] \xrightarrow{S(a)} A'[S]} \qquad\qquad \frac{A \xmapsto{w} A'}{A[S] \xmapsto{w} A'[S]}$$

$$\frac{A \xrightarrow{a} A' \quad X \stackrel{\mathrm{def}}{=} A}{X \xrightarrow{a} A'} \qquad\qquad \frac{A \xmapsto{w} A' \quad X \stackrel{\mathrm{def}}{=} A}{X \xmapsto{w} A'}$$

$$\frac{A \xrightarrow{a} A'}{\Theta(A) \xrightarrow{a} \Theta(A')} \qquad\qquad \frac{A \xmapsto{n\omega^i} A' \not\exists(j>i).A \xmapsto{m\omega^j} A''}{\Theta(A) \xmapsto{n} \Theta(A')}$$

Table 1. Operational rules for WSCCS

- $X \stackrel{\mathrm{def}}{=} A$ represents binding the process variable X to the expression A .

The semantics of WSCCS is transition based, defining the actions that a process can perform and the weight with which a state can be reached. The operational rules of WSCCS, presented in Table 1, formalise the descriptions above. In particular note the two different arrows which feature in the table: $\xrightarrow{a}$ represents a transition associated with the action a ; and $\xmapsto{w}$ represents a transition associated with a weight w . The auxiliary predicate $does_L(A)$, which denotes the ability of A to perform L after zero or more actions, is well defined since only finitely branching choice expressions are allowed.

2.2. Deriving Mean Field Equations from WSCCS Models

In McCaig's thesis [17] and the related report [18] a method (referred to hereafter as the Stirling method) is described to automatically derive Mean Field Equations from WSCCS models. We give an overview of the approach here to aid understanding of the following sections. Worked through derivations are given at the end of this section and in Sect. 3.2.1. The method provides equations for all other models in the paper.

Consider the simple model of population growth in Fig. 1. The N agents die with probability p_d, becoming the null agent 0, give birth with probability p_b, becoming the agent consisting of two N agents in parallel, or do neither with probability $(1 - p_d - p_b)$, remaining as a single agent N. The model can be simulated,

$$N \quad \stackrel{\text{def}}{=} \quad p_d.\sqrt{} : 0 + p_b.\sqrt{} : (N \times N) + (1 - p_d - p_b).\sqrt{} : N$$
$$Population \quad \stackrel{\text{def}}{=} \quad N\{n\}\lceil\{\sqrt{}\}$$

FIGURE 1. Naive population model

producing a single trace through the dynamics of the system. A second simulation may of course produce quite different behaviour since this is a stochastic process; therefore, of more interest is the *average* behaviour of the system as time progresses. This can be obtained by averaging the time series results of repeated simulations of the system. Clearly this becomes time-consuming and computationally expensive as the number of processes n and number of repetitions increases. An alternative is to generate the whole transition system for the model and to average the states produced, but as n increases the state space grows exponentially thereby also incurring considerable computational expense.

The Stirling method avoids generating the state space of the whole system, instead applying transformations to the WSCCS expression of the model, yielding an approximation (mean) of the transition system in the form of first-order mean field equations. The approximation is shown to be "good" (i.e. lies within the standard deviation when compared with repeated simulations) in McCaig's thesis. Further, when the system becomes infinitely large, the mean of the DTMC corresponding to the transition system is proved to be equivalent to the derived MFEs. Larger populations eliminate the stochastic effects associated with low copy numbers.

The advantages of our approach are: the computational expense of generating the state space and/or simulation is avoided (the Stirling method is $O(a^2c)$ where a is the number of agent types and c is the number of distinct actions in the WSCCS description); it is a symbolic approach (avoiding questions regarding the exploration of the parameter space); and the MFEs, being a different view of the system and amenable to further analysis, offer new insight to the system.

The Stirling method applies to models of the form $A1\{n_1\} \times ... \times Am\{n_m\}$ where the Ai communicate with each other (usually on a subset of actions). Models are limited in that steps involving probabilistic choice between actions must be separate from steps involving communication (which must have branches weighted 1). These restrictions on model form have not proved to be a barrier to expressing epidemiological models.

Independently, the PEPA group [5, 13] and Cardelli [7] have proposed methods for deriving ODEs from process algebra. Their work differs in that their process algebras are continuous, based on rates rather than probabilities. Two of the methods are based on a mass action assumption, and not tied to the standard process algebra semantics. In contrast, our goal has been to preserve this association, so that understanding individuals and their interactions translates automatically to population behaviour via process algebra semantics.

```
for each agent Ai {
  for each (w_j.a_j : Aj) ∈ transitions(Ai) {
    for each Ak ∈ components(Aj) {
      TT[(Ai, a_j),Ak] = TT[(Ai, a_j),Ak] + calculateTerm(Ai, w_j, a_j)
} } }
```

FIGURE 2. Constructing the transition table from a WSCCS model

Transition Table: Relating Actions to Overall System Evolution. The transition system may be viewed as the evolution of the state vector $A1\{n_1\} \times ... \times Am\{n_m\}$. For a particular Ai an action has three possible effects:

exit activity. Following the action, the process Ai evolves to some other agent Aj therefore the number of Ai agents is decreased.

entry activity. Following the action, the process Ai evolves to some other agent Aj therefore the number of Aj agents is increased.

none. The process becomes Ai and there is no change in number of Ai agents.

WSCCS is a synchronous calculus, therefore in each time step, for every agent in the system, one of the above activities will occur. Our method is based around construction and interpretation of a transition table TT noting these exit and entry activities (Fig. 2).

The rows of TT denote the agents Ai at time t and their enabled actions a_j. The columns of the transition table denote the agents Ak at the next time $t + 1$. The term in cell $[(Ai,\ a_j), Ak]$ is an expression describing the number of Ai_t agents performing a_j to become Ak_{t+1}. The derivation of this symbolically expressed term is fully determined (see below) by the context of the action carried out (e.g. part of a probabilistic choice, or part of a communication) and the composition of the population. Where Ai evolves to the same agent Ak irrespective of the action performed a single row labelled $[(Ai,\ *), Ak]$ is used for that agent. An example is the $Res1$ agent of Fig. 4. The mean field equation for Ak_{t+1} is obtained by summing the terms in the column Ak.

Some auxiliary definitions are required. Processes can be classified by syntactic features as: *communicating* (having an action enabled that is involved in a communication), *probabilistic* (having only actions enabled that are not involved in communication), and *priority* (communicating and using ω weights). Given a *serial* process $A = w_1.a_1 : A1 + w_2.a_2 : A2 + ... + w_n.a_n : An$ define $transitions(A) = \{w_1.a_1 : A1, w_2.a_2 : A2, ..., w_n.a_n : An\}$. Given a *parallel* process $A = A1 \times A2 \times ... \times An$ define $components(A) = \{A1, A2, ...An\}$. For a process communicating on action a, we define two groups of agents involved in the communication: *collaborators* are those processes with the matching action $\bar{a}$, and *competitiors* are those processes with the same action a.

The pseudo code to compute the terms in TT is given in Fig. 3. For probabilistic choice, the semantics of WSCCS (Table 1) specifies that over a number of experiments the different branches will be taken in numbers consistent with their

```
function calculateTerm (A, w, a): String {
   case A in {
      probabilistic(A): return w * At;
      communicating(A) and priority(A):
         term = (At * collaborators(A))/(At + competitors(A));
         if a equals √ return (A - term) else return term;
      communicating(A) and not priority(A):
         term = (At * collaborators(A))/(At + collaborators(A) + competitors(A));
         if a equals √ return (A - term) else return term;
}}
```

FIGURE 3. Calculating the number of agents at time $t + 1$

weights. For convenience, the weights in such choices sum to 1 in the models in this paper hence the term is simply wA_t. For communication, we enumerate the possible outcomes based on a classification of modes of communication (prioritised or not, single action a or multiple actions e.g. $a\#a\#a$). This results in complex formulae based on the weighted multinomial choice of those outcomes giving the average number of communications. For single actions, as used in this paper, these formulae can be simplified. These are the formulae used in the *calculateTerm* function of Fig. 3. The full version of the method [17, 18] assumes weights do not have to sum to 1, and also gives the formulae for multiple action communications.

2.2.1. Derivation of MFE for a Simple Population Growth Model. Consider again the simplistic model of population growth given in Fig. 1. The actions in Fig. 1 are simply $\sqrt{}$. That is, activities are of no interest, only the evolution of numbers of agents is significant. As in all of our models, the system as a whole is described by the system equation *Population*, comprising multiple copies of each kind of agent in parallel.

   The transition table for this system is as follows:

	0	N_{t+1}
$(N_t, \sqrt{})$	$p_d N_t$	$(1 - p_b - p_d)N_t + 2p_b N_t$

Each column leads to a MFE for that agent, but 0 is ignored here since this is not of interest to us. The Stirling method generates the following MFE:

$$N_{t+1} = (1 + p_b - p_d)N_t \, , \tag{1}$$

where N_{t+1} represents the number of N agents at time $t + 1$ expressed in terms of N_t, the number of N agents at time t. Since this model has no communication between agents, and a single step with probabilistic choice, the derived MFE can be easily verified manually. A further example of derivation of the MFE is given in Section 3.2.1, but otherwise the method is used to generate MFE for models of population growth without further explanation.

3. Density Dependent Growth

Equation (1) describes a simple recurrence relation. With $p_b > p_d$ the population will become infinitely large; $p_b < p_d$ will lead to the population dying out, while $p_b = p_d$ will lead to an equilibrium state for any initial population size, $N_0 = n$. The probabilities p_b and p_d are fixed, therefore the average behaviour of this model is similar to that of the simple exponential growth model described by Malthus [16]. Biologically, it is more realistic to consider a model in which the probability of birth and death vary depending on the size of the population at each instant in time (density dependence). For example, as the population grows, resources such as food and shelter become scarce, therefore individuals become weaker and are more likely to die. Alternatively this weakness may manifest itself as a reduced fecundity and a reduction in the birth rates. This section presents several ways of modelling these notions in WSCCS, obtaining more realistic models of population growth.

3.1. Functional Probabilities

What is required is the ability to modify p_b and/or p_d on the fly as the population changes. The first method described here is to add assumptions about how probability of birth and death depend on population size using *functional probabilities* [17]. Functional probabilities add considerable convenience and elegance of expression to complex models, while adding no new semantic concepts to WSCCS. Functional probabilities are implemented by encoding population size as part of the agent name, a technique [19] commonly used in process algebra. The size of the resultant model is much increased, and the translation itself is unremarkable: the interested reader is referred to [17] for the full details.

Instead of fixed probabilities, a functional definition is given. For example, probability p_x is a function f of the number of A agents (denoted $\lfloor A \rfloor$):

$$p_x \stackrel{\text{def}}{=} min(max(0, f(\lfloor A \rfloor)), p_L) \ .$$

The function may take any format required, since it appears directly in the MFEs and is often not computed numerically. The probability p_L is the upper limit for p_x, chosen to ensure that all probabilities in the model are always in the range $0 \le p \le 1$. The *min* and *max* expressions may be required to ensure that p_x is in the allowed range, but these terms make mathematical analysis of the MFEs more complex. If there is a low likelihood of reaching a state where *min* and *max* are not satisfied by f then it is reasonable to assert $p_x = f(\lfloor A \rfloor)$ in further analyses. The justification is that those states make little contribution to the *average* behaviour captured by the MFEs therefore can be ignored.

3.1.1. Density Dependent Birth.
Density dependent birth can be added to the model of Fig. 1 by making the probability of birth p_b inversely proportional to $\lfloor N \rfloor$.

$$p_b \stackrel{\text{def}}{=} min(max(0, p_{b_0} - k * \lfloor N \rfloor), p_L) \ ,$$

where p_{b_0} is the probability of birth in the absence of crowding and k is a measure of the strength of the effect of crowding, $0 < k \ll 1$.

Using the method of Sect. 2.2, the MFE derived is

$$
\begin{aligned}
N_{t+1} &= N_t + (p_{b_0} - kN_t - p_d)N_t \\
&= N_t + (p_{b_0} - p_d)N_t\left(1 - \frac{kN_t}{p_{b_0} - p_d}\right) \ .
\end{aligned}
\tag{2}
$$

This is our first realistic model of population growth, derived from an expression of individual behaviour. Compare this to the discrete time version of Verhulst's logistic equation

$$
N_{t+1} = N_t + rN_t\left(1 - \frac{N_t}{K}\right) \ .
\tag{3}
$$

where r represents reproductive rate and K the carrying capacity of the population. Simple substitution of $r = (p_{b_0} - p_d)$ and $K = (p_{b_0} - p_d)/k$ in (3) yields (2). The logistic equation is the most commonly used equation for describing population dynamics and is frequently included as a self limiting growth term in models of disease spread. This gives confidence in our approach, and endorses Verhulst's equation.

3.1.2. Density Dependent Death. Density dependent death can similarly be added to Fig. 1 by choosing probability of death p_d directly proportional to $\lfloor N \rfloor$ with

$$
p_d \stackrel{\text{def}}{=} min(max(0, p_{d_0} + k * \lfloor N \rfloor), p_L) \ ,
$$

where p_{d_0} is the probability of death in the absence of crowding. The MFE, derived once again using our method,

$$
\begin{aligned}
N_{t+1} &= N_t + (p_b - (p_{d_0} + kN_t))N_t \\
&= N_t + (p_b - p_{d_0})N_t\left(1 - \frac{kN_t}{p_b - p_{d_0}}\right) \ ,
\end{aligned}
$$

is equivalent to the logistic equation with $r = (p_b - p_{d_0})$ and $K = (p_b - p_{d_0})/k$.

3.1.3. Summary. The results above are pleasing: we have shown that it is possible to derive the logistic equation from an individual based model of population growth. This contradicts the findings of Brännström and Sumpter [4] who did not find the logistic equation for any of their models. Our results should not be surprising: in the functional probabilities we are making the probabilities linearly proportional to the population size, effectively encoding the same assumptions which lead to the logistic equation in the traditional population level models. It would have been more surprising if we had not derived the logistic equation.

$$
\begin{aligned}
N1 &\stackrel{\text{def}}{=} 1.get : (N2 \times N2) + 1.\surd : N2 \\
Res1 &\stackrel{\text{def}}{=} 1.\overline{get} : Res2 + 1.\surd : Res2 \\
N2 &\stackrel{\text{def}}{=} p_d.\surd : 0 + (1 - p_d).\surd : N1 \\
Res2 &\stackrel{\text{def}}{=} 1.\surd : Res1 \\
Population &\stackrel{\text{def}}{=} (N1\{n\} \times Res1\{f\})\lceil\{\surd\}
\end{aligned}
$$

FIGURE 4. Density dependence on births with non-prioritised communication

3.2. Explicit Resource

The advantage of individual based modelling techniques is that population level assumptions can be avoided, to be replaced by population level behaviours arising from the explicit individual interactions. To the models seen so far we add agents representing "resource", e.g. food, shelter, or space. The resource is required by individuals to survive. It is finite and individuals must compete for it. It does not last forever, therefore must be reacquired at regular intervals. Access to this resource can be used to determine the likelihood of either birth or death.

Acquiring a resource is modelled in WSCCS by communication between resource agents and individuals, requiring the use of more complex language features than seen in the models so far. Two forms of communication are available: prioritised and non-prioritised. Using prioritised communication between the resource agents and the population agents forces individuals to obtain the resource if it is available; however, in a population it is likely that some individuals may fail in this. For example, individuals foraging may fail to find food which is present. Using non-prioritised communication models the possibility that individuals do not obtain the resource, even when it is available, and is therefore more biologically plausible. As above, models exploring density dependence on births and density dependence on deaths are considered separately.

3.2.1. Density Dependence on Births.
The model given in Fig. 4 has individuals in the population competing for the available resource (the *get* action), giving birth after obtaining this resource, and dying probabilistically.

The agents $N1$ and $N2$ represent the members of the population at the different stages of the model. The $N1$ agents can obtain the resource and become the parallel agent $N2 \times N2$, representing birth. If they do not obtain the resource the $N1$ agents become a single $N2$ agent. In the second stage of the model the $N2$ agents make a probabilistic choice to die or survive. The total number of resource agents is constant therefore the agents $Res1$ and $Res2$ should be thought of as units of resource produced by the environment in a time step rather than, for example, discrete portions of food consumed by the population.

Deriving the terms of the MFEs for this model is more complex than for the previous examples: although the definition of $N1$ suggests the choice to *get* or not is

equally weighted, in fact this choice is also influenced by availability of $Res1$ agents with which to synchronise. This is reflected in the *calculate Term* function described in Sect. 2.2. For example, here it is possible that no individuals get the resource (with very low probability), or that all do (also with low probability, and assuming $\lfloor N1 \rfloor \leq \lfloor Res1 \rfloor$), or all of the possibilities inbetween. As explained earlier, the *calculate Term* function yields a formula based on the weighted multinomial choice of those possible outcomes. The Stirling method yields the following transition table. Note that the term for the communicating action (*get*) reflects that $N1$ *collaborates* with $Res1$ but has no *competitors* for the action.

	0	$N1_{t+1}$	$N2_{t+1}$	$Res1_{t+1}$	$Res2_{t+1}$
$(N1_t, get)$			$2\frac{N1_t * Res1_t}{N1_t + Res1_t}$		
$(N1_t, \checkmark)$			$N1_t - \frac{N1_t * Res1_t}{N1_t + Res1_t}$		
$(Res1_t, *)$				$Res1_t$	
$(N2_t, \checkmark)$	$p_d N2_t$	$(1 - p_d)N2_t$			
$(Res2_t, *)$				$Res2_t$	

From the table above, summing each column, the Stirling method generates MFE for all agents, i.e. $N1, N2, Res1, Res2$, where $N1$ is expressed in terms of $N2$ and vice versa. Similarly for $Res1$ and $Res2$. Generally we are interested only in a complete cycle of behaviour. That is, starting with agents $N1$, evolving to agents $N2$, then back to $N1$ (two stages here). We take the $N1$ equation, substitute to remove occurrences of $N2$ and obtain an equation only in $N1$ (and $Res1$). Finally, we rename $N1$ as simply N. The fact that the number of resource agents remains constant means that the derived MFE for $Res1$ can be simplified to f in the MFE for N. This leads to the MFE

$$N_{t+1} = (1 - p_d)N_t + \frac{(1 - p_d)fN_t}{f + N_t} \quad . \tag{4}$$

Here the term $(1 - p_d)N_t$ represents those individuals in the population which survive the probabilistic death stage. The term $fN_t/(f + N_t)$ represents the mean number of new births with the factor $(1 - p_d)$ representing the proportion of new births which survive the probabilistic death stage. We find the steady state of this model by setting $N_{t+1} = N_t = N^*$:

$$N^* = (1 - p_d)N^* + \frac{(1 - p_d)fN^*}{f + N^*} \quad .$$

Solving for N^* we get

$$N^* = \frac{(1 - 2p_d)f}{p_d} \quad .$$

For biological realism the steady state should be positive, therefore $p_d < 0.5$. Note that this fact is not obvious from the WSCCS model, but becomes clear in the MFE. The values of these probabilities can be observed in the field, but an important factor is the length of timestep. If we need to reduce p_d to meet

$$
\begin{aligned}
N1 &\stackrel{\text{def}}{=} 1.get : N2 + 1.\sqrt{} : 0 \\
Res1 &\stackrel{\text{def}}{=} 1.\overline{get} : Res2 + 1.\sqrt{} : Res2 \\
N2 &\stackrel{\text{def}}{=} p_b.\sqrt{} : (N1 \times N1) + p_d.\sqrt{} : 0 + (1 - p_b - p_d).\sqrt{} : N1 \\
Res2 &\stackrel{\text{def}}{=} 1.\sqrt{} : Res1 \\
Population &\stackrel{\text{def}}{=} (N1\{n\} \times Res1\{f\})\lceil\{\sqrt{}\}
\end{aligned}
$$

FIGURE 5. Density dependence on deaths with non-prioritised communication

the above requirement we can reduce the timestep represented by our models and adjust all other parameters accordingly.

Sumpter [26] developed a mechanism for describing self limiting growth in a population which made use of food as an agent. Using an heuristic he derived the following MFE

$$
N_{t+1} = (1 - p_d)N_t + \min[(1 - p_d)N_t, f] \ ,
$$

where p_d is the probability of death in any timestep and f is the number of food agents. The underlying assumptions of this model are undesirable biologically: individuals are guaranteed to find food if it is available because prioritised communication is used. Therefore, every member of the population will give birth at each step of time until the size of the population is larger than the number of food agents, after which the number of births will be equal to the number of food agents. This model has a stable steady state of $N^* = f/p_d$, when $p_d \leq 0.5$, which is larger than for our model.

3.2.2. Density Dependence on Deaths. In Fig. 5 the $N1$ agents can once again get the resource, becoming the agent $N2$, but here if they do not get the resource they die, becoming the null agent 0. The $N2$ agents then give birth probabilistically and, to be realistic, can also die probabilistically. That is, in each step of time a proportion of the population die, for instance, due to age and some die due to a lack of food. The MFE for this model is

$$
N_{t+1} = (1 + p_b - p_d)\frac{fN_t}{f + N_t} \ , \tag{5}
$$

where the term $f/(f + N_t)$ represents the proportion of the population who survive the competition for resource, with the factor $(1 + p_b - p_d)$ representing the increase in the population due to births and the decrease due to probabilistic death. Equation (5) can be rearranged to give

$$
N_{t+1} = \frac{aN_t}{1 + bN_t} \ , \tag{6}
$$

where $a = (1 + p_b - p_d)$ and $b = 1/f$. Equation 6 is the Beverton-Holt model [3], originally proposed as a model of salmon populations displaying density dependent

birth. Even though our model is based on density dependent death the interpretations of a and b here are similar to the original Beverton and Holt definitions. Parameter a corresponds to the proliferation rate per generation and parameter b corresponds to a measure of the maximal population size. Our derivation endorses the plausibility of the Beverton-Holt model, which is commonly used in models of plant populations but not so widely used for animal populations.

Setting $N_{t+1} = N_t = N^*$ in (5) and solving for N^* yields the steady state

$$N^* = (p_b - p_d)f \ .$$

In this case to ensure the steady state is positive we require $p_b > p_d$.

3.2.3. Summary. Communication is one of the most important language features of process algebra. Here, it has been used to synchronise the behaviour of processes and thus restrict population growth. This explicit resource acquisition does not result in logistic growth, as in the previous models, but does yield a growth model previously defined in the literature by Beverton and Holt. This gives confidence in this method of regulating population growth, and also in our modelling approach.

Of course, making as much as possible explicit in the model relies on a deep understanding of the behaviour described, and of the nature of synchronisation and parallelism in process algebra. Small changes to the process behaviour may have a large effect on population dynamics. This can be both an advantage and a disadvantage, and forces the modeller to think carefully about the biological interpretation of the model. For example, swapping the order of the step where resource is obtained and the step where individuals are born or die seems a small change, but the biological implication is that newborn individuals are now available to compete for the available resource and the individuals which probabilistically die are not. The population dynamics have altered. Biological interpretation of the model is paramount and this becomes more obvious when considering more complex models such as that in Sect. 4 in which infectious disease spread is added to population dynamics.

4. Population Dynamics and Disease

While population dynamics are interesting in their own right they are also crucial in developing realistic models of disease spread. In this section a model adding disease spread to population growth is presented and the model is then fitted to widely available data on HIV spread in the UK [12].

4.1. Model

The model of Fig. 6 adds infectious disease spread, based on the models of Norman and Shankland [20], to the density dependent death population dynamics of Fig. 5. In a typical disease model the population is divided into 3 groups: susceptibles (S) have never had the disease, infecteds (I) currently have the disease, and recovereds (R) have previously had the disease and are immune to future infection. This is

$$S0 \stackrel{\text{def}}{=} 1.get : S1 + 1.\sqrt{} : 0 \qquad\qquad S1 \stackrel{\text{def}}{=} \omega.infect : SI2 + 1.\sqrt{} : S2$$

$$R0 \stackrel{\text{def}}{=} 1.get : R1 + 1.\sqrt{} : 0 \qquad\qquad R1 \stackrel{\text{def}}{=} \omega.infect : R2 + 1.\sqrt{} : R2$$

$$I0 \stackrel{\text{def}}{=} 1.get : (I1 \times Trans) + 1.\sqrt{} : 0 \qquad I1 \stackrel{\text{def}}{=} \omega.infect : I2 + 1.\sqrt{} : I2$$

$$Res0 \stackrel{\text{def}}{=} 1.\overline{get} : Res1 + 1.\sqrt{} : Res1 \qquad Trans \stackrel{\text{def}}{=} \omega.\overline{infect} : 0 + 1.\sqrt{} : 0$$

$$Res1 \stackrel{\text{def}}{=} 1.\sqrt{} : Res2$$

$$Res2 \stackrel{\text{def}}{=} 1.\sqrt{} : Res0$$

$$S2 \stackrel{\text{def}}{=} p_b.\sqrt{} : (S0 \times S0) + (1 - p_b - p_d).\sqrt{} : S0 + p_d.\sqrt{} : 0$$

$$SI2 \stackrel{\text{def}}{=} p_b.\sqrt{} : (S0 \times S0) + p_a.\sqrt{} : I0 + (1 - p_a - p_b - p_d).\sqrt{} : S0 + p_d.\sqrt{} : 0$$

$$I2 \stackrel{\text{def}}{=} p_b.\sqrt{} : (I0 \times S0) + p_r.\sqrt{} : R0 + (1 - p_r - p_b - p_d).\sqrt{} : I0 + p_d.\sqrt{} : 0$$

$$R2 \stackrel{\text{def}}{=} p_b.\sqrt{} : (R0 \times S0) + (1 - p_b - p_d).\sqrt{} : R0 + p_d.\sqrt{} : 0$$

$$Population \quad \stackrel{\text{def}}{=} \quad \Theta((S0\{s\} \times I0\{i\} \times Res0\{f\})\lceil\{\sqrt{}\})$$

FIGURE 6. SIR model with density dependence on deaths

the classic SIR presentation of disease transmission, introduced by Kermack and McKendrick [15].

The first stage in the model is the foraging stage in which $S0, I0$ and $R0$ all compete for resource. Those that do not obtain the resource will die, as in the model of Fig. 5. The second stage is a contact stage in which infected agents come into contact with the population and potentially pass the disease to susceptibles. The infected individuals are represented by parallel agents with the $Trans$ agents passing on the disease and the $I1$ agents able to be contacted by a $Trans$ agent. Communication is prioritised so that all $Trans$ make contact. While this form of communication was inappropriate for obtaining a resource in the models of Section 3.2, here prioritised contact is biologically plausible for two reasons. First, contact is not guaranteed to result in infection (see $SI2$). Second, contact with the whole population is possible (so the $Trans$ are not magically seeking out the susceptibles). $S1$ that are contacted become $SI2$, while $I1$ and $R1$ agents are not affected by contact since infected and recovered individuals cannot become infected again. After the contact stage the $Trans$ agents all become the null agent 0 so that the infected individuals are once again represented by a single agent. The final stage is the probabilistic stage in which all individuals can give birth to a susceptible individual, with probability p_b, or die, with probability p_d. In addition the $SI2$ agents become infected with probability p_a and $I2$ agents can recover with probability p_r.

The system of MFEs derived from this model is:

$$
\begin{aligned}
S_{t+1} &= \frac{f}{f + N_t}\left((1 - p_d)S_t + p_b N_t - \frac{p_a S_t I_t}{N_t}\right), \\
I_{t+1} &= \frac{f}{f + N_t}\left((1 - p_d - p_r)I_t + \frac{p_a S_t I_t}{N_t}\right), \\
R_{t+1} &= \frac{f}{f + N_t}\left((1 - p_d)R_t + p_r I_t\right),
\end{aligned}
\tag{7}
$$

where $N_t = S_t + I_t + R_t$, the total population size at time t. These are similar to the standard SIR equations with frequency dependent transmission of disease [2], a form arising naturally from WSCCS models [20]. Here, however, there is an extra factor of $f/(f + N_t)$ on each equation describing the proportion of the population with appropriate resource. This is unconventional since in traditional models the transmission term (in this case $(p_a S_t I_t)/N_t$) is not affected by the density dependent birth or death term. We emphasise that the population dynamics of (7) come directly from explicit representation of individuals competing for resource rather than any population level assumptions imposed on the model. These equations are therefore candidates for modelling population dynamics in disease systems, despite the differences to traditional models.

In contrast, if we had taken the population dynamics from Sect. 3.1, with functional probability of birth, and added disease as above, we would merely add a logistic term to the equation for S with each group also dying probabilistically. This result would be closer to the traditional ODE models and would be simpler to analyse mathematically than (7) since the nonlinear density dependent term only appears in one equation (S). Despite this, the resource based approach to density dependence is preferable since this avoids implicit assumptions about population growth that may be incorrect.

4.2. Data Fitting

Although Fig. 6 and the derived MFE (7) describe a generalised disease they can still be useful as a first model for studying real disease systems, given appropriate simplifying assumptions. In this section we fit (7) to data for the spread of HIV in the UK. HIV was chosen for three reasons. First, data is available. The UK population statistics are published by the Office for National Statistics [23] and statistics on HIV diagnoses and deaths are published by Health Protection Agency [12]. The data we use is annual population data from 1997 to 2007 and numbers of people infected with HIV for the same period. Second, the model of Fig. 6 incorporates frequency dependent transmission of disease, i.e. individuals make a fixed number of contacts regardless of the size of the population. HIV, in common with other sexually transmitted diseases, is usually regarded as having frequency dependent transmission. Third, the model of Fig. 6 does not incorporate a term for death due to the disease. HIV can be viewed in this way if data for the period 1997-2007 is considered, which corresponds to the period after the introduction of highly active antiretroviral therapy (HAART) [8]. Since HAART greatly reduces the number of

deaths associated with infection the assumption that the disease does not increase the probability of death is reasonable.

The process of fitting to data is done in two steps. First we fit the data for overall population growth ignoring the disease, assuming values for the parameters p_b (probability of birth) and f (amount of resource), and fitting a value for p_d (probability of death) to get the best fit of the model to the data. Second we look at aspects of the disease, assuming a value for p_r (probability of recovery) and fitting p_a (probability of transmission following contact).

The time step we choose for the model is one year since we are fitting the model to annual data. This implies that infectious individuals make only one potentially infectious contact per year. Recall that this means unprotected sex between an infectious individual and a susceptible individual, and is an average for the whole population (not just the sexually active subgroup of the population). Note also that in our models, transmission is the product of c the contact rate and p_a the rate of infection following contact. Therefore, for a fixed transmission rate, as c increases, p_a decreases. Given that we are taking a simplified view of the disease, it is more straightforward to set $c = 1$ and to fit p_a to the data. We will take the simplifying assumptions into account when interpreting the results.

The underlying population growth expressed in Fig. 6 is

$$\begin{aligned} N_{t+1} & = & S_{t+1} + I_{t+1} + R_{t+1} \\ & = & (1 + p_b - p_d)\frac{fN_t}{f + N_t} \quad , \end{aligned}$$

i.e. the same as (5), the MFE for Fig. 5. This is because the disease has no effect on the overall size of the population. We fit (5) to the total population size [23].

The probability of birth was taken to be $p_b = 0.0119$. This corresponds to a crude birth rate of 11.90 per 1000 of the population, which was the mean birth rate in the UK between 1997 and 2006 [22, 9, 21]. The quantity of available resource was chosen to be very large, $f = 10^{12}$, because it was felt that competition for resources would have a small effect in a developed country such as the UK. This means that p_d is the only parameter to be fitted and the equation becomes

$$N_{t+1} = (1.0119 - p_d)\frac{10^{12}N_t}{10^{12} + N_t} \quad , \tag{8}$$

with initial conditions $N_{1997} = 58314249$ (the UK population in 1997 [23]). Fitting (8) to the data by the method of least squares [31] gives the probability of death as $p_d = 0.00782$.

Next the data for the numbers of HIV infections is fitted to (7). Since recovery from HIV infection is not possible we choose $p_r = 0$ and can eliminate the equation for R. Making use of the parameters which were already fitted above the equations
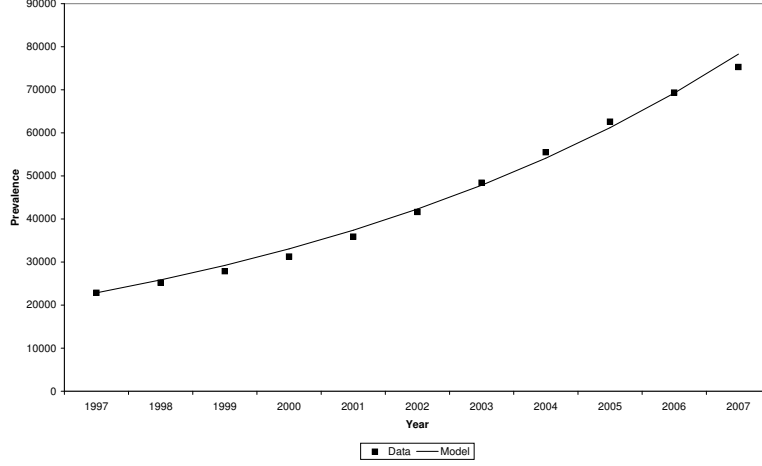
FIGURE 7. Infecteds (I) of (7) for $p_b = 0.0119, f = 10^{12}, p_d = 0.00782, p_r = 0, p_a = 0.138$ and initial population $S_{1997} = 58291402, I_{1997} = 15182, R_{1997} = 0$.

to be fitted to the data for the numbers of infected individuals are

$$
\begin{aligned}
S_{t+1} &= \frac{10^{12}}{10^{12} + S_t + I_t}\left(0.99218S_t + 0.0119(S_t + I_t) - \frac{p_a S_t I_t}{S_t + I_t}\right), \\
I_{t+1} &= \frac{10^{12}}{10^{12} + S_t + I_t}\left(0.99218I_t + \frac{p_a S_t I_t}{S_t + I_t}\right),
\end{aligned}
\tag{9}
$$

with $I_{1997} = 22847$ (the number of infected individuals in the UK in 1997 [12]) and $S_{1997} = N_{1997} - I_{1997} = 58291402$. This leaves us with one parameter, p_a, to be chosen. By fitting the equation for I to the data, again by least squares, we find $p_a = 0.138$.

In Fig. 7 the MFE for I is plotted alongside the data for the number of infections. We can see that the model follows the data closely for the entire period being considered. This shows us that, despite its simple, generalised nature, our model can be useful in describing a real disease system. To more accurately describe the spread of HIV a more specific model could be developed. Such a model would feature disease induced death - HAART does not entirely eliminate AIDS related deaths - and may also split the population into subpopulations since it is well known that some groups (e.g. intravenous drug users and gay men) are at increased

risk of infection. The population model itself would also be more realistic, including migration as well as births and deaths.

5. Conclusion

In this paper we have presented models of population dynamics in which the population will, over time, tend to some steady state and will not display unbounded growth. Two distinct mechanisms were used to achieve bounded growth: the implicit approach in which the effects of restricted resources are included by allowing more complex language features in the model (functional probabilities) and the explicit approach in which those resources are represented by agents. The introduction of functional probabilities allow us to succinctly take full advantage of the expressive capabilities of WSCCS. These models led naturally to the logistic equation [30], the classical expression used to describe population dynamics. This is in contrast to the results of Brännström and Sumpter [4] who found several other existing expressions could be derived from their individual based models but not the logistic equation. The logistic equation arises from our models because the assumptions used to introduce density dependence – functional probabilities which are linearly proportional to the population size – match the assumptions on which the logistic equation is based. If we use functional probabilities which are non-linearly proportional to the population size we would of course obtain different MFEs. It can be easily argued that adding functional rates is self-defeating for our objectives; if we allow inclusion of strong implicit assumptions, such as the nature of population growth, then we may as well simply write down the MFEs directly.

In order to reduce the number of population level assumptions in our models we have also developed models which feature agents to represent resource, with the dynamics in the population arising from the competition between individuals for that resource. With density dependent death this model leads to the Beverton-Holt model [3] which was proposed for the population dynamics of fish stocks. The fact that this equation has naturally arisen here from the competition between individuals means we can consider the Beverton-Holt model a serious candidate to be used when modelling population dynamics.

Lastly, our goal in population modelling is to incorporate models of disease to gain a more realistic individual based disease model. By adding a model of disease spread to population dynamics we have derived a system of equations (7) which differs from those which have previously been described in the literature. The population dynamics in our model naturally arise from the interactions between individuals and the environment, rather than any assumptions imposed at the population level. Therefore, we have well-founded reason to propose this model for a disease system featuring density dependence in deaths. The model was validated with respect to data on HIV in the UK population. Since a very simple model was used, a number of strong assumptions were required in fitting the model to

the data. Future work will include developing more complex realistic models and validating those models with disease data.

Of course it would have been possible, especially for these simple models, to write down plausible mean field equations, or to have written down the Markov Chain directly. The advantage of process algebra is that it gives a convenient and modular way of writing down individual behaviour. The contribution of our work is to then convert that individual based model into something *facilitating rigorous algebraic manipulation*. The resulting (automatically derived) population dynamics are based directly on explicit assumptions about the individual interactions which are fundamentally important in any biological system. Our approach therefore makes it straightforward to study the population level dynamics under different assumptions.

References

[1] E. Ahmed and A.S. Elgazzar. On some applications of cellular automata. *Physica A-Statistical Mechanics and its Applications*, 296:529–538, 2001.

[2] M. Begon, M. Bennet, R.G. Bowers, N.P. French, S.M. Hazel, and J. Turner. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology and infection*, 129:147–153, 2002.

[3] R.J.H. Beverton and S.J. Holt. *On the dynamics of exploited fish populations*, volume 19 of *Fisheries Investigations, Series 2*. H.M.S.O., 1957.

[4] A. Brännström and D.J.T. Sumpter. The role of competition and clustering in population dynamics. *Proceedings of the Royal Society of London Series B*, 272:2065–2072, 2005.

[5] M. Calder, S. Gilmore, and J. Hillston. Automatically deriving ODEs from process algebra models of signalling pathways. In *Proceedings of Computational Methods in Systems Biology (CMSB 2005)*, pages 204–215, 2005.

[6] M. Calder, S. Gilmore, and J. Hillston. Modelling the influence of RKIP on the ERK signalling pathway using the stochastic process algebra PEPA. *Transactions on Computational Systems Biology VII*, 4230:1–23, 2006.

[7] L. Cardelli. On process rate semantics. *Theoretical Computer Science*, 391:190–215, 2008.

[8] The CASCADE Collaboration. Survival after introduction of HAART in people with known duration of HIV-1 infection. *The Lancet*, 355:1158–1159, 2000.

[9] General Register Office for Scotland. Scotland's Population 2007: The Registrar General's Annual Review of Demographic Trends. Published online: retrieved September 2008. www.gro-scotland.gov.uk/files1/stats/scotlands-population-2007-the-register-generals-annual-review-153rd-edition/scotlands-population-2007-the-register-generals-annual-review-153rd-edition.pdf.

[10] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London Series B*, 115:513–585, 1825.

[11] M.P. Hassell. Density-dependence in single-species populations. *Journal of Animal Ecology*, 45:283–296, 1975.

[12] Health Protection Agency. New HIV Diagnoses Surveillance Tables: UK data to the end of December 2007. Published online: retrieved August 2008. www.hpa.org.uk/web/HPAwebFile/HPAweb_C/1208763421275.

[13] J. Hillston. Fluid Flow Approximation of PEPA models. In *QEST'05, Proceedings of the 2nd International Conference on Quantitative Evaluation of Systems*, pages 33–42. IEEE Computer Society Press, 2005.

[14] J. Joo and J.L. Lebowitz. Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercubic lattice. *Physical Review E*, 70, 2004.

[15] W.O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics i. *Proceedings of the Royal Society of London Series A*, 115:700–721, 1927.

[16] T. Malthus. *An Essay on the Principle of Population*. J. Johnson, London, 1798.

[17] C. McCaig. *From individuals to populations: changing scale in process algebra models of biological systems*. PhD thesis, University of Stirling, 2007.

[18] C. McCaig, R. Norman, and C. Shankland. Deriving mean field equations from large process algebra models. Technical Report CSM-175, Department of Computing Science and Mathematics, University of Stirling, March 2008. http://www.cs.stir.ac.uk/research/publications/techreps/pdf/TR175.pdf.

[19] R. Milner. *Communication and Concurrency*. Prentice Hall, 1989.

[20] R. Norman and C. Shankland. Developing the use of process algebra in the derivation and analysis of mathematical models of infectious disease. In *Computer Aided Systems Theory - EUROCAST 2003*, volume 2809 of *Lecture Notes in Computer Science*, pages 404–414. Springer-Verlag, 2003.

[21] Northern Ireland Statistics and Research Agency. Population and vital events, 1926 to 2006. Published online: retrieved September 2008. www.nisra.gov.uk/archive/demography/publications/annual_reports/2006/Table1.1_2006.xls.

[22] Office for National Statistics. Birth Statistics: Review of the Registrar General on births and patterns of family building in England and Wales, 2006. Published online: retrieved August 2008. www.statistics.gov.uk/downloads/theme_population/FM1_35/FM1_No35.PDF.

[23] Office for National Statistics. United Kingdom 1971 - 2081. Published online: retrieved August 2008. www.statistics.gov.uk/populationestimates/svg_pyramid/uk/data/data.xls.

[24] A. Regev, E.M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. Bioambients: an abstraction for biological compartments. *Theoretical Computer Science*, 325:141–167, 2004.

[25] W.E. Ricker. Stock and recruitment. *Journal of the Fisheries Research Board of Canada*, 11:559–623, 1954.

[26] D. Sumpter. *From Bee to Society: an agent based investigation of honeybee colonies*. PhD thesis, UMIST, 2000.

[27] C. Tofts. Using process algebra to describe social insect behaviour. *Transactions of the Society for Computer Simulation*, 9:227–283, 1993.

[28] C. Tofts. Processes with probabilities, priority and time. *Formal Aspects of Computing*, 6:536–564, 1994.

[29] J. Turner, M. Begon, and R.G. Bowers. Modelling pathogen transmission: the interrelationship between local and global approaches. *Proceedings of the Royal Society of London Series B*, 270:105–112, 2002.

[30] P.F. Verhulst. Notice sur la loi que la population suit dans son accroissement. *Corr. Math. et Phys.*, 10:113–121, 1838.

[31] C. R. Wylie. *Advanced Engineering Mathematics*. McGraw-Hill, 1966.

Chris McCaig
Department of Computing Science and Mathematics, University of Stirling,
Stirling, FK9 4LA, UK.
e-mail: `cmc@cs.stir.ac.uk`

Rachel Norman
Department of Computing Science and Mathematics, University of Stirling,
Stirling, FK9 4LA, UK.
e-mail: `ran@cs.stir.ac.uk`

Carron Shankland
Department of Computing Science and Mathematics, University of Stirling,
Stirling, FK9 4LA, UK.
e-mail: `ces@cs.stir.ac.uk`