

A rigorous approach to investigating common assumptions about disease transmission

Process algebra as an emerging modelling methodology for epidemiology

Chris McCaig · Mike Begon · Rachel Norman · Carron Shankland

Abstract Changing scale, for example the ability to move seamlessly from an individual-based model to a population-based model, is an important problem in many fields. In this paper we introduce process algebra as a novel solution to this problem in the context of models of infectious disease spread. Process algebra allows us to describe a system in terms of the stochastic behaviour of individuals, and is a technique from computer science. We review the use of process algebra in biological systems, and the variety of quantitative and qualitative analysis techniques available. The analysis illustrated here solves the changing scale problem: from the individual behaviour we can rigorously derive equations to describe the mean behaviour of the system at the level of the population. The biological problem investigated is the transmission of infection, and how this relates to individual interactions.

Keywords epidemiology, multiscale modelling, theoretical computer science, changing scale

Mathematics Subject Classification (2000) 92D30 · 68Q85

1 Introduction

Wing (2006) puts forward the notion that “*Computational thinking is a fundamental skill for everyone, not just for computer scientists.*” and advocates the application of computational approaches in other disciplines.

C. McCaig · R. Norman · C. Shankland
Department of Computing Science and Mathematics, University of Stirling,
Stirling, FK9 4LA, UK. E-mail: {cmc,ran,ces}@cs.stir.ac.uk

M. Begon
School of Biological Sciences, University of Liverpool,
Liverpool, L69 3BX, UK. E-mail: mbegon@liverpool.ac.uk

One discipline which might particularly benefit from the structured abstraction tools of computer science is biology. For example, in Systems Biology (Kitano 2002), biologists have embraced the computational approach to unravel the dynamics of cell behaviour (Bernardo et al. 2008; Cohen 2008; Priami 2006). Computational models are part of a virtuous cycle (Kitano 2002) consisting of: hypothesis generation, hypothesis testing *in silico* (which is cheap and fast), leading to hypothesis testing *in vivo* or *in vitro* (which is thus targeted more efficiently to interesting cases), providing more information to refine the original hypothesis. Thus far, other areas of biological research, such as ecology and epidemiology, have been slower to take up computational approaches to modelling and analysis.

The traditional approach to modelling ecological systems, such as those investigating the spread of infectious disease, has made use of systems of coupled ordinary differential equations (ODEs) which describe the system in terms of changes in the population as a whole (Anderson and May 1981; Kermack and McKendrick 1927). Such models of disease are often referred to as SIR models: where S are susceptible individuals who have never had the disease; I are infected individuals who can pass on the disease to susceptibles; and R are recovered individuals who are assumed to be immune from future infection. The advantage of these models is that they are amenable to a wide range of well established analytical techniques, for instance equilibrium analysis. Such SIR models are, however, based on assumptions about how individual behaviour, such as interactions between infected and susceptible individuals, affects the population as a whole. In particular, there is much debate on the correct form of the transmission term for disease spread (Fenton et al. 2001). The original SIR type models (Anderson and May 1981) use

density dependent transmission (βSI) which assumes that as the population size increases, the number of contacts an individual makes increases proportionally. This describes a well mixed population and might be suitable for some wildlife populations. Another commonly used alternative is frequency dependent transmission ($\beta SI/(S + I + R)$) which assumes that an individual makes a constant number of contacts no matter what the population size. This is often used to model sexually transmitted diseases, but is being used increasingly to model other human diseases under the assumption that people usually interact in relatively fixed social groups. In reality, transmission will be much more complicated than either of these, incorporating, for example: super-spreaders, where some individuals are more likely to transmit the disease than others; different mixes of spatially local and global interactions to take into account the fact that people travel more extensively; or dynamic interactions where individuals change their behaviour in response to the disease (e.g. increased handwashing in response to swine flu).

Our goal in this paper is to present a particular computational approach, process algebra (Baeten 2005), as an exciting method for modelling infectious disease spread. Process algebra gives ways to think about a model from several angles, including the behaviour of individuals, and the more abstract properties of the population. We will show that the computational style brings additional ways of thinking about a problem through abstraction and modular composition of sub-problems, and through various forms of qualitative and quantitative analysis. Although originally developed to study distributed computer systems, process algebras are increasingly used to study biological systems. Early examples include the behavioural models of task allocation and synchrony in ants of Tofts (1993), the model of the reproductive cycle of honeybee parasite *Varroa jacobsoni* of Sumpter and Broomhead (2001) and Regev et al.'s model of the RTK-MAPK signal transduction pathway (2001). More recently, Calder and Hillston (2010) have surveyed approaches to a number of signalling pathway models. The key advantages of the process algebra approach are illustrated in this paper by investigating the biological problem alluded to above: the link between different forms of individual level interaction and the resultant population level transmission terms for the disease. This is done through a rigorous method for deriving equations to describe the mean behaviour of an infectious disease system as a whole (McCaig et al. 2008; 2009).

The structure of the paper is as follows. We review the advantages of process algebras in Section 2, considering them in relation to the established mathe-

matical approaches. The biological problem to be addressed is described in Section 3. Novel technical work in the shape of process algebra models to investigate that problem are presented in Section 4. Finally, we conclude with a discussion on the utility of process algebra for epidemiology.

2 Process algebra

In the 1970's computers and computer networks were growing and becoming more complex. Such systems are by nature highly distributed (components in different locations), concurrent (components evolving separately but simultaneously), and involve communication between components to allow some coordinated activity to occur. Process algebras (Baeten 2005) were developed to describe such systems in an abstract way to allow computer scientists to analyse the structure of the network, the intercommunication between components, and the ways in which some greater goal could be achieved by many interacting components. Typically, process algebras are small languages with a limited set of operations, to make modelling simple, and mathematically based, to make rigorous analysis possible.

Process algebras describe complex systems by the composition of multiple parallel agents, called *processes*. A process can carry out *actions* or *events*, some of which may involve interactions with other processes. Choices between actions may be made deterministically, non-deterministically, or probabilistically. There are a wide range of process algebras; see Baeten (2005) for a historical review. In addition to the core features (processes, events, choice, parallelism and interaction), variants can include timed events, passing of data in interactions, alternative synchronisation mechanisms for events, and inclusion of spatial information.

Process algebras are therefore ideally suited to describing biological systems which may typically be viewed as networks of (many) interacting components, where the components themselves may have complex, nondeterministic, individual behaviour. Our group has pioneered the use of process algebra for epidemiology (Norman and Shankland 2003; McCaig 2007; McCaig et al. 2009). Process algebra is particularly suited for epidemiology and ecological systems, because these can be considered in terms of individual behaviours and interactions between individuals, or in terms of changes in the age- or stage- structure of a population. Moreover, these local changes may be fundamental to the way a population changes over time. Many of these behaviours and interactions can potentially be observed and measured in the field. For example, animals can be observed to be solitary or gregarious, or to interact

in small or large groups. In contrast, parameters such as transmission of disease can only ever be estimated since they come from a complex interaction of several elements.

Despite the range of different process algebras available, all share the same advantages both in constructing models, and in investigation of those models. Process algebra models depend on a notion of observation, which can be varied to suit the requirements of the modeller. This yields *abstraction*. Process algebra models are composed of smaller interacting parts. This yields *modularity*. Lastly, a process algebra model gives access to a range of investigatory techniques¹, each of which contributes to understanding of the system in a different way, allowing a fuller picture to be developed.

The modular and compositional nature of process algebra makes it straightforward to incrementally develop a model. The abstract nature of process algebra means that the modeller can decide how much information to include about a system, and what to ignore. And of course they can change their mind later. Further, only the behaviour of the individuals is considered explicitly; the complex state space of the combined activities of individuals is generated automatically using the process algebra semantics. Lastly, process algebra gives access to quantitative analysis and qualitative analysis:

Stochastic simulation. Gillespie simulation and related techniques (1977) can be used to simulate system behaviour. This is particularly appropriate for a small number of individuals, where stochastic variability is more influential in the system.

Markov analysis. The semantics of (stochastic) process algebras are given in terms of Markov Chains. This gives access to a range of established mathematical analysis techniques.

Algebraic manipulation. Process equivalences can be applied to transform the model (perhaps into one with known properties).

Verification. Logically expressed properties can be evaluated against the model. For example, properties relating to the size of population groups such as peak of infection, or infection dying out.

Derivation of Mean Field Equations (MFEs). The model can be translated to give a population-based model in terms of MFEs (McCaig et al. 2008; Hillston 2005; Cardelli 2008b). This also gives access to a range of established mathematical analysis techniques commonly used in mathematical biology.

While the others are useful, we believe this last to be the most exciting, and focus on this analysis technique in Section 4. The translation process is rigorous and automated, therefore this is a way of changing scale in the model from the level of individuals to the level of the population. Explicit features in the individual-based model, together with the underlying semantics of interaction in process algebra, are translated into high level population properties. These are often termed *emergent properties*. Such properties are hard or impossible to grasp by looking at individual components: they often only become apparent in the whole system model. Changing scale yields greater insights about the effect of individual interactions on population dynamics.

The computational approach allows us to use the huge amounts of experimental data being generated through wet lab experimentation and field observations. This is a constructive rather than reductionist approach: improving understanding of a system by building models capturing hypotheses about the functionality of its components and the way those components work together to achieve the functionality of the system as a whole. Most importantly, given a model we can test it (and therefore our hypotheses) in a variety of computational and analytical ways. Testing hypotheses through *in silico* methods is much cheaper and faster than through lab based *in vitro* or *in vivo* tests. Further, in the face of overwhelming quantities of data, biologists need an abstract functional view of the dynamic evolution of a system through time to better understand that system (Priami 2006). Through *in silico* testing, the range of interesting parameter values can be narrowed down, and more focussed lab or field based tests identified. For epidemiology this is particularly important: it may not even be possible to carry out *in vivo* tests due to practical and ethical constraints.

Process algebras have been strongly adopted for use in Systems Biology, e.g. (Bernardo et al. 2008; Calder and Hillston 2010; Priami 2006). The most common approach is to view molecules or species as processes, and biochemical reactions as the evolution of one process to another (possibly through interaction with other processes). Calder and Hillston (2010) review approaches to molecular biology, also considering process algebras which have been specially enhanced to facilitate biological modelling. Process algebras have also been used in other areas of biology. For example, the synapse model of Bracciali et al. (2008), the ant colonies of Tofts (1993) and Sumpter and Broomhead (2001), and the models of vertical parasite transmission of Hatcher et al (1995).

Some others have followed us in applying process algebra to epidemiology. For example, Cardelli (2008a)

¹ The process algebra community refers to these as analysis techniques, but the word is used in a far more general sense of investigation rather than the specific sense of mathematical analysis.

presents a small epidemiological model (based on the classic Kermack and McKendrick model (1927)) and Bio-PEPA is used in a model of avian flu (Ciocchetta and Hillston 2010). Bio-PEPA has been developed for biochemical reactions, and the H5N1 model includes spatial aspects and event-driven introduction of disease control. The modelling approach is population-based, and therefore essentially equivalent to writing an ODE description, so not all of the advantages mentioned above are exploited. In particular, they do not exploit the possibility of changing scale and assume the usual mass action transmission rate, instead of deriving it as we do.

2.1 How does process algebra differ from other techniques?

Process algebra allows us to develop probabilistic individual-based models and to carry out a range of analyses, including derivation of population level behaviour. There are several other approaches to both individual-based modelling and population-based modelling: what advantages can modellers obtain through using process algebra?

Cellular Automata Another approach to modelling ecological systems uses individual-based cellular automata models, which describe a system in terms of the individuals that make up the population (Joo and Lebowitz 2004; Turner et al. 2002). Individuals are modelled on a grid, interacting with their nearest neighbours. This approach captures the fact that properties emerging at the whole population level are the result of interactions at the subpopulation or individual level. Model analysis is typically carried out by performing simulations. As with process algebra simulations, an individual simulation considers only a single realisation of the stochastic model and it is only by performing a large number of simulations that we can comment on the average behaviour of the system. This becomes computationally infeasible as the number and range of parameters increases.

Some algebraic analysis is available through methods such as pair approximation. This method was introduced by Matsuda et al. (1992). It is a moment closure method which produces a system of ordinary differential equations for the frequency of each type of neighbouring site pairs. Higher order frequencies (e.g. triplets) are approximated by pair frequencies in order to obtain a closed system of equations (Ellner 2001). There are a number of ways in which this closure approximation can be done and it is not yet clear what the consequences of making each of these approximations

is, or when they are equivalent (Webb et al. 2007b). Although pair approximations have been shown to predict the behaviour of lattice models (Webb et al. 2007b) there are examples in which discrepancies between the equilibrium analysis of the pair approximations and the simulations arise. For example, the pair approximations sometimes predict cycles when these do not appear in the simulations (Boots and Sasaki 2001). Another limitation is that in most cases interactions are purely local; however, some authors have included longer range interactions e.g. (Ellner 2001; Webb et al. 2007a).

A key difference between cellular automata and many process algebras is the explicit inclusion of space. The simpler process algebras assume random mixing in the first instance. This assumption makes the process algebra more immediately comparable to SIR ODE models. Cellular automata and process algebra become more similar in the limits of global interactions in cellular automata or totally local interactions in the process algebra. It is possible to impose spatial relationships in a process algebra model through the topology of communications, and there are spatial variants of process algebra. For example, Bio-PEPA (Ciocchetta and Hillston 2010) includes compartments. Such extensions allow description of explicitly spatial models and meta-population models.

The final type of models to which process algebra models could be compared are individual-based models in which simulations are run and the fate of each individual is determined, e.g. (Railsback and Harvey 2002). However, these models are often complex and rely on simulation. Process algebra allows us to do the same thing, but also derive mean field equations.

Stochastic models There are a number of ways of formulating stochastic models. In ecological models stochasticity is often added to ODE models through either environmental noise which we add as a noise term to each equation, e.g. (Nowicki et al. 2009), or demographic noise in which we allow individual parameter values to vary randomly, e.g. (Cornell et al. 2004). Models with environmental noise cannot readily be compared to process algebra models. Models with demographic noise essentially allow for probabilistic variation between individuals. From such models MFE can be derived, therefore this is similar to the process algebra approach. However, models with demographic noise start with assumptions about the population level behaviour and then add in randomness. The behaviour of each individual is not explicitly modelled, therefore there can be no derivation of population equations from individual level interactions.

Stochastic reaction equations are a further alternative. In a stochastic reaction equation, the transition probabilities between states are explicitly modelled and mean field equations can be derived. Reaction equations can be simulated using techniques such as Gillespie’s algorithm (1977) and its derivatives. This is therefore rather similar to the rules describing how one process evolves to another in process algebra, especially for non-interaction evolutions such as birth and death rates. Process algebra models can also be simulated in the same way. The difference with process algebra lies in the approach to derived terms involving interaction between processes. Transmission rates for disease are a classic example. Such terms emerge from the process algebra semantics directly, whereas stochastic reaction equations resort to assumptions about density dependence in these terms.

Process algebras are closely related to stochastic models in that the same underlying theory, i.e. the work of Kurtz (1970), can be used to prove the relationship between the individual-based models and the mean field equations. Ultimately, process algebra is the only one of the techniques considered in which it is possible to model the interaction of individuals explicitly and rigorously determine the impact of that behaviour on the population.

Priami (2006) suggests that process algebra brings a fundamentally different, algorithmic, approach to describing biological systems. While ODEs describe the function “computed” by a system, i.e. the end result given a set of inputs, the process algebra model allows a more direct mechanistic description of the steps carried out moment by moment to effect that function. This captures causal, temporal and spatial information (depending on the process algebra used). We might capture this as the difference between “what” the result is (equations) and “how” the result is obtained (process algebra). This is particularly important in systems where the transient dynamics are of interest. Two systems may have the same initial conditions and equilibria, but have quite different transient dynamics. The appropriate process algebra model can capture that explicitly. Equations can only do so implicitly.

3 Biological Problem

Traditional ODE models for the spread of infectious diseases have utilised different terms to describe transmission of the disease (Anderson and May 1981; Begon et al. 2002; Briggs and Godfray 1995; Hochberg 1991). The classical model developed by Kermack and McKendrick (1927) introduced ideas which have formed the

basis of models up to the present day. In particular they introduced the transmission term

$$\beta SI, \quad (1)$$

where S and I are the numbers (or densities) of susceptible and infectious individuals respectively and β is the transmission rate. This term has been widely used, for example in the seminal models of Anderson and May (1981), and is referred to as density dependent or mass action transmission. Here, contacts between S and I are assumed to increase linearly with host density.

Several authors have suggested that density dependent transmission may not be the most suitable term to describe disease spread, proposing alternative transmission terms (Briggs and Godfray 1995; Fenton et al. 2001; Hochberg 1991; Knell et al. 1998). Most notable amongst these is frequency dependent transmission,

$$\frac{\beta' SI}{N} \quad (2)$$

where β' is dimensionally different from β in (1) (Begon et al. 2002) and N is the total number (or density) of individuals in the population. Frequency dependent transmission is used most commonly in diseases where contact saturation is assumed to have occurred. This means that only a fixed number of contacts are made per unit time, regardless of the total population size. The choice of which term to use is based on assumptions about how the behaviour of individuals affects the population as a whole.

Begon et al. (2002) propose a general transmission term from which both frequency dependent and density dependent transmission can be derived:

$$Scpv, \quad (3)$$

where S is the number of susceptibles, c is the rate of ‘appropriate’ contacts between individuals, p is the probability that contact for a susceptible host is with an infected host, and v is the probability that contact between an ‘infected’ and a susceptible host leads to successful transmission. The difference between the frequency and density dependent transmission terms is simply the form of c : if c is constant we have frequency dependent transmission and if c is proportional to the population size we have density dependent transmission. In the models presented in Section 4 we investigate the effect of varying the contact rate in this way.

Turner et al. (2002) employ cellular automata models that implement analogues for density and frequency dependent transmission at the level of the individual (one individual per cell). Their first model has susceptibles making contact with the immediate neighbours in the spatial grid, with the number of contacts reduced

if neighbouring grid squares are empty — analogous to density dependent transmission. In their second model, susceptibles make a fixed number of contacts, travelling outward to seek contacts if neighbouring grid squares are empty — analogous to frequency dependent transmission.

Fitting equations statistically to the results obtained from the cellular automata, Turner et al. found that irrespective of the individual-based behaviour implemented, the frequency dependent transmission term most accurately describes behaviour at the population level. This implies that whatever rules operate at the individual level, frequency dependent transmission will be the outcome at the population level. This is counter to normal assumptions and has serious implications for models of infectious disease spread.

More recently, Rhodes and Anderson (2008) developed a population level model including the movement of individuals, in which transmission depends on how close together susceptible and infected individuals are. By making differing assumptions about how individuals behave they found that both density dependent and frequency dependent transmission can accurately describe the behaviour of the population. The models of Turner et al. and Rhodes and Anderson contain spatial information but support the traditional terms of (1) and (2) in which random mixing is assumed. This suggests that random mixing might be a suitable simplifying assumption. In the next section this question is investigated using process algebra.

4 Models

The present work makes use of the process algebra Weighted Synchronous Calculus of Communicating Systems (WSCCS) (Tofts 1994), because it has proved particularly useful in studying a wide range of biological systems including the behaviour of social insects (Tofts 1993; Sumpter and Broomhead 2001), the spread of infectious disease (Norman and Shankland 2003), population growth controlled through limited resource, added to an HIV model (McCaig et al. 2009), and vertical parasite transmission in *Gammarus duebeni* (Hatcher et al. 1995). WSCCS is a discrete time process algebra with each agent (representing an individual in the population) performing an action (representing some aspect of individual behaviour) in each step of time. These time steps are not of a defined length and there is no notion of absolute time in WSCCS; however, choices are made probabilistically and the selection of values for these probabilities reflects a notion of time scale. For example, if a probability of contact between individuals is described, then it must be associated with

a time scale such as contacts per minute, per hour, or per day. We describe different types of individual in the syntax of WSCCS and place many individuals in parallel to represent a population. A brief summary of the syntax of WSCCS can be found in Appendix A.

The models considered in this section are process algebraic interpretations of standard SIR models and assume random mixing. We present the WSCCS syntax of our models (Figs. 1 and 2) and give a description of the individual behaviours captured, along with the mean field equations (MFEs) that are derived (making use of a rigorous algorithm described by the authors elsewhere (McCaig et al. 2008; 2009)). The derived equations depend on the individual behaviour captured in WSCCS and do not rely on population level assumptions (other than random mixing). In this way we hope to either validate existing population level equations of host-pathogen systems, which have previously relied on assumptions about the population, or propose new equations to describe such systems. As mentioned in Section 2 other forms of analysis are possible, but we do not present these here.

Sumpter and Broomhead (2001) used an unstated heuristic to derive equations to describe the mean behaviour at the population level of social insect colonies. Norman and Shankland (2003) used a similar method to derive MFEs for their models of infectious disease spread, in particular finding that frequency dependent transmission (Begon et al. 2002) is most naturally described in WSCCS. In McCaig's thesis (2007) and the related report (McCaig et al. 2008) a method is described to automatically derive Mean Field Equations from WSCCS models. Essentially, a model of the form $A1\{n_1\}|\dots|Am\{n_m\}$ will evolve to a model composed of the same agents, but in different numbers i.e. $A1\{n'_1\}|\dots|Am\{n'_m\}$. The translation is based on analysis of the potential actions of each agent in the context of the whole population, producing a term in the MFE capturing change in the number of that agent. The translation is purely syntax based: the classification of WSCCS operators relates directly to a range of possible terms to be added to the MFE. For example, given an unprioritised communicating action a , and processes $A1 = a.A2$ and $B1 = \bar{a}.B2$ then the method prescribes that the resulting change in numbers of $A2$ and $B2$ agents is *(the number of $A1 \times$ the number of $B1)/N$* where N is the total number of agents potentially able to carry out a or \bar{a} actions.

The translation from process algebra to mean field equations is based on a well known result for Markov chains by Kurtz (1970). The underlying semantics of a WSCCS model can be viewed as a Discrete Time Markov Chain (DTMC). McCaig shows in his thesis

that the translation from WSCCS syntax to MFE is the same one as would be obtained by converting the WSCCS syntax to a DTMC and using the Kurtz result to obtain a MFE. The models have also been simulated (not shown here), and in accordance with McCaig's result match the MFE time series exactly.

Due to the discrete time nature of WSCCS, our MFEs are difference equations, the discrete time analogues of ODEs. Another consequence of the discrete time nature of WSCCS is that behaviour in our model is separated into several stages. For instance, we separate probabilistic behaviours (e.g. birth, death, recovery) from those involving contact between individuals (e.g. transmission of infection). By doing this we can more easily reason about the overall behaviour of the system. The different stages happen sequentially during one iteration of the model which corresponds to the interesting unit of time for the modeller. The length of the time step depends on the units of the parameters.

Begon et al.'s general transmission term (2002) assumes that susceptibles can make contact with any type of individual but can only become infected by contact with an infectious individual. The nature of WSCCS means it is more natural to write models in which infecteds can make contact with any type of individuals, and it is only susceptibles that can become infected after contact. This means that for our purposes we consider a basic equation of the form

$$Icp'v, \quad (4)$$

where I is the number of infected individuals, p' is the probability that contact is with a susceptible, c is the rate of contacts and v is the probability that contact between a susceptible and an infected leads to transmission of the disease.

4.1 Frequency dependent transmission

The model presented in Fig. 1 is based on the disease models of Norman and Shankland (2003), with the addition of birth and death to the model (McCaig et al. 2009). In this model the rate of contacts is constant. The model consists of three stages. In the first stage all individuals can give birth (to a single newborn $Nb2$) with a density dependent probability ($p_b = p_{b_0} - kN_t$, where p_{b_0} is the probability of birth in the absence of crowding, k is a scaling constant and $N_t = S_t + I_t + R_t$). The infected individuals ($I1$) can also probabilistically become transmitters (becoming the parallel agents $Trans \times I2$ with fixed probability p_c), which make contact in the current iteration of the model.

$$\begin{aligned}
 p_b &\stackrel{\text{prob}}{=} p_{b_0} - k([S1] + [I1] + [R1]) \\
 S1 &\stackrel{\text{def}}{=} p_b.\sqrt{} : S2 \times Nb2 + (1 - p_b).\sqrt{} : S2 \\
 I1 &\stackrel{\text{def}}{=} p_b.\sqrt{} : I2 \times Nb2 + p_c.\sqrt{} : I2 \times Trans \\
 &\quad + (1 - p_b - p_c).\sqrt{} : I2 \\
 R1 &\stackrel{\text{def}}{=} p_b.\sqrt{} : R2 \times Nb2 + (1 - p_b).\sqrt{} : R2 \\
 S2 &\stackrel{\text{def}}{=} \omega.infect : SI3 + 1.\sqrt{} : S3 \\
 I2 &\stackrel{\text{def}}{=} \omega.infect : I3 + 1.\sqrt{} : I3 \\
 Trans &\stackrel{\text{def}}{=} \omega.infect : 0 + 1.\sqrt{} : 0 \\
 R2 &\stackrel{\text{def}}{=} \omega.infect : R3 + 1.\sqrt{} : R3 \\
 Nb2 &\stackrel{\text{def}}{=} 1.\sqrt{} : Nb3 \\
 SI3 &\stackrel{\text{def}}{=} p_i.\sqrt{} : I1 + p_d.\sqrt{} : 0 + (1 - p_d - p_i).\sqrt{} : S1 \\
 S3 &\stackrel{\text{def}}{=} p_d.\sqrt{} : 0 + (1 - p_d).\sqrt{} : S1 \\
 I3 &\stackrel{\text{def}}{=} p_r.\sqrt{} : R1 + p_d.\sqrt{} : 0 + (1 - p_r - p_d).\sqrt{} : I1 \\
 R3 &\stackrel{\text{def}}{=} p_d.\sqrt{} : 0 + (1 - p_d).\sqrt{} : R1 \\
 Nb3 &\stackrel{\text{def}}{=} 1.\sqrt{} : S1 \\
 Popn &\stackrel{\text{def}}{=} S1\{s\} \times I1\{i\} \times R1\{r\}[\{\sqrt{}\}]
 \end{aligned}$$

Fig. 1 Frequency dependent transmission model.

The second stage features interaction between individuals (through the complementary actions *infect* and *infect*), which can lead to disease transmission. The transmitters ($Trans$) each make a single contact with any member of the population, although it is only susceptible individuals ($S2$) that are affected by contact. The newborn individuals from the previous stage ($Nb2$) do not take part in this interaction.

In the final stage individuals make probabilistic choices, with all probabilities being fixed. The contacted susceptibles ($SI3$) can become infected (becoming $I1$ with probability p_i), die (becoming the null agent 0 with probability p_d) or return to being susceptible (becoming $S1$ with probability $1 - p_i - p_d$). The infected individuals ($I3$) either recover (becoming $R1$ with probability p_r), die (becoming 0 with probability p_d) or remain infected (becoming $I1$ with probability $1 - p_r - p_d$). The recovered ($R3$) and uncontacted susceptibles ($S3$) each die (becoming 0 with probability p_d) or remain in their current state (becoming $R1$ and $S1$ respectively, with probability $1 - p_d$). The newborns from the first stage all mature to be susceptible individuals ($S1$) in the next iteration of the model.

Through the three stages of the model we have moved from a population consisting of some numbers of $S1$, $I1$ and $R1$ agents (representing susceptible, infected

and recovered individuals), to a population consisting of these same types of agent but in differing numbers. The mean of this change, across an iteration of the model (in this case three stages), is what is captured in the derived MFEs.

The MFEs derived from this model are

$$\begin{aligned} S_{t+1} &= (1 - p_d)S_t - \frac{p_i p_c S_t I_t}{N_t} + (p_{b_0} - k N_t) N_t, \\ I_{t+1} &= (1 - p_d - p_r)I_t + \frac{p_i p_c S_t I_t}{N_t}, \\ R_{t+1} &= (1 - p_d)R_t + p_r I_t. \end{aligned} \quad (5)$$

By choosing $\beta' = p_i p_c$ the transmission term in these MFEs can be rewritten as $\beta' S_t I_t / N_t$ i.e. the standard frequency dependent transmission term, (2).

Relating the transmission term in (5) to the general term, (4): the probability that contact is with a susceptible (p' in (4)) is S_t / N_t ; the probability that contact leads to infection (v in (4)) is p_i ; and the contact rate (c in (4)) is the constant probability p_c .

We can see that we have derived the frequency dependent transmission term by making the same assumption about contact rate as Begon et al. (2002). The advantage of our approach is that our population level equations have been rigorously derived from the WSCCS model of Fig. 1. Our equations are therefore a direct consequence of the individual behaviour described above and captured in Fig. 1.

4.2 Density dependent transmission

This second model (Fig. 2) differs from the first only in the form of p_c , the infecteds' probability of making contact. Rather than a fixed probability we now consider a density dependent value, directly proportional to the size of the population (McCaig 2007; McCaig et al. 2009) ($p_c = \kappa N_t$ where κ is a scaling constant). The MFEs derived for this model are of the same general form as (5), with the substitution of $p_c = \kappa N_t$, which gives

$$\begin{aligned} S_{t+1} &= (1 - p_d)S_t - \frac{p_i (\kappa N_t) S_t I_t}{N_t} + (p_{b_0} - k N_t) N_t, \\ I_{t+1} &= (1 - p_d - p_r)I_t + \frac{p_i (\kappa N_t) S_t I_t}{N_t}, \\ R_{t+1} &= (1 - p_d)R_t + p_r I_t. \end{aligned} \quad (6)$$

If we simplify and choose $\beta = p_i \kappa$ the transmission term in these MFEs becomes $\beta S_t I_t$. By once again making the same assumption about contact rate as Begon et al. (2002) we have rigorously derived the standard density dependent transmission term, (2).

$$\begin{aligned} p_b &\stackrel{\text{prob}}{=} p_{b_0} - k([S1] + [I1] + [R1]) \\ p_c &\stackrel{\text{prob}}{=} \kappa([S1] + [I1] + [R1]) \\ S1 &\stackrel{\text{def}}{=} p_b.\sqrt{} : S2 \times Nb2 + (1 - p_b).\sqrt{} : S2 \\ I1 &\stackrel{\text{def}}{=} p_b.\sqrt{} : I2 \times Nb2 + p_c.\sqrt{} : I2 \times Trans \\ &\quad + (1 - p_b - p_c).\sqrt{} : I2 \\ R1 &\stackrel{\text{def}}{=} p_b.\sqrt{} : R2 \times Nb2 + (1 - p_b).\sqrt{} : R2 \\ S2 &\stackrel{\text{def}}{=} \omega.infect : SI3 + 1.\sqrt{} : S3 \\ I2 &\stackrel{\text{def}}{=} \omega.infect : I3 + 1.\sqrt{} : I3 \\ Trans &\stackrel{\text{def}}{=} \omega.infect : 0 + 1.\sqrt{} : 0 \\ R2 &\stackrel{\text{def}}{=} \omega.infect : R3 + 1.\sqrt{} : R3 \\ Nb2 &\stackrel{\text{def}}{=} 1.\sqrt{} : Nb3 \\ SI3 &\stackrel{\text{def}}{=} p_i.\sqrt{} : I1 + p_d.\sqrt{} : 0 + (1 - p_d - p_i).\sqrt{} : S1 \\ S3 &\stackrel{\text{def}}{=} p_d.\sqrt{} : 0 + (1 - p_d).\sqrt{} : S1 \\ I3 &\stackrel{\text{def}}{=} p_r.\sqrt{} : R1 + p_d.\sqrt{} : 0 + (1 - p_r - p_d).\sqrt{} : I1 \\ R3 &\stackrel{\text{def}}{=} p_d.\sqrt{} : 0 + (1 - p_d).\sqrt{} : R1 \\ Nb3 &\stackrel{\text{def}}{=} 1.\sqrt{} : S1 \\ Popn &\stackrel{\text{def}}{=} S1\{s\} \times I1\{i\} \times R1\{r\}[\{\sqrt{}\}] \end{aligned}$$

Fig. 2 Density dependent transmission model, in which p_c is proportional to N_t .

5 Conclusions

In this paper we have presented process algebra as an emerging modelling methodology for epidemiology. The approach provides benefits both in modelling and analysis of such systems. In particular, the ability to change scale has the possibility to transform the way we think about disease models. This particular form of analysis was illustrated through models of infectious disease spread developed in the stochastic process algebra WSCCS that make use of the same assumptions that Begon et al (2002) used in producing biological justification for the commonly used transmission terms: density- and frequency-dependence. We found that the resulting MFEs reflect the individual-level assumptions in the expected manner. This is in contrast to the results of Turner et al. (2002), who found that irrespective of the individual level behaviour modelled the population was most accurately described by the frequency dependent transmission term. One key difference between the models, likely to account for this, is the treatment of space. The cellular automata developed by Turner et al. are inherently spatial, with individuals located on a grid. Individuals do not move so no mixing occurs.

In that case clustering of infection is likely, which will restrict the opportunities for some infected individuals to spread the disease. In contrast our WSCCS models contain no spatial information, which amounts to an assumption of random mixing. While contact is likely to have some heterogeneity, which is not captured by random mixing, it is unlikely to be a strict grid as used by Turner et al. In fact Rhodes and Anderson (2008) found that random mixing is a reasonable assumption for sufficiently large populations, which gives us confidence in our approach.

The technique presented here, then, allows us to relate the dynamics of the population as a whole to the behaviour of individual members of the population. Models developed in other ways generally describe the population at one of these levels, and attempts to relate the two rely, for instance, on fitting equations to simulation results. By rigorously deriving population level equations from the description of the population in terms of individuals we can be sure that the equations we obtain are a direct consequence of the individual level behaviour described. This is an important step forward in our understanding of disease transmission dynamics.

The work presented here is at the proof of concept stage. Work is under way to explore the consequences, at the population level, of introducing more complex interaction rules at the individual level, where those population-level consequences cannot be predicted with confidence in advance. For example, transmission within spatially heterogeneous populations, transmission with superspreaders (Kemper 1980) and transmission between subgroups generally. It may be expected that as more biological detail is added to the individual level descriptions of the system, the resulting equations will become more complex than the simple equations of Section 4.

At the heart of our work is the rigorous derivation of population level system dynamics from the individual-based process algebra model. This automatic conversion between scales allows each perspective to be used in the strongest way possible. In particular, individual observations can inform the individual-based model. Use of process algebra brings additional benefits. The model is composed in a modular fashion of smaller agents. The model can be explored directly through stochastic simulation. Algebraic analysis of the model via standard process algebra manipulations is also possible. At the population level, the emergent behaviour of the system is revealed. The equations are well justified since they are derived from local interactions and therefore based on explicit rather than implicit assumptions. The MFE

are amenable to further standard mathematical numerical and algebraic analyses.

Moving between scales is a challenging general theoretical problem. It is important because we wish to understand how large scale properties emerge from small scale interactions. Using process algebra we can experiment with different interactions and observe their impact on the system dynamics.

Acknowledgements This work was supported by the Engineering and Physical Sciences Research Council through a Doctoral Training Grant (CM, from 2004-2007), and through *System Dynamics from Individual Interactions: A process algebra approach to epidemiology* (EP/E006280/1, all authors, 2007-2010).

References

- Anderson RM, May RM (1981) The population-dynamics of micro-parasites and their invertebrate hosts. *Philos T R Soc B* 291:451–524
- Baeten JCM (2005) A brief history of process algebra. *Theor Comput Sci* 335(2/3):131–146
- Begon M, Bennet M, Bowers RG, French NP, Hazel SM, Turner J (2002) A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiol Infect* 129:147–153
- Bernardo M, Degano P, Zavattaro G, editors (2008) *Formal Methods for Computational Systems Biology*, volume 5016 of *Lect Notes Comput Sci*. Springer Berlin
- Boots M, Sasaki A (2001) Parasite-Driven Extinction in Spatially Explicit Host-Parasite Systems. *Am Nat* 159:706–713
- Bracciali A, Brunelli M, Cataldo E, Degano P (2008) Synapses as stochastic concurrent systems. *Theor Comput Sci* 408:66–82
- Briggs CJ, Godfray HCJ (1995) The dynamics of insect-pathogen interactions in stage-structured environments. *Am Nat* 145:855–887
- Calder M, Hillston J (2010) Process algebra modelling styles for biomolecular processes. In *Transactions on Computational Systems Biology XI*, volume 5750 of *Lect Notes Comput Sci*, pages 1–25, Springer Berlin
- Cardelli L (2008a) From processes to ODEs by chemistry. In *5th IFIP International Conference on Theoretical Computer Science (TCS 2008)*, volume 273 of *International Federation for Information Processing*, pages 261–281, Springer Boston
- Cardelli L (2008b) On process rate semantics. *Theor Comput Sci* 391:190–215
- Ciocchetta F, Hillston J (2010) Bio-PEPA for epidemiological models. In *Fourth International Workshop on Practical Applications of Stochastic Modelling (PASM09)*, volume 261 of *Electronic Notes in Theoretical Computer Science*, pages 43–69, Elsevier
- Cohen J (2008) The crucial role of CS in systems and synthetic biology. *Commun ACM* 51:15–18
- Cornell SJ, Isham VS, Grenfell BT (2004) Stochastic and spatial dynamics of nematode parasites in farmed ruminants. *P R Soc, B* 271:1243–1250
- Ellner A (2001) Pair approximation for lattice models with multiple interaction scales. *J Theor Biol* 210:435–447
- Fenton A, Fairbairn JP, Norman RA, Hudson PJ (2001) Parasite transmission: reconciling theory and reality. *J Anim Ecol* 71:893–905
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical-reactions. *J Phys Chem* 81(25):2340–2361
- Hatcher M, Tofts C, Dunn A (1995) The effect of the embryonic bottleneck on vertically transmitted parasites. *Proceedings of the 1st Conference on Information Processing in Cells* 339-353
- Hillston J (2005) Fluid Flow Approximation of PEPA models. In *QUEST’05, Proceedings of the 2nd International Conference on*

- Quantitative Evaluation of Systems, pages 33–42, IEEE Computer Society Press, Torino
- Hochberg ME (1991) Non-linear transmission rates and the dynamics of infectious disease. *J Theor Biol* 153:301–321
- Joo J, Lebowitz JL (2004) Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercubic lattice. *Phys Rev E*, 70:036114
- Kemper JT (1980) Identification of superspreaders for infectious-disease. *Math Biosci* 48:111–127
- Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics I. *P R Soc A*, 115:700–721
- Kitano H (2002) Systems Biology: A Brief Overview. *Science* 295:1662–1664
- Knell RJ, Begon M, Thompson DJ (1998) Transmission of *Plodia interpunctella* granulosis virus does not conform to the mass action model. *J Anim Ecol* 67:592–599
- Kurtz TG (1970) Solutions of ordinary differential equations as limits of pure jump markov processes. *J Appl Probab* 7:49–58
- Matsuda H, Ogita N, Sasaki A, Sato K (1992) Statistical mechanics of population - the lattice Lotka Volterra model. *Prog Theor Phys* 88:1035–1049
- McCaig C (2007) From individuals to populations: changing scale in process algebra models of biological systems. PhD thesis, University of Stirling <http://hdl.handle.net/1893/398>
- McCaig C, Norman R, Shankland C (2008) Deriving mean field equations from large process algebra models. Technical Report CSM-175, Department of Computing Science and Mathematics, University of Stirling <http://hdl.handle.net/1893/1584>
- McCaig C, Norman R, Shankland C (2009) From individuals to populations: A symbolic process algebra approach to epidemiology. *Math Comp Sci* 2(3):139–155
- Milner R (1980) A Calculus of Communicating Systems, volume 92 of *Lect Notes Comput Sci*. Springer-Verlag
- Norman R, Shankland C (2003) 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 *Lect Notes Comput Sci*, pages 404–414, Springer Berlin
- Nowicki P, Bonelli S, Barbero F, Balletto E (2009) Relative importance of density-dependent regulation and environmental stochasticity for butterfly population dynamics. *Oecologia*, 161:227–239
- Priami C (2006) Process calculi and life science. *Electronic Notes in Theoretical Computer Science*, 162:301–304
- Railsback SF, Harvey BC (2002) Analysis of habitat-selection rules using an individual-based model. *Ecology*, 83:1817–1830
- Regev A, Silverman W, Shapiro E (2001) Representation and simulation of biochemical processes using the pi-calculus process algebra In: *Pacific Symposium of Biocomputing 2001 (PSB2001)*, vol. 6, 459–470
- Rhodes CJ, Anderson RM (2008) Contact rate calculations for a basic epidemic model. *Math Biosci* 216:56–62
- Sumpter DJT, Broomhead DS (2001) Relating individual behaviour to population dynamics. *P R Soc B*, 268:925–932
- Tofts C (1993) Using process algebra to describe social insect behaviour. *T Soc Comput Simul*, 9:227–283
- Tofts C (1994) Processes with probabilities, priority and time. *Form Asp Comput*, 6:536–564
- Turner J, Begon M, Bowers RG (2002) Modelling pathogen transmission: the interrelationship between local and global approaches. *P R Soc B*, 270:105–112
- Webb SD, Keeling MJ, Boots M (2007a) Host-parasite interactions between the local and the mean-field: How and when does spatial population structure matter? *J Theor Biol*, 249(1):140–152
- Webb SD, Keeling MJ, Boots M (2007b) Spatially extended host-parasite interactions: The role of recovery and immunity. *Theor Popul Biol* 71, 251–266
- Wing J (2006) Computational thinking. *Commun ACM*, 49:33–35

A WSCCS

The Weighted Synchronous Calculus of Communicating Systems (WSCCS) is one of many process algebras based on Milner’s seminal Calculus of Communicating Systems (CCS) (1980). For our work, the main difference between CCS and WSCCS is the addition of weighted, or probabilistic, choice. A full introduction to the notation of WSCCS can be found in Tofts’ introductory paper (1994). Here we use examples from our models to give a brief overview of the WSCCS notation. Biological models in WSCCS utilise various different types of agents. Complex behaviours within the system can be constructed by having agents interact with each other and by allowing different simple behaviours to happen at different stages.

Consider the agent $S3$, which is written

$$S3 \stackrel{\text{def}}{=} p_d.\checkmark : 0 + (1 - p_d).\checkmark : S1 .$$

This says that $S3$ will become the agent $S1$ with weight $(1 - p_d)$, or become the null agent 0 with weight p_d . In each case it performs the action \checkmark , which can be thought of as representing a step of time, though this is not equal to any particular unit of time and tick actions need not be uniformly distributed in time. It is convenient to consider agents which can perform only tick actions as representing a probabilistic choice between the different outcomes. In general weights in WSCCS define the relative frequencies of the different outcomes but for probabilistic agents such as $S3$ we choose the weights to be between 0 and 1 with the sum of the weights equal to 1 .

Another type of agent which must be considered is a parallel agent. The probabilistic agent

$$I1 \stackrel{\text{def}}{=} p_b.\checkmark : I2 \times Nb2 + p_c.\checkmark : I2 \times Trans + (1 - p_b - p_c).\checkmark : I2$$

can become the agent $I2$ and also the parallel agents

$$I2 \times Nb2 , \tag{7}$$

and

$$I2 \times Trans . \tag{8}$$

The agent in (7), which consists of an $I2$ agent in parallel with an $Nb2$ agent, represents an infected individual giving birth to a newborn individual. On the other hand the agent in (8), which consists of an $I2$ agent in parallel with a $Trans$ agent, represents two different aspects of an infected individual’s behaviour (absorbing an infectious contact and making an infectious contact). In the next stage the $Trans$ agent will always become the null agent 0 so that the infected individual is once more represented by a single agent.

The agent

$$Popn \stackrel{\text{def}}{=} S1\{s\} \times I1\{i\} \times R1\{r\}[\{\checkmark\}]$$

is another type of parallel agent, which defines an initial population with which to analyse the system, consisting of s $S1$ agents, i $I1$ agents and r $R1$ agents. The $[\{\checkmark\}]$ forces communicating agents to cooperate on all actions other than \checkmark .

Interaction between agents forms an important part of agent behaviour in WSCCS. For example the agent $Trans$,

$$Trans \stackrel{\text{def}}{=} \omega.\overline{\text{infect}} : I3 + 1.\checkmark : I3,$$

will either perform the $\overline{\text{infect}}$ action or perform a tick, in either case becoming $I3$. The special weight ω (which is larger than any real number weight) means that this agent must perform the $\overline{\text{infect}}$ action when available. However, the restrictions on the actions which are possible in $Popn$ mean that $Trans$ can only

perform the \overline{infect} action if an agent is available to perform a complementary action. For example the agent $S2$,

$$S2 \stackrel{\text{def}}{=} \omega.infect : SI3 + 1.\sqrt{} : S3$$

can perform the action $infect$ which would allow $Trans$ to perform \overline{infect} . If this happens the $S2$ agent will become an $SI3$ agent, indicating that it has made an infectious contact. Otherwise the $S2$ agent will perform a tick and become an $S3$ agent. The action $infect$ is thought of as the input action while \overline{infect} is the output action.

The probabilities with which agents interact is influenced by the number of agents of the same type in the population, as well as the number of agents available to perform a complementary action. For example if there are no agents in the population available to perform an \overline{infect} action then an $S2$ agent would be unable to perform the input action and would be forced to perform a tick action.

Our models are designed in such a way that interaction and probabilistic choice happen in different stages. This means that behaviours in the models happen over multiple consecutive stages with behaviour being averaged over these stages to represent one time step in the MFEs. By separating probabilistic choice and communication it is easier to write the models and correctly determine the MFEs for the system.

Previous work introduced notation for functional parameters in WSCCS (McCaig 2007; McCaig et al. 2009). The models considered here make use of this notation to incorporate density dependent behaviour. For example the probability of birth in both models is density dependent:

$$p_b \stackrel{\text{prob}}{=} p_{b_0} - k(\lfloor S1 \rfloor + \lfloor I1 \rfloor + \lfloor R1 \rfloor), \quad (9)$$

where $\lfloor S1 \rfloor$, $\lfloor I1 \rfloor$ and $\lfloor R1 \rfloor$ are respectively the numbers of $S1$, $I1$ and $R1$ agents present in the population. When deriving MFEs functional parameters are incorporated by substituting the functional form of the parameter into the equations. For (9) this means substituting for

$$\begin{aligned} p_b &= p_{b_0} - k(S1_t + I1_t + R1_t), \\ &= p_{b_0} - kN_t. \end{aligned}$$