Developing the Use of Process Algebra in the Derivation and Analysis of Mathematical Models of Infectious Disease

R. Norman and C. Shankland

Department of Computing Science and Mathematics, University of Stirling, UK.
{ces,ran}@cs.stir.ac.uk

Abstract. We introduce a series of descriptions of disease spread using the process algebra WSCCS and compare the derived mean field equations with the traditional ordinary differential equation model. Even the preliminary work presented here brings to light interesting theoretical questions about the "best" way to defined the model.

1 Motivation

Moving from individual to population level processes is a key challenge of biological theory. Within the life sciences community there is an increasing need to model various systems at a level where individuals (whether cells or whole organisms) interact and from this to be able to determine the behaviour of a group or population of those individuals. Individual variation is likely to be important in, for example, predicting disease spread, explaining social insect behaviour or predicting the effect of new drugs at the molecular level. Systems of non-linear coupled differential equations are often used to describe the spread of infectious diseases from a population level [1]. Mathematical techniques are available to analyse algebraically the long-term behaviour of these models. However, once the level of realism in these models increases the mathematical techniques currently in use often cannot give a full picture of the system dynamics. Further, this approach often assumes population level behaviour is understood without taking individual behaviour, and hence variation, into account. Developing an analytical framework for individual based models is necessary if we are to understand the contribution of the individual to population level behaviours. This is particularly important in situations where the observation of the biological system takes place at the level of the individual.

One commonly used modelling approach that explicitly simulates a collection of individuals is the use of probabilistic cellular automata [2, 3]. These are often grid-based models in which each cell within the grid consists of an individual that moves and interacts with its neighbours according to a strict set of rules. If we then look at the individuals summed across the whole grid we can infer population level behaviour. This method does, therefore, fit the desired biological criteria stated above. However, the method does have some limitations as it is usually simulation based with no algebraic analysis possible. Although simulation is useful, it is not computationally feasible to explore the whole parameter space using simulation therefore there is a risk that some important behaviour has been missed. Recent innovative work on cellular automata has allowed some algebraic progress to be made. The method of pair approximations has been used to allow some rules of general behaviour of the system to be determined [4]; however, these *approximate* the *spatial* relationships within a population and do not include the individual processes explicitly.

A method of modelling biological problems which is gaining increasing recognition is the application of theoretical computing science techniques, such as Process Algebra. Process algebra allows description of the rules of behaviour and interaction for an individual. Individuals are loosely coupled together to describe a system. Typically, process algebras are used to describe concurrent distributed computer systems, such as communications protocols, but some have already been applied to biological systems. For example, the discrete time process algebra Weighted Synchronous Calculus of Communicating Systems (WSCCS) [5] has been useful in describing dynamics of social insect populations [6, 7], supported by use of the tool Probability Workbench (PW). Process algebra allows creation of an individual-based model, but with the advantage of a formal mathematical semantics, thereby allowing formal analysis. Both long and short term behaviours of the population as a whole may be rigorously explored by examining the Markov chain resulting from the description, or by derivation of a mean field model (MFE) [7, Chapter 3]. The tool PW supports these activities to some extent, as well as providing simulation facilities, which are useful for initial confirmation that the model behaves as expected. The fact that several different and powerful mathematical analyses are possible means that process algebra has a huge potential for allowing novel and innovative approaches to the modelling of biological systems.

The question of how those process algebra models relate to the traditional population level models has not yet been fully explored. In this paper we will concentrate on the use of process algebra, WSCCS in particular, to model systems of disease spread. We carry out numerical analyses of these systems using the Probability Workbench and manual probabilistic analysis. We compare the standard Ordinary Differential Equations (ODEs) to mean field approximations for population level behaviour derived from the WSCCS descriptions. This allows us to better understand the correlation between individual and population level behaviour. Simply concentrating on the average behaviour of these systems can answer important biological questions and bring to light interesting questions about using process algebra to express biological systems. By beginning with such simple, well-understood systems we also gain a better understanding of the relationship between the rules of individual behaviour as expressed in process algebra, and the derived MFEs. A particular biological problem we plan to address in future work is described in Section 4, and our conclusions are made in Section 5.

2 Background

2.1 Population level model

Consider a system of disease spread in which the population is split into three groups of individuals: susceptible (S), infected (I), and recovered (R). In this simplest case we assume no births or deaths, and no deaths due to the disease. We are therefore considering an epidemic in a closed population in which the only things that can happen are transmission of the disease from I to S at rate β, and recovery from I to R at rate γ. The most commonly used discrete time form of the equations which describe the system dynamics are as follows:

$$\begin{aligned} S_{t+1} &= S_t - \beta S_t I_t \\ I_{t+1} &= I_t + \beta S_t I_t - \gamma I_t \\ R_{t+1} &= R_t + \gamma I_t \end{aligned} \tag{1}$$

See the biological problem presented in Section 4 for more detail on this particular formalisation. This model will be used as a reference point for the process algebra descriptions of Section 3.

2.2 WSCCS notation

For a full introduction to the notation of WSCCS see Tofts' introductory paper [5]. We give here an overview of the notation, using excerpts of the examples of Section 3. The biological implications of these will be discussed in full in the next section. The notation used is the ASCII variant of WSCCS accepted by the tool Probability Workbench.

As with other process algebras, *agents* describe behaviours. Agents are composed using the key operations of action prefix, choice, and parallelism. Actions are named events that we wish to observe, for example, "send a message", "receive a message", "toss a coin and get heads", "toss a coin and get tails", "pass on an infection", and so on. The special action *tick*, written `t`, simply allows time to pass. There is no notion of real time in WSCCS, but there is a notion of ordering. For example, the following PW input describes an agent `S2`:

```
bs S2 1.infect:I1 + 1.t:S1
```

`S2` can perform the action `infect` and subsequently behave like `I1` (i.e. this individual has become infected), or `S2` can perform the action `t` and subsequently behave like `S1` (i.e. this individual stays susceptible). `bs` is an instruction to PW to direct it to interpret what follows as a basic sequential agent definition.

Since WSCCS is a probabilistic process algebra, the choice can be affected by *weights*. Above, the weights before the actions are both 1, meaning that each branch is equally likely (although this can be modified by communication between agents, see below). In the descriptions of Section 3 weights less than

1 are used, adding up to 1 across a choice, so weights can be thought of as probabilities.

For example, the following PW input describes an agent `I1`:

```
bs I1 pr.t:R2 + pa.t:T2 + (1-pr-pa).t:I2
```

`I1` can evolve in three ways. With probability `pr`, a tick action occurs and `I1` subsequently behaves like the agent `R2`. With probability `pa`, a tick action occurs and `I1` subsequently behaves like the agent `T2`. Finally, with probability `(1-pa-pr)`, a tick action occurs and `I1` subsequently behaves like the agent `I2`.

More interesting behaviour arises when a number of agents are composed in parallel, allowing interaction between agents. The following expression introduces an agent `W2`:

```
bpa W2 I2|Trans
```

Agent `W2` is composed of the two agents `I2` and `Trans` running in parallel. The term `bpa` is an instruction to PW that this is a basic parallel process.

Communication between agents arises when an action and its *inverse* occur:

```
bs Trans 1.infect^-1:I1 + 1.t:I1
```

The `infect^-1` action is a partner to the `infect` action of `S2` above. The composition of `infect` and `infect^-1` is a tick action. Communication is strictly one to one.

The following expression introduces an agent `Population`:

```
btr Population S1|S1|S1|S1|I1/L
```

Population is composed of four instances of the agent `S1` and one `I1`. The `/L` at the end is *restriction*. This means that only actions in the set `L` are permitted to occur. In the models of Section 3 `L` is defined to contain only the action `t`. By only allowing tick actions to occur we disallow solitary `infect` and `infect^-1` actions. In Section 3 only the agents are described explicitly. To complete each model a parallel expression describing `Population` as above is required. The term `btr` is an instruction to PW that this is a agent consisting of processes in parallel, with restriction, and *priority* occuring.

Priority behaves like an infinite weight on actions. For example:

```
bs T2 1@1.infect^-1:I1 + 1.t:I1
```

Here the communication action has priority over the tick action (signified by the `1@`). Practically, this means that if there is another process with which to communicate, then `T2` must perform the `infect^-1` action. The alternative branch can only be taken if there are no other processes with which to communicate.

The features of WSCCS are common to many probabilistic process algebras. We have chosen to work initially with WSCCS because of its historical association with biological problems, and the availability of tool support.

2.3 Overall Picture of our Research Plan

Our long-term research programme addresses three open questions: Can we improve the mathematical models which describe infectious diseases by comparing and combining the individual and population approaches? Can process algebras yield more tractable analyses than the differential equation based models for realistic models of disease spread? How does the ability to model explicit rules of behaviour and interaction of individuals, permitted by the process algebra, affect model results? This paper describes our initial investigations into these questions.

Our approach is illustrated in Figure 1. The focus of our work is developing methods of deriving accurate Mean Field Equations (MFEs) from the process algebra (the dashed arrow), the comparison of the two sets of MFEs obtained (the thick double headed arrow), and the comparison of the results with real world data.

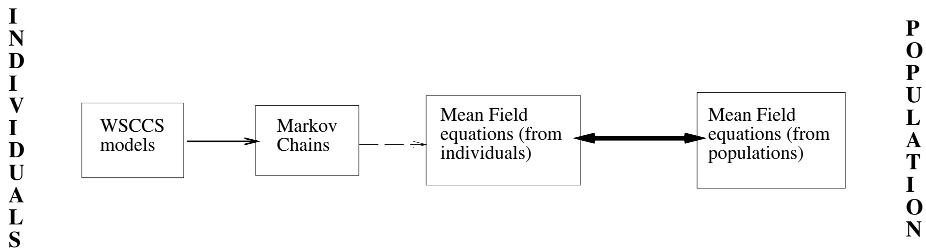

Fig. 1. The Relationship between the Process Algebra and PDE Approach

Sumpter [7, Section 3.4] described a heuristic for deriving MFEs from the process algebra specification which we adopt here. We expect that, in general, more realistic rules of individual behaviour will result in different population level models to those traditionally used. This may additionally enhance our understanding of disease systems, and allow us to reformulate the assumptions made in modelling at the population level.

3 Individual based models in WSCCS

We will now describe the development of an individual based model using WSCCS which gives mean field equations equivalent to those derived at the population level (1). We start with an existing model [7], and then develop it to increase its biological realism and to make its MFE more similar to the traditional model (1). At this point we are refining our method and do not expect that the individual based model should be different from the population based model. At each stage interesting theoretical and biological questions arise.

In the initial model [7, Section 3.4] agents are either infected, susceptible or infected. The WSSCS description of the model is in Figure 2. Here, as in all the models, there are two steps. First the choice to either recover, actively try to infect, or do nothing is taken. Second the actual communication which produces infection is made. This separation of concerns is essential. A naive model might have both of these in one step; however, the communication mechanism of WSSCS may distort the effect of the weights on the choice, thereby giving an incorrect model.

In Figure 2 at each tick an infected individual recovers with probability p_r, is infectious with probability p_a, or simply remains infected with probability $1 - p_r - p_a$. Susceptible individuals are available to be infected with probability p_c and remain susceptible with probability $1 - p_c$. The use of priorities on the infect actions means that the maximum number of contacts will occur between those agents who are available to infect ($p_a I_t$, where I_t is the number of `I1` processes at time t) and those who are available to be infected ($p_c S_t$, where S_t is the number of `S1` processes at time t). Therefore, after two ticks (one time step in the MFE) there will be n new infected individuals, where n is the minimum of $p_a I_t$ and $p_c S_t$.

```
bs I1 pr.t:R + pa.t:T2 + (1-pr-pa).t:I2
bs I2 1.t:I1
bs T2 1@1.infect^-1:I1 + 1.t:I1
bs S1 pc.t:P2 + (1-pc).t:S2
bs S2 1.t:S1
bs P2 1@1.infect:I1 + 1.t:S1
bs R1 1.t:R2
bs R2 1.t:R1
```

Fig. 2. Version 1: Sumpter 2000: some susceptibles available for infection

In this case the mean field equations are as follows. We use the method of derivation described by Sumpter in his thesis [7, Chapter 3].

$$
\begin{aligned}
S_{t+1} &= (1 - p_c)S_t + max\{p_c S_t - p_a I_t, 0\} \\
I_{t+1} &= (1 - p_r)I_t + min\{p_a I_t, p_c S_t\} \\
R_{t+1} &= R_t + p_r I_t
\end{aligned}
\tag{2}
$$

These are clearly not in the same form as the population level equations (1) because of the maximum and minimum terms. In addition, the model of Figure 2 is not very realistic biologically since it is unlikely that susceptibles will independently become available for infection.

Our first attempt to make the model more realistic was to make *all* susceptibles available for infection every time step. Clearly this means a change in interpreting the role of p_a, since now this is the only probability contributing to infection. The WSCCS notation for this model is given in Figure 3. The probability p_r is as described for the previous case.

```
bs I1 pr.t:R2 + pa.t:T2 + (1-pr-pa).t:I2
bs I2 1.t:I1
bs T2 1@1.infect^-1:I1 + 1.t:I1
bs S1 1.t:S2
bs S2 1@1.infect:I1 + 1.t:S1
bs R1 1.t:R2
bs R2 1.t:R1
```

Fig. 3. Version 2: all susceptibles available for infection

The mean field equations are as follows:

$$\begin{aligned} S_{t+1} &= S_t - min\{S_t, p_a I_t\} \\ I_{t+1} &= (1-p_r)I_t + min\{p_a I_t, S_t\} \\ R_{t+1} &= R_t + p_r I_t \end{aligned} \tag{3}$$

Depending on the value selected for p_a and the number of susceptibles and infecteds, the system evolves by infecting exactly $p_a I_t$ at each step, until the final step when the min terms become relevant and the last few susceptibles are infected. However, this model is very unsatisfactory biologically since many infections are guaranteed every time step, whereas in reality, as more individuals in the population become infected and recovered, it will be more difficult to find susceptibles and so infection should become more unlikely.

Biologically, we therefore require some rules that mean that infectious individuals can "waste" the chance to infect by communicating with infected or recovered individuals in a given time step as well as susceptibles. Our third model is given in Figure 4.

```
bs I1 pr.t:R2 + pa.t:T2 + (1-pr-pa).t:I2
bs I2 1@1.infect:I1 + 1.t:I1
bs T2 1@1.infect^-1:I1 + 1.t:I1
bs S1 1.t:S2
bs S2 1@1.infect:I1 + 1.t:S1
bs R1 1.t:R2
bs R2 1@1.infect:R1 + 1.t:R1
```

Fig. 4. Version 3: allowing infection attempts to be "wasted" on infecteds and recovereds

The ability to communicate on the `infect` action has been added to both `I2` and `R2` processes. This means that, as desired, there is now some probability that when a `T2` process makes its `infect^-1` action, the agent it is communicating with will not be a susceptible, and no new infection will result.

The mean field equations are:

$$\begin{aligned} S_{t+1} &= S_t - (p_a I_t S_t)/(S_t + (1-p_a)I_t + R_t) \\ I_{t+1} &= (1-p_r)I_t + (p_a I_t S_t)/(S_t + (1-p_a)I_t + R_t) \\ R_{t+1} &= R_t + p_r I_t \end{aligned} \tag{4}$$

An obvious problem with this model is that although infections can be passed to more of the population than in the model of Figure 3, one class is missed out from the communication, that of the transmitting infecteds (`T2` processes).

There are at least two ways to add communication between infectious individuals to this model using WSCCS. Both sets of notation give the same simulation results and the same mean field equations; however, what is interesting is whether there are other theoretical implications to the representation, what the advantages and disadvantages of each notation are, and whether one is a more efficient representation for calculation, for example in PW, than the other. Other differences may appear when other forms of analysis, such as Markov chain analysis, are carried out, but this is beyond the scope of the present work.

In the fourth version, given in Figure 5, all individuals may communicate with an infectious individual, regardless of whether they are susceptible, already infected, or recovered (and therefore immune). In particular, we want to express that even other infectious individuals can communicate with each other. The mechanism for this is slightly clumsy. An infectious individual is represented by a pair of processes in parallel: one to do the `infect^-1` action (`I2`), and the other to possibly do the `infect` action (`Trans`). Note that this merely presents an additional opportunity to "waste" an infection; we do not enforce that the `Trans` process *has* to communicate with the particular `I2` agent it has been paired with. Recall that this agent will also run in parallel with all the other individuals in the systems, therefore it could easily communicate with any of the others, including the susceptibles. This notation introduces an extra agent `Trans` into the system. These are removed (by becoming the null agent `T` which is automatically absorbed) after a single `infect^-1` or tick action. They are not counted twice. This notation, given in Figure 5, is more explicit since we are determining what happens to each individual.

The alternative way to describe this is by taking the model of Figure 4 and relinquishing the use of priority in communications. This exploits the way that actions occur in WSCCS and the underlying semantics of the language. The notation, seen in Figure 6, is somewhat more concise and elegant. Instead of forcing communication to occur wherever possible, we allow an equally weighted choice between communicating and not communicating. As above, new infections are not guaranteed, and depend on p_a, and the number of `S2`, `I2`, and `R2` processes.

```
bs I1 pr.t:R2 + pa.t:T2 + (1-pr-pa).t:I2
bs I2 1@1.infect:I1 + 1.t:I1
bpa T2 I2|Trans
bs Trans 1@1.infect^-1:T + 1.t:T
bs S1 1.t:S2
bs S2 1@1.infect:I1 + 1.t:S1
bs R1 1.t:R2
bs R2 1@1.infect:R1 + 1.t:R1
```

Fig. 5. Version 4: Explicit notation: forced communication, all available to infect

```
bs I1 pr.t:R2 + pa.t:T2 + (1-pr-pa).t:I2
bs I2 1.infect:I1 + 1.t:I1
bs T2 1.infect^-1:I1 + 1.t:I1
bs S1 1.t:S2
bs S2 1.infect:I1 + 1.t:S1
bs R1 1.t:R2
bs R2 1.infect:R1 + 1.t:R1
```

Fig. 6. Version 5: Implicit notation: all available to infect, except `T2`

As stated above the two sets of notation give the same simulation results. The first one contains more agents (although temporary) which may be important when models become more complicated, but is more explicit and the probabilistic rules of interaction are clearer in this case. The second set of notation relies more on the underlying rules of WSCCS and may be more efficient both to write down and to simulate but relies on a deeper understanding of WSCCS and so may be more open to errors, particularly if non-experts start to use these methods.

Given both sets of notation above the mean field approximation is as follows:

$$\begin{aligned} S_{t+1} &= S_t - p_a I_t S_t/(S_t + I_t + R_t) \\ I_{t+1} &= (1 - p_r)I_t + p_a I_t S_t/(S_t + I_t + R_t) \\ R_{t+1} &= R_t + p_r I_t \end{aligned} \tag{5}$$

We have now lost the minimum and maximum terms and this model is now equivalent to the population level model (1) with $\gamma = p_r$ and $\beta = p_a/(S_t+I_t+R_t)$ since $S_t + I_t + R_t$ is constant in this model.

This is an important breakthrough in the development of these models since we have now made the connection between the population models and the individual-based models. This allows us a strong base from which to start addressing other important biological questions.

4 Future Directions: A Biological Problem

Models of infectious disease spread have been used successfully for a number of years to predict the dynamics and control of a variety of both human and animal diseases. In particular they have been used increasingly to determine government policy in the control of diseases such as BSE, and Foot and Mouth disease. A large body of work in this area has built up, particularly since the ground breaking work of Anderson and May in the early 1980s [1]. However, the majority of this work is based on population level behaviour and assumes a priori that we know what is happening to the population as a whole.

One common assumption of these models that has come under question recently is the way in which the spread of infection is modelled. The majority of these models assume that the rate at which susceptible individuals become infected is proportional to the density of susceptible and infectious individuals and is written as βSI where S is the density of susceptibles, I is the density of infectious individuals and β is the transmission rate. When transmission is written in this form it is known as *density dependent transmission*.

An alternative term which is beginning to be used is *frequency dependent transmission*. This assumes that the number of contacts an individual has per day is independent of the density of individuals around it, and has traditionally been used to describe sexually transmitted diseases or vector borne diseases. There is however an increasing body of evidence that it might be more appropriate than density dependent transmission in systems in which animals live in social groups and mainly interact with their own groupings, for example, rabbits in a warren. Frequency dependent transmission appears in the equations as the term $\beta' SI/N$ where N is the total population density.

As stated above, these transmission terms assume that we know what is happening at the population level whereas in reality it is more likely that, experimentally, we have observations on how individuals interact. We are interested, firstly in what rules of individual behaviour would give rise to density or frequency dependent transmission terms at the population level, and ultimately what realistic rules of individual behaviour would mean in terms of population level transmission.

In this paper we have laid the ground work to allow these questions to be investigated. Here, we examined a simple disease system in which the population is closed, i.e. there are no births and deaths. It should be noted that in a closed population, density and frequency dependent transmission are just scalings of one another since N is constant. The WSCCS descriptions of Figures 5 and 6 seem to capture frequency dependent transmission.

5 Conclusions and Future Work

Mathematical models of infectious disease systems are used to address specific and important problems in the dynamics, persistence and control of those systems. Here we have brought together two contrasting modelling approaches: a population based approach from mathematics and an individual based approach from computing science, and applied them to epidemiological problems.

We have examined a simple SIR system (with a closed population), comparing the use of process algebra (WSCCS) with the use of differential equations. Our main aim was to test the use of process algebras in describing infectious disease systems since they may provide an accessible method of deriving individual based models that may be analysed algebraically to yield population level results. Initial studies have shown process algebra to be an intuitive and useful way to describe disease systems. To fully investigate the questions posed above, we need to consider more complex models with decreasing levels of algebraic tractability (from the viewpoint of traditional analyses). We are also considering the use of other process algebras for this work; in particular a continuous time model could be more realistic, and certainly be more comparable with traditional models. A suitable formalism in which to write such models would be the Performance Evaluation Process Algebra (PEPA) [8].

We have shown here that it is possible to chose rules of individual behaviour that give us the same population level behaviour as the established ODE approach. At this point we are establishing the theoretical basis for future work, which will throw into contrast the different assumptions made in each modelling exercise as more complex systems are considered. Our studies have raised theoretical questions about the best way to write the rules down. For example, the need to separate choice from communication is clear, while the use of the priority operator to force communication seems at present to be optional (depending on the form of the other rules). These questions are likely to become more important as we try to add more realism.

The next step is to add births and deaths to this simple system. This is where the interesting dynamics lie. This will require formulation of explicit rules giving density dependent transmission since the total population will no longer be constant. We also intend to carry out further analysis, for example exploring the Markov chain behaviour.

References

1. Anderson, R., May, R.: The population dynamics of microparasites and their invertebrate hosts. Philosophical transactions of the Royal Society **291** (1981) 451–524
2. Rand, D., Keeling, M., Wilson, H.: Invasion, stability and evolution to criticality in spatially extended, artificial host-pathogen ecology. Proceedings of the Royal Society of London Series B **259** (1995) 55–63
3. Sato, K., Matsuda, H., Sasaki, A.: Pathogen invasion and host extinction in lattice structured populations. Journal of Mathematical Biology **32** (1994) 251–268
4. Boots, M., Sasaki, A.: 'Small worlds' and the evolution of virulence: infection occurs locally and at a distance. Proceedings of the Royal Society of London Series B **266** (1999) 1933–1938
5. Tofts, C.: Processes with probabilities, priority and time. Formal Aspects of Computing **6** (1994) 536–564
6. Tofts, C.: Describing social insect behaviour using process algebra. Transaction of the Society for Computer Simulation (1993) 227–283

7. Sumpter, D.: From Bee to Society: an agent-based investigation of honeybee colonies. PhD thesis, UMIST (2000)
8. Hillston, J.: A Compositional Approach to Performance Modelling. Cambridge University Press (1996)