Measles Epidemics and PEPA: An Exploration of Historic Disease Dynamics Using Process Algebra

Soufiene Benkirane, Rachel Norman, Erin Scott, and Carron Shankland

University of Stirling, Stirling UK
ces@cs.stir.ac.uk
http://www.cs.stir.ac.uk/SystemDynamics/

Abstract. We demonstrate the use of the process algebra PEPA for realistic models of epidemiology. The results of stochastic simulation of the model are shown, and ease of modelling is compared to that of Bio-PEPA. PEPA is shown to be capable of capturing the complex disease dynamics of the historic data for measles epidemics in the UK from 1944–1964, including persistent fluctuations due to seasonal effects.

1 Introduction

According to the World Health Organization [26], in 2002, about 19.1% of worldwide deaths, and 52.7% of deaths in Africa were caused by infectious and parasitic diseases. Understanding a given disease, to prevent, cure or reduce its impact, is inherently a multidisciplinary endeavour, involving medicine, geography, sociology and biology, but also, through modelling, mathematics and computer science. Most epidemiological modelling has been mathematical; for example, using Ordinary Differential Equations (ODEs) [1]. Formal methods, traditionally used for computer science, are beginning to be more widely used to construct computational models of disease [3,17,18,7]. Process algebras are designed to describe a system of interacting autonomous agents and allow study of emerging collective dynamics (the epidemic). This is in contrast to the typical mathematical biology approach which is forced to make assumptions about how interaction leads to population-level effects. Using process algebra, those effects are generated by the underlying semantics of interaction.

We present a novel case study in using PEPA (Performance Evaluation Process Algebra) [14] for epidemiology. The discussion of general principles of modelling disease spread is focussed through application to the specific example of measles dynamics in England and Wales between 1944 and 1964. The emergent behaviour of measles is complex, involving recurrent outbreaks in small populations and cyclic outbreaks in larger populations [4]. Adequate modelling requires a number of features common to other diseases, including transmission, population growth, seasonality and immigration. Moreover, a large data set is freely available of the number of reported cases of measles in England and Wales over more than 20 years. This allows the model to be validated.

D. Giannakopoulou and D. Méry (Eds.): FM 2012, LNCS 7436, pp. 101–115, 2012.

Our group has long experience of applying process algebra to epidemiology. Initially WSCCS [24], a CCS inspired process algebra, was used by Norman and Shankland [21] to model basic transmission mechanisms. Further work with McCaig incorporated the essential features of population growth [18] and showed the advantages of the approach over traditional styles of epidemiological modelling [17]. PEPA has also been used by our group [3]. Few other groups have made extensive study of epidemiology using process algebra beyond simple examples. A notable exception is the study of contact network structure and avian influenza by Ciocchetta and Hillston [7] using Bio-PEPA [6]. Their models are on a closed population, and do not include seasonal behaviour.

This paper is organised as follows. A brief introduction to PEPA is given in Section 2 and to measles at the start of Section 3. Modelling of measles dynamics raises generic issues for modelling epidemiology: these are discussed in the context of PEPA in Section 3. The complete model appears at the end of that section. No new language features are introduced: our contribution is to test the expressivity of PEPA for realistic models of epidemiology. Stochastic simulations of the model are compared in Section 3.2 with data available on the University of Cambridge website [25]. Since Bio-PEPA has features especially designed for biology, Section 3.3 gives details of how to model measles dynamics in Bio-PEPA and a comparison with PEPA. Conclusions are drawn in Section 4 regarding the benefits and limitations of PEPA as a modelling tool for epidemiology.

2 PEPA

PEPA [14] has been used to study the performance of a wide variety of systems [15]. PEPA has a small set of combinators, allowing system descriptions to be built up as the concurrent execution and interaction of simple sequential components which undertake actions. We informally introduce the syntax required for the model of Section 3 below. More detail can be found in [14].

Prefix: $(\alpha, r).P$ carries out action α at rate r, behaving subsequently as P. In PEPA actions have a duration, or delay. Thus the expression $(\alpha, r).P$ denotes a component which can undertake an α action, at rate r defining an exponential distribution (where rate is 1/delay) to evolve into a component P.

Choice: $P + Q$ represents a system which may behave either as P or as Q.

Constant: $X \stackrel{def}{=} E$ assigns the name X to the pattern of behaviour E.

Cooperation: $P \underset{L}{\bowtie} Q$ denotes cooperation between P and Q over L. The *cooperation set* L determines those activities on which the *cooperands* are forced to synchronise. For action types not in L, the components proceed independently and concurrently with their enabled activities. $P \parallel Q$ abbreviates $P \underset{\{\}}{\bowtie} Q$.

Unlike some other stochastic process algebras, PEPA assumes *bounded capacity*: a component cannot be made to perform an activity faster by cooperation, so the rate of a shared activity is the minimum of the rates of the activity in the cooperating components. In some cases, when an activity is known to be carried out in cooperation with another component, a component may be *passive* with

respect to that activity. This means that the rate of the activity is left unspecified (denoted $\top$) and is determined upon cooperation by the rate of the activity in the other component. All passive actions must be synchronised.

3 Modelling Measles Dynamics in PEPA

Despite the worldwide efforts to vaccinate children against it, measles is still the vaccine-preventable disease of childhood that causes the most deaths [9]. Without vaccination, measles infects 95-98% of children before they turn eighteen [22], and an infectious person will infect 75%-90% of susceptible household contacts. The incubation period lasts for six to nine days [4], then measles symptoms begin with increasing fever, cough, coryza and conjunctivitis [22]. This is when the infectivity is highest. A rash then appears on the face and the neck, and may spread to the rest of the body. Usually, the individual recovers after six to seven days and then has lifelong immunity to the disease.

Measles has been extensively studied, giving an opportunity to show that formal methods can perform as well as long-established techniques of mathematical biology. A detailed set of data is freely available for measles in England and Wales between 1944 and 1964[1] giving an excellent opportunity to validate results from modelling. Certain key features are required to model disease transmission in general. We demonstrate that PEPA can be used to capture these, specifically in relation to the benchmark disease measles, despite not being designed for this purpose. Benkirane's thesis [2] identified the following key characteristics of measles epidemics:

Transmission of Disease. McCaig [19, page 137-138] showed in his work on epidemiology and WSCCS that in process algebra all forms of transmission can be reduced to either *direct* or *indirect* transmission. Direct transmission indicates that the disease is passed through host-to-host contact, as in measles. Indirect transmission uses an intermediary (e.g. air, surface, or via a vector). For human diseases it is common to assume the number of contacts an individual makes is constant and independent of population size. In ODE models this is known as frequency dependent transmission.

Births and Deaths. For anything other than a short timescale epidemic, the population must have births and deaths, immigration and emigration. In the case of measles, if these are not included the model shows a single initial epidemic, after which the whole population becomes resistant to the disease and there are no further outbreaks, in contradiction with observed behaviour.

Timed Events. Timed events are essential in many biological and epidemiological systems, whether these are one-off events (e.g. control measures such as

[1] In 1944, national notification of measles patients was made mandatory in England and Wales [11]. This provides the number and location of measles cases, with a reporting rate of over 50% [4]. Mass vaccination was introduced in 1968 [23], changing the landscape of the disease entirely. We use an earlier cut-off for data, due to some changes in administrative regions.

vaccination), or recurring events (e.g. circadian clock and seasonality). Seasonality is an essential feature of measles epidemics: disease dynamics are strongly dependent on transmission rate, which is in turn influenced by the aggregation of children at school [4]. In addition, immigration provides new infectious individuals to a city where the disease had faded out, especially in the case of small and isolated cities.

Structured Populations. The capacity to allocate the population to different categories can be essential in describing certain diseases. The structuring category can be diverse: age, social, hobbies, space, etc., depending on which feature has an important influence on the behaviour of the disease. In the case of measles dynamics, the spatial location of cities and their degree of connection impacts the overall behaviour of the disease. Smaller cities usually experience fade-outs of the disease, but still experience regular outbreaks, often within weeks of those in neighbouring cities.

Benkirane shows it is possible to code all of these features using PEPA [2], but due to limited space, only the first three will be illustrated here. The model presented is a representation of the city of Leeds from 1944 to 1964. The city initially has a population of 508,010, ten of whom are assumed to be infectious. The model is based on Bjørnstad et al. [4], apart from the immigration mechanism and parameters, which are inspired by Finkenstädt et al. [12]. In the following sections, modelling of the elements above will be discussed in turn.

3.1 Presentation of the Model

Transmission of Disease. The classic mathematical model of disease transmission is the SIR model, first described by Kermack and McKendrick [16] in 1927. SIR corresponds to *Susceptible*, *Infectious* and *Recovered* as follows:

Susceptibles represent the people that never had the disease, and may acquire it after exposure to the infection.

Infectious are the people who carry the disease, and may pass it (directly or indirectly) to *susceptible* individuals.

Recovered (or Removed) are immune to the disease. This might be because they have been infectious and recovered from it, because they have been vaccinated, or because they are naturally immune to the disease.

Additional classes may be necessary for particular diseases. For example, measles requires an *Exposed* class:

Exposed are infected *susceptibles* who are not yet infectious, i.e. they are undergoing an incubation period.

This general model of disease spread has been successfully applied to a wide range of different diseases. It is straightforward to encode these behaviours as separate process algebra agents: see the model of Figure 1, where the agents S, E, I and R respectively represent the *Susceptible*, *Exposed*, *Infectious* and

$$
\begin{aligned}
S &\stackrel{def}{=} (contact, \top).E \\
E &\stackrel{def}{=} (infected, ir).I \\
I &\stackrel{def}{=} (contact, \top).I + (recover, rr).R \\
R &\stackrel{def}{=} (contact, \top).R + (lose\_immunity, li).S
\end{aligned}
$$

$$
\begin{aligned}
I' &= (contact, cr).I' + (recover, \top).Rest; \\
Rest &= (infected, \top).I';
\end{aligned}
$$

$$
(S[990] \parallel I[10]) \underset{\{infected, contact, recover\}}{\bowtie} (I'[10] \parallel Rest[990])
$$

Fig. 1. Simple direct transmission in PEPA

Recovered individuals. Activities occur at rates controlled by the exponential variables *ir* (incubation), *rr* (recovery), *li* (loss of immunity), and *cr* (contact). Exponential rates are highly suitable for the first three, since these take place at a constant rate, and provide a reasonable approximation for contact behaviour.

The additional agents I' and *Rest* are required to implement direct transmission[2]. With standard PEPA syntax it is possible to have all I agents communicate with all S agents, or none, but not a single I communicating with a single S. One solution would be to introduce new operators capturing the desired behaviour. Instead, we wish to operate within the constraints of standard PEPA. To achieve direct transmission, it is necessary to have a main population of S and I who do not communicate directly with each other ($\parallel$) but who can communicate one-to-one with the "mirror group" of I' and *Rest*. The addition of the mirror group splits the *Infectious* functionality between two agents. On one hand, agent I, the infectious individual who can be passively contacted, or recover. On the other hand, agent I', the infectious individual actively contacting other individuals to pass on infection. This idea of where the driver of functionality lies is reflected in the choice to make *contact* passive in I but not in I'. To guarantee that the model remains consistent, the number of I and their mirror I' must be equal at all times. The mirror *Rest* agents have been added so that the mirror I' group can grow and shrink with the I population correctly. Note that *Rest* does not have to model all the behaviour of agents S, E and R: it only has to capture the movement from exposed to infectious. A more detailed discussion of direct transmission and the mirror group may be found in Benkirane's thesis [2].

Births and Deaths. PEPA is limited when expressing births and deaths because agents cannot be created or deleted: a more inventive approach must be adopted. Three different approaches to births and deaths are considered in Benkirane's thesis [2]: only the one adopted for his measles model is described here.

[2] Indirect transmission is straightforward. To agents S, E, I and R of Figure 1 add agents for the environment and a suitable system equation describing interaction [3].

A method to describe births and deaths arises naturally from the introduction of a reserve pool of *dormant* agents, as shown in this simple example:

$$Active \stackrel{def}{=} (death, death\_rate).Dormant$$
$$Dormant \stackrel{def}{=} (birth, birth\_rate).Active$$

The pool corresponds to available agents, ready to be activated when needed. Also, when an individual dies, it returns to the pool. The initial size of the pool has to be chosen carefully: too many means longer processing times for the model, too few and the pool might run out and the number of births will be blocked as long as it is empty, leading to unexpected behaviour.

In the complete measles model of Figure 2 a similar *Dormant* population is used, while the *Active* agents correspond to the *S*, *Exp*, *Inf* and *R* agents. All agents *S*, *Exp*, *Inf* and *R* can give birth and die naturally. Newborns are susceptible. This new behaviour is added to the mirror group via the *DS* agents.

Although not done for this reason here, the introduction of *Dormant* also gives a way to regulate population size. If the birth and death rate are roughly constant, the number of agents initially in *Dormant* can be chosen such that $Dormant + Active = K$ with K the carrying capacity of the population. This way, the population can never increase past the carrying capacity, and the overall number of births naturally decreases with the number dormant. Although in general this solution is biologically unrealistic, it may be useful for cases where the population does not fluctuate much. This is true for measles dynamics: the disease is not usually deadly, and the number of immigrants over the period is low compared to the total population. The number of *Dormant* here is selected to ensure that as many births or immigrations as required can take place.

Timed Events. Infectious immigrants to the city can start a new outbreak, if the timing is right[3]. Immigration is represented by a subgroup formed by a single component type, *Immigration*. The action *immigration* fires every $1/imrate$ time step and one agent in *Dormant* moves to the *Inf* state. This models the arrival of one infectious immigrant in the city. Note the asymmetry between births and immigration. Births are driven from the main group, while immigration is driven from its own subgroup. In PEPA, given $P = (\alpha, r).Q$ the rate r follows the cumulative distribution function of the exponential distribution $F_\alpha(t) = 1 - e^{-rt}$. That is, on average, α will be fired after $1/r$ time steps. The actual moment at which the intervention takes place varies. This is very suitable for immigration: the timing of immigration is not precise. It is not suitable for seasonality, where more control over the timing of events is required.

Still considering $P = (\alpha, r).Q$, an action firing at rate 1 has actually only a 38% probability of happening between 0.5 and 1.5 time steps. In order to reduce this variability, one solution is to split the action into several steps. In other

[3] Susceptible or recovered immigrants and emigrants do not influence the dynamics of the disease and are not modelled here. Similarly, infectious or exposed emigrants do not have any influence on the dynamics of the disease in the studied city, and their number is sufficiently low to have a negligible impact on population size.

words, an expression $(\alpha, r).Q$ is replaced by $(\alpha', r \times n).(\alpha', r \times n).....(\alpha', r \times n).Q$ with n the number of steps, that we will denote $(\alpha', r \times n)^n.Q$ for readability purposes. The distribution of the resulting chain of actions can be calculated using the following theorem [10]:

Theorem 1. *If X_1,X_2,...X_n are independent following an exponential distribution $Exp(\alpha)$ then $\sum_{i=1}^{n} X_i$ follows a Gamma distribution $Gamma(n, 1/\alpha)$*

As the rate of each of the activities is actually rn, the resulting probability density function is:

$$f_{n,r}(x) = \frac{x^{n-1}.e^{-rnx}.(rn)^n}{(n-1)!} \tag{1}$$

The cumulative density function can be simplified, in the special case where $n \in \mathbb{N}$ to the following expression:

$$F_{n,r}(x) = \sum_{i=n}^{\infty} \frac{(rnx)^i}{i!} e^{-rnx} \tag{2}$$

This formula allows us to estimate the probability of an event happening between time $t = a$ and $t = b$ (with $b > a$) as:

$$F_{n,r}(b) - F_{n,r}(a) = \sum_{i=n}^{\infty} \frac{(rnb)^i}{i!} e^{-rnb} - \sum_{i=n}^{\infty} \frac{(rna)^i}{i!} e^{-rna} \tag{3}$$

Thus, the number of steps n can be chosen to provide the modeller with what she deems an acceptable probability of an action occurring within the desired time. We use this technique to model seasonality using the following agent:

$$Summer \stackrel{def}{=} (go\_winter, n/summer\_duration)^n.Winter + (insummer, big).Summer$$
$$Winter \stackrel{def}{=} (go\_summer, n/winter\_duration)^n.Summer + (inwinter, big).Winter$$

The season agent has two roles: performing the chain of actions leading to a change in season (*go_winter* and *go_summer*), or broadcasting the current season (*inwinter* or *insummer*) to all agents for the whole season. The rate *big* is introduced as a practical proxy for $\top$ in simulations.

Seasonality affects measles dynamics through a varying contact rate. The average age of the infected individual according to data is low [20, Table II]: the disease is very infectious, and getting infected grants lifelong immunity. The average number of contacts a child makes change significantly depending on season, as more contacts are made when children go to school (in winter). The two seasons only affect Inf', the mirror component of Inf, which has been divided into Inf'_s for the summer, and Inf'_w for the winter. The difference between the two is contact rate (*crs* and *crw*). The model is composed of two seasons: a four month summer, and an eight month winter.

To explain the seasonality mechanism further, for example, once the season changes from *Summer* to *Winter*, *Winter* cooperates with the agents in Inf'_s (of

Figure 2) over the action *inwinter*, in order for them all to move to Inf'_w (also of Figure 2). The rate at which the cooperation is performed has to be very large compared to the other parameters of the model in order for the process to be considered instantaneous. In this model, this rate is at least $big/crs \approx 1.66 \times 10^8$ times bigger than any other rate in the model. The number of infectious individuals is always under 1000, so the whole operation takes less than $1000/big \approx 10^{-6}$ time step to be performed. Note that any immigrating individuals move to Inf'_s initially, but if the season is winter they will move almost instantaneously to Inf'_w. After the action *go_summer* has been fired n times, the agent describing the season moves from *Winter* to *Summer*, and the agents in Inf'_w are forced to move to Inf'_s. The choice of the value of n depends on the precision required by the modeller, as well as the processing time of the model.

Complete Model and Parameters. The measles model is presented in Figure 2, and the parameters used presented in Figure 3 (taken from the literature [4,12,25] except n, srs, srw) . In the final measles model of Figure 2 there are agents for S, *Exp*, *Inf*, R and *Dormant*, as above. There is not a direct mapping between agents in the main group and and the mirror group (S', Inf'_s, Inf'_w and DS). In particular, DS only models movement of births and $DS \neq Dormant$. Other *Dormant* behaviour is captured in S'. The model obeys two invariants concerning agent numbers: $Inf = Inf'_s + Inf'_w$, and $S + Exp + R + Dormant + Immi = S' + DS$.

$$
\begin{aligned}
Summer &\stackrel{def}{=} (go\_winter, srs)^n.Winter + (insummer, big).Summer \\
Winter &\stackrel{def}{=} (go\_summer, srw)^n.Summer + (inwinter, big).Winter \\
S &\stackrel{def}{=} (contact, \top).Exp + (birth, br).S + (die, dr).Dormant \\
Exp &\stackrel{def}{=} (contact, \top).Exp + (incubation, ir).Inf + (die, dr).Dormant + (birth, br).Exp \\
Inf &\stackrel{def}{=} (contact, \top).Inf + (recover, rr).R + (dieI, dr).Dormant + (birth, br).Inf \\
R &\stackrel{def}{=} (contact, \top).R + (birth, br).R + (die, dr).Dormant \\
Dormant &\stackrel{def}{=} (born, big).S + (immigration, \top).Immi \\
Immi &\stackrel{def}{=} (gotoInf, big).Inf \\
S' &\stackrel{def}{=} (incubation, \top).Inf'_s + (birth, \top).DS + (gotoInf, \top).Inf'_s \\
Inf'_s &\stackrel{def}{=} (contact, crs).Inf'_s + (recover, \top).S' + (inwinter, \top).Inf'_w + (dieI, \top).S' \\
Inf'_w &\stackrel{def}{=} (contact, crw).Inf'_w + (recover, \top).S' + (insummer, \top).Inf'_s + (dieI, \top).S' \\
DS &\stackrel{def}{=} (born, \top).S' + (timeout, 100.0).S' \\
Immigration &\stackrel{def}{=} (immigration, imrate).Immigration
\end{aligned}
$$

$$
\begin{aligned}
((S[508000] \parallel Inf[10] \parallel Dormant[100000] &\underset{\{born, birth, incubation, recover, contact, dieI, gotoInf\}}{\bowtie} \\
S'[608000] \parallel Inf'_w[10]) &\underset{\{immigration\}}{\bowtie} Immigration) \underset{\{insummer, inwinter\}}{\bowtie} Winter
\end{aligned}
$$

Fig. 2. PEPA measles model for Leeds

According to Bjørnstad et al. [4, p. 171], the critical community size, in order for the virus not to go extinct, lies between 300,000 and 500,000 in England and Wales. For this reason, the city of Leeds has been chosen to test this model: its

Parameter	Rate (per day)	Description
popn	508010	Total population size [25]
big	999999999	The fast rate used for immigration and seasonality
ir	1/7.5	Incubation rate is 1/incubation period [4]
rr	1/6.5	Recovery rate is 1/infectious period [4]
crw	39.1/6.5	Winter contact rate [4]
crs	19.8/6.5	Summer contact rate [4]
br, *dr*	0.017/360	Birth rate and Death rate [4]
imrate	$0.02 * \sqrt{popn}/360$	Immigration rate of infectious individuals [12]
n	96	Number of iterations of the change of season action
srw	$1/(8 \times 30) \times n$	Rate of one iteration of change season (winter)
srs	$1/(4 \times 30) \times n$	Rate of one iteration of change season (summer)

Fig. 3. The parameters used in the model in Figure 2

population in 1944 was about 508,000 inhabitants. The model starts with 100% susceptible individuals, and evolves naturally towards its steady state susceptible proportion of between 3.5 and 9%. This number of susceptibles is consistent with biological studies [4, p. 180]. The model displays transient behaviour while establishing the susceptible population. This has been empirically determined to correspond to the first thousand steps, and has been removed from the results shown in Figure 4 as it does not relate to the observed behaviour of measles.

Finkenstädt et al [12, p. 755] give immigration of infectious individuals as:

$$\text{average number of imports per year} = 0.02\sqrt{\text{population size}}$$

In the case of our model, it results in $0.02\sqrt{508000} = 14.25$ infectious imports per year. This results in a total of 285 immigrants across the twenty-one years studied, who increase the overall population by 0.056%. The birth and death rates are assumed to be constant by taking average figures for the period. While this was not the case in reality (the maximum number of births recorded was 10821 in 1947, and the minimum was 7584 in 1954), the difference can be considered negligible for the scope of this analysis.

The contact rate has been chosen based on the value of R_0 given in the paper by Bjørnstad et al. [4, p. 180]. R_0 is the number of successful contacts an infectious individual would make in an entirely susceptible population. According to that paper, its maximum value is 39.1 in December, and its minimum value is 19.8 in August. In the absence of an average value over the course of a season, or a monthly value, these values are used for the winter and summer season respectively. The daily number of contacts are derived in a standard way from R_0 by dividing it by the infectious period.

Finally, the parameters related to the seasons are *srw*, *srs* and n. Seasonality has been simplified by assuming that a month lasts 30 days, and a year 360 days. Winter has been assumed to last eight months. The choice of the value of n lies in the hands of the modeller: it must be chosen in order to give an acceptable variability in the season length, while not increasing the length of the simulations

too much. With $n = 96$, and using equation (3), the probability that the winter lasts between 7 and 9 months, and that the summer lasts between 3.5 and 4.5 months, is 78%, which is deemed sufficient for the scope of this study.

3.2 Results

Analysis of the model is performed through a series of single stochastic simulations. Due to variability in the timing of season change, stochastic simulations cannot be meaningfully averaged. Moreover, the analysis will only be performed on semi-quantitative factors, such as the length of the cycle between two consecutive outbreaks and the average size of the peak of each epidemic. Although Benkirane developed a tool to derive ODEs from PEPA models [2], this cannot be used here as the hypothesis behind the derivation, that the number of agents is large, is not met by the subgroups for seasonality and immigration (one instance each). Their effect of the main group would not be correctly captured in the derived ODEs. A benefit of process algebra (not explored here) is that additional analyses are possible via the PEPA plugin [13].

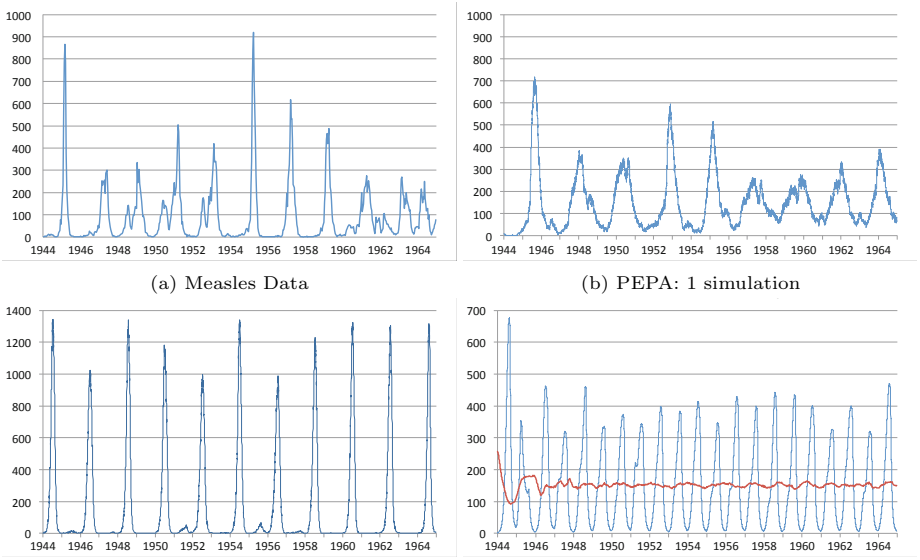

Fig. 4. Graphical results for Measles epidemics in Leeds between 1944 and 1964. The horizontal axis is year, and the vertical axis is number of infectious individuals.

Comparing the model with measured data is not straightforward: this is one of the problems of carrying out a realistic case study. The data to which the model is compared (city of Leeds only) is available from University of Cambridge [25]. The data must be normalised to match the format of the simulated data. The

simulated data reports number of infected individuals every day, while the field data is reported every 14 days. Each data point in the original data has been divided by 14 (to get the number of reported infectious individuals per day) and multiplied by 6.5 (to reflect the average infectious period of 6.5 days). Bjørnstad et al. [4, p. 172] mention a reporting rate across England and Wales of just over 50%. We assume every individual is infected at some point in her life, therefore, the number of births should correspond to the number of cases. For Leeds between 1944 and 1964 (inclusive), 181,539 births were recorded, while 109,730 cases of measles were reported. Assuming the birth rate was constant (as above) gives an average reporting rate of 60.4%. A constant reporting rate will be assumed. The number of infectious individuals at each data point are corrected by multiplying the reported number of cases by (number of births / reported cases). Figure 4(a) shows the corrected data.

A sample simulation of the model of Figure 2 is shown in Figure 4(b). It corresponds to 8560 days of simulations, where the first 1000 days have been removed, as described previously. The two graphs exhibit comparable behaviour. In both cases, measles outbreaks occur on a regular cycle. The length of the cycle varies between the collected data and the stochastic simulation. In the case of the data, the cycle lasts exactly two years, with very little variation over the length of the studied period, apart from a single one year cycle between the 1963 and the 1964 epidemics. The simulation on the other hand, has biennial cycles most of the time, but sometimes exhibits 2.5 year cycles. Across the studied period, the collected data experiences 11 outbreaks, for 9 in the simulation. As detailed by Bjørnstad et al. [4], the cycles are a consequence of the addition of seasonality to the model. The presence of a summer season where the contact rate is lower allows more time for the pool of susceptibles to increase in size, until the contact rate increases again in the winter. The number of infectious individuals at the peak of infection in both cases can vary a lot, between 260 and 920 in the case of the collected data, and 260 and 710 in the simulation. The average of the number of individuals at the peak of each outbreak confirms this difference: 479 infecteds in the case of the collected data, for 424 in the case of the simulations.

3.3 Comparison with Bio-PEPA

It can be seen from the discussion above that some epidemiological features, while they can be modelled in PEPA, are not naturally expressed. For example, neither direct transmission, nor births and deaths, are very elegantly expressed in PEPA. An obvious question arises: are these features better modelled in Bio-PEPA, which was designed for biological systems? Brief details of repeating the modelling exercise with Bio-PEPA are given here. The model was based on a combination of standard mathematical biology techniques [1,11] and the model of Figure 2. The Bio-PEPA model is shown in Figure 5. The syntax used here is the syntax of the Bio-PEPA tool [8]. Parameter values are as in Figure 3.

Transmission of Disease. Control is given to the modeller via the kinetic laws: any rate expressed via arithmetic and trigonometric functions can be

$$
\begin{array}{ll}
endWinter \stackrel{def}{=} 4; startWinter \stackrel{def}{=} 9; month \stackrel{def}{=} floor(time/30); & \\
season\_time & \stackrel{def}{=} H(((month - 12 * floor(month/12)) - endWinter) \\
 & \quad *(startWinter - (month - 12 * floor(month/12)))); \\
kineticLawOf\, birth & : br * (S + Exp + Inf + R); \\
kineticLawOf\, dieS & : dr * S; \\
kineticLawOf\, dieExp & : dr * Exp; \\
kineticLawOf\, dieInf & : dr * Inf; \\
kineticLawOf\, dieR & : dr * R; \\
kineticLawOf\, contact & : ((crw * S * Inf)/(S + Exp + Inf + R)) * (1 - season\_time) + \\
 & \quad ((crs * S * Inf)/(S + Exp + Inf + R)) * (season\_time); \\
kineticLawOf\, incubation & : ir * Exp; \\
kineticLawOf\, recover & : rr * Inf; \\
kineticLawOf\, immigration & : imrate;
\end{array}
$$

$$
\begin{array}{rl}
S = & (contact, 1) \ll +(birth, 1) \gg +(dieS, 1) \ll; \\
Exp = & (contact, 1) \gg +(incubation, 1) \ll +(dieExp, 1) \ll +(birth, 1)(.); \\
Inf = & (contact, 1)(.) + (incubation, 1) \gg +(dieInf, 1) \ll +(birth, 1)(.) \\
 & +(recover, 1) \ll +immigration \gg; \\
R = & (recover, 1) \gg +(birth, 1)(.) + (dieR, 1) \ll;
\end{array}
$$

$$Inf[10] <*> S[508000] <*> Exp[0] <*> R[0]$$

Fig. 5. Bio-PEPA measles model for Leeds

given. This is rather similar to the way in which mathematical biologists choose terms in their ODE models: the link with interacting processes is decreased. Thus incorporating direct transmission is no longer about designing the right sort of interaction, it is simply a matter of writing the commonly used term for frequency-dependent direct transmission in the kinetic law for *contact*.

Births and Deaths. In Bio-PEPA the style is to describe change to species numbers, where species are similar to agents in PEPA. In Figure 5 the species are *S*, *Exp*, *Inf* and *R*. For example, increasing population through birth is described by the event $(birth, 1) \gg$ and decreasing population through death by the event $(dieS, 1) \ll$ (increase or decrease being indicated by the direction of the arrows). The 1 in these events describes the change to the number of susceptibles, but not the rate. The rate is described by the appropriate kinetic laws for *birth* and *death*.

Timed Events. Time can be used explicitly in the model through the variable *time*, and thus can influence variables and hence kinetic laws. In Figure 5 the variable *season_time* switches between 0 and 1 (with the use of the built-in Heaviside function *H*) to indicate winter or summer respectively. This is then used in the kinetic law for *contact*.

Bio-PEPA, like PEPA, gives access to a range of analysis techniques through a tool: the Bio-PEPA plugin [8], which offers, for example, stochastic simulations,

interpretation as ODEs, translation to SBML, invariant inference and model-checking. The result of a single simulation of the Bio-PEPA model is given in Figure 4(c): the pattern of outbreaks is similar to the collected data of 1944–1964, although the peaks are significantly higher. In Bio-PEPA the switch between seasons happens on the same day every year, in every simulation. For interest, Figure 4(d) shows the average of one hundred simulations for both the seasonally switching contact rate (lighter line) and a fixed (crw) contact rate (heavier line). For the former, the average tends to peak lower, and annually: the period between epidemics varies from one to three years. For the latter, the long-term pattern shows a more steady rate of infection of around 150 individuals. Both of these simulations are shown starting after 1720 steps: the Bio-PEPA simulation takes slightly longer to stabilise than the PEPA simulations.

Bio-PEPA overcomes some of the feature-capturing problems of PEPA. Arguably the model of Figure 5 is simpler and more elegant: the species are no longer confused with modelling artefacts to handle one-to-one communication. Bio-PEPA also has limitations. The formulation of kinetic laws means rates no longer depend on interaction and semantics: they come from implicit assumptions the modeller has made about population-level dynamics. This is therefore rather similar to the standard mathematical biology approach.

4 Conclusions

PEPA and Bio-PEPA models have been constructed to reflect cyclic epidemics of measles, and their output compared to collected data. While PEPA is not ideally suited to capturing all features of disease progression, suitable approximations can be made. An important feature is that the population dynamics emerge from the specified individual behaviour. In contrast, the Bio-PEPA model may be simpler, but required high-level assumptions to be made about population dynamics. In both cases, existing well-developed tools [13,8] were used for analysis. The results are promising: the simulated results for both PEPA and Bio-PEPA are comparable to the collected data and would allow meaningful exploration of patterns of epidemics under different parameter regimes. The PEPA results are closer to the collected results than the Bio-PEPA results, which demonstrate too much regularity. Differences between our results and the data may be associated with the granularity of modelling. For example, birth and death rates have the same value throughout the simulation. Seasonality has been approximated by splitting the year into two seasons each with a single contact rate which does not change throughout the season. The value used by Bjørnstad et al. [4, Fig. 7, p. 178] varies noticeably within each season. The models presented here could be altered to reflect these changes, with varying degrees of difficulty.

One of the difficulties encountered in this study, and encountered in any realistic modelling exercise, is the problem of parameter values. For example, the main source for this model was Bjørnstad et al. [4] who give incubation rate and infectious period as 7.5 and 6.5 days respectively. As shown in Section 3.2 this gives a good match to the measured data. Our models have also been tested

with data from Bolker et al. [5] who propose an incubation period of ten days, and infectious period of 3.7 days. This gives fewer disease outbreaks than shown by the data (3-4 years between outbreaks, with peaks between 600-1500 cases). The beauty of modelling is that the parameter choices can be easily explored. A further difficulty of dealing with collected data for measles is estimating the reporting rate. The approximation used might impact the number of infectious individuals at a given time, but would not influence cycle duration.

Due to lack of space, the influence of the nearby cities have been completely ignored in this study. Benkirane [2] has developed a novel extension to PEPA to allow structured populations to be easily expressed, and demonstrates its use through a more complex model of measles in the linked cities of Cardiff, Newport, Bristol and Bath. Similarly, Bio-PEPA has compartments which allow spatial elements of epidemiology to be modelled.

Process algebra has been shown here to be useful in modelling quite complex infectious disease systems. Determining which approach is suitable for a given problem depends on which questions we wish to answer about that problem; that is, the sort of analysis we wish to carry out. An advantage of process algebra over traditional mathematical biology is the range of automated analyses available.

Acknowledgments. This work was carried out under the EPSRC award *System Dynamics from Individual Interactions: A process algebra approach to epidemiology* (EP/E006280/1, 2007-2010), in consultation with Mike Begon, School of Biology and Biological Sciences, University of Liverpool. The authors thank the PEPA and Bio-PEPA Plug-in development team at the University of Edinburgh, for help, particularly Mirco Tribastone, Adam Duguid and Allan Clark. Finally, we thank the anonymous reviewers for their helpful comments.

References

1. Anderson, R.M., May, R.M.: The population-dynamics of micro-parasites and their invertebrate hosts. Philosophical Transactions of the Royal Society of London Series B 291, 451–524 (1981)
2. Benkirane, S.: Process algebra for epidemiology: evaluating and enhancing the ability of PEPA to describe biological systems. Ph.D. thesis, University of Stirling (2011), `http://hdl.handle.net/1893/3603`
3. Benkirane, S., Hillston, J., McCaig, C., Norman, R., Shankland, C.: Improved Continuous Approximation of PEPA Models through Epidemiological Examples. In: From Biology to Concurrency and Back, FBTC 2008. ENTCS, vol. 229, pp. 59–74. Elsevier (2008)
4. Bjørnstad, O.N., Finkenstädt, B.F., Grenfell, B.T.: Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological Monographs 72(2), 169–184 (2002)
5. Bolker, B., Grenfell, B.: Space, persistence and dynamics of measles epidemics. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences 348(1325), 309–320 (1995)
6. Ciocchetta, F., Hillston, J.: Bio-PEPA: A framework for the modelling and analysis of biological systems. Theor. Comput. Sci. 410(33-34), 3065–3084 (2009)

7. Ciocchetta, F., Hillston, J.: Bio-PEPA for epidemiological models. Electronic Notes in Theoretical Computer Science 261, 43–69 (2010); Proceedings of Practical Application of Stochastic Modelling (PASM 2009)
8. Duguid, A., Gilmore, S., Guerriero, M.L., Hillston, J., Loewe, L.: Design and development of software tools for Bio-PEPA. In: Proc. of Winter Simulation Conference 2009, pp. 956–967 (2009)
9. Duke, T., Mgone, C.S.: Measles: not just another viral exanthem. The Lancet 361, 763–773 (2003)
10. Durrett, R.: Probability: Theory and Examples. Cambridge Series in Statistical and Probabilistic Mathematics (2010)
11. Fine, P.E., Clarkson, J.A.: Measles in England and Wales–I: An analysis of factors underlying seasonal patterns. Int. Journal of Epidemiology 11(1), 5–14 (1982)
12. Finkenstädt, B.F., Keeling, M., Grenfell, B.T.: Patterns of density dependence in measles dynamics. Proceedings of the Royal Society B 265, 753–762 (1998)
13. Gilmore, S., Tribastone, M., Duguid, A., Clark, A.: PEPA plug-in for eclipse (2008), homepages.inf.ed.ac.uk/mtribast/plugin/
14. Hillston, J.: A Compositional Approach to Performance Modelling. Cambridge University Press (1996)
15. Hillston, J.: Tuning systems: From composition to performance. The Computer Journal 48(4), 385–400 (2005); The Needham Lecture Paper
16. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics. Proceedings of the Royal Society of London A 115, 700–721 (1927)
17. McCaig, C., Begon, M., Norman, R., Shankland, C.: A rigorous approach to investigating common assumptions about disease transmission: Process algebra as an emerging modelling methodology for epidemiology. Theory in Biosciences 130, 19–29 (2011); special issue on emerging modelling methodologies
18. McCaig, C., Norman, R., Shankland, C.: From individuals to populations: A symbolic process algebra approach to epidemiology. Mathematics in Computer Science 2(3), 139–155 (2009)
19. McCaig, C.: From individuals to populations: changing scale in process algebra models of biological systems. Ph.D. thesis, University of Stirling (2008), http://hdl.handle.net/1893/398
20. Miller, D.L.: Frequency of complications of measles, 1963. British Medical Journal 2, 75–78 (1964)
21. Norman, R., Shankland, C.: Developing the Use of Process Algebra in the Derivation and Analysis of Mathematical Models of Infectious Disease. In: Moreno-Díaz Jr., R., Pichler, F. (eds.) EUROCAST 2003. LNCS, vol. 2809, pp. 404–414. Springer, Heidelberg (2003)
22. Perry, R., Halsey, N.: The clinical significance of measles: A review. Journal of Infectious Diseases 189(1), S4–S16 (2004)
23. The Medical News: Measles history, http://www.news-medical.net/health/Measles-History.aspx
24. Tofts, C.: Processes with probabilities, priority and time. Formal Aspects of Computing 6, 536–564 (1994)
25. University of Cambridge: Pathogen population dynamics (2002), http://www.zoo.cam.ac.uk/zoostaff/grenfell/measles.htm
26. World Health Organization: The world health report 2004 (2004), http://www.who.int/whr/2004/annex/topic/en/annex_2_en.pdf