Large-Scale Meta-Population Patch Models of Infectious Diseases on Cray machines

Iain Barrass, Health Protection Agency
Kevin Roy, Manchester Computing
Steve Leach, Health Protection Agency

ABSTRACT: Estimating the spatial spread of an infectious disease through a mobile human population is an important aspect in guiding public health policy in the event of a large scale outbreak. In this paper we consider ensembles of large stochastic meta-population patch models for the spread of an infectious disease in the United Kingdom parameterised by realistic population size and movement data. In particular, we consider the use of a Cray XT3 for these types of model at varying levels of model detail. For each case we conclude that high performance computing can be a significant tool in assessing the impact of a newly emerging disease and any available mitigation strategies.
KEYWORDS: Infectious diseases and public health; message passing; scalability; meta-population; compartmental.

1 Introduction

Infectious diseases can have a significant impact on the health of large numbers of people. The possible emergence of a pandemic strain of human influenza, for example, is currently of great concern. Pandemic influenza has been estimated to be responsible for the deaths of between 20 and 100 million people in the last century alone (in 1918–19, 1956–57 and 1968–69) [1, 2]. The introduction of the SARS virus in East Asia emphasised the possibility of rapid worldwide spread of an infectious agent [3]. In addition, there are fears of the deliberate re-introduction of previously eradicated diseases such as smallpox [4].

Modelling the spread of an infectious disease is an important step in guiding public health policy with regards to the mitigation of their effects. Commonly used models describe the transmission of a disease by considering the contacts of individuals. With the assumption of *homogenous mixing* any one person is as equally likely as any other to make contact with—and be subsequently infected by—an infected person. These models, through their simplicity, can quickly give an indication of the likely toll of such a disease when certain parameters are known.

However, particularly when considering large populations such as occur when modelling the global spread of a disease, or even spread within a country, the assumption of homogenous mixing can quickly break down. Models which incorporate spatial structure thus become important at such scales. Recently developed models have been agent-based, such as in [5], where detailed information is stored about all of the individuals in the population including their disease state and interactions, or consider meta-populations [6]. In this paper we consider a meta-population patch model for the spread of infectious disease. At a local scale the homogenous mixing assumption is retained, but consideration is paid to the long-range interactions that are presented. This model formulation has the advantage of being more easily parameterised and more computationally tractable than corresponding agent-based models. The model structure chosen is described in Section 2.

All of these model types rapidly develop complications beyond those of simple models, and attention must be paid to developing methods to solve them. In Section 3 we describe a parallel algorithm suitable for use on distributed memory machines. This method is implemented and run on a Cray XT3. In addition, we target an Opteron cluster and compare the performance against that of the Cray machine.

To give a more reliable indication of the likely behaviour of transmission it is necessary to include stochastic effects. There is likely random activity in the contacts made by individuals, infection between individuals, as well as in the disease progression in a single host. With

these stochastic models it is necessary to take an ensemble of realisations large enough to limit the effect of extreme, but rare, events. In addition, it is necessary to in some way quantify the "average" behaviour of the dynamic system. Typically, the median of the results at large spatial scale is used to provide indicative behaviour, but a detailed description of the output analysis is beyond the scope of this paper.

2 The model

In this paper we consider a meta-population patch model of the spatial dynamics of an infectious disease in the United Kingdom [7]. The patches we consider are taken to be administrative regions of the constituent countries with population sizes as given by the 2001 Census [8, 9, 10]. Connections between the patches are determined by the *Travel to Work* statistics of the Census, which reflects the movement of people living in one patch and working in another. At each level the geographic coverage of the constituent countries (England, Wales, Scotland and Northern Ireland) is complete. Here we use the administrative levels of electoral wards, a total of 10608 patches, and districts (426 patches).

Each day of the model is split into two parts, *daytime* and *nighttime*. During daytime steps individuals are taken to be active, with regard to the disease dynamics, in the patch in which they work; during nighttime steps individuals are active in their home patch. Each patch, therefore, has a *working population* during the daytime stage, and a *resident population* during the nighttime stage. Those individuals who live and work in the same patch will contribute to both populations of that patch.

Through this population movement there is the possibility of the spatial spread of an infectious disease: a person may be infected through contact with an infectious person in their work patch and take this infection with them to their home patch.

The dynamics of the disease progression are taken to be a compartmental SEIR-type structure [11]. That is, an individual could be described initially as *susceptible* to the disease. On infection they would become *exposed* (also known as being in the latent period of the disease progression), in which state they remain for a period of time known as the latent period of the disease[1]. Upon conclusion of the latent period, transition is undertaken to an infectious state, whereupon there is the possibility of infection of those in their own susceptible state. After the infectious period of the disease, the individual will transition to a *removed* state, where they are deemed to no longer be infectious or liable to reinfection.

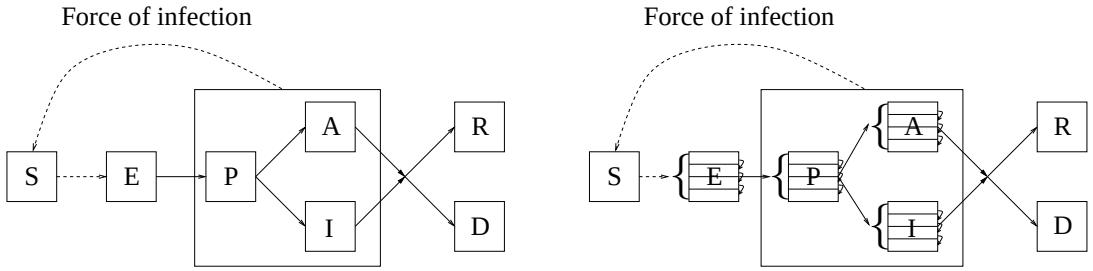
During the infectious period of the disease an individual would make contact with others and the number of resulting secondary cases would, on average in an otherwise completely susceptible population, be basic reproduction number, R_0, of the disease [12].

In this paper we use a slight variant of this compartmental structure which utilises more compartments. Firstly, we create an extended SEIR-type model, introducing a *prodromal*, P, stage, which occurs after the latent period but before the infectious period. During this prodromal period the individual may be infectious, but possibly less so than those individuals in the classical infectious stage, and will express no symptoms or less severe symptoms than later in the disease progression. In addition, we split the infectious compartment in two, to create a symptomatic infectious compartment, I, and an asymptomatic infectious compartment, A, where one individual would be either symptomatic or asymptomatic during their infectious stage. Finally, the removed state is split into a *recovered*, R, state and a *dead*, D, state. Disease progression in this framework is characterised as shown in Figure 1 and the full population model is created by tracking the proportions of the population in each disease compartment and examining transition rates for the populations. Each compartment is therefore described by a single number and indicative equations for the transitions can be found in, for example, [7].

Secondly, we consider a *pseudo-individual* model by extending the concept of a compartment to chart the history of disease progression for each individual in the E, P, I or A states. Upon transition to any one of these compartments the length of time the individual is to spend in that state is taken as a sample from a distribution [13]. Individuals are then grouped by their compartment and the departure time of that compartment. The number of steps for a given compartment is determined by the disease kinetic parameters. For smallpox the latent, prodromal and infectious periods are taken to be 12, 2.5 and 8.6 days, respectively, with an R_0 value of 5 [4], whereas for pandemic influenza we take periods 2, 1 and 1.5 days with $R_0 = 1.8$ [14].

Each of these extensions to the basic SEIR-type model can be further complicated by additional compartments to reflect an increasing complexity of disease kinetics. For example, compartments could be added to detail possible intervention strategies such as vaccination and hospitalisation [7]. Such scenarios, and an analysis of the simple and pseudo-individual models, and their

[1] The latent period of a disease, and other such parameters, can be approximately determined by epidemiological studies.



(a) Simple (b) Pseudo-individual

Figure 1: The compartment based progression of disease for each individual in the basic SEIR-type model and the pseudo-individual model. The boxed compartments contribute to the force of infection acting on the susceptible population. Compartmental transition is indicated by the solid arrows and in the case of the pseudo-individual model transition to the compound E, P, I and A compartments is governed by a probability distribution.

comparison, form part of ongoing work and, for the sake of clarity, are not further discussed here.

The full meta-population patch model contains many groups of these population SEIR-type structures. Each patch is described by a number of subpatches, with a subpatch of a patch describing the resident population of that patch commuting to distinct working patches (including the population living and working in the same patch). In this way, a patch of an n-patch system connected by commuter movements to every other patch will be described by n subpatches, whereas a patch connected to only one other will be described by 2 patches. The overall size of the meta-population patch model is therefore $\mathcal{O}(n^2 \sum C_i)$ where $\{C_i\}_{i=1}^{N}$ are the sizes of the N compartments in the SEIR-type structure.

In practice, it is not the case that every patch is connected to every other patch—at large geographic distances commuting between patches decreases—and so the number of subpatches per patch varies. As can be seen in Section 3.2 there can be significant variation in the connectedness of individual patches. The total number of subpatches at the level of electoral wards (10608 patches) is $1,550,819$, rather than $112,529,664$ (1.4%). For the district level there are 82899 subpatches (45.7%).

3 A parallel algorithm

At large problem sizes, the implementation of the model becomes impractical to be run on desktop machines. For a disease such as smallpox with long latent, prodromal, and infectious periods, the pseudo-individual model will require a large number of steps in those compartments. With a timestep of a quarter day the pseudo-individual model we use has the equivalent of around 300 compartments for each of the $1,550,819$ subpatches at the electoral ward level. As the timestep decreases, or the number of compartmental disease states increases, the total number of pseudo-individual compartments increases.

The greater availability of cluster or high performance computing in a public health policy setting leads naturally to a consideration of an implementation of this model with distributed memory. With this approach, the distribution of patches, and the passing of messages related to their interaction, becomes key to the efficiency with which the model can be run. Even when the problem size is not very large—so that the model implementation may be run on a desktop machine, or a single node of a cluster—the requirement to address multiple stochastic realisations to obtain representative behaviour means that it may be desirable to split each individual realisation over many processes. Instead of using many processes to run an ensemble of realisations (one model run per process, for example) simultaneously, splitting the model over many processes allows for an accumulation of results over time. The increasing availability of results over time may be of great help in a public health setting at a time of a response to a developing situation, such as will be the case with a new introduction of an infectious agent, in preference to a wait for a final, more accurate answer. The final cost of the overall answer in this manner will be greater than that obtained by splitting the ensemble. The relative costs will be related, again, to the efficiency of the message passing and the load balance obtained. In this paper we therefore consider the splitting of smaller problems to both elucidate the message passing ideas incorporated into the model implementation and to conclude that a many-process approach may be obtainable without too much cost to the benefit of those in the public health policy arena.

3.1 Disease transmission and message passing

We begin splitting the meta-population patch model described above by allocating individual patches to processes. For fine spatial scales, the number of patches in the model will generally be much greater than the number of processors available in the computer. Further, as we shall see in Section 3.2, our load balancing requirements are such that we will wish to assign many patches to each process, even when the number of patches is of the same order as the number of processors available.

It is important to note that, aside from the external effect of disease importation to a patch, all steps in the calculation for the disease progression in a patch are determined completely by information held on the process owning the patch. This external effect is characterised by a *force of infection* and is determined by the population sizes of the P, A, and I compartments of all of the subpatches within a patch (the contributing subpatches to a given patch will differ during the daytime and nighttime steps of the model).

A naïve message passing implementation will have each process tracking the movement of subpatches from patch to patch across the daytime–nighttime change and reconstruct patches each time. However, it is clear that, for many-compartment models, the description of the subpatches is of much larger size than that of the force of infection that is required to describe the interaction between the various patches.

More efficient is to pass only information regarding the force of infection. Here, all subpatches are retained by the process holding the home patch to which they belong. During nighttime steps the dynamics progress using the aggregation of the force of infection of subpatches within their own home patch. As all of these subpatches are held locally, no remote communication is required.

During daytime steps, however, information is required from remote processes regarding the contribution from patches to the force of infection to which each subpatch is exposed. The overall force of infection is retrieved by the passing of two messages. In the first round of message passing, each process sends details of the contribution to the force of infection for all subpatches to the corresponding host processes of the daytime patches. The forces of infection of all subpatches in the same working patch are then aggregated and this information then distributed to the process holding the subpatches affected by this force. A schematic for these messages is shown in Figure 2. Following this aggregation, the dynamics on the various patches can be computed.

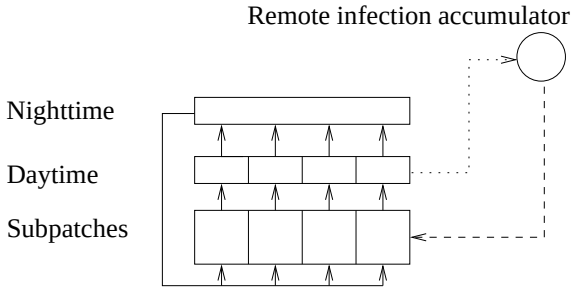


Figure 2: Dynamics of the meta-population patch model in a single patch of a four-patch system showing the differing interaction of subpatches during daytime and nighttime steps. Only one remote aggregation of subpatches is shown with outgoing messages indicated by dotted lines and incoming messages by dashed.

This aggregation of the force of infection acts as a synchronisation barrier in the computation. The alternative implementation of passing subpatch information so that a process works on alternatively day and night patch populations contains two such synchronisations and with a significantly greater volume of passed information.

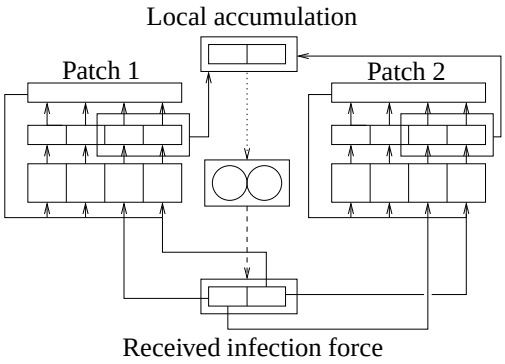


Figure 3: Message passing with many patches per process. In this four-patch example two processes own two patches each and messages are passed between them. Shown is the local action by one process only with the local aggregation of the forces of infection for subpatches 3 and 4 from each patch.

3.2 Patch distribution and load balancing

The final step in the construction of the model implementation is to assign home patches to processes. The current implementation of the meta-population patch model described above uses only static distribution of patches amongst processes. That is, each patch is assigned to a process prior to the commencement of the simulation where it remains throughout. The layout of these patches

is key to the efficiency with which the simulation runs as it directly relates to the load balancing of the system.

The only significant communication between processes during the simulation arises during the population movement. Assignment of patches to processes can focus on optimising this communication or optimising the load balance of the computation stages. The relative cost of the communication is decreased as the number of compartments in the subpatches is increased. To this end we consider the strategy of arranging patches so that the number of subpatches allocated to each process is close to that allocated to all the others. This is motivated by the understanding that the amount of computation each process is required to undertake is roughly proportional to the number of subpatches. The distribution of patches in this way is a bin packing problem and in general it will not be possible that each process is assigned exactly the same number of subpatches: for 10608 patches and $1,550,819$ subpatches the range of subpatches-per-patch is 3 to 445, and assignment to 16 processes sees the range of subpatches-per-process to be $193,736$ to $193,872$. As the process count varies, so does the load imbalancing as measured by the difference between the maximum and the minimum of the subpatch allocation size. Further, there is additional variation in this distribution as the structure of the connection matrix changes.

When there are multiple patches per process, message passing can be made more efficient by local accumulation of forces of infection where a process holds information regarding the same subpatch on a remote process. That is, the forces of infection for two subpatches (in distinct home patches) targeting the same working patch may be aggregated by the owning process before communication. In addition, only one copy of the return global aggregation is required. This local aggregation is demonstrated in Figure 3.

To examine the efficiency of this splitting of the model and message passing we create scalability curves for the model at the level of electoral wards using both the simple SEIR-type compartments and the pseudo-individual compartments. These curves are shown in Figure 4 for runs on Palu—a Cray XT3 with 1664 dual-core Opteron processors and SeaStar interconnect, owned by the Swiss National Supercomputing Centre—and Iluvatar—a 88 core Opteron cluster with Gigabit Ethernet interconnect, owned by the Health Protection Agency.

For the simple SEIR-type model scaling is quite poor for increasing CPU utilisation. However, as the computational demands increase with the use of the pseudo-individual model, the scaling becomes good for high processor counts. Further, for the smallpox model, with the highest computational load, the scaling is best. In all cases, the scaling for Palu is in line with Iluvatar. However, Palu retains good scaling for process counts higher than are possible on the smaller machine.

Results (not shown) for the very small problem having 426 district patches show very poor scaling for all process counts. As stated before, this is mostly a result of the limitations arising from the small number of patches: using 128 processes to share 426 patches, for example, leads to a very high load imbalance, even with optimal packing.

Similar scalability curves can be seen when using an Origin 3800.

4 Discussion and further work

In this paper we have examined a compartmental SEIR-type meta-population patch model for the spread of infectious disease. We have shown that with large pseudo-individual models the scalability of the algorithms described is good, and this holds on a range of machines. As these models increase in complexity in the future, by the incorporation of further compartments or more patches (either by increasing spatial resolution or the encompassing of greater areas, such as for the US, Europe or the whole world), our results suggest that increased scalability will be attained.

For small problems, the advantage of good scalability may not be immediately apparent. Indeed, the ensemble nature of the problem from a public health viewpoint suggests that the most efficient usage of many processes would be to spread the individual realisations across those processes. For very small problems, such as the district level models, this would be the optimal approach. However, with the efficiency of this algorithm, we have seen that even for the moderately sized problems presented here the costs of splitting the realisations so that there is more rapid attainment of early, indicative, results is outweighed, from an emergency response viewpoint, by the availability of these results. This is more keenly felt when the runtime of a realisation becomes significant.

However, the model can be extended in ways which complicate the message passing that is required. At present, the only interaction between the patches is in the movement of population. Model details which increase the dynamic interaction between patches, such as consideration of policies which act on groups of patches, can be introduced and these could serve to reduce the efficiency of splitting meta-population patch models in this way.

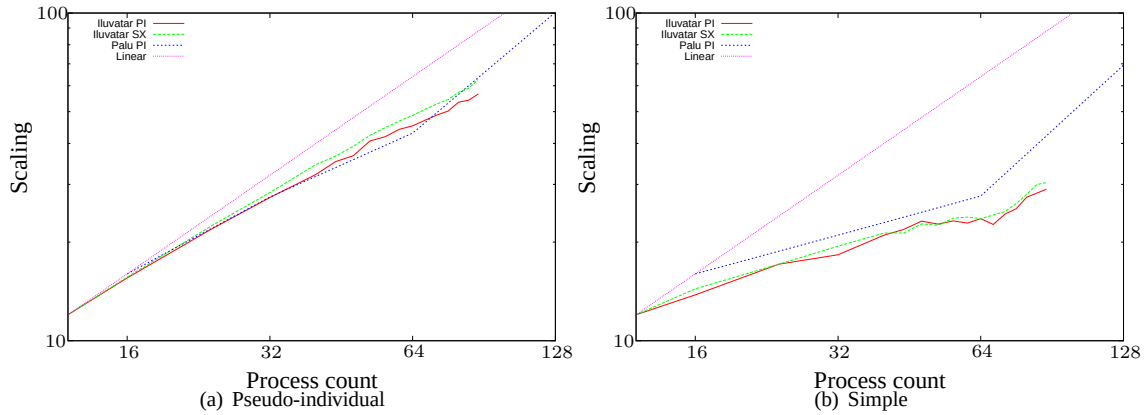


Figure 4: Scalability curves for the meta-population patch model at the level of 10608 electoral wards (1, 550, 819 subpatches) for (a) the pseudo-individual and (b) the simple cases.

Finally, the spread of infectious disease is a spatially explicit dynamic process. In this paper we have only considered static load balancing. As models become more complicated there are likely to arise significant differences in the computation load of patches over time (as disease activity passes through the patches). In this situation dynamic load balancing must be considered and this forms part of ongoing work.

In conclusion, the spread of infectious disease in a human population continues to present problems in biology, public health policy and in the field of high performance computing. The development of efficient methods for solving these systems will remain important, and facilities such as the Cray XT3 and the upcoming XT4 offer valuable means beyond those of smaller clusters.

Acknowledgements

We acknowledge the support of EPSRC under research grant GR/S42314/01. We also thank CSCS for the use of their Cray XT3 Palu and Ian Hall for helpful comments.

About the authors

Iain Barrass is a mathematical modeller within the Microbial Risk Assessment unit of the UK's Health Protection Agency based at the Centre for Emergency Preparedness and Response. His primary interests are in the spatio-temporal dynamics of infectious diseases and the application of high performance computing. Email: iain.barrass@hpa.org.uk

Kevin Roy is an HPC consultant at Manchester Computing.

Steve Leach is Scientific Head of the Microbial Risk Assessment programme.

References

[1] Johnson, N.P., Mueller, J.: Updating the accounts: global mortality of the 1918–1920 "Spanish" influenza pandemic. Bull. Hist. Med. (2002) **76**:105–15

[2] Nguyen-Vam-Tam, J.S., Hampson, A.W.: The epidemiology and clinical impact of pandemic influenza. Vaccine (2003) **21**:1762–8

[3] Anderson, R.M., Fraser, C., Ghani, A.C., Donnelly, C.A., Riley, S., Ferguson, N.M., Leung, G.M., Lam, T.H., Hedley, A.J.: Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. Phil. Trans. R. Soc. Lond. B (2004) **359**:1091–1105

[4] Gani, R., Leach S.: Transmission potential of smallpox in contemporary populations. Nature (2001) **414**:748–751

[5] Ferguson, N.M., Cummings, D.A.T., Fraser, C., Cajka, J.C., Cooley, P.C., Burke, D.S. Strategies for mitigating an influenza pandemic. Science (2006) **442**(7101):448–452

[6] Keeling, M.J., Bjornstad, O.N., Grenfell, B.T., Metapopulation dynamics of infectious diseases. In: Hanski, I., Gaggiotti, O. eds. *Ecology, Genetics, and Evolution of Metapopulations*. Elsevier, 2004, *pp.* 415–445

[7] Hall, I.M., Egan J.R., Barrass, I., Gani, R., Leach, S.: Spatio-temporal modelling for a potential smallpox outbreak: Comparison of control strategies using SEIR-type meta-population models. Epidemiol. Infect. *In press*

[8] Office for National Statistics: 2001 Census. http://www.statistics.gov.uk/census2001

[9] 2001 Census, data supplied by the General Register Office for Scotland

[10] 2001 Census, data supplied by the Northern Ireland Statistics and Research Agency

[11] Kermack, W.O., McKendrick, A.G.. A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. Lond. A (1927) **115**:700–721

[12] Anderson, R.M., May, R.M.: Infectious Disease of Humans: Dynamics and Control. OUP, 1992

[13] Ma, J., Earn, D.J.D. Generality of the final size formula for an epidemic of a newly invading infectious disease. Bull. Math. Biol. (2006) **68**:679–702

[14] Flahaut, A., Letrait, S., Blin, P., Hazout, S., Menares, J., Valleron, A.J. Modelling the 1985 influenza epidemic in France. Stat. Med. (1988) **7**:1147-55