MRC Biostatistics Unit

UNIVERSITY OF CAMBRIDGE

WORKSHOP ON AGENT-BASED MODELS

17th – 18th December 2024

Elton-Bowring Room,
Memorial Court, Clare College

Durham University

NIHR | National Institute for Health and Care Research

The Alan Turing Institute

Timetable

Tuesday 17th December

9:30–10:00	**Tea/Coffee**	
10:00–10:10	**Welcome remarks – Paul Birrell (UKHSA, University of Cambridge)**	
10:10–11:00	**Session 1**	*Chair: Paul Birrell*
10:10–10:35	**Christophe Fraser** University of Oxford	TBA
10:35–11:00	**Jasmina Panovska-Griffiths** University of Oxford	TBA
11:00–11:25	**Frank Krauss** University of Durham	Introduction to JUNE
11:25–11:45	**Coffee Break**	
11:45–12:30	**Session 2**	*Chair: Paul Birrell*
11:45–12:30	**David Banks** Duke University	Statistical Issues in Agent-Based Models
12:30–13:45	**Lunch**	
13:45–15:00	**Session 3**	*Chair: André Charlett*
13:45–14:10	**Ronni Bowman** dstl	Government Digital Twin Definition to Aid Development and Procurement
14:10–14:35	**Thomas Finnie** UK Health Security Agency	Modelling telecoms data to understand human movement patterns
14:35–15:00	**Fernando Benitez-Paez** University of St Andrews	Building Synthetic Populations for Agent-Based Models: Expanding Epidemic Scenario Simulations with SPC
15:00–15:30	**Coffee Break**	
15:30–16:55	**Session 4**	*Chair: Ewan Coleman*
15:30-15:55	**Keiran Suchak** University of Leeds	Real-time Prediction and Uncertainty in Agent-Based Models
15:55-16:20	**TJ McKinley** University of Exeter	Efficient Bayesian modelling of infectious diseases in wildlife: an application to bovine tuberculosis in wild badgers
16:20-16:55	**Simon Spencer** University of Warwick	Inference for agent based models of infectious diseases using coupled hidden Markov models
19:30	**Workshop Dinner**	

Wednesday 18th December

9:30–10:00	**Tea/Coffee**	
10:00–10:50	**Session 5**	*Chair: Kathryn Bowers*
10:00 – 10:25	**Elizabeth Hunter** University of Galway	Combining Agent-Based and Equation-Based Models: Lessons Learned from COVID-19 and Measles Modelling
10:25–10:50	**Arnau Quera-Bofarull** University of Oxford	Differentiable Agent-Based Models
10:50–11:20	**Coffee Break**	
11:20–12:10	**Session 6**	*Chair: Frank Krauss*
11:20-11:45	**Kevin Fong** University College London Hospital	Surge Capacity, Data, Information and Insight during the Covid-19 Pandemic.
11:45–12:10	**Jan Sheppard** ESR, New Zealand	Simulating infectious diseases in New Zealand – The ideal laboratory.
12:10–13:30	**Lunch**	
13:30–14:20	**Session 7**	*Chair: Birgitte Freiesleben di Blasio*
13:30–13:55	**Francesco Di Ruscio** Norwegian Institute of Public Health	Agent-Based Modelling for Infection Control and Public Health Response in Norway
13:55–14:20	**Jørgen Midtbø** Norwegian Institute of Public Health	In-silico exploration of vaccine effectiveness estimation using an ABM
14:20–14:50	**Break**	
14:50–15:40	**Session 8**	*Chair: TJ McKinley*
14:50–15:15	**Andy Iskauskas** University of Durham	Calibration of Stochastic Simulators using Emulation and History Matching
15:15–15:40	**Nicky McCreesh** London School of Hygiene and Tropical Medicine	Calibration of a complex tuberculosis IBM using history matching with model calibration
15:45–17:00	**Discussion, Future Directions and Challenges**	*Chair: Daniela De Angelis*

List of Abstracts – Talks

Tuesday 17th December

TBA

Christophe Fraser

University of Oxford

TBA

TBA

Jasmina Panovska-Griffiths

University of Oxford

TBA

Introduction to JUNE

Frank Krauss

University of Durham

In this talk I will introduce the individual-based model JUNE, focusing mainly on its design principles, input data, and some example applications.

Statistical Issues in Agent-Based Models

David Banks

Duke University

Agent-based models have become an ubiquitous and essential methodology in hard and soft sciences. But statistical inference on them is underdeveloped, and too few statisticians are working on this area. There are clear challenges in validating such models, in making quantitative statements of uncertainty, and in comparing different models with each other. This talk sketches some of these issues, and particularly points up the need for estimation of local dimension when working with complex models. It also considers the relative merits of emulators and Approximate Bayesian Computation when trying to make inferences with agent-based models.

Government Digital Twin Definition to Aid Development and Procurement

Ronni Bowman

dstl

This talk will present the new government definition for digital twins, commissioned by the national digital twin programme. It will cover the recommended general principles and how this could be met by particular digital twin programmes.

Modelling telecoms data to understand human movement patterns

Thomas Finnie

UK Health Security Agency

Understanding how people move is one of the key pieces of information when considering ABMs for disease. For those diseases that require no specific environment for transmission, think respiratory pathogens, the epidemiologically relevant patterns of gross human movement are likely well approximated by the movement patterns exhibited by the mobile phones carried by people and captured in the administrative data of the network providers. The data captured is however only a single realisation, the actualisation, of all the possible movements that could have been made by people at the point at which it occurs. For the purposes of planning and preparedness this is unlikely to be sufficient, especially as this observational data provides no mechanistic understanding of the generative processes of the patterns. In this talk I will show work where we have used mobile phone administrative data to inform a pair of mechanistic movement models which then allow multiple, realistically constrained realisations of human movement to be generated in a form suitable for use in ABMs.

Building Synthetic Populations for Agent-Based Models: Expanding Epidemic Scenario Simulations with SPC

Fernando Benitez-Paez

University of Saint Andrews

The Synthetic Population Catalyst (SPC) is a tool designed to simulate the daily activities of the population of Great Britain at a fine scale, providing critical insights for epidemic preparedness through agent-based models (ABMs). By leveraging a spatial microsimulation model, synthetic individuals are created to reflect the socio-economic, health, and time-use characteristics of the British population. This enables local or regional authorities to simulate social demographic distributions and, crucially, the interactions between people in various settings such as workplaces, schools, and large-scale events. These simulations facilitate researchers to create "what-if" scenarios modelling to shape a more nuanced understanding of disease transmission dynamics while preserving individual privacy due to the synthetic nature of the data.

The SPC forms part of a broader Urban Analytics toolkit developed by the Alan Turing Institute's Urban Analytica Programme. In this presentation, I will discuss how SPC can be integrated within ABMs to enhance epidemic modelling by offering fine-scaled insights into how diseases may spread across different social and spatial contexts. I will also demonstrate a web-based application that allows urban researchers to explore and analyse these interactions in detail, moving beyond traditional census-based demographic analysis. Researchers can test different behavioural and disease-spread scenarios, making the SPC an invaluable resource for modelling potential outbreaks and informing policy decisions to safeguard public health.

Real-time Prediction and Uncertainty in Agent-Based Models

Keiran Suchak

University of Leeds

Agent-based models (ABMs) are a valuable tool for studying complex systems. By simulating the actions and interactions of heterogeneous agents, ABMs enable researchers to explore how individual behaviours shape broader system dynamics. Their flexibility makes them particularly suitable for evaluating interventions across diverse populations and scenarios, such as those found in epidemiology and human mobility.

Human mobility plays a central role in how infections spread, as many infectious diseases rely on individuals being co-located — that is, in the same place at the same time — for transmission to occur. Consequently, the movement of individuals affects contact rates, transmission pathways, and spatial diffusion, whether at the scale of pedestrian dynamics or larger migration patterns.

In such contexts, researchers may wish to make estimates and predictions in real-time, enabling the development of data-driven policies that can adapt at short notice — a need amplified by the growing availability of high-velocity data. However, applying ABMs in real-time contexts presents notable challenges. Even with a well-calibrated model, predictions generated by an ABM are likely to diverge from the real system state over time, particularly when operating in real-time, accompanied by growing levels of uncertainty. Frequent recalibration of ABMs using historical data can be costly and impractical. To address this, there is a need for novel methods to incorporate real-time observations into ABMs in a more efficient and dynamic manner.

This talk will therefore examine recent developments in ABMs, focusing on how data assimilation techniques can improve their real-time predictive accuracy and reduce uncertainty. Drawing on examples ranging from global policy responses during the COVID-19 pandemic to urban crowd dynamics, we will discuss how these methods align simulations with real-world observations. Additionally, we will consider the interplay between human mobility and infection spread, exploring how definitions of co-location impact the detection of contact events and how uncertainty in mobility data influences these processes.

Efficient Bayesian modelling of infectious diseases in wildlife: an application to bovine tuberculosis in wild badgers

TJ McKinley

University of Exeter

To better understand transmission and control of infectious diseases, it is crucial to be able to fit dynamic transmission models to observed data in a robust and efficient way. In practice, epidemiological events are at best only partially observed, and as such it is necessary to infer missing information alongside the model parameters as part of any inference routine, typically requiring large computational resources.

With this in mind, we implement a recently proposed individual forward filtering backward sampling algorithm to fit a complex individual-based stochastic epidemic model to data from a large-scale longitudinal study of bovine tuberculosis in wild badgers. This data set, from Woodchester Park in south-west England, comprises >2,300 badgers across 34 social groups over a 40+ year period. We deal with many complexities typical to endemic wildlife disease systems: incomplete sampling of individuals over time (through capture-mark-recapture events), the use of multiple imperfect diagnostic tests, spatial meta-population structures, and non-Markovian demographic aspects such as age-dependent mortality rates (with censoring). The method produces full posterior distributions for the parameters, and predictive distributions for the hidden states over time for each individual, and fits in just a few hours on a desktop machine.

We discuss some of the key epidemiological insights from the model, and also propose a novel individual-level reproduction number, from which we provide quantitative evidence for the presence of superspreader badgers in Woodchester Park. The inference framework is very flexible, and could be applied to other individual-level disease systems, and we discuss future extensions to explore further important epidemiological questions.

Inference for agent based models of infectious diseases using coupled hidden Markov models

Simon Spencer

University of Warwick

Agent based models provide a huge amount of flexibility to describe individual-level heterogeneities in infectious disease transmission. However, for such models to become grounded in reality, they must be successfully fitted to available data. Model fitting often takes the form of Bayesian inference using Markov chain Monte Carlo (MCMC) algorithms. These algorithms involve imputing the infection time of each individual infected during the epidemic alongside the model parameters, and quickly become computationally intensive when the infected population is large.

In this talk, I will outline a flexible modelling framework for describing infectious disease transmission with individual-level heterogeneities. I will describe the advantages of viewing such models as coupled hidden Markov models. In particular, I will suggest efficient approaches to performing inferences that can be scalable to realistic population sizes.

Wednesday 18th December

Combining Agent-Based and Equation-Based Models: Lessons Learned from COVID-19 and Measles Modelling

Elizabeth Hunter

University of Galway

Agent-based models allow us to simulate the spread of an infectious disease through a population and can be used as a tool to help understand different social and geographical factors that influence an outbreak. Often these social or geographical factors cannot be modelled with a more traditional equation-based model. Agent-based models, however, have a major disadvantage, the larger the agent-based model and the more complexity added to the model, the more computing power and time is required to run the model. Combining agent-based models with other modelling types to create hybrid models can help to reduce some of the computationalrequirements. However, when hybridizing an agent-based model we can lose some of the complexity and detail that is needed to answer questions often asked of agent-based models. The question then remains, if we want to make useful models, how can we combine agent-based and equation-based models to help us learn more about disease spread and prepare for the next outbreak.

Differentiable Agent-Based Models

Arnau Quera-Bofarull

University of Oxford

Automatic differentiation has been fundamental to the success of deep learning, enabling the training of neural networks with billions of parameters. This talk explores extending these differentiation techniques to Agent-Based Models (ABMs), presenting a framework for differentiable ABMs that computes simulation gradients with minimal computational overhead. By making ABMs differentiable, we enable gradient-based inference methods such as variational inference for parameter estimation. Using the classical SIR epidemiological model as a case study, I will demonstrate how this approach enables efficient Bayesian calibration of ABMs with large parameter spaces, advancing our ability to develop scalable, data-driven agent-based models.

Surge Capacity, Data, Information and Insight during the Covid-19 Pandemic

Kevin Fong

University College London Hospital

Professor Kevin Fong is a consultant anaesthetist at University College London Hospital, with joint specialist accreditation in anaesthesia and critical care medicine. In March 2020 he was seconded to NHS England as National Clinical Adviser in Emergency Preparedness Resilience and Response. In this role he worked with teams across the national response including NHSE's Severe Covid Response Cell, the Critical Care Capacity Panel, the Critical Care Transfer Cell, UKHSA's Joint Modelling Team, Joint Biosecurity Centre's Alert Level Update team and the Durham based JUNE modelling team.

Between the Spring of 2020 and Summer of 2021 he established a survey tool, designed to study of the impact of pandemic surge on the mental health of frontline NHS staff. In 2020 he organised NHSE's system of in-reach visits which took clinical teams into the hardest hit hospitals and ICUs across England to deliver Peer Support and Rapid Review. In 2021 he developed the joint military and civilian Rapid Evaluation Team capability. Throughout the pandemic he took on clinical shifts in anaesthesia, ICU and critical care transfer.

Simulating infectious diseases in New Zealand – The ideal laboratory

Jan Sheppard

ESR, New Zealand

New Zealand is an island nation with a population 5.2 million people, and a land area of slightly more than the UK. These characteristics make New Zealand the ideal laboratory for developing and running models simulating how infectious diseases spread to inform policy settings when the next pandemic emerges, be that measles, Avian Bird flu or something new. The team at the Institute of Environmental Science and Research (ESR) in New Zealand have been collaborating with the University of Durham, building off the June Model to create a model specific for New Zealand and driven by a sophisticated synthetic population dataset. This synthetic population dataset was developed from a wealth of government data about people combined with a range of census data, survey data, transport data and more to mimic the entire population using agents with 80 attributes each. Jan Sheppard is the Chief Data & Analytics Officer at ESR and will share their modelling journey, where that has taken New Zealand in terms of future pandemic preparedness and also the wider benefits to the country the modelling is providing.

Agent-Based Modelling for Infection Control and Public Health Response in Norway

Francesco Di Ruscio

Norwegian Institute of Public Health

In this talk, I will present Norway's approach to leveraging agent-based models (ABMs) to tackle public health challenges. Beginning with a model designed to study the spread of methicillin-resistant Staphylococcus aureus (MRSA) in hospitals and communities, I will show how this foundational work was adapted during the COVID-19 pandemic to simulate disease spread and support critical public health decisions. During the crisis, the model became an important tool for analyzing vaccination deployment strategies, assessing the impact of non-pharmaceutical interventions, and addressing uncertainties related to the emergence of new variants, such as Omicron. Beyond its application to COVID-19, I will highlight ongoing efforts to address challenges faced during the pandemic, such as ABM calibration, and discuss current research directions and applications of ABMs in public health in Norway.

In-silico exploration of vaccine effectiveness estimation using an ABM

Jørgen Midtbø

Norwegian Institute of Public Health

Effect estimates for vaccines (VE) are complicated by the complex dynamics of infectious disease transmission, where unvaccinated are protected indirectly by others around them being vaccinated. Using an agent-based model to perform in-silico "randomized controlled trials" or observational studies, we explore the landscape of parameters that shape VE estimates, and discuss how the vaccine parameters in the model relate to measured VE.

Calibration of Stochastic Simulators using Emulation and History Matching

Andy Iskauskas

University of Durham

ABMs are extremely well-suited to modelling infectious diseases, providing a granular means of tracking the natural history of a disease through a population that cannot be readily replicated in deterministic approaches. Their depth and complexity, however, can lead to difficulties in their use: high-dimensional parameter spaces required for an accurate simulator are hard to explore robustly; and the stochastic nature of the simulator precludes an analysis of a 'true', underlying behaviour or identification of trajectories through the output space without performing large numbers of repetitions at input points. These two problems are exacerbated by the (often high) computational expense of such models, making the process of matching the output of such a model to observational data (loosely termed 'calibration') extremely difficult.

I will present an efficient means by which such calibration tasks can be achieved, via stochastic emulation and history matching. The emulator, a statistical surrogate for the simulator, allows simulator evaluations to be estimated across the whole parameter space with a limited ensemble of true simulator runs; history matching allows us to find the full parameter space of combinations that could give rise to observed reality. Via a hierarchy of emulators we may also characterise the underlying stochasticity of the model, account for the inadequacies in finite repetition number at design points, and provide a robust process for matching a stochastic simulator's output to data. I will motivate the methodology via application to a complex model of human papillomavirus — HPVsim — and demonstrate the understanding one can gain about the input and output spaces of such a modelling framework via these techniques.

Calibration of a complex tuberculosis IBM using history matching with model calibration

Nicky McCreesh

University of Durham

An estimated 270K people developed TB in South Africa in 2023, with 56K dying of the disease. Active case finding (ACF), the systematic screening of high-risk groups or populations, is one way of reducing transmission, by detecting people with TB sooner. We need to better understand the potential impact and cost-effectiveness of different screening approaches – e.g. testing everyone or symptomatic people only – and screening in different settings – e.g. clinics vs the community. This requires a complex model, for instance to capture variation in clinic visiting behaviour. It also requires the model to be comprehensively calibrated, to capture the uncertainty resulting from high levels of uncertainty in factors such as TB natural history.

We developed an individual-based model (IBM) of TB infection transmission, disease development, progression, and treatment, incorporating individual-level variation in clinic visiting rates. The model was calibrated to 43 calibration targets, varying the values of 61 input parameters. It was calibrated using the R package hmer, which implements history matching and model emulation, a calibration method that iteratively cuts out implausible parameter space over multiple waves, using fast emulators to estimate the values of model outputs for parameter sets where the model has not been run. A number of techniques were used to improve model fit, including transforming input parameters, running the first few waves using widened ranges for some parameters and targets, and using full fitting points identified during initial calibration attempts to validate the emulators during the final calibration.

Calibration found over 1000 full fitting parameter sets, successfully incorporating the wide ranges of uncertainty that exist in the values of many input parameters and calibration targets. History matching and model emulation using hmer makes possible the comprehensive calibration of complex IBMs, varying large numbers of input parameters and fitting to large numbers of calibration targets.

List of Posters

A small number of posters will be exhibited throughout the workshop. Though there is no formal poster session, there are plenty of breaks which should, hopefully, give plenty of

Poster Abstracts

Modelling with SPEED: A Stochastic Predictor of Early Epidemic Detection

Kathryn Bowers

University of Cambridge

Emerging infectious diseases pose a significant threat to public health. This work focuses on quantifying the risk of sustained transmission following the initial detection of a new pathogen through routine surveillance. This poster introduces the SPEED model, a Stochastic Predictor of Early Epidemic Detection. SPEED is an adaptation of the classic Susceptible-Infected-Recovered (SIR) framework that simulates stochastic early-stage disease transmission dynamics using a Gillespie-like algorithm. The model incorporates individual-level testing and dynamically adjusts public health responses by increasing testing probabilities and reducing detection times once a single case has been identified. SPEED serves two main functions. First, as a statistical inference tool that takes a prior specification for the reproduction number and refines this estimate following the detection of a small number of cases. Second, to simulate epidemic scenarios under specified values for the reproduction number which are used to construct a distribution of the time to detection of a second case. Our results demonstrate SPEED applied to a single case of influenza A(H1N2)v, detected through routine flu surveillance in Yorkshire. Comparisons with simulations under heightened surveillance scenarios demonstrate the model's utility in assessing response efficacy on the initial outbreak spread.

JUNE: Generating a High Resolution UK Population for Simulating Epidemics

Callum Brown

University of Durham

JUNE is an individual-based epidemiology simulation designed to predict the spread of infectious diseases within large-scale populations. Leveraging modern high-performance computing enables simulations of large populations, such as that of England, at high resolution and encompassing over 56 million individuals (as of 2021). The simulation's predictive accuracy relies on its ability to realistically recreate population demographics, including living arrangements (and conditions) commuting patterns, and daily activities. This poster highlights the methods used to construct the population and assign individuals to households or other communal residences, offering a glimpse into the framework's capability to support precise epidemic forecasting.

Optimising interventions against HIV and Hepatitis C in people who inject drugs using an agent-based model

Ewan Colman, Josephine Walker, Mercy Nyakowa, Hannah Manley, Lindsey Riback, Peter Vickerman, Matthew Akiyama, Jack Stone

University of Bristol

People who inject drugs (PWID) are a high-risk group for blood borne viruses. Programmes to reduce needle sharing help slow the spread of these diseases, but when resources are limited, it is important that interventions are targeted effectively. To address this, we developed an agent-based model to study the effects of interventions to prevent the spread of HIV and Hepatitis C in PWID. The model incorporates the initiation and cessation of injecting drug use and the forming and breaking of injecting partnerships that can span multiple injecting sites (known informally as “dens”). We calibrated the model to data from a respondent-driven sampling survey of PWID in Kenya, specifically the distribution of the number of injecting partners, the number of dens used, how long PWID have been injecting for, prevalence of HIV and Hepatitis C, and treatment uptake for both diseases. We then evaluated the impact on HCV and HIV incidence of targeting interventions based on the network properties of individuals compared to random targeting. Two heuristic strategies in particular yield a substantial improvement: selecting individuals based on how many injecting partners they have, and selecting individuals who have injecting partners across multiple injecting sites. Our future work will explore how our model can be used to guide public health decision-making for this at-risk population.

JUNE: A Flexible Framework for Simulating Diverse Epidemics from Measles to Medieval Plagues

Martha Correa Delval

University of Durham

We present JUNE, an open-source, individual-based epidemiological simulation framework designed to model the spread of infectious diseases through detailed social interactions within virtual populations. Constructed using geographically granular census data that reflect demographics such as age, sex, ethnicity, and socio-economic indicators, JUNE simulates interactions in various social settings—including households, schools, workplaces, and community activities—using adaptable social mixing patterns. Originally applied to COVID-19 in England, we have extended JUNE to simulate a range of infectious diseases, including measles, monkeypox, and historical epidemics like the Black Death in medieval England. This poster outlines the adaptable framework of JUNE and showcases its applications across different epidemiological scenarios.

Agent-Based Models with Contact Networks Based On Contact Tracing Data

Ida-Marie Johansson

University of Oslo

The COVID-19 contact tracing data from Oslo municipality provides valuable insights into the contact patterns among various age groups and districts within the city. Utilizing this data, we developed an agent-based model to simulate the first two years of the COVID-19 pandemic. Our goal is to evaluate the effectiveness of the contact tracing strategy in Oslo. In this poster, we will introduce our agent-based model, outline the key assumptions underlying our simulations and how we will proceed with model calibration and scenario analysis.

Partner Institutions and Sponsors

This workshop is generously funded by the NIHR Health Protection Research Unit in Behavioural Science and Evaluation at the University of Bristol, a partnership with the UK Health Security Agency, in collaboration with the MRC Biostatistics Unit, University of Cambridge and the University of the West of England. Support from the MRC Biostatistics Unit and the Alan Turing Institute Foundational funding into Probabilistic Programming Languages is also gratefully acknowledged

Sponsors

NIHR | National Institute for Health and Care Research

Durham University

The Alan Turing Institute