Bayesian Inference of Epidemics

Book of Abstracts

Levi (Finland), 12-14 March 2023

Contents

1 Introduction

The idea of having a workshop for statisticians in epidemics has been in our minds for quite some time. The challenges that arise when analysing epidemic data are complex and have often led us to the borders of the knowledge in Bayesian Inference and Computational Methods.

In the last 4 years, many of the people involved in this workshop had the opportunity to gather frequently for events related to the EPSRC Bayes4Health grant in the UK. Stimulating discussion around open problems have benefitted all the participants and started synergistic collaborations from which initiatives such as this workshop started.

Many researchers in statistical methods for epidemics have served in national and international advisory boards to soundly inform the response to the Covid pandemic. While following tight deadlines to provide real-time results, models have been improved and new computational methods have been developed. Three years after the beginning of the Covid pandemic we hope that this workshop would allow these researchers to share their achievements and discuss with colleagues, as well as with experts in Bayesian Computation. For this reason, this event is a satellite of the conference BayesComp 2023.

When drafting the initial programme, we wanted this workshop to be accessible to novices to the field of statistical modelling of epidemics and, at the same time, we wanted to propose some topics that differ from traditional approaches to epidemic models, so that even people with substantial expertise could learn something new.

Thanks to the positive response from our keynote and invited speakers and an enthusiastic participation by many researchers that submitted proposals for talks and posters we have completed this programme. We very much hope that you will enjoy this workshop, learn from the works presented by the speakers, and find new collaborators to address the many open questions in the field.

We foresee this would be the first of many appointments for our growing community.

Nauttikaa!

The Scientific Board

2 Concise Schedule

Day 1: Sun 12th of March

Time	Session	Speaker	Title	Length
19:30 - 21:00	Satellite Opening	SEF Spencer	*Introduction to Epidemic Models and their statistical analysis*	80 min
		A Mira	*Introduction to PERISCOPE project*	10 min

Day 2: Mon 13th of March

Time	Session	Speaker	Title	Length
9:20 - 10:10	Tutorial	TJ McKinley	*A tutorial on history matching with emulation for epidemic models*	50 min
10:10 - 10:30	Coffee			20 min
10:30 - 12:00	Keynotes	V Minin	*Fitting stochastic epidemic models to noisy surveillance data: are we there yet?*	45 min
		C Fuchs	*Integrative modelling of infections in a corona virus cohort study*	45 min
12:00 - 13:30	Lunch			90 min
13:30 - 15:00	Invited session	T Kypraios	*Bayesian nonparametric inference for stochastic infectious disease models*	30 min
		PJ Birrell	*An approximate diffusion process for environmental stochasticity in infectious disease transmission modelling*	30 min
		L Guzmann-Rincon	*Bayesian estimation of the instant growth rate of SARS-CoV-2 positive cases in England, using Gaussian processes*	30 min
15:00 - 15:30	Coffee			30 min
15:30 - 17:00	Contributed session	R Deardon	*Identifying behavioural change mechanisms in epidemic models*	30 min
		E Semoneva	*Spatial statistics with deep generative modelling: flexible and efficient disease mapping with MCMC and deep learning*	30 min
		A Beloconi	*Malaria, climate variability and the effect of interventions: modelling transmission dynamics*	30 min
17:00 - 18:00	Break			30 min
18:00 - 21:00	Poster session			180 min

Day 3: Tuesday the 14th of March

Time	Session	Speaker	Title	Length
9:10 - 10:10	Invited session	C Pooley	*Fast inference and model selection on epidemiolgical models using model-based proposals*	30 min
		J Xu	*Efficient Branching Process Proposals and Data-Augmented MCMC for the Stochastic SIR Model*	30 min
10:10 - 10:30	Coffee			20 min
10:30 - 12:00	Keynotes	S Cauchemez	*Bayesian data augmentation methods applied to infectious disease epidemiology*	45 min
		P Nouvellet	*Understanding and quantifying pathogen transmission, power of Bayesian approaches*	45 min
12:00 - 15:00	Lunch and Ski			180 min
15:00 - 16:30	Invited session	L Rimella	*Approximating optimal SMC proposal distributions in individual-based epidemic models*	30 min
		M Whitehouse	*Consistent and fast inference in compartmental models of epidemics using Poisson Approximate*	30 min
		J Wheeler	*Informing policy via dynamic models: Cholera in Haiti*	30 min
16:30 - 17:00	Coffee			30 min
17:00 - 18:30	Invited session	P Marttinen	*A Bayesian model of acquisition and clearance of bacterial colonization incorporating within-host variation*	30 min
		J Koskela	*Bayesian inference of recombinant ancestries*	30 min
		A Gill	*Bayesian Inference of Reproduction Number from Genomic and Epidemic Data using MCMC Methods*	30 min

3 Talks on the 13th March

3.1 Keynote session I

Mon 13th from 10:30 to 12:00. Chair: Gareth O. Roberts.

Fitting stochastic epidemic models to noisy surveillance data: are we there yet?
Volodymyr Minin (University of California Irvine)

Stochastic epidemic models describe how infectious diseases spread through a population of interest. These models are constructed by first assigning individuals to compartments (e.g., susceptible, infectious, and recovered) and then defining a stochastic process that governs the evolution of sizes of these compartments through time. I will review multiple lines of attack of a challenging and not fully solved problem of fitting these models to noisy infectious disease surveillance data. These solutions involve a range of mathematical techniques: particle filter Markov chain Monte Carlo algorithms, approximations of stochastic differential equations, and Poisson random measure-based Bayesian data augmentation. Importantly, many of these computational strategies open the door for integration of multiple infectious disease surveillance data streams, including less conventional ones (e.g., pathogen wastewater monitoring and genomic surveillance). Such data integration is critical for making key parameters of stochastic epidemic models identifiable. I will illustrate the state-of-the-art statistical inference for stochastic epidemic models using Influenza, Ebola, and SARS-CoV-2 surveillance data and will conclude with open problems and challenges that remain to be addressed.

Integrative modelling of infections in a corona virus cohort study
Christiane Fuchs (Universität Bielefeld)

Since the outbreak of the corona pandemic, many mathematical and statistical approaches have been used to identify drivers of infection and disease, to advise decision makers, and to predict the further course. However, models that worked well in idealized contexts often encountered limitations in real-world situations; these include unreliable information on numbers of infected individuals, and a rapidly changing environment regarding control measures, varying testing and vaccination capacities and strategies, weather conditions, and new virus variants.

I will report from our work in the KoCo19 Consortium, which has established a representative cohort for the Munich population in private households and has collected and analyzed laboratory and questionnaire data in several rounds since April 2020. We estimate the prevalence of SARS-CoV-2 in the population, evaluate the reliability of diagnostics, investigate the role of infection within households, and combine different data sources to reliably estimate the effectiveness of non-pharmaceutical intervention methods. This is done using a variety of statistical methods, not least Bayesian modelling.

3.2 Invited session on Environmental Stochasticity

Mon 13th from 13:30 to 15:00. Chair: Raiha Browning.

Bayesian nonparametric inference for stochastic infectious disease models

Theodore Kypraios (University of Nottingham)

Infectious disease transmission models require assumptions about how the pathogen spreads between individuals. These assumptions may be somewhat arbitrary, particularly when it comes to describing how transmission varies between individuals of different types or in different locations, and may in turn lead to incorrect conclusions or policy decisions. In this talk, we will present a novel and general Bayesian nonparametric framework for transmission modelling which removes the need to make such specific assumptions with regards to the infection process. We use multi-output Gaussian process prior distributions to model different infection rates in populations containing multiple types of individuals. Further challenges arise because the transmission process itself is unobserved, and large outbreaks can be computationally demanding to analyse. We address these issues by data augmentation and a suitable efficient approximation method. Simulation studies using synthetic data demonstrate that our framework gives accurate results. Finally, we use our methods to enhance our understanding of the transmission mechanisms of the 2001 UK Foot and Mouth Disease outbreak.

An approximate diffusion process for environmental stochasticity in infectious disease transmission modelling

Paul J Birrell (UK Health Security Agency)

Throughout the course of the SARS-CoV-2 pandemic we fulfilled a requirement both from within the UK Health Security Agency (including Public Health England) and externally through SPI-M, to produce pandemic nowcasts and medium-term 'projections' in real-time. This was achieved through the PHE-Cambridge real-time model (RTM), a deterministic compartmental model designed to assimilate an array of pandemic data streams to produce epidemic nowcasts and forecasts, whilst estimating key epidemic quantities. Such models need to accurately capture fluctuations in transmission arising from external determinants (e.g. changing government advice, the emergence of new variants). Ignoring this "environmental stochasticity" can lead to mis-calibrated models that underestimate uncertainty and produce biased predictions. This stochasticity was injected in the current RTM by modelling the infection hazard, β_t, as a stochastic process, a discrete-time random-walk with coarse time-steps (Birrell et al, 2021).

This, however, was a pragmatic choice, designed, above all, to alleviate computational burden. Here we will present an alternative representation for β_t, exploiting an approximating (deterministic) pathwise-expansions of diffusion processes to replace the stochastic random-walk. The method is shown to capture the dynamics equally well with the potential for curbing the dimensionality of the posterior as data accumulate.

Bayesian estimation of the instant growth rate of SARS-CoV-2 positive cases in England, using Gaussian processes

Laura Guzmann-Rincon (University of Warwick)

The growth rate estimation of SARS-CoV-2 positive cases is crucial for understanding the evolution of the pandemic. We propose a method for estimating the growth rate of the proportion of positive cases in England and its local authorities. The proposed Bayesian model incorporates a Gaussian process as a latent effect, employed to compute the growth rate and higher derivatives. This method does not make assumptions about generation times and can be adapted to different spatial geographies and population subgroups.
Preprint: https://www.medrxiv.org/content /10.1101/2022.01.01.21268131v1.full.

3.3 Contributed session on Informing Policy

Mon 13th from 15:30 to 17:00. Chair: Simon E. F. Spencer.

Identifying behavioural change mechanisms in epidemic models

Rob Deardon (University of Calgary)

The COVID-19 pandemic has illustrated both the utility and limitation of using epidemic models for understanding and forecasting disease spread. One of the many difficulties in modelling epidemic spread is that caused by behavioural change in the underlying population. This can be a major issue in public health since, as we have seen during the COVID-19 pandemic, behaviour in the population can change drastically as infection levels vary, both due to government mandates and personal decisions. Such changes in the underlying population result in major changes in transmission dynamics of the disease, making the modelling challenges. However, these issues arise in agriculture and public health, as changes in farming practice are also often observed as disease prevalence changes. We propose a model formulation where time-varying transmission is captured by the level of alarm in the population and specified as a function of the past epidemic trajectory. The model is set in a data-augmented Bayesian framework as epidemic data are often only partially observed, and we can utilize prior information to help with parameter identifiability. We investigate the estimability of the population alarm across a wide range of scenarios, using both parametric functions and non-parametric Gaussian process and splines. The benefit and utility of the proposed approach is illustrated through an application to COVID-19 data from New York City.

Spatial statistics with deep generative modelling: flexible and efficient disease mapping with MCMC and deep learning

Elizaveta Semoneva (University of Oxford)

Hierarchical Bayesian models are the current state-of-the-art approach to disease mapping. When working with areal type of data (e.g. district-level aggregates), routinely used models rely on the adjacency structure of areal units to account for spatial correlations. This approach ignores the continuous nature of spatial processes and is very rigid with respect to the change of support problem, i.e. when administrative boundaries change or when mapping needs to be done at a different administrative level. We present a novel, practical and easy to implement solution relying on a methodology combining deep generative modelling and fully Bayesian inference. We apply a recently proposed method of encoding spatial priors with Variational autoencoders (VAEs) to the change of support problem and malaria mapping. As malaria control programs continue to create novel control strategies, district level disease mapping remains a fundamental surveillance tool for analysing present and historical distribution of the disease. We encode realisations of the Gaussian Process prior over a fine artificial spatial grid, aggregated to the level of administrative boundaries, and use these VAE-priors at the inference stage. We demonstrate that the new method is faster and more efficient than state-of-the-art Bayesian models estimated via Markov Chain Monte Carlo algorithms.

Malaria, climate variability and the effect of interventions: modelling transmission dynamics

Anton Beloconi (Swiss TPH)

Assessment of the relative impact of climate change on malaria dynamics is a complex problem. Climate is a well-known factor that plays a crucial role in driving malaria outbreaks in epidemic transmission areas. However, its influence in endemic environments with intensive malaria control interventions is not fully understood, mainly due to the scarcity of high-quality, long-term malaria data. The demographic surveillance systems in Africa offer unique platforms for quantifying the relative effects of weather variability on the burden of malaria. Here, using a process-based stochastic transmission model, we show that in the lowlands of malaria endemic western Kenya, variations in climatic factors played a key role in driving malaria incidence during 2008–2019, despite high bed net coverage and use among the population. Bayesian computation and inference based on pMCMC were compared to iterated filtering. The model accounts for the main mechanisms related to malaria dynamics, including immunity, infectivity, and human migration, and opens the possibility to forecast malaria in endemic regions, taking into account the interaction between future climatic conditions and intervention scenarios.

4 Talks on the 14th March

4.1 Invited session on Sampling from the Hidden states

Tue 14th from 9:10 to 10:10. Chair: Christopher Jewell.

Fast inference and model selection on epidemiolgical models using model-based proposals

Chris Pooley (Biomathematics and Statistics Scotland)

Stochastic process-based models are used in a variety of fields, from epidemiology to biochemistry to quantitative genetics to finance. Performing Bayesian inference on these models has proved challenging, especially when the number of model parameters is large or the posterior is highly correlated. A standard approach has been to use Data Augmentation Markov chain Monte Carlo (DA-MCMC). This, however, can suffer from three major shortcomings: 1) a poor initial chose for the chain can lead to the system becoming stuck in local minima, 2) a high degree of correlation between successive samples can lead to long computational times, and 3) reliable model selection is challenging. Widely used – and highly parallelisable - alternatives are Approximate Bayesian Computation (ABC) and sequential Monte Carlo variants (SMC-ABC), and in principle exact particle filtering methods such as particle MCMC (PMCMC). However, each these approaches has limitations which in practice can result in poor estimation of the posterior.

To move beyond these limitations this talk introduces two new inference algorithms: Approximate Bayesian Computation using MBPs (ABC-MBP) and Particle Annealed Sampling using MBPs (PAS-MBP). Both of these methods combine, in a novel way, two pre-existing ideas: firstly, using many "particles" (combinations of model parameters and system state) which, over successive generations, pass from the prior to the posterior, and secondly so-called "model-based proposals" (MBPs) originally designed for speeding up DA-MCMC mixing.

ABC-MBP and PAS-MBP differ in how they treat the observation model: ABC-MBP assumes a simple cut-off in an error function, whereas PAS-MBP allows for a flexible mechanistic observation model to be specified. In this talk PAS-MBP and ABC-MBP are compared against other methodologies in the literature (standard ABC, ABC-SMC, PMCMC and MC3) using a number of benchmark epidemiological models (ranging in complexity and level of stochasticity). The new methods are found to come out fastest, as well as being easily parallelisable and posing fewer problems in terms of optimisation. ABC-MBP is applied to an age-stratified compartmental model using publicly available COVID-19 for England and Wales. This represents a challenging problem (fitting 150 model variables using 9000 observations) which yielded improved estimation of the age-stratified contact matrix as well as time variation in the reproduction number.

Joint work with Andrea B. Doeschl-Wilson (The Roslin Institute, The University of Edinburgh) and Glenn Marion (Biomathematics and Statistics Scotland)

Efficient Branching Process Proposals and Data-Augmented MCMC for the Stochastic SIR Model

Jason Xu (Duke University)

We propose a novel data-augmented Markov Chain Monte Carlo algorithm for exact Bayesian inference under the stochastic susceptible-infected-removed model, given only discretely observed counts of infection. Incidence data in our setting present challenges to inference due to only partially informing us of the underlying continuous-time proces. To account for the missing data while targeting the exact posterior of model parameters, we make use of latent variables that re jointly proposed from a surrogate process carefully designed to closely resemble the SIR model. This allows us to efficiently generate epidemics consistent with the observed data, and extends to non-Markovian settings. Our Markov chain Monte Carlo algorithm is shown to be uniformly ergodic, and we find that it mixes significantly faster than existing single-site samplers on several real and simulated data applications.

4.2 Keynote session II

Tue 14th from 10:30 to 12:00. Chair: Daniela De Angelis.

Bayesian data augmentation methods applied to infectious disease epidemiology
Simon Cauchemez (Institute Pasteur)

In this talk, I will discuss how Bayesian data augmentation approaches have been used to analyze complex infectious disease datasets and gain key insights on transmission dynamics, correlates of protection, identification of unobserved infections, evolution of key biomarkers, etc... The talk will be illustrated with analyses that considered different types of epidemiological datasets including outbreak investigations, cross-sectional serological studies and longitudinal cohorts, as well as different pathogens (influenza, dengue, COVID-19).

Understanding and quantifying pathogen transmission, power of Bayesian approaches
Pierre Nouvellet (University of Sussex)

Bayesian approaches have become increasingly popular in epidemiology for estimating key parameters of infectious disease transmission, such as the reproduction number. The basic reproduction number, commonly denoted as R0, measures the average number of secondary infections caused by a single infected individual. R0 can be generalised to an effective reproduction number, noted Rt, estimated at any specific point in time, given the current conditions (e.g., population immunity, control measures in place). In this presentation, I will show how, using a Bayesian approach, we have extended a popular framework to estimate Rt to 1) disentangle the impact of importation from local transmission, 2) understand and quantify the drivers of transmission, 3) estimate the potential transmission advantage of emerging variants. Overall, Bayesian approaches offer a powerful framework for measuring the reproduction number of pathogens and understanding the complex dynamics of infectious disease transmission. These methods can help public health officials make informed decisions and develop effective strategies for controlling and preventing the spread of disease.

4.3 Invited session on Inference of nonlinear dynamics

Tue 14th from 15:00 to 16:30. Chair: Paul J. Birrell.

Approximating optimal SMC proposal distributions in individual-based epidemic models

Lorenzo Rimella (University of Lancaster)

Many epidemic models are naturally defined as individual-based models: where we track the state of each individual within a susceptible population. Inference for individual-based models is challenging due to the high-dimensional state-space of such models, which increases exponentially with population size. We consider sequential Monte Carlo algorithms for inference for individual-based epidemic models where we make direct observations of the state of a sample of individuals. Standard implementations, such as the bootstrap filter or the auxiliary particle filter are inefficient due to mismatch between the proposal distribution of the state and future observations. We develop new efficient proposal distributions that take account of future observations, leveraging the properties that (i) we can analytically calculate the optimal proposal distribution for a single individual given future observations and the future infection rate of that individual; and (ii) the dynamics of individuals are independent if we condition on their infection rates. Thus we construct estimates of the future infection rate for each individual, and then use an independent proposal for the state of each individual given this estimate. Empirical results show order of magnitude improvement in efficiency of the sequential Monte Carlo sampler for both SIS and SEIR models.

Preprint: ArXiv [2206.05161] Approximating optimal SMC proposal distributions in individual-based epidemic models

Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods

Michael Whitehouse (University of Bristol)

Addressing the challenge of scaling-up epidemiological inference to complex and heterogeneous models, we introduce Poisson Approximate Likelihood (PAL) methods. In contrast to the popular ODE approach to compartmental modelling, in which a large population limit is used to motivate a deterministic model, PALs are derived from approximate filtering equations for finite-population, stochastic compartmental models, and the large population limit drives the consistency of maximum PAL estimators. Our theoretical results appear to be the first likelihood-based parameter estimation consistency results applicable across a broad class of partially observed stochastic compartmental models concerning the large population limit. Compared to simulation-based methods such as Approximate Bayesian Computation and Sequential Monte Carlo, PALs are simple to implement, involving only elementary arithmetic operations and no tuning parameters; and fast to evaluate, requiring no simulation from the model and having computational cost independent of population size. Through examples, we demonstrate how PALs can be: embedded within Delayed Acceptance Particle Markov Chain Monte Carlo to facilitate Bayesian inference; used to fit an age-structured model of influenza, taking advantage of automatic differentiation in Stan; and applied to calibrate a spatial meta-population model of measles.

Informing policy via dynamic models: Cholera in Haiti

Jesse Wheeler (University of Michigan)

Policy decisions related to an infectious disease outbreak are often informed by incidence data and scientifically motivated dynamic models. The development of useful models requires addressing the tradeoff between biological fidelity and model simplicity, and the reality of misspecification for models at all levels of complexity. As a case study, we consider the 2010-2019 cholera epidemic in Haiti. We study three dynamic models developed by expert teams to advise on vaccination policies. We assess previous methods used for fitting and evaluating these models, and we develop data analysis strategies leading to improved statistical fit. Specifically, we present approaches to diagnosis of model misspecification, development of alternative models, and computational improvements in optimization, in the context of likelihood-based inference on nonlinear dynamic systems. Our workflow is reproducible and extendable, facilitating future investigations of this disease system.

4.4 Invited session on Phylogenetic inference

Tue 14th from 17:00 to 18:30. Chair: David Helekal.

A Bayesian model of acquisition and clearance of bacterial colonization incorporating within-host variation

Pekka Marttinen (Aalto University)

Bacterial populations that colonize a host can play important roles in host health, including serving as a reservoir that transmits to other hosts and from which invasive strains emerge, thus emphasizing the importance of understanding rates of acquisition and clearance of colonizing populations. Studies of colonization dynamics have been based on assessment of whether serial samples represent a single population or distinct colonization events. With the use of whole genome sequencing to determine genetic distance between isolates, a common solution to estimate acquisition and clearance rates has been to assume a fixed genetic distance threshold below which isolates are considered to represent the same strain. However, this approach is often inadequate to account for the diversity of the underlying within-host evolving population, the time intervals between consecutive measurements, and the uncertainty in the estimated acquisition and clearance rates. Here, we present a fully Bayesian model that provides probabilities of whether two strains should be considered the same, allowing us to determine bacterial clearance and acquisition from genomes sampled over time. Our method explicitly models within-host variation using population genetic simulation, and the inference is done using a combination of Approximate Bayesian Computation (ABC) and Markov Chain Monte Carlo (MCMC). We validate the method with multiple carefully conducted simulations and demonstrate its use in practice by analyzing a collection of methicillin resistant Staphylococcus aureus (MRSA) isolates from a large longitudinal clinical study.

Reference: Järvenpää et al. (2019). A Bayesian model of acquisition and clearance of bacterial colonization incorporating within-host variation. PLoS Computational Biology.

Bayesian inference of recombinant ancestries

Jere Koskela (Univeristy of Warwick)

DNA sequences are correlated due to common ancestry among individuals. In most cases ancestral relations and DNA sequences cannot be observed, necessitating a mathematical model for latent ancestries. The so-called coalescent with recombination, or CwR, is a gold-standard model in the genome-scale setting, where ancestral trees of different genes differ due to a biological process called recombination. However, imputing missing ancestries under the CwR is notoriously computationally expensive. I will introduce the CwR model, and show how recent progress in data structures for storing CwR realisations clarifies the exact reason for the computational bottleneck. I'll also demonstrate that despite their cost, MCMC algorithms for the CwR provide useful benchmarks for assessing biases and uncertainty in more scalable methods.

Bayesian Inference of Reproduction Number from Genomic and Epidemic Data using MCMC Methods

Alicia Gill (University of Warwick)

The reproduction number R(t) represents the average number of new infections caused by a single infected individual at time t. Estimation of the reproduction number R(t) is of vital importance during an epidemic outbreak, for example, to decide whether to implement control measures and to determine their effects once implemented. Typically, the reproduction number R(t) is inferred using only epidemic data, such as prevalence per day. However, prevalence data is often noisy, partially observed and biased. Genomic data is therefore increasingly being used to understand infectious disease epidemiology.

We take a Bayesian approach to this problem to find the trajectory of R(t) given a dated phylogeny and partial prevalence data using particle Markov chain Monte Carlo methods. We have implemented a particle marginal Metropolis–Hastings algorithm with backward simulation to jointly infer the hyper-parameters of the model, the latent epidemic and the trajectory of R(t). The performance of this approach is analysed using simulated data. These simulations show that incorporating genomic data as well as epidemic data improves inference in a variety of cases.

5 Posters

Within-host modeling to measure dynamics of antibody responses after natural infection or vaccination: A systematic review

Irene Garcia-Fogeda (Centre for Health Economics Research and Modelling Infectious Diseases (CHERMID), Vaccine & Infectious Diseases Institute (VAXINFECTIO), University of Antwerp, Antwerp, Belgium)

Background: Within-host models describe the dynamics of immune cells when encountering a pathogen, and how these dynamics can lead to an individual-specific immune response. This systematic review aims to summarize which within-host methodology has been used to study and quantify antibody kinetics after infection or vaccination. In particular, we focus on mechanistic models that study the evolution of antibodies and the underlying mechanisms that are responsible for their production. Materials: PubMed and Web of Science databases were used to identify eligible papers published until May 2022. Eligible publications included those studying mathematical models that measure antibody dynamics as the primary outcome (ranging from phenomenological to mechanistic models). Results: We identified 78 eligible publications, of which 8 relied on an Ordinary Differential Equations (ODEs)-based modelling approach to describe the antibody dynamics after vaccination, and 12 studies used such models in the context of humoral immunity induced by natural infection. Mechanistic modeling studies were summarized in terms of type of study, sample size, measurements collected, antibody half-life, compartments and parameters included, inferential or analytical method, and model selection. Conclusions: Despite the importance of investigating antibody kinetics and underlying mechanisms of (waning of) the humoral immunity, few publications explicitly account for antibody processes in a mathematical model. In particular, most research focuses on phenomenological rather than mechanistic models. The limited information of the age groups or other risk factors that might impact antibody dynamics, as well as a lack of experimental or observational data remain important concerns regarding the interpretation of mathematical modeling results. We gathered differences between vaccination and infection, and emphasize that it might be worthwhile to translate some of the characteristics from one setting to the other. However, we also stress that some of the biological mechanisms must be distinguished. We found that data-driven mechanistic models tend to be more simplistic, and theory-driven approaches lack representative data to validate the model results.

A flexible, random histogram kernel for discrete-time Hawkes processes

Raiha Browning (University of Warwick)

Hawkes processes are a self-exciting stochastic process used to describe phenomena whereby past events increase the probability of the occurrence of future events. This work presents a flexible approach for modelling a variant of these, namely discrete-time Hawkes processes. Most standard models of Hawkes processes rely on a parametric form for the function describing the influence of past events, referred to as the triggering kernel. This is likely to be insufficient to capture the true excitation pattern, particularly for complex data. We develop a bespoke reversible-jump Markov chain Monte Carlo sampler, whereby our proposed model for the triggering kernel can take the form of any step function, affording significantly more flexibility than a parametric form. We first demonstrate the utility of the proposed model through a comprehensive simulation study. This includes univariate scenarios, and multivariate scenarios whereby there are multiple interacting Hawkes processes. We then apply the proposed model to the interaction between Italy and France during the early to middle stages of the COVID-19 pandemic as an exemplar of two countries that are likely to have an interaction due to spatial proximity. We recover excitation patterns between these countries that agree with what occurred historically and provides useful insights into the transmission between these countries.

Integrated nested Laplace approximations for extended latent Gaussian models, with application to the Naomi HIV model

Adam Howes (Imperial College London)

Naomi (Eaton et al, 2021) is a spatial evidence synthesis model used to produce district-level HIV epidemic indicators in sub-Saharan Africa. Multiple outcomes of interest, including HIV

prevalence, HIV incidence and treatment coverage are jointly modelled using both household survey data and routinely reported health system data. The model is provided as a tool for countries to input their data to and generate estimates. In this setting, computationally intensive inference methods like MCMC are impractical. We propose a new inference method which combines the simplified integrated nested Laplace approximation approach of Wood (2020) with adaptive Gauss-Hermite quadrature to enable fast and accurate inference for Naomi and other extended latent Gaussian models. Using data from Malawi, our method provides substantially more accurate inferences than the empirical Bayes Gaussian approximation approach used currently, and is comparable to Hamiltonian Monte Carlo with the No-U-Turn sampler. By extending the aghq package (Stringer, 2021) we facilitate flexible and easy use of our method when provided a TMB C++ template for the log-posterior.

Bayesian Inference for the Structured Coalescent

Ian Roberts (University of Warwick)

The structured coalescent models the common ancestry of individuals drawn from a spatially structured population. A realisation of the process consists of a phylogenetic tree relating the samples alongside a migration history identifying the geographic location of each ancestor. Current inference methods either simultaneously infer the phylogenetic tree and plausible migration histories, which is computationally expensive, or rely on approximations of the structured coalescent, which introduces bias. I will present a Reversible Jump Markov Chain Monte Carlo (RJMCMC) scheme which strikes a balance between these extremes by sampling migration histories (along with governing static parameters) for a fixed phylogenetic tree under the full structured coalescent model, with demonstrations of its application to pathogen populations causing infectious diseases.

Bayesian Inference of Fitness Cost of Antimicrobial Resistance from Genomic Data

David Helekal (University Of Warwick)

Antimicrobial resistance of bacterial pathogens is a pressing public health problem. Resistance has a clear benefit for pathogen transmission in the presence of antimicrobial use. Increased antimicrobial use therefore leads to an increased net value of adaptions that confer resistance to it. However, often adaptations that confer resistance to antimicrobials come at a fitness cost in the absence of sufficient antimicrobial use. Quantifying the fitness cost of antimicrobial resistance as a function antimicrobial usage through time, along with the uncertainty associated is essential to understanding the emergence and dispersal of antimicrobial resistance as well for public health decision making. Using genomic data for this purpose has several advantages over traditional data. One such advantage is much greater resolution, avoiding averaging over lineages with distinct fitness parameters. Another advantage is possibility of using the coalescent process as an observational process, as the coalescent process is less sensitive to sampling considerations. It however also comes with additional challenges, such as the complex interplay between different lineages being only partially observed. We investigate how fitness cost and benefit of antimicrobial resistance can be inferred from genomic data in a Bayesian framework.

A Bayesian approach to integrate epidemiological and whole genome sequence data for analysing infectious disease outbreaks

Joseph Marsh (University of Nottingham)

Advances in sequencing technology and the reduction in associated costs have enabled scientists to obtain highly detailed genomic data on disease-causing pathogens on a scale never seen before. Combining genomic data with traditional epidemiological data (e.g. incidence data) provides a unique opportunity to determine the actual transmission pathway of the pathogen through a population. Despite recent advances, existing approaches have their own limitations, such as simplifications to the underlying biological processes, arbitrary phenomenological models or approximations to the likelihood function, to name a few.

We present a modelling framework for integrating epidemiological and whole genome sequence data to overcome the above limitations where (i) we use the matrix of pairwise horizontal distances between sequences as a summary statistic for the genetic data and (ii) explicitly derive

the joint probability distribution of pairwise genetic distances under the assumption of a mutation model. We develop computationally efficient data-augmentation MCMC algorithms to infer the transmission network and the unobserved pathogen distances at the time of transmission as well as the times of transmission. Finally, we demonstrate the performance of our framework on simulated data and also analyse an outbreak of S. aureus in an intensive care unit in Brighton during 2011-2012.

PDMP samplers for Bayesian models in epidemics.

Sebastiano Grazzi (University of Warwick)

Sampling the high-dimensional latent space of infection times of individuals in the SIR model with notifications and estimating relevant parameters for such models have been proven to be difficult for MCMC methods, as both the mixing of the underline processes and the complexity of the algorithms often scale poorly with the population size.

In this talk, I will present a novel Monte Carlo approach based on non-reversible piecewise deterministic Markov processes (PDMPs). The method proposed shows promise as it automatically sets individuals from 'susceptible' to 'infected' (and viceversa) based on the likelihood and explores efficiently the latent space by continuously moving all infection times according to the likelihood. The algorithm also exploits the local dependencies of the infection times to scale well with the population size.

The method presented here extends [1] and is joint work with J. Bierkens, G. Roberts, and M. Schauer.

[1] Sticky PDMP samplers for sparse and local inference problem; Bierkens J., Grazzi S., van der Maulen F., Schauer M.

Bayesian combination of longitudinal studies to estimate the duration of SARS-CoV-2 PCR positivity in the general population

Joshua Blake (MRC Biostatics Unit, University of Cambridge)

The duration of SARS-CoV-2 PCR positivity is vital to inferring the incidence of infection from prevalence data and informing the public health response. Current estimates do not provide a full distribution and are from subsets of the population (eg: healthcare workers or hospitalised patients). We utilise two longitudinal studies: the first study (the Assessment of Transmission and Contagiousness in Contacts of COVID-19, ATACCC) provides granular data (daily testing), however, suffers from short follow-up and an inadequate sample size; the second study (the Coronavirus Infection Survey, CIS) provides coarse data, however, has a large sample size and long-term follow-up. We aim to combine these datasets to provide unbiased estimates of the duration distribution in the general population.

We use different statistical models for each dataset, and then use transformed posterior estimates of the ATACCC analysis as a prior distribution in the CIS analysis. Using different models initially is desirable due to the differing structures of the studies and needing to analyse the CIS data within a compute-constrained trusted research environment. To enable the combination, we develop a hierarchical prior which respects the posterior correlations from the first analysis and allows incorporation of subjective model uncertainty to accurately represent our prior beliefs.

A new approach for the estimation of the contagiousness ratio between two viral strains

Giulia Della Croce (Politecnico di Torino)

The present work proposes a new method for the estimation of the ratio between the basic reproduction numbers of a new emerging variant and the one currently dominating within the population under analysis. Our method, which takes as input incidence data from random samples and the epidemic curves of total infections and recoveries, is based on a discrete-time SIR model with two strains. The SIR model is considered both in a deterministic and in a stochastic setting and, in this perspective, two different methodologies are presented. In the deterministic setting we apply the method of Maximum Likelihood. In the stochastic setting, instead, we need to reconstruct the missing information about the incidence and prevalence of the new variant within a hierarchical Bayesian model. Both methods are applied to the ISS

quick surveys data for the Piedmont Italian region in the period of December-January 2022, that is when the first Omicron variant of the SARS-CoV-2 virus started to be observed and quickly became prevalent. We show how it is possible to obtain an estimate of the contagiousness ratio between the two variants from public data which is consistent with other studies specifically designed to the aim.

Bayesian Individual-level Infectious Disease Modelling

Hannah Bensoussane (University of Warwick)

Fitting mathematical models to epidemic data is challenging because the transmission process is largely unobserved. Add into the mix that the characteristics of an individual can affect their ability to catch and transmit infectious diseases and resulting models can be both complex and computationally costly. Here we develop a model that allows us to make inference about underlying parameters that drive epidemics whilst accounting for individual-level covariates (e.g. age) and shared covariates – living in the same house, for example. The use of an adaptive MCMC algorithm featuring novel approaches to updating unknown infection times allows us to estimate model parameters and identify characteristics that affect transmission in a fashion that is computationally viable. A dataset providing a wealth of information on an outbreak of swine flu in a language school setting is used to both motivate model development and test model proficiency (1). Model performance is additionally assessed through a variety of simulation work.

(1) N. Arinaminpathy, N. Raphaely, L. Saldana, C. Hodgekiss, J. Dandridge, K. Knox, and N. D. McCarthy. Transmission and control in an institutional pandemic influenza A (H1N1) 2009 outbreak. Epidemiology and Infection, 140(6):1102–1110, 2012. ”

Inferring the impact of recent local rabies cases on human responses to exposed dogs and exploring the consequences for outbreak sizes

Elaine Ferguson (University of Glasgow)

Canine rabies typically circulates at low levels in the dog populations of endemic low- and middle-income countries; a characteristic difficult to reconcile with the presence of large free-roaming dog populations, an absence of acquired immunity, and limited dog vaccination. We hypothesised that large outbreaks may be prevented by human populations becoming more sensitised to rabies when there have been recent cases in the local area; so that exposed or rabid dogs become more likely to be restrained before they can transmit. Using 18 years of contact tracing data from the Serengeti District of Tanzania, we fitted a model in Stan (Hamiltonian Monte Carlo) to determine the impact of prior rabies cases on the probability that a rabid dog is restricted prior to biting, and how this impact changes with varying spatiotemporal distance from the focal dog. We then ran simulations of rabies transmission in Serengeti District using versions of a spatially-explicit individual-based model both with and without the fitted dog restriction mechanism. The time series of rabies cases and the outbreak sizes generated with incidence-dependent restriction were much closer to those observed in the data; indicating a key role of human behaviour in maintaining low rabies incidence.

Statistical Communication with Decision Makers During COVID Response

Cathal Walsh (University of Limerick)

Statistical Communication with Decision Makers During COVID Response

We reacted quickly to COVID in Ireland and formed a team of statisticians and mathematicians to inform the governmental response to the pandemic. This ensured early interventions at a policy level. Despite substantial resistance early shutdown was implemented. This poster highlights how we discussed probability and risk with decision makers and their response to it.

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

Evandro Konzen (University of Exeter)

To better understand the dynamics of infectious diseases of wildlife, it is crucial to be able to fit dynamic transmission models to observed data in a robust and efficient way. As key epidemiological events are at best only partially observed, it is necessary to infer the missing

information alongside the model parameters as part of the inference routine. The main challenge is that the computational burden of inference algorithms often increases non-linearly with population size and with increased dimensionality of the hidden states. With this in mind, we implement a recently proposed individual forwards-filtering backwards-sampling algorithm to a large-scale longitudinal study of bovine tuberculosis in badgers (the Woodchester Park study, comprising ¿ 2,700 badgers in 44 social groups over a 45-year period). We deal with many typical complexities of wildlife disease systems: incomplete sampling of individuals over time (capture-mark-recapture events), the use of multiple diagnostic tests, meta-population structures, and non-Markovian demographic aspects such as age-dependent mortality rates (with censoring), all alongside a hidden stochastic compartmental model of disease spread. The method produces full posterior distributions for the parameters, and predictive distributions of the hidden states over time for each individual, and fits in just a few hours.

Bayesian Poisson Regression and Decomposition Model for Learning Mortality Pattern Changes during COVID-19 Pandemic

Wei Zhang (Università della Svizzera italiana)

COVID-19 has led to excess deaths around the world, however it remains unclear how the mortality of other causes of death has changed during the pandemic. Aiming at understanding the wider impact of COVID-19 on other death causes, we use an Italian data set that consists of monthly mortality counts of different causes of death staring from pre-COVID-19 era to June 2020. Due to the high dimensional nature of the data, we develop a model which combines the conventional Poisson regression with tensor train decomposition to explore the lower dimensional structure of the data. We take a Bayesian approach and assume priors on model parameters. The posterior inference is made using an efficient Metropolis-Hastings within Gibbs algorithm. Our method provides informative interpretation in addition to the Poisson regression that conforms to our hypothesis of the relationship between COVID-19 and other causes of death.

Modelling and Bayesian Sequential Learning of Latent Epidemic Dynamics

Molly Cui (King's College London)

The global outbreak of the COVID-19 pandemic and its variants have severely affected the world health system and population. Sequential real-time disease surveillance models based on a Hidden Markov structure have played a prominent role in the evaluation and forecasting of epidemic infectious dynamics over time.

In this paper, we consider a novel Susceptible-Exposed (including Asymptomatic)-Infected-Recovered-Death (SE(A)IRD) deterministic epidemic compartment model with stochastic transmissions rate, in which we take into consideration the number of asymptomatic individuals that are still infectious. We also incorporate the effects of both pharmaceutical and non-pharmaceutical interventions by adopting a mean-reverting Ornstein-Uhlenbeck process with embedded lockdown and vaccination factors for the transmission rate time-varying parameter.

Bayesian inference is performed through the particle Markov chain Monte Carlo (p-MCMC) algorithm to simultaneously estimate parameters and latent epidemic status states. Model validation and performance checks are conducted on synthetic data in different model scenarios, followed by inference of real UK and EU regional daily COVID data. Our model is able to learn the emerging disease transmission dynamics only through very few simple and routine surveillance data inputs.

As a result, the inferred model output can be agilely generalised worldwide, removing the dependence of predictions on regional features.

Spatial non-parametric Bayesian clustered regression coefficients

Wala Areed (Queensland university of technology)

A subject of interest in population health research is defining similarities between regions and estimating their shard coefficients. In this study, we propose a spatial Dirichlet process clustered heterogeneous regression model with auxiliary variables that account for similarities in educational and demographic characteristics throughout the spatial domain. Our Bayesian non-parametric methodology provides an estimate for each cluster parameter as well as an inference

for the number of clusters and clustering configurations. The proposed approach is presented using simulated data and further used to investigate the influential variables for children's health development domain data in Queensland.

Index