Statistics and Nonlinear Dynamics in Biology and Medicine
July 28 - August 1, 2014

MEALS

*Breakfast (Buffet): 7:00–9:30 am, Sally Borden Building, Monday–Friday
*Lunch (Buffet): 11:30 am–1:30 pm, Sally Borden Building, Monday–Friday
*Dinner (Buffet): 5:30–7:30 pm, Sally Borden Building, Sunday–Thursday
Coffee Breaks: As per daily schedule, in the foyer of the TransCanada Pipeline Pavilion (TCPL)
***Please remember to scan your meal card at the host/hostess station in the dining room for each meal.**

MEETING ROOMS

All lectures will be held in the lecture theater in the TransCanada Pipelines Pavilion (TCPL). An LCD projector, a laptop, a document camera, and blackboards are available for presentations.

SCHEDULE

Sunday

16:00	Check-in begins (Front Desk - Professional Development Centre - open 24 hours)
17:30–19:30	Buffet Dinner, Sally Borden Building
20:00	Informal gathering in 2nd floor lounge, Corbett Hall (if desired) Beverages and a small assortment of snacks are available on a cash honor system.

Monday

7:00–8:45	Breakfast
8:45–9:00	Introduction and Welcome by BIRS Station Manager, TCPL
9:00–9:50	Hulin Wu *Scale-Up Parameter Estimation and Variable Selection for High-Dimensional Dynamic Models with Applications to Multi-Level Systems Biology Research*
9:50–10:40	Simon Wood *Simple Statistical Methods for Non-linear Ecological Models*
10:40–11:10	Coffee Break
11:10–12:00	Susanne Ditlevsen *Partially Observed Stochastic Models in Neuroscience*
11:30–13:00	Lunch
13:00–14:00	Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall
14:00	Group Photo; meet in foyer of TCPL (photograph will be taken outdoors so a jacket might be required).
14:00-15:00	4 Minute Talks
15:00-15:30	Coffee Break
15:30-17:30	4 Minute Talks
17:30–19:30	Dinner

Tuesday

7:00–9:00	Breakfast
9:00–9:40	Eberhard Voit *Identification of Metabolic Pathway Models*
9:45–10:30	Jens Timmer *Uncertainty Analysis in Systems Biology*
10:30–11:00	Coffee Break, TCPL
11:00–11:20	Itai Dattner *Statistical Inference for Systems of Ordinary differential Equations Linear in the Parameters*
11:20–11:40	Giles Hooker *Robustness, Inference and Gradient Matching*
11:40–12:00	Oksana Chkrebtii *Inference for Differential Equation Models with Discretization Uncertainty*
12:00–15:00	Lunch and Leisure
15:00–15:30	Coffee Break, TCPL
15:30–16:30	Aaron King *Forecasting Cholera Using Mechanistic Models*
16:30–17:00	Vanja Dukic *Bayesian Inference in Structured Epidemics*
17:00–17:30	Joon Ha Park *Inference for Disease Dynamics in Multiple Cities using Sequential Monte Carlo: A Case Study in Measles*
17:30–18:00	Jiguo Cao *Ordinary Differential Equation Models Selection*
18:00–19:30	Dinner

Wednesday

7:00–9:00	Breakfast
9:00–9:45	Junling Ma *Disease Dynamics on Random Contact Networks*
9:45–10:15	Matteo Fasiolo *An Extended Empirical Saddlepoint Approximation for Intractable Ecological Models*
10:45–11:15	Coffee Break, TCPL
11:15–12:00	Lea Popovic *Stochastic Dynamics in Intracellular Systems*
12:00–12:30	Darren Wilkinson *Likelihood-free Algorithms for Intractable Markov Processes*
12:30–13:30	Lunch
13:30–17:30	Free Afternoon, possible trip to Lake Louise
17:30–19:30	Dinner

Thursday

7:00–9:00	Breakfast
9:00–9:45	Ed Ionides *A New Iterated Filtering Algorithm*
9:45–10:30	Michael Dowd *Data Assimilation for Ocean Biology*
10:30–11:00	Coffee Break, TCPL
11:00–11:30	Alexandre Bouchard-Cote *Divide-and-Conquer Sequential Monte Carlo*
11:30–12:00	Simon Preston *Piecewise Approximate Bayesian Computation*
12:00–15:00	Lunch and Leisure
15:00–15:30	Coffee Break, TCPL
15:30–16:15	Greg Dwyer *Using Stochastic Models to Make Inferences About Pathogen Epidemics in Insect Populations*
16:15-17:00	Perry de Valpine *Programming with models using NIMBLE*
17:00–17:30	Steve Ellner *Markov Chain Models for Individual Life Paths and their Population-Level Consequences*
17:30–18:00	Scott McKinley *Sensing and Decision Making in Random Search*
18:00–19:30	Dinner

Friday

7:00–9:00	Breakfast
9:00 - 11:00	Dicussion *Stochastic versus Deterministic Models* Jim Ramsay, Cindy Greenwood, Ed Ionides Coffee break at 10:30.
11:30–13:30	Lunch
Checkout by 12 noon.	

** 5-day workshop participants are welcome to use BIRS facilities (BIRS Coffee Lounge, TCPL and Reading Room) until 3 pm on Friday, although participants are still required to checkout of the guest rooms by 12 noon. **

Statistics and Nonlinear Dynamics in Biology and Medicine
July 28 - August 1, 2014

ABSTRACTS
(in alphabetic order by speaker surname)

Speaker: **Alexendre Bouchard** University of British Columbia
Title: *Divide-and-Conquer Sequential Monte Carlo*
Abstract: I will describe Divide-and-Conquer Sequential Monte Carlo (D&C SMC), a method for performing inference on a collection of auxiliary distributions organized into a tree. In contrast to standard SMC samplers, D&C SMC exploits multiple populations of weighted particles, while still being an exact approximate method.

I will describe an application of this method to the problem of approximating intractable stochastic processes on a phylogenetic tree. Addressing this problem is in turn motivated by the gap between the complex models developed by evolutionary biologists and the much simpler models at the limit of what can currently be used to reconstruct trees and ancestral sequences.

Speaker: **Jiguo Cao** Simon Fraser University
Title: *Ordinary Differential Equation Models Selection*
Abstract: We consider model selection and estimation in a context where there are competing ordinary differential equation (ODE) models, and all the models are special cases of a "full" model. We propose a computationally inexpensive approach that employs statistical estimation of the full model, followed by a combination of a least squares approximation (LSA) and the adaptive Lasso. We show the resulting method, here called the LSA method, to be an (asymptotically) oracle model selection method. The finite sample performance of the proposed LSA method is investigated with Monte Carlo simulations, in which we examine the percentage of selecting true ODE models, the efficiency of the parameter estimation compared to simply using the full and true models, and coverage probabilities of the estimated confidence intervals for ODE parameters, all of which have satisfactory performances. Our method is also demonstrated by selecting the best predator-prey ODE to model a lynx and hare population dynamical system among some well-known and biologically interpretable ODE models.

Speaker: **Oksana Chkrebtii** Simon Fraser University
Title: *Inference for differential equation models with discretization uncertainty*
Abstract: Exact inference for differential equation models requires the ability to evaluate states explicitly for given parameter values. However, model solutions are rarely available in closed form and many existing inferential tools therefore rely on time discretization and the resulting approximate likelihood. In the first part of this talk, I will show that ignoring discretization uncertainty often results in biased parameter estimates, even for apparently simple ODE and PDE models. The second part of this talk will introduce a new formalism for modelling and propagating this uncertainty through the Bayesian inferential framework, allowing exact inference and uncertainty quantification for discretized differential equation models.

This is joint work with David Campbell (Simon Fraser University), Mark Girolami (University of Warwick), and Ben Calderhead (Imperial College London).

Speaker: **Itai Dattner** University of Haifa
Title: *Statistical inference for systems of ordinary differential equations linear in the parameters*
Abstract: The inverse problem of parameter estimation from noisy observations is a major challenge in statistical inference for dynamical systems. The focus of this talk will be on the fairly general and often applied class of systems of ordinary differential equations linear in the parameters. A new estimation methodology will be introduced. The estimation approach bypasses numerical integration, avoids the

estimation of slopes, and does not require a search over the parameter space. Furthermore, it can be used in both cases of fully or partially observed systems. Theoretical results, simulation studies and application to real data will be presented.

Speaker: **Susanne Ditlevsen** University of Copenhagen
Title: *Partially observed stochastic models in neuroscience*
Abstract: When constructing a model for a given system under study, decisions about characteristics and levels of detail of the model have to be taken. Which choices are appropriate depend on the questions, one wants to answer. It should also depend on available data, such that the model can exploit the information that can be extracted and not suffer too much by what cannot. I will present some examples where a simple model extracted from more biophysical based models can answer specific questions of interest, as long as the simple model is interpreted and used in a suitable way.

Speaker: **Michael Dowd** Dalhousie University
Title: *Data Assimilation for Ocean Biology*
Abstract: There has been an observation revolution for the ocean in recent decades. New technologies, such as satellites and autonomous underwater vehicles, complement the more traditional measures obtained from moored instruments and water sampling/surveys. In parallel, dynamic numerical models for oceanic systems have improved dramatically with increased computing power. A major challenge is to develop statistical approaches that efficiently and effectively combine these high dimensional spatio-temporal dynamic models and data, a problem that is termed data assimilation in the ocean sciences. In this talk, I explore the data assimilation problem as applied to ocean biology, focusing mainly on lower trophic levels (the planktonic ecosystem or marine biogeochemistry), but also mention approaches for higher trophic levels (fish and marine mammals). This talk is illustrated with various ongoing collaborative works. I argue that state space models and Bayesian approaches provide a unifying framework for state and parameter estimation for such systems, including the treatment of model identification and sampling design. Challenges and new statistical directions are emphasized.

Speaker: **Vanja Dukic** University of Colorado-Boulder
Title: *Bayesian Inference in Structured Epidemics*
Abstract: This talk will discuss Bayesian state space modeling and inference suitable for on-line epidemic surveillance, using flu as an example.

Speaker: **Greg Dwyer** University of Chicago
Title: *Using Stochastic Models to Make Inferences About Pathogen Epidemics in Insect Populations*
Abstract: Many forest insects undergo outbreaks, in which their densities rise from extremely low levels to levels at which defoliation is severe. Outbreaks damage economically valuable timber and can turn forests from carbon sinks to carbon sources, exacerbating climate change, and are therefore the subject of intensive research, in the Canadian and US Forest Services, and among academic researchers. Fortunately, peak populations usually crash because of epizootics (= epidemics in animal populations) of fatal, directly transmitted diseases known as baculoviruses. The severe, density-dependent mortality caused by baculoviruses strongly suggests that these pathogens help drive the long-period, large-amplitude cycles typical of many forest insect populations. Given the severity of the effects of the disease, however, an additional key question is, what is the role of natural selection in insect outbreaks?

Work in my lab attempts to answer this question by using field data to choose between competing models of baculovirus epizootics and forest insect outbreak cycles. For many infectious diseases, a major obstacle to direct application of mathematical models is that parameter inference is often very difficult. Fortunately, however, the transmission biology of baculoviruses is sufficiently simple that parameter estimation is relatively straightforward. Transmission occurs when insect larvae accidentally consume foliage contaminated with virus particles released from infectious cadavers of conspecifics (individuals of the same species). Because adults do not feed, they do not become infected, and because there is often only one

generation per year, epizootics are discrete events within the year. It is therefore possible to describe baculovirus epizootics with models of single epizootics.

We have therefore fit stochastic epizootic models to data on baculovirus infection rates in two insects, the gypsy moth (*Lymantria dispar*) and the Douglas-fir tussock moth (*Orgyia pesudostugata*). The model is a standard SEIR epidemic model, but it additionally assumes that transmission rates are drawn each day from a log-normal distribution with unknown mean and standard deviation, to be jointly inferred from the data. We then calculate an integrated likelihood, which we approximate as an average across stochastic realizations, using a Monte Carlo MISER algorithm to reduce the variance across realizations. Using MCMC with vague priors produced times between infection and death (incubation times) that were far too short and far too variable, and so we additionally used experimental data to elicit informative priors on incubation times. This latter approach then allowed us to show that a model that includes variability in host infection risk provides a much better explanation of the epizootic data than does a model that neglects such variability.

Variability in infection risk is important because it allows for a refuge from the disease, such that high variability leads to the unrealistic prediction of stable population dynamics in outbreak models without natural selection. We therefore used the posterior distribution of the parameters from fitting the epizootic data as informative priors on parameters of competing long-term models, such that one of the models includes host evolution and the other does not. Calculating likelihoods from long-term data on outbreaks in gypsy moth populations and Douglas-fir tussock moth populations shows that in both cases the evolutionary model has a lower value of the Deviance Information Criterion (DIC), supporting the conclusion that host evolution plays an important role in driving insect outbreaks.

Speaker: **Stephen P. Ellner** Cornell University
Title: *Some get lucky: Markov chain models for individual life paths and their population-level consequences*
Abstract: Some live, some die; some grow, some shrink; some have few offspring, some many. Add those all up, and you predict the dynamics of a population. But the rules of the game are specified at the level of individuals, and in a structured population model the rules describe changes in individual state (size, age, disease state, etc.) and their consequences. I will talk about models where the individual-level rules are a general state-space Markov chain in discrete time, leading to integrodifference equations at the population level. As background I will present some ecological applications (e.g., at what age and size should a thistle flower? how large is the random variation in total lifetime reproductive output?). But I aim to focus on things I don't yet know how to do, such as: (1) combining deterministic and stochastic individual-level state dynamics; (2) model selection with many potential environmental covariates; (3) how can we identify the key events or attributes that determine which individuals live long and have many offspring: is it determined by state at birth, by events early or later in life, by one stroke of good luck or by many, and so on.

Speaker: **Matteo Fasiolo** University of Bath
Title: *An Extended Empirical Saddlepoint Approximation for intractable ecological models*
Abstract: The use of simulation-based inferential approaches is widespread in computational biology and ecology. Here we focus on one such approach: Synthetic Likelihood (SL) (Wood, 2010). This method reduces the observed and simulated data into a set of features or summary statistics, and quantifies the discrepancy between them through a synthetic likelihood function. While requiring less tuning than some alternative approaches (such as Approximate Bayesian Computation), SL has the drawback of relying on the summary statistics being approximately normally distributed. We will describe how this shortcoming can be addressed by using a flexible density estimator: the Extended Empirical Saddlepoint Approximation (E-ESA). This estimator has the advantage of being scalable to high dimensions and fast to compute, while being able to capture large departures from normality. In addition, we alleviate the computational burden of SL by using the stochastic optimization routine of Ionides et al. (2006) to maximize the synthetic likelihood.

References:

Simon N Wood. Statistical inference for noisy nonlinear ecological dynamic systems. Nature, 466(7310):1102?1104 2010.

EL Ionides, C Breto, and AA King. Inference for nonlinear dynamical systems. Proceedings of the National Academy of Sciences, 103(49):18438?18443, 2006.

Speaker: **Giles Hooker** Cornell University
Title: *Robustness, Inference and Gradient Matching*
Abstract: Considerable recent statistical attention has been given to indirect methods for estimating parameters in ordinary differential equations (ODEs). Gradient matching – also known as two-stage least squares – involves first obtaining a non-parametric smooth of the data and then choosing the parameters to match the smooth of the derivative to the right-hand side function of the ODE. This procedure is often justified by removing the computational cost of repeatedly solving the ODE numerically and as resulting in less complex optimization problems.

In this talk I suggest that one of the most important justifications for gradient matching is robustness to system mis-specification. Empirically, parameter estimates from perturbed systems obtained by gradient matching exhibit less bias than those obtained by fitting ODE solutions directly, although this is difficult to quantify formally. More importantly, if system disturbances are regarded as being random, gradient matching allows us to incorporate this source of variation into inference about parameters. I suggest some means to do this and demonstrate that they yield more nearly correct coverage than ignoring these effects.

Speaker: **Ed Ionides** University of Michigan
Title: *A New Iterated Filtering Algorithm*
Abstract: Iterated filtering algorithms recursively combine parameter pertur- bations with latent variable reconstruction, providing stochastic op- timization procedures for latent variable models. Previously, theoretical support for these algorithms was based on using conditional moments of the perturbed parameters to approximate derivatives of the log likelihood function. A new theoretical approach is intro- duced based on the convergence of an iterated Bayes map. A new algorithm supported by this theory leads to substantial numerical im- provements on a computational challenge, inferring parameters of a partially observed Markov process.

Speaker: **Aaron King** University of Micigan
Title: *Forecasting cholera using mechanistic models*
Abstract: Mechanistic models have long been used to great effect in epidemiology to both ground and expand scientific understanding of basic processes of transmission, immunity, and disease and as tools for developing public health policy. By contrast, only rarely have such models been used to make true forecasts and their utility as forecasting instruments remains largely unexplored. Indeed, there is no strong reason to expect such models to perform better at forecasting than more unconstrained schemes. In this talk, I describe recent work aimed at developing practical forecasting technology for cholera in its endemic home, Bangladesh.

Speaker: **Junling Ma** University of Victoria
Title: *Disease dynamics on random contact networks*
Abstract: Contact networks can represent contact heterogeneity and fixed partners. On random contact networks without clustering (triangles), contact heterogeneity implies that the basic reproduction number is determined by the average degree of a node found by following a random edge. Fixed partners imply that a node cannot reinfect its neighbors before its neighbours recover, and thus SIS and SIR epidemic have different basic reproduction numbers on the same contact network. In addition, the serial interval of a node depends on the number of susceptible neighbors, and thus shortens as its neighbours are infected. Hence, it is difficult to impute the basic reproduction number from the exponential growth rate. When the average degree becomes large, the dynamics of network models approach that of the homogeneous mixing models. These results are illustrated by the Miler network SIR model and our new network SIS model.

Speaker: **Scott McKinley** University of Florida
Title: *Sensing and Decision Making in Random Search*
Abstract: Ecologists and wildlife conservationists use animal tracking data to characterize home territories, to locate nesting regions and favored resource locations; and to discover daily and annual activity cycles. A primary step in each of these tasks is to develop a quantitative description of individual paths. Researchers have repeatedly reported power-law tails in inter-observation relocation distributions, but the consequent use of scale-free movement models belies the fact that biological search is intrinsically a multi-scale process. At large scales animals seek out land regions that seem to have favorable conditions for containing their target of interest. At shorter length scales, animals engage in active local movement in pursuit of their targets, likely responding to a rapid succession of prey signal cues. Interestingly, when multi-scale models for animal movement are introduced to classical ecological frameworks, there are novel consequences for population-level processes. In this talk we will consider the impact on assumptions about encounter rates and functional response for theoretical predators that have various degrees of sensing and response to targets.

Speaker: **Joon Ha Park** University of Michigan
Title: *Inference for disease dynamics in multiple cities using Sequential Monte Carlo: A Case Study in Measles*
Abstract: Inference for nonlinear dynamic systems of high dimension has been regarded as computationally heavy problems. As a result, little efforts have been made to jointly estimate disease epidemics in multiple cities. In this presentation, we propose a variation of Sequential Monte Carlo (SMC) method for estimating latent states which can provide a computationally feasible solution to the joint estimation of dynamic trajectories of interacting systems. This method, which we name Space-Time Filter (STF), is expected to reduce the computational cost by a huge amount when the sub-systems are weakly interacting, while achieving the desired property that the sampled latent states form a proper sample from its true distribution according to the underlying model. We apply this method to the measles epidemic in the UK from 1950 to 1953. The analysis performed on the five largest cities by population shows that the proposed Space-Time Filter yields a reasonable estimate of epidemic history with relatively low computational cost. The conventional SMC method applied on the same set of data could not generate any result. We also estimate key epidemic parameters using the Iterated Filtering method (Ionides et. al, 2011), and show that the transmission rate is substantially different between during school term and during school holidays.

Speaker: **Lea Popovic** Concordia University
Title: *Stochastic dynamics in intracellular systems*
Abstract: Cellular functions in biological organisms comprise of complex interactions of different proteins, DNA, mRNA molecules, and others. Sources of stochasticity in cells are multiple: some are due to inherent randomness of biochemical reactions between the species, while others are due to variations in cellular composition, cellular division mechanisms, etc. I will present mathematical results characterizing some phenomena observed in intracellular systems for which stochasticity is responsible. These include: stochastic switching, diffusion moderated sensitivity, and sharpening of spatial patterns.

Speaker: **Simon Preston** University of Nottingham
Title: *Piecewise Approximate Bayesian Computation*
Abstract: I will discuss Piecewise Approximate Bayesian Computation (PW-ABC), an inference approach for discretely observed Markov models that involves dividing the dataset into subsets and using ABC within each subset. Such an approach has the benefits of being easy to parallelise and of reducing the dimension of the data used by the ABC. The reduced dimension aids in avoiding use of summary statistics and large tolerances both of which are ordinarily needed in ABC but which result in the ABC posterior being different from the true. The main challenge in PW-ABC is in putting together the collection of ABC samples to estimate the full posterior density, a task for which I will discuss two strategies, one using Gaussian approximations and another using kernel density estimates. I will discuss behaviour in the limit

as the number of ABC samples tends to infinity, and present some numerical results to illustrate practical performance.

I will also briefly mention how an analogous piecewise approach for inference in ODE models relates to, but is distinct from, the multiple shooting method of Bock, and can be thought of as interpolating between non-linear least squares and gradient matching.

Speaker: **Jens Timmer** University of Freiburg
Title: *Uncertainty Analysis in Systems Biology*
Abstract: Dynamical models of biological systems usually contain unknown parameters that have to estimated from time-resolved data. This involves at least four types of uncertainties: (i) uncertainty about finding the global optimum, (ii) uncertainty of the estimated parameters due to uncertainty in the data, (iii) uncertainty of model predictions due to uncertainty of the estimated parameters, (iv) uncertainty about the model structure. We discuss reasons for and a simple procedure to deal with uncertainty (i), show how the profile likelihood can deal with uncertainties (ii) and (iii), and will briefly touch uncertainty (iv).

Literature

A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelker, D. Kaschek, S. Hug, C. Kreutz, B.D. Harms, F.J. Theis, U. Klingmuller, J. Timmer. Lessons learned from quantitative dynamical modeling in systems biology. PLoS ONE 8, 2013, e74335

C. Kreutz, A. Raue, D. Kaschek, J. Timmer. Profile likelihood in systems biology. FEBS Journal 280, 2013, 2564-2571

A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmuller, J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25, 2009, 1923-1929

Speaker: **Perry de Valpine** University of California-Berkeley
Title: *Programming with models using NIMBLE*
Abstract: I will introduce a new software package called NIMBLE that we have just released in version 0.1 at R-nimble.org. First I will motivate the need to be able to write algorithms that can adapt to different model structures, such as MCMC, particle filters, Gaussian quadrature, approximate Bayesian computation, bridge sampling, and many more. These will come from ecological population dynamics problems such as state-space models and cohort development models, but the software is not specific to ecological models. Then I will show three components of NIMBLE. First, it allows models written in the BUGS language for model specification to be turned into native R objects that can be used in programming. Doing so can include inspecting the model graph to determine model structure and controlling the calculation and simulation of sets of nodes. Under the hood the model computations can be compiled via C++ for fast execution. Second, it provides a small language, embedded within and with similar syntax to R, for writing algorithms that can operate on BUGS models and can also be compiled via C++. And third, it provides the beginnings of an algorithm library, including MCMC. Examples will be given of the utility of keeping high-level processing in R and having low-level processing compiled via C++ and interfaced through R functions and objects. For example, an MCMC specification is first obtained as an R object, which can then be customized before generating and compiling the working MCMC functions.

Speaker: **Eberhard Voit** Georgia Institute of Technology
Title: *Identification of Metabolic Pathway Models*
Abstract: Over the past decade, time series data have become available in biology at an increasing rate. The trend is to be welcomed, as these data contain enormous information, which however is implicit and needs to be extracted with computational means. Time series data are particularly beneficial for analyses of metabolic pathway systems, because these are strongly constrained by stoichiometric and other intrinsic features, which effectively bound the space of admissible parameter values that need to be specified in order to translate the pathway system into a computable structure. Even within these bounds, the estimation

of suitable parameter values is a major bottleneck in the harvesting of information from time series data. The overriding quality criterion for the result of this step is usually the sum of squared residual errors between data and model. While this criterion is a natural starting point, there are various instances where a good fit alone is not sufficient. I will discuss several examples of such instances. Beyond the estimation of parameter values, an even greater challenge is the fact that the most appropriate functional forms for describing biological processes are not even known when a system is to be estimated. This structural uncertainty clearly complicates any estimation strategy. I will show that the problem can be ameliorated to some degree if the right types of time series data are available.

Voit, E.O.: What if the fit is unfit? Criteria for biological systems estimation beyond residual errors. In: M. Dehmer, F. Emmert-Streib and A. Salvador (Eds.): Applied Statistics for Biological Networks. J. Wiley and Sons, New York, pp. 183-200, 2011.

Goel, G., I-C. Chou, and E.O. Voit: System estimation from metabolic time series data. Bioinformatics 24, 2505-2511, 2008.

Iwata, M., F. Shiraishi, and E.O. Voit: Coarse but efficient identification of metabolic pathway systems. Int. J. Syst. Biol. 4(1), 57-72, 2013.

Voit, E.O.: Characterizability of metabolic pathway systems from time series data. Math. Biosc. 246(2):315-25, 2013.

Speaker: **Darren Wilkinson** Newcastle University
Title: *Likelihood-free algorithms for intractable Markov processes*
Abstract: Inferring the parameters of continuous-time Markov process models using partial discrete-time observations is an important practical problem in many fields of scientific research. Such models are very often "intractable", in the sense that the transition kernel of the process cannot be described in closed form, and is difficult to approximate well. Nevertheless, it is often possible to forward simulate (exact) realisations of trajectories of the process using stochastic simulation. There have been a number of recent developments in the literature relevant to the parameter estimation problem, involving a mixture of approximate, sequential and Markov chain Monte Carlo methods. This talk will compare some of the different "likelihood free" algorithms that have been proposed. Particular emphasis will be placed on the problem of Bayesian parameter inference for the rate constants of stochastic biochemical network models, using noisy, partial high-resolution time course data, such as that obtained from single-cell fluorescence microscopy studies.

Speaker: **Simon Wood** University of Bath
Title: *Simple statistical methods for non-linear ecological models*
Abstract: This talk will present some of the difficulties inherent in attempting statistical inference using highly non-linear models for relatively short ecological time series. The main issues are that i) useful models often do not attempt to get every aspect of the dynamics right and ii) chaotic or near chaotic behaviour presents inferential difficulties. The talk will offer some comparisons of the alternative inferential approaches that can be taken, in particular contrasting state space methods (such as direct MCMC and filtering), with information reduction approaches (such as ABC of Sythetic Likelihood).

Speaker: **Hulin Wu** University of Rochester
Title: *Scale-Up Parameter Estimation and Variable Selection for High-Dimensional Dynamic Models with Applications to Multi-Level Systems Biology Research*
Abstract: Many systems in engineering and physics can be represented by differential equations, which can be derived from well-established physics laws and theories. However, currently no laws or theories exist to deduce exact quantitative relationships/interactions among elements in a biological system. Even it is unclear whether the biological systems follow a mathematical representation such as differential equations, similar to that for a man-made physics or engineering system. Fortunately, recent advances in cutting-edge biomedical technologies allow us to generate intensive high-throughput data to gain insights into biological systems. However, a biological system is often comprised of multi-level and high-dimensional elements

such as genes, proteins, molecules, and cells, which is very challenging to develop a mathematical model to describe their complex relationships and interactions. Thus, it is badly needed to develop novel statistical methods to construct high-dimensional mathematical models such as differential equation models based on high-throughput experimental data. In this talk, I will present and discuss how to construct data-driven ordinary differential equations (ODE) to describe biological systems, in particular for dynamic gene and protein regulatory network systems. The ODE models allow us to quantify both positive and negative regulations as well as feedback effects. We propose to combine the high-dimensional variable selection approaches and ODE model estimation methods to construct the ODE models based on experimental data. We will also discuss the possibility to scale up the proposed methods to handle more than 1000 ODEs with more than one million unknown parameters in the era of Big Data. We apply the proposed approaches to study immune response to influenza infections as illustrations.