

# BISP7

## Seventh Workshop on BAYESIAN INFERENCE IN STOCHASTIC PROCESSES

### Abstracts of Talks

**Antónia Amaral Turkman**<sup>1,2</sup>, Sandra Ramos<sup>2,3</sup>, Marília Antunes<sup>1,2</sup> and Valeska Andreozzi<sup>2</sup>

<sup>1</sup>Faculty of Sciences, DEIO, University of Lisbon, <sup>2</sup>Center of Statistics and Applications, University of Lisbon, <sup>3</sup>Polytechnic Institute of Oporto, ISEP

#### Optimal Bayesian semiparametric alarm system for epidemiological data

The objective of screening is to decide whether an individual is a success (a response  $Y$  belongs to a known region  $C_Y$ ) or not, based on the information of a continuous  $d$ -dimensional feature vector  $X$  informative about  $Y$ . Specifically, the screening procedure involves the construction of a specification region  $C_X$  in the  $d$ -dimensional space such that future individuals with a feature vector in  $C_X$  are more likely to be a success. Within the Bayesian predictive framework, the optimal region  $C_X$  is constructed such that the predictive probability of success, conditional on the feature vector being in  $C_X$ , is increased to a desired level, optimizing certain operational characteristics. Suppose that, at a certain time  $t$ , the successful event corresponds to the process of interest  $Y$  exceeding a threshold  $u$  and let  $X$  be a vector of time-dependent variables. In such case, screening methodology can be developed as an optimal alarm system with an alarm being given at time  $t$ -lag, when the observed values of  $X$  belong to the optimal region  $C_X$ . In this talk, a Bayesian semi-parametric solution will be proposed for this problem and will be applied to predict the critical values of climatic conditions during the wet season which trigger an increase above a certain threshold in the incidence of dengue in the city of Rio de Janeiro.

**Keywords:** epidemics; screening; alarm system.

**R. Argiento, S. Galli and A. Pievatolo**

CNR IMATI, Milan

Bayesian analysis of alternating renewal processes with time-varying covariates

We model the on/off state of a residential combined heat and power generator as an alternating renewal process with time varying covariates (climatic variables and thermostat settings). We estimate the curve of the electricity produced by a large number of cogenerators during a typical day, taking parameter estimation uncertainty into account. This is made possible by an approximation of the stationary distribution of a characteristic of the process.

**Keywords:** electricity production; renewal processes

**Sergio Bacallado**

Stanford University

Bayesian analysis of variable-order reversible Markov chains, with applications to molecular dynamics

Reversible Markov chains are a mainstay of statistical modeling and simulation, as well as a common characterization of physical dynamics with microscopic reversibility or detailed balance. A discrete function of a reversible Markov chain, such as a lumping, does not in general satisfy the Markov property, but in many cases the extent of memory in the process becomes negligible after some time, motivating the use of higher-order and variable-order Markov models. When the underlying chain is reversible, the higher-order models satisfy certain constraints which we characterize as "trajectorial reversibility" [1].

We define a family of conjugate priors for variable-order Markov chains with trajectorial reversibility. The priors are defined via random walks with reinforcement, which are recurrent and partially exchangeable, and by de Finetti's theorem for Markov chains [2], a mixture of variable-order Markov chains with trajectorial reversibility. The priors facilitate Bayesian inference and model comparison. They are also endowed with a natural sampling scheme based on reinforced random walks. We show the utility of the priors in the analysis of molecular dynamics simulations.

**Keywords:** reversible Markov chains; trajectorial reversibility.

**References**

[1] Bacallado, S. Bayesian analysis of variable-order, reversible Markov chains. *Ann. Stat.*, to appear, (2011).

[2] Diaconis, P. and Freedman, D. de Finetti's theorem for Markov chains. *Ann. Probab.*, **8-1**, p. 115 (1980).

## **Toros Caglar and Refik Soyer**

The George Washington University

Markov modulated queues have previously been addressed in areas needing to account for varying arrival and service rates dependent on a randomly changing environment. Discrete time queues and their Markov modulated extensions have also been studied previously, with an emphasis on the computer and communication system applications. In this work, we will be focusing on the Bayesian analysis of the Markov modulated discrete time queues. Our objective is to apply discrete time queues to systems that are not inherently discrete (e.g. emergency departments). In order to achieve this objective, we consider batch arrivals into and departures from the system. Specifically, we are interested in the Markov modulated  $Geo^X/Geo/C$  queue. We will mostly concentrate on the case where the service and arrival processes modulate based on a common environment. We will deal with label-switching problems through dependent prior analysis and discuss estimation of the number of states in the environment. Lastly, predictive distributions of certain performance measures as well as possible extensions of the inferential methods to queueing networks will be discussed.

**Keywords:** Discrete time queues; hidden Markov models; dependent prior analysis; label switching; hidden environment size estimation

## **Vanja Dukic**

University of Colorado

### Bayesian Inference in Structured Stochastic Epidemics

This talk will present an overview of the mean-field theory in structured epidemic processes, and some of the recent findings for epidemics with two levels of mixing. Bayesian inference based on underlying inhomogeneous Poisson process models, as well as extensions to processes with an unknown structure and an unknown number of levels of mixing will be discussed.

**Keywords:** epidemics; inhomogeneous Poisson processes

## Paul Fearnhead

Lancaster University

Inference for multivariate diffusion processes avoiding time-discretisation error

Bayesian inference for diffusion processes modelled by multivariate stochastic differential equations (SDEs) is a very challenging task due to the intractability of the dynamics of the process. Most methods rely on high frequency imputation and discrete-time approximations of the continuous-time model, thus leading to biased inference. Recently, methods that are able to perform inference for univariate diffusions which avoid time-discretisation errors have been developed. However these approaches cannot be applied to general multivariate diffusions.

We present a novel, continuous-time Importance Sampling method that allows inference for general continuous-time Markov processes whilst avoiding time-discretisation errors. The method can be derived as a limiting case of a discrete-time sequential importance sampler, and uses ideas from random-weight particle filters, retrospective sampling and Rao-Blackwellisation. We show how this method can be applied to perform inference for general multivariate diffusions.

**Keywords:** multivariate diffusion processes; continuous time importance sampling

**Thais C O Fonseca**<sup>1</sup> and Mark F J Steel<sup>2</sup>

<sup>1</sup>Universidade Federal do Rio de Janeiro, <sup>2</sup>Warwick University

On flexible modelling of spatiotemporal processes

In this work we construct flexible nonseparable covariance functions for processes that vary continuously in space and time. Stochastic modeling of phenomena over space and time is important in many areas of application. But choice of an appropriate model can be difficult as one needs to ensure that valid covariance structures are used. Although separable covariance functions are guaranteed to be valid, they are severely limited, since they do not allow space-time interactions. In this work, we propose a general and flexible class of valid nonseparable covariance functions based on mixture of separable models. The proposed model allows for different degrees of smoothness across space and time and long-range dependence in time. Moreover, the proposed class has as particular cases several popular covariance models proposed in the literature such as the Matern and the Cauchy Class. Some relevant extensions are developed such as nongaussianity, asymmetry and nonstationarity modelling. These extensions allow for more realistic modelling of complex spatiotemporal systems. The potential of these proposals is illustrated with applications to real and simulated data sets.

**Keywords:** spatiotemporal process; covariance models.

## **Flávio Gonçalves**

Universidade Federal de Juiz de Fora

Exact Bayesian inference for jump-diffusion processes with stochastic jump rate

Jump-diffusion processes have considerable appeal as flexible families of stochastic models and therefore have been used in numerous existing and emerging applications.

This gives rise to the challenging statistical problem of making inference for these processes. The infinite-dimensional nature of the problem has required from existing inference methodologies the use of discrete approximations that, naturally, represent a considerable source of error in the inference process. We propose a simulation-based method to make exact Bayesian inference for discretely observed jump-diffusions. The method is exact in the sense that there is no discretisation error involved.

The exactness of the method comes with a price though. Our methodology is, so far, restricted to univariate processes which exclude, for example, stochastic volatility models. Nevertheless, we also propose a new class of models that allows the jump rate itself to be stochastic and to evolve according to a diffusion process. The main contribution of such models is to incorporate a stochastic component to the total volatility of the model and still make inference without discretisation error.

**Keywords:** jump-diffusion process; volatility

**Brett Houlding**<sup>1</sup>, Simon Wilson<sup>1</sup> and Mark Costello<sup>2</sup>

<sup>1</sup>Discipline of Statistics, Trinity College, Dublin, Ireland, <sup>2</sup>Dept. of Marine Sciences, University of Auckland, Leigh Marine Laboratory, New Zealand

Bayesian estimation of the number of unknown species: incorporating a model for the discovery process.

We consider the problem of estimating the number of species within an ecosystem. The usual approach is to estimate the number of distinct species by considering a sample of individuals and noting how many species are included and their relative abundances. Nevertheless, species discovery records do not usually contain information relating to abundance, so estimating global species numbers requires

an alternative approach that does not rely on either sample size, or sample species abundances, being available.

Here we assume that sampling occurs sequentially via a temporal process, but that only the number of discovered species and their first observation times are available. To perform estimation of the number of unknown species using such data, the first observation times are linked to the unknown sample abundances via a latent 'effort' process which is modeled as a stochastic process. Furthermore, noisy information related to this process is assumed available through annual taxonomic description rates. A series of constraints are then derived which relate sample abundances with known first time of species observation. Finally, with estimates of the number of samples taken between species discoveries, inference is performed on the number of unknown species through traditional sampling theory approaches to species estimation.

**Keywords:** Species estimation; stochastic discovery effort; temporal sampling process.

## Janine Ilian

University of St Andrews

### A class of non-separable spatio-temporal point process models

Due to rapidly improving technology that facilitates data collection, and a growing awareness of the importance and relevance of small scale spatial information, spatial and spatio-temporal point pattern data sets have become increasingly available. In this talk we will focus on spatio-temporal log-Gaussian Cox models, that is, models in which the logarithm of the intensity surface is modelled by a Gaussian random field.

The practical difficulty with these models is that inference is extremely computationally intensive and therefore model simplifications are required in order to make inference feasible. One particularly common simplification is to assume that the covariance function of the latent Gaussian field factors into purely spatial and purely temporal components, however, it is tricky to justify this separability assumption in applications. In this talk we consider a class of computationally efficient non-separable models for the latent field that are (marginally) Matérn fields by extending the stochastic partial differential equation models recently introduced by Lindgren, Rue and Lindström (2011). The key point is that the latent Gaussian field can be represented as a Gaussian Markov random field, allowing for fast inference using tools like INLA.

**Keywords:** space time process; Cox model; Gaussian random field

Nicholas Chopin<sup>1</sup>, **Pierre Jacob**<sup>1</sup> and Omiros Papaspiliopoulos<sup>2</sup>

<sup>1</sup>CREST-ENSAE, <sup>2</sup>UPF, Barcelona

## SMC<sup>2</sup>: A Sequential Monte Carlo Algorithm with Particle Markov Chain Monte Carlo Updates

We consider the generic problem of performing sequential Bayesian inference in a state-space model with observation process  $y$ , state process  $x$  and fixed parameter  $\theta$ .

An idealized approach would be to apply the iterated batch importance sampling (IBIS) algorithm of Chopin (2002). This is a sequential Monte Carlo algorithm in the  $\theta$ -dimension, that samples values of  $\theta$ , reweights iteratively these values using the likelihood increments  $p(y_t|y_{1:t-1}, \theta)$ , and rejuvenates the  $\theta$ -particles through a resampling step and a MCMC update step. In state-space models these likelihood increments are intractable in most cases, but they may be unbiasedly estimated by a particle filter in the  $x$ -dimension, for any fixed  $\theta$ . This motivates the SMC<sup>2</sup> algorithm proposed in this article: a sequential Monte Carlo algorithm, defined in the  $\theta$ -dimension, which propagates and resamples many particle filters in the  $x$ -dimension. The filters in the  $x$ -dimension are an example of the random weight particle filter as in Fearnhead et al. (2010). On the other hand, the particle Markov chain Monte Carlo (PMCMC) framework developed in Andrieu et al. (2010) allows us to design appropriate MCMC rejuvenation steps. Thus, the  $\theta$ -particles target the correct posterior distribution at each iteration  $t$ , despite the intractability of the likelihood increments. We explore the applicability of our algorithm in both sequential and non-sequential applications and consider various degrees of freedom, as for example increasing dynamically the number of  $x$ -particles. We contrast our approach to various competing methods, both conceptually and empirically through a detailed simulation study, included here and in a supplement.

**Keywords:** Iterated batch importance sampling; particle filtering; particle Markov chain Monte Carlo; sequential Monte Carlo; state-space models

**Fabrizio Leisen**

Department of Statistics, Universidad Carlos III de Madrid

Conditionally identically distributed sequences and Bayesian nonparametrics

Conditional identity in distribution (Berti et al. (2004)) is a new type of dependence for random variables, which generalizes the well-known notion of exchangeability. In Bassetti, Crimaldi, Leisen (2010, *Advances in Applied Probability*) a class of random sequences, called *Generalized Species Sampling Sequences*, is defined

and a condition to have conditional identity in distribution is given. In particular, a class of generalized species sampling sequences that are conditionally identically distributed is introduced and studied: the *Generalized Ottawa sequences* (GOS). Species sampling sequences is an important tool in Bayesian non parametrics. Indeed, many popular Bayesian nonparametric priors can be characterized in terms of exchangeable species sampling sequences. One example is the Dirichlet Process prior that has been increasingly used for modeling purposes in mixture of DP hierarchical models. However, in some applications, the implied exchangeability assumption may not be considered appropriate. In the GOS family it is available a species sampling sequence, characterized by a tractable predictive probability function, and with weights driven by a sequence of independent Beta random variables. We discuss some of the properties that can be useful in applications, and we compare our findings with well known properties of the DP and the two parameter Poisson-Dirichlet process. We detail on Markov Chain Monte Carlo posterior sampling, and illustrate the behavior of such priors.

**Keywords:** nonparametrics; species sampling; Dirichlet process.

**Ioanna Manolopoulou**, Melanie P. Matheu, Michael D. Cahalan, Mike West and  
Thomas B. Kepler

Duke University

Semi-parametric Bayesian modelling of inhomogeneous tactic fields in single-cell motility

We develop dynamic models of single cell motion involving nonparametric representations of nonlinear spatial fields that guide cellular motility. Assuming a discretized diffusion model for the cell motion, the tactic field is flexibly modelled using radial basis kernel regression. Our methods are motivated by the temporal dynamics of lymphocytes in the lymph nodes, critical to the immune response. The primary goal is learning the structure of the tactic fields that fundamentally characterize the immune cell motion. We develop Bayesian analysis via customized Markov chain Monte Carlo methods for single cell models, and multi-cell hierarchical extensions for aggregating models and data across multiple cells. Our implementation explores data from multi-photon vital microscopy in murine lymph node experiments, and we use a number of visualization tools to summarize and compare posterior inferences on the 3-dimensional tactic fields.

**Keywords:** cellular motility; kernel regression; diffusion models.

Joaquín Míguez<sup>1</sup> and Dan Crisan<sup>2</sup>

<sup>1</sup>Universidad Carlos III de Madrid, <sup>2</sup>Imperial College, London

## Particle Approximation of the Filtering Density and Its Derivates in General State Space Models

Sequential Monte Carlo (SMC) methods, also known as particle filters, have become a very popular tool for (numerical) Bayesian filtering in state space models. The asymptotic convergence of this class of algorithms (as the number of *particles*, or samples in the state space, grows) has been well studied. Most of these results, however, refer to the approximation of the filtering measure (the posterior probability distribution of the state of the system), i.e., they ensure that a posteriori expectations of functions of the system state can be accurately approximated by weighted sums.

In many cases of practical interest, the filtering measure has an associated probability density function (pdf). In the talk, we will analyze the approximation of this density, and its derivates, using particles and sequences of arbitrary kernels. In particular, we will show how to obtain uniform (over the space of the system states) convergence rates for the absolute error between the true pdf and its particle-kernel approximation. These results can be extended to the approximation of derivates of the density and hold under very mild assumptions regarding both the SMC algorithms and the kernel functions.

In the second part of the talk, we explore some applications of the proposed particle-kernel approximations. First, we prove that the approximation error of the filtering density converges to 0 almost surely (a.s.) in an  $L_1$  norm and this, in turn, implies in a straightforward way that the particle approximation of the filtering measure converges a.s. to the true measure in total variation distance. Then we consider the numerical implementation of maximum a posteriori (MAP) estimators and, finally, study the approximation of functionals of the filtering pdf. This includes, in a first approach, bounded and Lipschitz continuous functionals. In a second approach, we extend the analysis to include the approximation of the Shannon entropy. The latter result is particularly interesting because it involves, as an auxiliary result, a proof of the a.s. convergence of the standard particle filter for the approximation of integrals of unbounded functions w.r.t. the filtering measure, a result not previously available in the literature.

**Keywords:** particle filters; state space models.

**Vladimir Minin**<sup>1</sup>, Todd Oakley<sup>2</sup> and Marc Suchard<sup>3</sup>

<sup>1</sup>University of Washington, Seattle, <sup>2</sup>University of California, Santa Barbara, <sup>3</sup>UCLA

## A Bayesian approach to testing the independent origin hypothesis

Estimating the number of times a discrete evolutionary trait changed its state is one of the fundamental questions in evolutionary developmental biology. Often, researchers are interested in testing a so called independent origin hypothesis that asserts that the number of changes of a certain type exceeds a predefined threshold. Testing such a hypothesis in a formal statistical framework requires a model of trait evolution and a phylogenetic tree along which the evolutionary trait evolves.

We use a fairly unrealistic, but mathematically and computationally convenient, Markov model of trait evolution, parameterized in terms of rates at which the trait changes its state. Next, we assume that the phylogenetic tree of organisms under study can be inferred using molecular data. However, such estimation remains imprecise and the resulting phylogenetic uncertainty must be accounted for when testing the independent origin hypothesis. Both, the phylogenetic tree and rates of the Markov evolutionary model are nuisance parameters for our purposes.

We propose to test the independent origin hypothesis in a Bayesian framework, because the Bayesian paradigm naturally allows for integration over nuisance parameters. We demonstrate how to construct and to calculate a Bayes factor to perform the desired test. We evaluate our test via simulations and apply the developed method to study evolution of a compound eye in arthropods. We conclude by discussing extensions of our Bayes factor test to non-Markovian models of trait evolution.

**Keywords:** phylogenetic trees; independent origin hypothesis; Markov models.

Krzysztof Latuszyński<sup>1</sup>, **Jan Palczewski**<sup>2</sup> and Gareth Roberts<sup>1</sup>

<sup>1</sup>University of Warwick, <sup>2</sup>University of Leeds

## Exact MCMC Inference for Markov Switching Diffusion Models

Consider a stochastic differential equation for  $X = X_t$  and  $t \in [0; T]$ ,

$$dX_t = b(X_t, Y_t, \theta)dt + \sigma(X_t, \theta)\gamma(Y_t, \theta)dB_t$$

where  $B = B_t$  is a Brownian motion,  $\theta \in \Theta \subset \mathbb{R}^n$  is a multi-dimensional parameter and the switching process  $Y = Y_t; t \in [0; T]$  is a continuous-time Markov process on the state space  $\mathcal{Y} = \{1, \dots, m\}$  with the intensity matrix  $\Lambda = \{\lambda_{ij}\}$ . Models of this type naturally arise in finance and economics, where  $Y$  represents unobservable factors that describe the state of the economy (e.g. a business cycle).

We assume that  $X$  is observed at discrete time instances and propose an exact Bayesian inference method for  $\theta$  and  $\Lambda$  given their prior distributions  $\pi_\theta$  and  $\pi_\Lambda$ : Let  $X_D$  be the observed data and  $X_M$  the missing path of  $X$ : Our MCMC algorithm samples from the exact full posterior

$$\pi(X_M, Y, \theta; \Lambda | X_D) \propto \pi_\theta(\theta) \pi_\Lambda(\Lambda) \pi(Y | \Lambda) \pi(X_M, X_D | Y, \theta)$$

avoiding any discretization error. The approach relies on a generalization of the exact algorithm of [1, 2].

**Keywords:** Markov switching; diffusion process; exact algorithm

## References

- [1] A. Beskos, O. Papaspiliopoulos, G.O. Roberts (2008). A factorisation of diffusion measure and finite sample path constructions. *Methodol. Comput. Appl. Probab*, **10**, 85-104.
- [2] A. Beskos, O. Papaspiliopoulos, G.O. Roberts, and P. Fearnhead (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *JRSS B*, **68**, 333-382.

**Trond Reitan**

University of Oslo

Layered continuous time processes in biology

The evolution of a biological trait (phenotype) is thought to be a stochastic process tracking an optimum. The optimum may not be constant, and can also be a stochastic process tracking another process, such as climate. So in evolutionary biology, there may often be several layers of hidden processes influencing the measurements we have on the top. Arbitrary fossil measurement ages means handling these processes in continuous time. We analyze such data using Bayesian inference on hidden linear stochastic differential equations (SDEs), where the interactions between processes are assigned a causal structure. As long as the SDEs are linear, we are able to use the Kalman filter for likelihood

calculations and process inference. However, with non-linear causal influence, other methods need to be contemplated.

**Keywords:** evolutionary biology; stochastic differential equations; Kalman filter.

Omiros Papaspiliopoulos<sup>1</sup> and **Matteo Ruggiero**<sup>2</sup>

<sup>1</sup>Universitat Pompeu Fabra, <sup>2</sup>University of Pavia

Filtering the Fleming-Viot process with applications to Bayesian nonparametric inference in continuous time

We tackle the problem of filtering an unobserved measure-valued diffusion process which is assigned a Fleming-Viot process as prior distribution on the paths. While updating dynamically the marginal posterior based on the current observations simply exploits the conjugacy property of the Dirichlet process, prediction is difficult to implement since the transition probability function does not have a density and its explicit representation involves infinite sums.

In this work we obtain analytical results on the transition function of a Fleming-Viot process which leads to designing an exact filter, allowing arbitrary-far-in-time prediction of future states of the unobserved measure-valued diffusion by means of a finite computation, thus excluding the need for approximations or truncations.

**Keywords:** filtering; inference for diffusions; hidden Markov model; Fleming-Viot process; Dirichlet process.

**Alexandre J. Santos** and Alvaro Faria

Open University, UK

Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) model

The main goal of this paper is to introduce a Bayesian formulation of a special class of nonlinear time series models known as the Smooth Transition Autoregressive (STAR) model.

This proposal consists of writing the STAR model in a Dynamic Linear Model (DLM) form, called Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) model.

Compared to the classical STAR model, the proposed DBSTAR model has the advantage of not requiring extensive data for the parameter estimation, as it estimates both the linear and nonlinear parameters in analytical closed form sequentially in time. Furthermore, the DBSTAR model allows formal inputs and interventions from experts, where appropriate.

Unlike existing Bayesian formulations of the STAR model which estimate the parameters using approximation methods like Markov Chain Monte Carlo, our model allows for fast estimation of the model's parameters based on the closeness.

Our methodology is extensively studied with some simulated time series and also illustrated with real examples.

**Keywords:** STAR model, DLM, Nonlinear time series, Bayesian model

Paul Fearnhead, Vasileios Giagos and **Chris Sherlock**

Lancaster University

Bayesian inference for stochastic kinetic models of biological systems using the linear noise approximation

The interactions between species such as proteins and RNA in a biological system can be modelled as Poisson processes with rates dependent on the current state of the system. When the number of molecules of each species is very large the evolution of the system is determined by an ordinary differential equation, but in many of the systems under investigation today there are fewer molecules of each species, the stochasticity is important, and the evolution of the system is approximated by a stochastic differential equation (SDE). We investigate inference on the reaction rate parameters for biological systems where only a subset of the species are actually observed and where these are observed with error at a discrete set of time points. Using the linear noise approximation to the governing SDE and a variation on the extended Kalman filter we provide a computationally efficient method for inference and investigate its accuracy.

**Keywords:** biological systems; Poisson processes; Kalman filter.

**Mark Steel<sup>1</sup>** and **Jim Griffin<sup>2</sup>**

<sup>1</sup>University of Warwick, <sup>2</sup>University of Kent

## Stick-Breaking Autoregressive Processes

This paper considers the problem of defining a time-dependent nonparametric prior for use in Bayesian nonparametric modelling of time series. A recursive construction allows the definition of priors whose marginals have a general stick-breaking form. The processes with Poisson-Dirichlet and Dirichlet process marginals are investigated in some detail. We develop a general conditional Markov Chain Monte Carlo (MCMC) method for inference in the wide subclass of these models where the parameters of the marginal stick-breaking process are nondecreasing sequences. We derive a generalized Pólya urn scheme type representation of the Dirichlet process construction, which allows us to develop a marginal MCMC method for this case. We apply the proposed methods to financial data to develop a semi-parametric stochastic volatility model with a time-varying nonparametric returns distribution. Finally, we present two examples concerning the analysis of regional GDP and its growth.

**Keywords:** time series; nonparametrics; Dirichlet processes.

**<sup>1</sup>Joe Wheatley, <sup>1</sup>Paul Blackwell, <sup>2</sup>N.J. Abram and <sup>2</sup>E.W. Wolff**

<sup>1</sup>Department of Probability and Statistics, University of Sheffield UK, <sup>2</sup>British Antarctic Survey, Cambridge, UK

## Layer-counting in ice cores: reconstructing the time scale of multivariate signals

Layer-counting is a vital technique for determining the age-depth relationship within high-resolution ice cores, which is in turn a necessary step in interpreting the other climatic and environmental information preserved in such cores. The available data generally consist of multiple signals - concentrations of particular isotopes, chemicals or particles - which are all believed to have annual cycles, though typically out of phase. Each multivariate observation relates to a known depth within the core, and to an unknown date, although clearly the ordering of the dates is known. Layer-counting uses the periodicity of the signals to literally count back in time, year by year; currently, it is mostly done by eye. In this work, we look at some Bayesian approaches to automating the layer-counting process, while allowing for the lack of stationarity in the process, and the frequent clusters of missing values in one or more signals at a time. Using simple, flexible models for the underlying signal, we use a Markov chain Monte Carlo approach to reconstruct the underlying periodic process in sufficient detail to estimate the number of years

represented by a given section of core, and given a measure of uncertainty on the estimate - something that is hard to do coherently with current methods.

**Keywords:** layer counting; MCMC.

**Dawn Woodard**<sup>1</sup>, Ciprian Crainiceanu<sup>2</sup> and David Ruppert<sup>1</sup>

<sup>1</sup>Cornell University, <sup>2</sup>Johns Hopkins University

### Hierarchical Adaptive Regression Kernels for Regression with Functional Predictors

We introduce a new approach to regression with functional predictors. We represent the predictors using a parsimonious model designed for data that exhibit features such as dips, bumps, and plateaus whose frequency, location, size, and shape varies stochastically across subjects. Summaries of the representation, such as the frequency of bumps, or their average height or width, can have intuitive scientific interpretation; we regress the outcome on these functional summaries. This approach does not require alignment or even a common domain for the subject-specific functions, and naturally handles missing or co-located data. In order to account for uncertainty in the functions when estimating the regression coefficients, we introduce a joint hierarchical model. We show that Bayesian inference in our model is computationally feasible even for large datasets, using an approximation to the posterior distribution obtained via a technique called modularization. We contrast our approach with existing state-of-the-art methods for regression with functional predictors, and show that our method is more effective for data that include features occurring at varying locations. We apply our methodology to a large and complex dataset from the Sleep Heart Health Study, in order to better understand the relationship between sleep characteristics and health outcomes.

**Keywords:** regression; functional data; hierarchical models; modularization.