

# List of Talks

## **Title: Bayesian nonparametric mixtures inconsistency for the number of clusters**

**Speaker: Louise Alamichel**

CT

Université Grenoble Alpes

Bayesian nonparametric mixture models are often employed for modelling complex data. While these models are well-suited for density estimation, their application for clustering has some limitations. Miller and Harrison, JMLR (2014), proved posterior inconsistency in the number of clusters when the true number of clusters is finite for Dirichlet process and Pitman–Yor process mixture models. In this work, we extend this result to additional Bayesian nonparametric priors such as Gibbs-type processes and finite-dimensional representations of them. The latter include the Dirichlet multinomial process and the recently emerged Pitman–Yor and normalized generalized gamma multinomial processes. We show that mixture models based on these processes are also inconsistent in the number of clusters. However, upon establishing new posterior contraction rates in the Wasserstein divergence, we also obtain positive consistency results for the number of clusters by applying the post-processing procedure of Guha et al, Bernoulli (2021).

## **Title: Bayesian Computing in the Undergraduate Statistics Curriculum**

**Speaker: Jim Albert**

IS

Bowling Green State University

Advances in Bayesian methodology and software have made Bayesian techniques more accessible to applied statisticians and, in turn, have potentially transformed Bayesian education at the undergraduate level. We provide an overview of the various options for implementing Bayesian computational methods motivated to achieve particular learning outcomes. For each computational method, we propose activities and exercises, and discuss each method's pedagogical advantages and disadvantages based on our experience in the classroom. The goal is to present guidance on the choice of computation for the instructors who are introducing Bayesian methods in their undergraduate statistics curriculum.

## **Title: Hierarchical stochastic block model for community detection in multiplex networks**

**Speaker: Arash Amini**

IS

University of California, Los Angeles

Multiplex networks have become increasingly more prevalent in many fields, and have emerged as a powerful tool for modeling the complexity of real networks. There is a critical need for developing inference models for multiplex networks that can take into account potential dependencies across different layers, particularly when the aim is community detection. We add to a limited literature by proposing a novel and efficient Bayesian model for community detection in multiplex networks. A key feature of our approach is the ability to model varying communities at different network layers. In contrast, many existing models assume the same communities for all layers. Moreover, our model automatically picks up the necessary number of communities at each layer (as validated by real data examples). This is appealing, since deciding the number of communities is a challenging aspect of community detection, and especially so in the multiplex setting, if one allows the communities to change across layers. Borrowing ideas from hierarchical Bayesian modeling, we use a hierarchical Dirichlet prior to model community labels across layers, allowing dependency in their structure. Given the community labels, a stochastic block model (SBM) is assumed for each layer. We develop an efficient slice sampler for sampling the posterior distribution of the community labels as well as the link probabilities between communities. In doing so, we address some unique challenges posed by coupling the complex likelihood of SBM with the hierarchical nature of the prior on the labels. An extensive empirical validation is performed on simulated and real data, demonstrating the superior performance of the model over single-layer alternatives, as well as the ability to uncover interesting structures in real networks.

## **Title: Modelling species' responses to climate change using a flexible nonlinear Bayesian approach**

**Speaker: Marti Anderson**

CT

Massey University & PRIMER-e

Species are disappearing faster than they are being discovered. Effects of climate change on biodiversity require urgent assessment. Physiological responses to environmental and spatial gradients (such as temperature, altitude, depth, latitude, etc.) are non-linear – typically modal. Field data documenting species' abundances (typically discrete counts) also show zero-inflation and variance-mean relationships. Asymmetries and/or flat or peaked responses also hamper effective modelling. We propose a flexible hierarchical nonlinear mathematical model that caters handsomely to ecological data, and apply this to broad-scale datasets in time and space. A Bayesian approach prevents the model from heading towards illogical outcomes and provides a natural framework for updating. Species' modal positions along gradients can be estimated rigorously, and responses to changing environments can be tracked at global scales. We demonstrate our technique on North American Breeding Birds, sampled over the past 50+ years.

## **Title: Improved Inference for Doubly Robust Estimators of Heterogeneous Treatment Effects**

**Speaker: Joseph Antonelli**

IS

University of Florida

We propose a doubly robust approach to characterizing treatment effect heterogeneity in observational studies. We utilize posterior distributions for both the propensity score and outcome regression models to provide valid inference on the conditional average treatment effect even when high-dimensional or nonparametric models are used. We show that our approach leads to conservative inference in finite samples or under model misspecification, and provides a consistent variance estimator when both models are correctly specified. In simulations, we illustrate the utility of these results in difficult settings such as high-dimensional covariate spaces or highly flexible models for the propensity score and outcome regression. Lastly, we analyze environmental exposure data from NHANES to identify how the effects of these exposures vary by subject-level characteristics.

## **Title: Clustering grouped data via hierarchical mixture models**

**Speaker: Raffaele Argiento**

IS

Università degli Studi di Bergamo

In this work we propose a new approach to cluster grouped data by considering a mixture model with a hierarchical structure. In particular, the weights of the mixing measures are obtained by normalization of independent finite point processes, and the dependence across groups is ensured by assuming that the random probability measures share the same support. Our model induces both a local (within groups) and a global (between groups) clusters. We characterize the law and study the properties of this group-dependent clustering structure by providing a closed form expression for the partially exchangeable partition probability function. Furthermore, we provide a posterior characterization of the vector of random mixing measures and design a conditional Gibbs sampler to perform full Bayesian inference. The performances of the proposed approach are illustrated via simulated data and a real application.

## **Title: Inferring the Causal Impact of Online Advertising on Multiple Treated Units Using Synthetic Controls**

**Speaker: Tim Au**

IS

Google LLC

Inferring the causal impact of online advertising has become progressively more important as advertisers increasingly seek to better understand their marketing strategies. The synthetic control method has emerged as a popular technique for analyzing comparative case studies, where it has generally been assumed that the intervention of interest only affects a single unit (e.g., an entire country). However, marketing interventions frequently affect multiple units (e.g., several countries), and applying the synthetic control method in situations such as these has often entailed either applying the method separately to each treated unit or applying the method once to the aggregate of all the treated units. In this talk, we instead consider jointly modeling all of the treated units together in a multivariate linear Gaussian state space model to more efficiently estimate the synthetic controls, where a fully Bayesian approach for inference facilitates the evaluation of different quantities of interest.

## **Title: Bayesian strategies for multi-study integration using biological hierarchies**

**Speaker: Veera Baladandayuthapani**

IS

University of Michigan

Modern precision medicine endeavors are at an inflexion point – facing the fundamental challenge of assimilating, organizing, analyzing and interpreting multi-domain data types to make individualized health decisions. To this end, we propose an integrative hierarchical Bayesian framework for modeling multi-platform genomics and clinical outcome data. The key feature of our model is the use of the fundamental biological and mechanistic hierarchies, mapping them onto a series of hierarchical regression models which are used to quantify the mechanistic evidence and facilitate their use in predicting clinical outcome(s) of interest. For the latter, we propose a novel calibrated Bayesian variable selection scheme that allows incorporation of evidence from different modeling layers. This allows simultaneous high-dimensional variable selection and gives model enough flexibility to estimate and incorporate the different intrinsic structures of biological relationships for different high-throughput platforms. Using synthetic datasets, we illustrate how integrative methods have higher power to detect disease related markers than non-integrative approaches and exemplify our approaches using several case studies in oncology. (joint work with Rupam Bhattacharyya and Nicholas Henderson)

## **Title: Bayesian Nonparametric Analysis of Spatial Variation with Discontinuities**

**Speaker: Cecilia Balocchi**

IS

University of Edinburgh

Spatial data often display high levels of smoothness but can simultaneously present abrupt discontinuities, especially in urban environments. We model neighborhood crime trends over time in the City of Philadelphia by combining a spatial local shrinkage model with spatial partitions of areal units to allow for discontinuities. Two main challenges arise in this setting. First, the vast space of spatial partitions makes typical stochastic search techniques computationally prohibitive. We introduce an ensemble optimization procedure that summarizes the posterior by simultaneously targeting several high probability partitions. Second, the data are organized in a hierarchical structure with multiple resolution levels. We introduce a model combining the Nested Dirichlet Process with the Hierarchical Dirichlet Process to allow for flexible partitions of multi-resolution data and sharing of information between the partitions at different resolutions. Both our methods are demonstrated on synthetic data and on real data in Philadelphia.

## **Title: Simulation-based Bayesian Sample Size Calculations Using Design and Analysis Priors in Contemporary Clinical Trials**

**Speaker: Sudipto Banerjee**

IS

University of California, Los Angeles

Power and sample size analysis comprises a critical component of clinical trial study design. There is an extensive collection of methods addressing this fundamental problem from diverse perspectives. This talk builds upon a two-stage framework with different priors in the design and analysis stages that has been effectively used for cost-effectiveness analysis. We offer a general Bayesian framework for simulation-based sample size determination that can be easily implemented on modest computing architectures. We further qualify the need for different priors for the design and analysis stage. We work primarily in the context of conjugate Bayesian linear regression models and consider different settings with varying levels of information regarding the population variances. Throughout, we draw parallels with classical solutions, which arise as special cases, and alternate Bayesian approaches with an emphasis on how the numerical results from existing methods can arise as special cases of our framework.

## **Title: Bayesian estimation and forecasting of migration in Europe: From hierarchical models to panel VARs**

**Speaker: Emily Barker**

IS

University of Southampton

Migration flows are important for the demography, economy, and society in every European country, whether net receiver or sender of migrants. Still, there is a large amount of uncertainty around the size of migration flows due to differences in reporting methods and definitions. In this paper, we show Bayesian hierarchical models for estimating harmonised migration flows developed within the QuantMig project. The estimates are subsequently used in a forecasting exercise. Given the different migratory and macroeconomic profiles within Europe, one migration model cannot fit all situations. To forecast migration and examine the effects of migration shocks, we analyse links between migration and the macroeconomy, government finances, and labour markets by using Bayesian panel VAR models. Selected European countries are grouped based on macroeconomic and migration characteristics. The results show that, while migration predictions are on the whole highly uncertain, the impact of migration on the economy can vary between countries.

## **Title: Bayesian Divide-and-Conquer Propensity Score Based Approaches for Leveraging Real World Data in Randomized Control Trials**

**Speaker: Eric Baron**

IS

University of Connecticut

There has been a considerable increase in the usage of real world data (RWD) to supplement the control arm in a randomized control trial (RCT). Propensity score methods such as stratification have been used to balance baseline characteristics and prognostic factors between historical control patients and RCT patients to improve the estimation of a treatment effect. This paper merges propensity score methodology and Bayesian inference to estimate a treatment effect in the presence of substantial historical control data as follows: (i) match historical control patients and RCT patients by strata using the percentiles of the RCT patients' propensity scores, (ii) apply a prior within each stratum to leverage RWD to then estimate the stratum-specific treatment effect, and (iii) use a weighted average to combine the stratum-specific treatment effect estimates to estimate the treatment effect. In stage (ii), an extension of the borrowing by parts prior is used. An extensive simulation study is carried out to evaluate the performance of the proposed approaches and a case study is examined based on the Alzheimer's Disease Neuroimaging Initiative (ADNI) data base.

## **Title: Bayesian Joint Modeling of Within- and Between-Period Information in Capture-Recapture Studies of Migrating Populations**

**Speaker: Audrey Béliveau**

CT

University of Waterloo

We contribute a unified Bayesian approach for estimating both short- and long-term trends in the abundance of migrating animal populations from capture-recapture data. Conventional approaches to analysis, which model within-year information separately in each year, cannot reliably analyze sparse data, which commonly arise due to unfavorable environmental conditions, design flaws or other unforeseen circumstances. Moreover, conventional single-year approaches miss the opportunity to borrow strength across years. To address these issues, we develop a Bayesian hierarchical capture-recapture joint modeling framework which shares information across years without imposing a parametric form on the latent abundance and capture probability time series. Notably, flexibility is achieved by introducing Bayesian penalized splines of day and year latently while accommodating between-year variations using interaction terms, such as tensor products. We demonstrate the approach on data on outmigrating sockeye salmon smolts collected daily at the Koeye river since 2014.

## **Title: Hierarchical Modeling of Heterogeneous Networks for Animal Production Systems**

**Speaker: Nora M. Bello**

IS

The Ohio State University

Understanding the interconnections between performance outcomes in a system is increasingly important for integrated management. Structural equation models (SEMs) are a type of multiple-variable modeling strategy that allows investigation of directionality in the interconnections between outcome variables, thereby providing insight into links defining a functional network. A key assumption underlying SEMs is that of a homogeneous network structure, whereby the structural coefficients defining functional links are assumed homogeneous and impervious to environmental conditions or management factors. This assumption seems questionable as systems are regularly subjected to explicit interventions to optimize the necessary trade-offs between outcomes. Using a Bayesian approach, we propose methodological extensions to hierarchical SEMs that accommodate structural heterogeneity of the network by explicitly specifying structural coefficients as functions of systematic and non-systematic sources of variation. We validate the proposed approach using simulation and apply it to a dataset from commercial swine production. Results indicate that explicit modeling of network heterogeneity enhances understanding of complex systems in animal production agriculture.

## **Title: Beyond CRMs: normalized random measures with atoms' interaction for Bayesian mixture models**

*Speaker: Mario Beraha*

IS

Politecnico di Milano and Università di Bologna

Discrete random probability measures (RPMs) are one of the founding blocks of Bayesian non-parametrics (BNP). They are routinely used as mixing measures in mixture models. Normalized completely random measures are the prominent class of priors in BNP, thanks to their analytical tractability. In the context of mixture models, such priors assume that the cluster-specific parameters are i.i.d from some base distribution. This might come with drawbacks such as inconsistency for the number of detected clusters and lack of robustness to model misspecification. A large class of RPMs based on the normalization of non-completely-random measures will be described in this talk. In particular, by assuming a point process as the joint distribution for the number of support points of the measure and their location, the model could favor repulsiveness or attractiveness among the clusters. Several theoretical properties such as the posterior distribution of the underlying random measure and the marginal distribution of a sample will be discussed, together with numerical illustrations on simulated and real-world applications.

## **Title: A comparison of response adaptive randomization and arm dropping features in the presence of temporal drift**

*Speaker: Lindsay Berry*

IS

Berry Consultants

The implementation of adaptive platform trials shortly after the COVID-19 outbreak led to important findings of both life-saving treatments and ineffective treatments. This talk focuses on two topics in the ongoing conversation about adaptive platform trials: the use of response adaptive randomization (RAR) and the effect of time trends. First, RAR and arm-dropping designs are compared in the setting of multi-arm trials aiming to identify a best active arm and compare that arm to a control arm. Second, the effect of time trends is evaluated for both allocation methods under consideration. We show that, within the context of finding the best arm, standard arm-dropping techniques based on p-values have worse performance than methods based on the posterior probability each arm is optimal. Additionally, the performance of arm-dropping is slightly worse than a well-tuned RAR design. Finally, we show both methods are robust to time trends when appropriate time modeling is employed. We relate these findings to the REMAP-CAP platform trial for COVID-19 that utilizes Bayesian methods for RAR and modeling of time trends.

## **Title: MCMC and SMC algorithms for estimating parameters in Thermogravimetric Analysis.**

**Speaker: Matthew Berry**

IS

University of Wollongong

Thermogravimetric Analysis is an experimental procedure where a small sample is heated and the change of mass is recorded. These changes are the result of chemical reactions or phase transitions. This process is modelled using a system of differential equations, which must be solved numerically in order to evaluate the likelihood. In this talk we introduce a combined Runge-Kutta and Metropolis-Hastings algorithm to estimate the kinetic parameters in our model. For complex reaction schemes, traditional methods are difficult to use, due to issues such as multimodality and parameter identifiability. Our algorithm allows us to reliably estimate posterior distributions for the parameter values in these cases. Our algorithm combines multimodal data from both Thermogravimetric Analysis and fractional weight change observations. We utilise simulated data and experimental data as a test case for this algorithm. We then introduce a Sequential Monte-Carlo algorithm that allows us to integrate data from multiple experiments in succession to conduct our inference.

## **Title: Priors for microclustering based on allelic partitions**

**Speaker: Brenda Betancourt**

CT

University of Florida

In contrast to traditional clustering some applications, such as record linkage, require a large number of clusters with a few observations per cluster. We introduce a new class of prior distributions based on allelic partitions suited for the small cluster setting. Our approach makes it straightforward to introduce prior information about the cluster size distribution at different scales, and naturally enforces sublinear growth of the maximum cluster size i.e. the microclustering property. We also introduce a set of novel microclustering conditions in order to impose further constraints on the cluster sizes a priori. We evaluate the performance of our proposed class of priors using simulations and official statistics data, and show that our models provide competitive results compared to other microclustering models. Moreover, we compare the performance of different loss functions for optimal point estimation of the partitions using decision-theoretical based approaches recently proposed in the literature.

## Title: Graphical Evidence

*Speaker: Anindya Bhadra*

IS

Purdue University

Marginal likelihood, also known as model evidence, is a fundamental quantity in Bayesian statistics. It is used for model selection using Bayes factors or for empirical Bayes tuning of prior hyperparameters. Yet, the calculation of evidence has remained a longstanding open problem in Gaussian graphical models. Currently, the only feasible solutions that exist are for special cases such as the Wishart or G-Wishart, in moderate dimensions. We present an application of Chib's technique that is applicable to a very broad class of priors under mild requirements. Specifically, the requirements are: (a) the priors on the diagonal terms on the precision matrix can be written as gamma or scale mixtures of gamma random variables and (b) those on the off-diagonal terms can be represented as normal or scale mixtures of normal. This includes structured priors such as the Wishart or G-Wishart, and more recently introduced element-wise priors, such as the Bayesian graphical lasso and the graphical horseshoe. Among these, the true marginal is known in an analytically closed form for Wishart, providing a useful validation of our approach. For the general setting of the other three, and several more priors satisfying conditions (a) and (b) above, the calculation of evidence has remained an open question that this talk seeks to resolve.

## Title: Multiple latent clustering model for the inference of RNA life-cycle kinetic rates from sequencing data

*Speaker: Enrico Bibbona*

IS

Politecnico di Torino

We here propose a hierarchical Bayesian model to infer RNA synthesis, processing, and degradation rates from sequencing data, based on an ordinary differential equation system that models the RNA life cycle. We parametrize the latent kinetic rates, that rule the system, with a novel functional form, and estimate their parameters through 6 Dirichlet process mixture models. Owing to the complexity of this approach, we are able to simultaneously perform inference, clustering, and model selection. We apply our method to investigate transcriptional and post-transcriptional responses of murine fibroblasts to the activation of proto-oncogene Myc. Our approach uncovers simultaneous regulations of the rates, which had not previously been observed in this biological system.

## Bounding Wasserstein distance with couplings

*Niloy Biswas*

IS

Harvard University

Markov chain Monte Carlo (MCMC) provides asymptotically consistent estimates of intractable posterior expectations as the number of iterations tends to infinity. However, in large data applications, MCMC can be computationally expensive per iteration. This has catalyzed interest in sampling methods such as approximate MCMC, which trade off asymptotic consistency for improved computational speed. In this article, we propose estimators based on couplings of Markov chains to assess the quality of such asymptotically biased sampling methods. The estimators give empirical upper bounds of the Wasserstein distance between the limiting distribution of the asymptotically biased sampling method and the original target distribution of interest. We establish theoretical guarantees for our upper bounds and show that our estimators can remain effective in high dimensions. We apply our quality measures to stochastic gradient MCMC, variational Bayes, and Laplace approximations for tall data and to approximate MCMC for Bayesian logistic regression in 4500 dimensions and Bayesian linear regression in 50000 dimensions.

### **Title: Bayesian modeling, prior compatibility and selection for extreme value models**

*Speaker: Nicolas Bousquet*

CT

EDF

The presented work proposes a new Bayesian modeling of the parameters of the extreme laws (Fréchet, Weibull, Gumbel), by separating the basins of attraction, by an Approximate Posterior Prior approach. By its explicability, this approach is likely to allow a better adoption of the Bayesian approach. The compatibility between Bayesian models, aiming at allowing a fair selection of the best basin of attraction, is defined by variational approaches between marginal laws and cross calibrations of imaginary sample sizes, which are natural hyperparameters of these models.

## **Title: Comparative Study of Differentially Private Data Synthesis Methods**

**Speaker: Claire Bowen**

IS

Urban Institute

When sharing data among researchers or releasing data for public use, there is a risk of exposing sensitive information of individuals in the data set. Data synthesis is a statistical disclosure limitation technique for releasing synthetic data sets with pseudo individual records. Traditional data synthesis techniques often rely on strong assumptions of a data intruder's behaviors and background knowledge to assess disclosure risk. Differential privacy (DP) formulates a theoretical approach for a strong and robust privacy guarantee in data release without having to model intruders' behaviors. Efforts have been made aiming to incorporate the DP concept in the data synthesis process. In this paper, we examine current Differentially Private Data Synthesis (DIPS) techniques for releasing individual-level surrogate data for the original data, compare the techniques conceptually, and evaluate the statistical utility and inferential properties of the synthetic data via each DIPS technique through extensive simulation studies. Our work sheds light on the practical feasibility and utility of the various DIPS approaches, and suggests future research directions for DIPS.

## **Title: Bypassing Markov Chains for a Broad Class of Bayesian Generalized Linear Mixed Effects Models**

**Speaker: Jonathan Bradley**

CT

Florida State University

Markov chain Monte Carlo (MCMC) has become a standard in Bayesian statistics that allows one to generate dependent replicates from a posterior distribution for general Bayesian hierarchical models. However, convergence issues, tuning, and the effective sample size of the MCMC are nontrivial considerations that are often overlooked or can be difficult to assess. This motivates us to consider finding closed-form expressions of the posterior distribution that are computationally straightforward to sample from directly. We focus on a broad class of Bayesian generalized linear mixed-effects models (GLMM) that allows one to jointly model data of different types. We derive a class of distributions that allows one to specify the prior on fixed and random effects to be any conjugate multivariate distribution. The expression of the posterior distribution is given, and direct simulations have an efficient projection form. An analysis of an environmental dataset is presented.

## Title: An Automatic Finite-Data Robustness Check for Bayes and Beyond: Can Dropping a Little Data Change Conclusions?

Speaker: *Tamara Broderick*

IS

Massachusetts Institute of Technology

One hopes that Bayesian methods will be used to analyze data and make beneficial decisions regarding people's health, finances, and well-being. But the data fed to a statistical analysis may systematically differ from the data where these decisions are ultimately applied. For instance, suppose we analyze data in one country and conclude that microcredit is effective at alleviating poverty; based on this analysis, we decide to distribute microcredit in other locations and in future years. We might then ask: can we trust our conclusion to apply under new conditions? If we found that a very small percentage of the original data was instrumental in determining the original conclusion, we might expect the conclusion to be unstable under new conditions. So we propose a method to assess the sensitivity of data analyses to the removal of a small fraction of the data set. Analyzing all possible data subsets of a certain size is computationally infeasible even for small data sets, so we provide an approximation. We call our resulting method the Approximate Maximum Influence Perturbation. Our approximation is automatically computable, theoretically supported, and works with common schemes using or approximating a Bayesian posterior — including (but not limited to) variational Bayes and MAP. We show that any non-robustness our metric finds is conclusive. Empirics demonstrate that while some applications are robust, in others the conclusions be changed by dropping less than 0.1% of the data— even in simple models and even when posterior credible intervals are small.

## **Title: Nonparametric discrete-time Hawkes processes in practice**

**Speaker: Raiha Browning**

IS

Queensland University of Technology

A wide range of real-world self-exciting phenomena are modelled using continuous time Hawkes processes with a parametric form for the triggering process. Recent interest has focused on alternative discrete-time, nonparametric Hawkes processes that may capture the data acquisition and excitation patterns encountered in practice.

In this presentation we explore the use of trans-dimensional MCMC inference techniques for discrete-time Hawkes processes that allow the model for the triggering kernel to take the form of any step function, thereby affording much more flexibility than a parametric form. We illustrate situations in which the proposed model performs well or otherwise, and describe how the balance between prior informativeness and the informativeness of the data affects inference in this high-dimensional parameter space. For higher-dimensional problems, this is of particular importance.

We apply the proposed model to real-world data for several case studies. First, we evaluate the interaction between selected European countries in the COVID-19 pandemic and recover excitation patterns between these countries that align with what occurred historically. We then attempt to capture the interaction of terrorist activity in targeted South-East Asian countries.

This work is joint with Professors Judith Rousseau and Kerrie Mengersen.

## **Title: The Geometry of Continuous Latent Space Models for Network Data**

**Speaker: Catherine Calder**

IS

University of Texas at Austin

This talk will focus on the class of continuous latent space models for network data, paying particular attention to the role of the geometry of the latent space. In these models, the presence/absence of network dyadic ties is assumed to be conditionally independent given the dyads' unobserved positions in a latent space. In this way, these models provide a probabilistic framework for embedding network nodes in a continuous space equipped with a geometry that facilitates the description of dependence between random dyadic ties. Specifically, these models naturally capture homophilous tendencies and triadic clustering, among other common properties of observed networks. In addition to providing an overview of continuous latent space models from a geometric perspective, I will highlight the important role the geometry of the latent space plays on properties of networks arising from these models via intuition and simulation. In addition, I will discuss results from spectral graph theory that allow us to explore the role of the geometry of the latent space, independent of network size. Finally, I will discuss a Bayesian model-based approach for estimating latent space curvature. This presentation is based on joint research with Anna Smith and Dena Asta.

## **Title: Scaled process priors for Bayesian nonparametric estimation of the unseen genetic variation**

**Speaker: Federico Camerlenghi**

CT

University of Milano-Bicocca

The estimation of the number of unseen features is an important problem in biological sciences. We first emphasize advantages/disadvantages of the popular stable-Beta process prior in Bayesian nonparametric inference for the unseen-features problem: i) the disadvantage lies in the limited use of the sampling information in the posterior distributions, which depend on the observable sample only through the sample size; ii) the advantage lies in its analytical tractability. We then introduce an alternative nonparametric prior, called the stable-Beta scaled process prior: this is the first prior that allows to enrich the posterior distribution of the number of unseen features, through the inclusion of the additional sampling information on the number of distinct features in the observable sample, while preserving analytical tractability. We apply our approach to cancer genomics, showing that it outperforms available competitors in terms of estimation accuracy. (Joint work with S. Favaro, L. Masoero and T. Broderick)

## **Title: Scalable Gaussian-process regression and variable selection using Vecchia approximations**

**Speaker: Jian Cao**

CT

Texas A&M

Gaussian process (GP) regression is a flexible, nonparametric approach to regression that naturally quantifies uncertainty. In many applications, the number of observations and variables are both large, and a goal is to select variables that are related to the response. For this setting, we propose a novel, scalable method, which optimizes a penalized GP loglikelihood based on the Vecchia GP approximation from spatial statistics. We travel the solution path from strong to weak penalization, sequentially adding candidate variables based on the gradient of the loglikelihood and deselecting irrelevant variables through a new constrained quadratic coordinate descent algorithm. We propose Vecchia-based mini-batch subsampling, which provides unbiased gradient estimators. The resulting procedure is scalable to millions of observations and tens of thousands of variables. Theoretical analysis and numerical studies demonstrate the improved scalability and accuracy relative to existing methods.

## **Title: Simultaneous Reconstruction of Spatial Frequency Fields and Sample Locations via Bayesian Semi-Modular Inference**

**Speaker: Chris U. Carmona**

CT

University of Oxford

Traditional methods for spatial inference estimate smooth interpolating fields based on features measured at well-located points. When the locations of a subset of instances is missing, joint inference of the fields and locations is required. Conventional Bayesian Inference is inadequate for this task if the spatial model is mildly misspecified and the number of missing locations is considerable, this is mainly due to the feedback between the estimated fields and the imputed locations. Semi-Modular Inference (SMI) offers a solution by modulating the feedback between the different probabilistic assumptions in the model. Motivated by linguistic studies on a large corpus of Medieval English texts, we propose a probabilistic model to simultaneously learn English dialect fields and estimate missing locations of the texts on a map of England. We provide a (flow-based) Variational approximation to the SMI posterior for our model, which allows us to efficiently explore multiple rates of feedback between modules. A comparison with the MCMC-based approximation is also discussed.

## **Title: Shrinkage Bayesian Causal Forests for Heterogeneous Treatment Effects Estimation**

**Speaker: Alberto Caron**

CT

University College London

Shrinkage Bayesian Causal Forests are a sparsity-inducing version of Bayesian Causal Forests, a recently proposed nonparametric causal regression model that employs Bayesian Additive Regression Trees and is designed to estimate heterogeneous treatment effects under observational data. The sparsity-inducing component we introduce is motivated by empirical studies where not all the available confounders are relevant, leading to different degrees of sparsity underlying the surfaces of interest. Shrinkage BCF is equipped with a pair of priors allowing the model to adjust the weight of each covariate through the corresponding number of splits in the tree ensemble. These priors improve the model's adaptability to sparse data generating processes and allow to perform fully Bayesian feature shrinkage in treatment effects estimation, and thus to uncover the moderating factors driving heterogeneity. The method also allows prior knowledge about relevant confounders and the magnitude of their impact on the outcome to be incorporated in the model.

## **Title: Latent Association Graph Inference for Binary Transaction Data**

**Speaker: Luis Carvalho**

CT

Boston University

We discuss a novel approach for modeling multivariate binary transaction data and inferring co-purchase patterns in market basket data. To this end, we exploit a latent graph capturing these purchase associations, where each transaction is a clique, and set meaningful priors based on expected transaction sizes and frequency. We present a MCMC sampling procedure that handles large datasets and conclude that this model provides sparser representations of inferred associations compared to traditional frequent itemset mining (FIM) approaches, without sacrificing predictive accuracy. This is joint work with David Reynolds.

## **Title: Bayesian nonparametric panel Markov-switching GARCH models**

**Speaker: Roberto Casarin**

CT

University Ca' Foscari of Venice

This paper introduces a new model for panel data with Markovswitching GARCH effects to deal with regime changes and temporal clustering of conditional volatility and expected return in a large group of US financial assets. The model incorporates a series-specific hidden Markov chain process that drives the GARCH parameters. To cope with the high-dimensionality of the parameter space, the paper exploits the cross-sectional clustering of the series by first assuming a soft parameter pooling through a hierarchical prior distribution with a two-step procedure, and then introducing clustering effects in the parameter space through a Bayesian nonparametric prior distribution. A numerical analysis is carried out to evaluate the performance of the new model. More specifically, some simulation experiments are run along with an empirical application to financial returns data in the US.

## Title: Bayesian sample size determination for network structure learning

Speaker: *Federico Castelleti*

IS

Università Cattolica del Sacro Cuore (Milan)

Graphical models based on Directed Acyclic Graphs (DAGs) are widely used to model dependence relations among variables. Observational data cannot distinguish in general between DAGs representing the same conditional independence assertions (Markov equivalent DAGs), which implies a limited ability to learn causal relationships among variables. Interventional data produced after exogenous manipulations of the variables can greatly improve the structure learning process, because they can distinguish among equivalent DAGs. Since interventions are typically expensive, an optimal design aiming at minimizing the number of variables to be manipulated is crucial. Most importantly however, the accuracy behind the DAG learning process depends on the number of collected interventional data, whose sample size should be determined before the intervention experiment is performed. We tackle this problem from a Bayesian experimental design perspective by adopting the Bayes Factor as a measure of evidence between competing causal structures. For any candidate sequence of manipulated variables, our method determines the corresponding optimal sample size which guarantees a fixed degree of assurance that the intervention experiment will produce compelling evidence in favor of the true causal-model hypothesis.

## Title: Approximate General Bayesian Inference via Semiparametric Variational Bayes

Speaker: *Cristian Castiglione*

CT

Università degli Studi di Padova

We present a new efficient variational algorithm to approximate the general posterior distribution for Bayesian models that combine subjective prior beliefs with an empirical risk function. In particular, we consider the class of loss functions with a piecewise polynomial behavior, which includes, as remarkable cases, support vector machines, quantile regression, and expectile regression. Our iterative procedure lies in the class of semiparametric variational Bayes and enjoys closed-form updating formulas along with an analytic integration of the evidence lower bound. We require neither conjugacy nor elaborate data augmentation strategies. Structured prior distributions, e.g., cross-random effects, spatio-temporal processes, or inducing shrinkage priors, can be easily accommodated into such a framework without additional effort since the modularity of mean field variational Bayes is preserved. The performances of our algorithm are assessed through a simulation study, where we compare the proposed method with MCMC and conjugate MFVB.

## **Title: On complex dependence structures in Bayesian nonparametrics: a distance-based approach**

**Speaker: Marta Catalano**

IS

University of Warwick

Random measures are the main building block of Bayesian nonparametric models. The introduction of infinite-dimensional parameter spaces guarantees remarkable flexibility and generality to the models but makes the investigation of theoretical and inferential properties more demanding. This applies to the interpretation, prior elicitation, posterior sampling, robustness and frequentist properties of the models. We address many of these issues with a key unifying principle: they can all be framed in terms of model comparison and they can be quantified by relying on appropriate distances.

## **Title: Orthogonal calibration via projected posteriors**

**Speaker: Antik Chakraborty**

CT

Purdue University

We address the problem of statistical calibration when field observations and a computer model simulator is available within the framework originally proposed by Kennedy & O'Hagan (2001). A sufficient condition for identifiability of the calibration parameter was derived by Plumee(2017). Our proposed method accommodates this condition by a novel posterior projection technique. We demonstrate the advantages of the proposed method through simulation experiments and an application on astrophysics. Results are compared with state-of-the-art competitors.

## **Title: Bayesian Scalable Precision Factor Analysis for Massive Sparse Gaussian Graphical Models**

*Speaker: Noirrit Kiran Chandra*

IS

The University of Texas at Austin

We propose a novel approach to estimating the precision matrix of multivariate Gaussian data that relies on decomposing them into a low-rank and a diagonal component. Such decompositions are very popular for modeling large covariance matrices as they admit a latent factor based representation that allows easy inference. The same is however not true for precision matrices due to the lack of computationally convenient representations which restricts inference to low-to-moderate dimensional problems. We address this remarkable gap in the literature by building on a latent variable representation for such decomposition for precision matrices. The construction leads to an efficient Gibbs sampler that scales very well to high-dimensional problems far beyond the limits of the current state-of-the-art. The ability to efficiently explore the full posterior space also allows the model uncertainty to be easily assessed. The decomposition crucially additionally allows us to adapt sparsity inducing priors to shrink the insignificant entries of the precision matrix toward zero, making the approach adaptable to high-dimensional small-sample-size sparse settings. Exact zeros in the matrix encoding the underlying conditional independence graph are then determined via a novel posterior false discovery rate control procedure. A near minimax optimal posterior concentration rate for estimating precision matrices is attained by our method under mild regularity assumptions. We evaluate the method's empirical performance through synthetic experiments and illustrate its practical utility in a data sets from immunology.

## **Title: New partition based measures for data compatibility and information gain**

*Speaker: Ming-Hui Chen*

IS

University of Connecticut

It is of great practical importance to compare and combine data from different studies in order to carry out appropriate and more powerful statistical inference. We propose a partition based measure to quantify the compatibility of two data sets using their respective posterior distributions. We further propose an information gain measure to quantify the information increase (or decrease) in combining two data sets. These measures are well calibrated and efficient computational algorithms are provided for their calculations. We use examples in a benchmark dose toxicology study, a six cities pollution data and a melanoma clinical trial to illustrate how these two measures are useful in combining current data with historical data and missing data.

## **Title: Bayesian Graph Neural Networks**

**Speaker: Mark Coates**

IS

McGill University

Graph neural networks (GNNs) have proven successful for multiple graph learning tasks ranging from image captioning to recommendation and molecular property prediction. Despite this success, many of the methods fail to adequately account for uncertainty in the graph structure and cannot provide confidence intervals for the predictions they generate. In this talk, we will present several approaches for Bayesian graph neural networks, where we treat the observed graph as a sample from a random graph model and incorporate prior distributions over the graph. We then target joint inference of the graph topology and the weights of the neural network model. We illustrate how such an approach can be applied to perform spatio-temporal traffic prediction and improve the performance of large scale recommender systems.

## **Title: Block structured Gaussian graphical models for spectrometric functional data**

**Speaker: Alessandro Colombi**

IS

University of Milan - Bicocca

Motivated by the analysis of spectrometric data, we introduce a Gaussian graphical model for learning the dependence structure among frequency bands of the infrared absorbance spectrum. The spectra are modeled as continuous functional data through a B-spline basis expansion and a Gaussian graphical model is assumed as a prior specification for the smoothing coefficients. The structure of the graph is unknown and is learned from the data. Inference is carried out to simultaneously smooth the curves and to estimate the conditional independence structure between portions of the functional domain. To improve the interpretability of such relationships, we introduce a prior distribution that imposes a block structure in the adjacency matrix of the graph. Conditionally on this choice, the structure of the precision matrix is constrained by the graph through a G-Wishart prior distribution. Finally, we develop a Double Reversible Jump Monte Carlo Markov chain that avoids any G-Wishart normalizing constant calculation when comparing graphical models. The novel algorithm looks for block structured graphs, hence proposing moves that add or remove not just a single link but an entire group of them. The proposed model is applied to the analysis of infrared absorbance spectra of strawberry purees.

## **Title: A Bayesian Approach to Modeling Multivariate Multilevel Insurance Claims in the Presence of Unsettled Claims**

**Speaker: Marie-Pier Côté**

IS

Université Laval

A Bayesian model for individual insurance claims is proposed. It accounts for the multivariate multilevel features of the claims, including multiple claimants for the same event, each of whom may receive benefits under different coverages. A Bayesian approach makes it possible to account for missing values in the covariates and partial information contained in open files, thereby avoiding sampling bias induced when unsettled claims are ignored. For a given claim, the combination of coverages under which payments are made is modeled as a type with multinomial regression. The presence of legal and expert fees follows a logistic regression given the type. The non-zero claim amounts are then modeled with log skewed normal regressions linked by a Student t copula. The Bayesian framework yields a predictive distribution for the amounts paid, including parameter risk and process risk, while handling missing covariates and open files. The approach is illustrated with Accident Benefits car insurance claims from a Canadian company.

## **Title: Discussion of “Bayesian methods for network data”**

**Speaker: Forrest Crawford**

IS

Yale University

As the discussant of this session, I will lead an in-depth critique of “Bayesian methods for network data”

## **Title: Tensorised Rosenblatt Transport for High-Dimensional Stochastic Computation**

**Speaker: Tiangang Cui**

IS

Monash University

Characterising intractable high-dimensional random variables is one of the fundamental challenges in stochastic computation. It has broad applications in statistical physics, machine learning, uncertainty quantification, econometrics, and beyond. The recent surge of transport maps offers a mathematical foundation and new insights for tackling this challenge. In this talk, we present a functional tensor-train (TT) based order-preserving construction of inverse Rosenblatt transport in high dimensions. It characterises intractable random variables via couplings with tractable reference random variables. By integrating our TT-based approach into a nested approximation framework inspired by deep neural networks, we are able to significantly expand its capability to random variables with complicated nonlinear interactions and concentrated density functions. We demonstrate the efficacy of the TT-based inverse Rosenblatt transport on a range of applications in statistical learning and uncertainty quantification, including parameter estimation for dynamical systems, PDE-constrained inverse problems, and Bayesian filtering.

## **Title: Bayesian Inference and Partial Identification in Multi-Treatment Causal Inference with Unobserved Confounding**

*Speaker: Alexander D'Amour*

IS

Google Research

In causal estimation problems, the parameter of interest is often only partially identified, implying that the parameter cannot be recovered exactly, even with infinite data. Here, we study Bayesian inference for partially identified treatment effects in multi-treatment causal inference problems with unobserved confounding. In principle, inferring the partially identified treatment effects is natural under the Bayesian paradigm, but the results can be highly sensitive to parameterization and prior specification, often in surprising ways. It is thus essential to understand which aspects of the conclusions about treatment effects are driven entirely by the prior specification. We use a so-called transparent parameterization to contextualize the effects of more interpretable scientifically motivated prior specifications on the multiple effects. We demonstrate our analysis in an example quantifying the effects of gene expression levels on mouse obesity.

## **Title: Shrinking a Partition Distribution Towards a Baseline Partition, With Applications to Dependent Partitions**

*Speaker: David Dahl*

IS

Brigham Young University

The shrinkage partition distribution (SPD) takes any random partition distribution and pulls it towards a baseline partition. The baseline partition may be formulated from prior knowledge, but it can also be the basis for inducing dependence among partitions. The degree of dependence is governed by a shrinkage parameter. The normalizing constant is tractable and standard MCMC techniques are readily applied. We show that the SPD can hierarchically model a collection of random partitions and can also model time-dependent random partitions.

## Title: Graphical models for multivariate Gaussian Processes

Speaker: *Abhirup Datta*

IS

Johns Hopkins University

For multivariate spatial Gaussian process models, customary specifications of cross-covariance functions do not exploit relational inter-variable graphs to ensure process-level conditional independence between the variables. This is undesirable, especially in highly multivariate settings, where popular cross-covariance functions, such as multivariate Matérn functions, suffer from a curse of dimensionality as the numbers of parameters and floating-point operations scale up in quadratic and cubic order, respectively, with the number of variables. We propose a class of multivariate graphical Gaussian processes using a general construction called stitching that crafts cross-covariance functions from graphs and ensures process-level conditional independence between variables. For the Matérn family of functions, stitching yields a multivariate Gaussian process whose univariate components are Matérn Gaussian processes, and which conforms to process-level conditional independence as specified by the graphical model. For highly multivariate settings and decomposable graphical models, stitching offers massive computational gains and parameter dimension reduction. We demonstrate the utility of the graphical Matérn Gaussian process to jointly model highly multivariate spatial data using simulation examples and an application to air-pollution modelling.

## Title: Bayesian screening via mixtures of shrinkage priors with applications to light-sheet fluorescence microscopy in brain imaging

Speaker: *Francesco Denti*

IS

Università Cattolica del Sacro Cuore - Milan

The majority of classical screening methods partition the units of interest (e.g., brain regions) into two classes: significant and non-significant. However, such binary grouping methods may lead to overly simplistic results, filtering out weak but important signals that might have been degraded by the noise present in the data. To overcome this limitation, we introduce a new Bayesian approach that can classify the brain regions into several tiers with varying degrees of relevance. Our approach is based on a combination of shrinkage priors – used in regression and multiple hypothesis testing problems – and mixture models – employed in model-based clustering. Our model provides a general setting for Bayesian sparse estimation, reduces the number of shrinkage parameters needed, and creates a framework for sharing information across units of interest. This approach leads to more biologically meaningful and interpretable results in our brain imaging problem: it allows for the discrimination between active and inactive regions while ranking the discoveries into clusters representing tiers of similar importance. Additionally, we investigate an extension of the classical two-group model to exploit the hierarchical structure of the brain regions within our multiple hypothesis testing framework, accounting for each parent region's contribution to the overall brain activities.

## **Title: Beyond axis-aligned decisions: a new BART prior for structured categorical predictors**

**Speaker: Sameer Deshpande**

IS

University of Wisconsin-Madison

Default implementations of Bayesian Additive Regression Trees (BART) represent categorical predictors using several binary indicators, one for each level of each categorical predictor. Axis-aligned decision rules are well-defined with these indicators; they send one level of a categorical predictor to the left and all other levels to the right (or vice versa). Regression trees built with these rules partition the set of all levels of a categorical predictor by recursively removing one level at a time. Unfortunately, most partitions of the levels cannot be built with this “remove one at a time” strategy, meaning that default implementations of BART are extremely limited in their ability to “borrow strength” across groups of levels.

We overcome this limitation with a new class of decision rules and prior partitioning process that can send multiple levels of a categorical variable to each child of a decision node in a regression tree. We then extend these rules to accommodate structured categorical variables for which there are a priori preferences to cluster certain levels together. In spatio-temporal applications, such variables are frequently used to encode an observation’s membership in spatial units displaying natural adjacency structures like census tracts or counties. In these applications, our new decision rules are able to produce spatially-contiguous partitions of the areal units. Our new categorical decision rules can be integrated seamlessly within the standard Bayesian backfitting procedure used to fit BART, often yielding improved out-of-sample predictive performance without much additional computational burden. Despite their conceptual simplicity, our new decision rules open the door for performing Bayesian treed regression over complex discrete spaces like networks.

## **Title: Data Science to Address the Health Impacts of Climate Change**

**Speaker: Francesca Dominici**

KL

Harvard University

In this talk, I will provide an overview of data science methods, including methods for Bayesian analysis, causal inference, and machine learning, to inform environmental policy. This is based on my work analyzing a data platform of unprecedented size and representativeness. The platform includes more than 500 million observations on the health experience of over 95% of the US population older than 65 years old linked to air pollution exposure and several confounders. Finally, I provide an overview of studies on air pollution exposure, environmental racism, wildfires, and how they also can exacerbate the vulnerability to COVID-19.

## **Title: Graph based Gaussian processes on restricted domains**

**Speaker: David Dunson**

IS

Duke University

Gaussian processes (GP) are used routinely for modeling of functional data and surfaces. In such settings, it is common for the inputs to fall in a restricted subset of Euclidean space. Typical kernel-based methods that do not take into account the intrinsic geometry of the domain across which observations are collected may produce sub-optimal results. We focus on solving this problem in the context of Gaussian process (GP) models, proposing a new class of Graph Laplacian based GPs (GL-GPs), which learn a covariance that respects the geometry of the input domain. As the heat kernel is intractable computationally, we approximate the covariance using finitely-many eigenpairs of the Graph Laplacian (GL). The GL is constructed from a kernel which depends only on the Euclidean coordinates of the inputs. Hence, we can benefit from the full knowledge about the kernel to extend the covariance structure to newly arriving samples by a Nyström type extension. We provide substantial theoretical support for the GL-GP methodology, and illustrate performance gains in various applications.

## **Title: Bayesian nonparametric factorizations of matrix-valued parameters**

**Speaker: Daniele Durante**

IS

Bocconi University

Matrix-valued parameters are often encountered in several statistical models. In modern applications, such matrices are increasingly high-dimensional and often sparse, thus motivating lower dimensional representations based on latent variables. A common problem within this context is the lack of knowledge on the dimension of the latent space, which has to be learned from the observed data. To address this issue, it is possible to rely either on over-complete representations in combination with suitable shrinkage priors, or to consider a prior distribution also for the unknown dimension of the latent space. In this talk, I will discuss how Bayesian nonparametric methods can be effective in both directions. More specifically, the focus will be mainly on presenting a novel increasing shrinkage prior, named cumulative shrinkage process, which is based on a sequence of spike-and-slab distributions that assigns increasing mass to the spike as model complexity grows via a cumulative sum of stick-breaking probabilities. Using factor analysis as an illustrative example, I will show that this formulation has theoretical and practical advantages, including an improved ability to recover the model dimension. Towards the end of the talk, I will also briefly mention the problem of learning block-structures among groups of nodes in network data, while recovering the unknown number of clusters. Motivated by criminal network studies, I will discuss a general framework for stochastic block modeling, named extended stochastic block model, that infers groups of nodes via Gibbs-type priors on the partition process. Within this unified formulation, I will focus in particular on the Gnedin process as a realistic prior that allows the number of groups to be finite, random and subject to a reinforcement process coherent with the modular structures in organized crime. The performance is illustrated in an application to an Italian Mafia network, where the proposed model unveils complex block structures, mostly hidden from state-of-the-art alternative solutions.

## **Title: Bayesian sensitivity analysis for a missing data model**

**Speaker: Bart Eggen**

IS

Delft University of Technology

In the field of causal inference, sensitivity analysis is very important to assess the robustness of study conclusions when certain assumptions are not satisfied. We will look at the missing data model and perform sensitivity analysis under the assumption that missing outcomes are missing completely at random. Scharfstein et. al. (2003) introduced a fully Bayesian methodology and conjectured a Bernstein-von Mises theorem for the posterior of the outcome, by incorporating prior beliefs about non-identifiable, but interpretable parameters. We prove their conjecture by providing two Bernstein-von Mises theorems using a Dirichlet process prior and a normalized extended gamma process prior. We also provide a simulation study, showing the performance of the methods in finite sample scenarios.

## **Title: Lower-dimensional Bayesian Mallows model for rank-based unsupervised transcriptomic analysis**

**Speaker: Emilie Eliseussen**

IS

University of Oslo

The Bayesian Mallows model for analysing ranking data has been previously introduced, showing very good accuracy when using different types of preference data. However, inference with the Bayes Mallows model is computationally intensive when data are high-dimensional: for instance, the model cannot scale to the data dimension typical of -omics applications (thousands of patients and hundreds of thousands of omic features). We describe one possible way of performing inference on a reduced-dimensional Mallows model, when handling high-dimensional data. This is relevant for some reasons: (i) computational advantages, since by reducing the dimension of the parameter space, inference is much easier and faster; (ii) better modeling, since the reduced-dimensional Mallows allows for automatically selecting the items that are “relevant enough” to be included in the distance computation; (iii) better results interpretation, since a model with fewer parameters enforces a better distinction between noisy and relevant items. All these aspects have a quite substantial importance in -omics applications.

## **Title: Advances in Bayesian Inference for Binary and Categorical Data**

**Speaker: Augusto Fasano**

IS

European Commission, Joint Research Centre (JRC), Ispra and Collegio Carlo Alberto

Bayesian binary probit regression and its extensions to time-dependent observations and multi-class responses are popular tools in binary and categorical data regression due to their interpretability and non-restrictive assumptions. Although the theory is well established in the frequentist literature, these models still face a florid research in the Bayesian framework to overcome computational issues or inaccuracies in high dimensions as well as the lack of a closed-form expression for the posterior distribution of the model parameters in many cases. We develop a novel variational approximation for the posterior distribution of the coefficients in high-dimensional probit regression with binary responses and Gaussian priors, resulting in a unified skew-normal (SUN) approximating distribution that converges to the exact posterior as the number of predictors increases. Moreover, we derive closed-form expressions for posterior distributions arising from models that account for correlated binary time-series and multi-class responses, developing computational methods that outperform state-of-the-art routines. Finally, motivated by links to the Bayesian Nonparametrics literature, we introduce a novel partially-exchangeable species-sampling model with the twofold goal of predicting the classes (or species) of the future observations and of testing for homogeneity among the different available populations.

## **Title: Against all odds: enhancing statistical monitoring schemes using new sources of data**

**Speaker: Edgar Santos Fernandez**

IS

Queensland University of Technology

Modern monitoring schemes are increasingly relying on new sources of data such as those produced by sensor technology and citizen science programs. However, making inferences is challenging since these datasets are generally messy and observed with error. These complications are exacerbated when dealing with high-dimensional and big datasets and complicated spatial and temporal dependence.

This presentation will focus on two real-world monitoring challenges in stream networks and coral reefs, and the statistical methods that have been developed to correct bias, detect anomalies, and produce more meaningful and trustworthy outcomes. Attention will also be paid to computationally efficient implementations of these methods.

This work is joint with Professors Erin Peterson, Jay ver Hoef, James McGree and Kerrie Mengersen.

## **Title: Bayesian analysis of GLMMs with nonlocal priors for genome-wide association studies**

**Speaker: Marco Ferreira**

IS

Virginia Tech

We present a novel Bayesian method to find single nucleotide polymorphisms (SNPs) associated with particular phenotypes measured as discrete data from genome-wide association studies (GWAS). This is a regression problem with  $p$  two to three orders of magnitude larger than  $n$ , the subjects are correlated, and the SNPs regressors are highly correlated. To deal with these challenges, we propose nonlocal priors specifically tailored to GLMMs and develop related fast approximate computations for Bayesian model selection. To search through hundreds of thousands of possible SNPs, we use a two-step procedure: first, we screen for candidate SNPs; second, we perform model search that considers all screened candidate SNPs as possible regressors. A simulation study shows favorable performance of our Bayesian method when compared to other methods widely used in the GWAS literature. We illustrate our method with applications to the analysis of real GWAS datasets from plant science and human health.

## **Title: Posterior contraction for deep Gaussian process priors**

**Speaker: Gianluca Finocchio**

CT

University of Vienna

We study posterior contraction rates for a class of deep Gaussian process priors applied to the nonparametric regression problem under a general composition assumption on the regression function. It is shown that the contraction rates can achieve the minimax convergence rate (up to log factors), while being adaptive to the underlying structure and smoothness of the target function. The proposed framework extends the Bayesian nonparametrics theory for Gaussian process priors.

## **Title: Updating Variational Bayes: Fast sequential posterior inference**

**Speaker: Catherine Forbes**

IS

Monash University

Variational Bayesian (VB) methods produce posterior inference in a time frame considerably smaller than traditional Markov Chain Monte Carlo approaches. Although the VB posterior is an approximation, it has been shown to produce good parameter estimates and predicted values when a rich classes of approximating distributions are considered. In this paper, we propose the use of recursive algorithms to update a sequence of VB posterior approximations in an online, time series setting, with the computation of each posterior update requiring only the data observed since the previous update. We show how importance sampling can be incorporated into online variational inference allowing the user to trade accuracy for a substantial increase in computational speed. The proposed methods and their properties are detailed in two separate simulation studies. Additionally, two empirical illustrations are provided, including one where a Dirichlet Process Mixture model with a novel posterior dependence structure is repeatedly updated in the context of predicting the future behaviour of vehicles on a stretch of the US Highway 101.

## **Title: On the Importance of Priors in Bayesian Deep Learning**

**Speaker: Vincent Fortuin**

IS

University of Cambridge

While Bayesian deep learning has been a popular field of research in recent years, most of the work has focused on improving inference methods for better performance and lower computational costs. Conversely, the priors have often been ignored and merely chosen to be isotropic Gaussian, for mathematical and computational convenience. In this talk, I will review recent work that calls this popular practice into question and highlights pathologies that can arise from prior misspecification in Bayesian neural networks. I will then present different methods that can aid the selection of better priors and I will discuss the advantages of function-space priors over weight-space ones.

## **Title: Bayesian covariance structure modeling of clustered data**

**Speaker: Jean-Paul Fox**

IS

University of Twente

A Bayesian multivariate model with a structured covariance matrix for multi-way clustered data is proposed. This flexible modeling framework allows for positive and for negative associations among clustered observations, and generalizes the well-known dependence structure implied by random effects. A conjugate shifted-inverse gamma prior is proposed for the covariance parameters which ensures that the covariance matrix remains positive definite under posterior analysis. A numerically efficient Gibbs sampling procedure is defined for (balanced) designs. For categorical outcome data, a data augmentation procedure is proposed which facilitates sampling latent variables from (truncated) univariate normal distributions, and avoids numerical computation of the inverse of the structured covariance matrix. In the multivariate approach, the dependence structure can represent negative, no (zero), or positive associations, and therefore avoids issues with the lowerbound of zero of a random effect variance. This has the advantage that our general modeling approach does not show testing or estimation issues, when the within-cluster and/or between-cluster (co)variance approaches zero. The Bayesian covariance structure model is applied to interval-censored event times to analyze differences in adverse events between different groups of patients (analyze treatment effects) and between patients receiving the same treatment (analyze individual differences in treatment effects).

## **Title: Multi-Task Gaussian Process Models for NBA Production Curves**

**Speaker: Alexander Franks**

IS

University of California, Santa Barbara

Understanding how athlete skill changes with age is of immense importance across a range sports, in large part because it can be used to forecast changes future player value. To date, most existing work on aging curves in sports focuses on modeling changes in aggregate measures of player ability, like Wins Above Replacement (WAR). However, different aspects of player skill and ability evolve differently with age. For example, metrics that demand more athleticism and agility tend to degrade more quickly with age, whereas metrics that are less dependent on athleticism (e.g. free throw shooting in basketball) are relatively invariant to aging (or even improve with practice). Different metrics for player ability/skill are also not independent, but rather are a reflection of a combination of different latent player attributes. In this work, we jointly model how multiple metrics evolve with time using a multi-task Gaussian Process. We demonstrate the utility of our method with an analysis of NBA data.

## **Title: Nonparametric priors for multi-sample data: dependence and borrowing of information**

**Speaker: Beatrice Franzolini**

IS

Agency for Science, Technology and Research

When data come from different, though related, sources, they typically are homogeneous within and heterogeneous across groups and the Bayesian paradigm requires a prior law over a collection of distributions. From a modelling point of view, it is essential to study how this structure reflects on the observables, especially in nonparametric models. We explore the implications on the induced dependence structure and borrowing of information for a general class of nonparametric priors, which encompasses several popular dependent nonparametric priors and has the merit of highlighting their core features. We show how to compute the correlation between observables and show how its sign depends from the joint specification of the underlying random probabilities. Finally, we propose a novel class of dependent nonparametric priors, which may induce either positive or negative correlation across samples.

## **Title: Asymptotic Properties of Bayesian Synthetic Likelihood**

**Speaker: David Frazier**

CT

Monash University

Synthetic likelihood is a common approach for carrying out Bayesian inference when the likelihood is intractable, but it is straightforward to simulate from the model. The method constructs an approximate likelihood by assuming that the distribution of a vector summary statistic is multivariate normal, with the unknown mean and covariance matrix estimated by simulation for any given parameter value. We explore the theoretical behavior of the Bayesian synthetic likelihood (BSL) posterior and find several interesting conclusions. In correctly specified models, BSL delivers inferences that are asymptotically equivalent to those obtained by approximate Bayesian computation (ABC), however, we show that BSL is more computationally efficient than ABC. In contrast to ABC, we find that the BSL posterior can display a wide range of behaviors depending on the level of model misspecification, including being asymptotically mixed-Gaussian. We document all theoretical results using a simple running example.

## **Title: When it counts – Econometric Identification of Bayesian Factor Models with Sparse Loading Matrices**

**Speaker: Sylvia Frühwirth-Schnatter**

IS

WU Vienna University of Economics and Business

Factor analysis is a popular method to obtain a sparse representation of the covariance matrix of multivariate observations. In many contexts, it is also relevant to obtain inference regarding the number of factors and the structure of the loading matrix. In this talk we show that this challenging problem can be resolved within the class of Bayesian factor models with a generalized lower triangular (GLT) structure in combination with spike and slab priors that introduce sparsity in the loading matrix. Sparse GLT structures not only resolve rotational identification, but allow verifying variance identification through a simple counting rule that also yields a mathematically justified estimator of the number of factors. Applications to financial and economic data will serve as an illustration.

## **Title: Partially-exchangeable multilayer stochastic block models**

**Speaker: Francesco Gaffi**

CT

Bocconi University

There is an increasing availability of complex network data encoding connectivity information among a set of nodes, often belonging to different layers. For example, in bill co-sponsorship networks, the nodes correspond to political candidates, whereas layers denote the party of affiliation, and the goal is to infer grouping structures made by candidates with similar behaviors in co-sponsoring bills. Although it could be reasonable to expect such blocks to coincide with layers, this assumption is too strong and fails to learn sub-blocks within each layer as well as across-layer grouping structures. To incorporate these mixed architectures while accounting for layer information in a principled manner, we propose a new generation of partially-exchangeable multilayer stochastic block models relying on a hierarchical random partition prior for the node allocations driven by the urn scheme of a hierarchical normalized completely random measure (H-NRMI) or a hierarchical Pitman-Yor process (H-PYP). The partial exchangeability assumption among nodes according to layer partitions allows to infer both within- and across-layer blocks, while preserving probabilistic coherence, principled uncertainty quantification and formal inclusion of prior information from layer membership. The mathematical tractability of such priors further allows to analytically derive and compare predictive within- and across-layer co-clustering probabilities, thereby providing conditions on hyperparameters to enforce interpretable features on the grouping structures. The applied potentials of this new class of Bayesian nonparametric models are illustrated in political and criminal network studies.

## Title: Structue Learning of Mixed Graphical Models

Speaker: *Chiara Galimberti*

CT

Università degli Studi di Milano-Bicocca

Mixed data are variables of heterogeneous nature, with both discrete and continuous measurements, and they can be found in various areas, such as genomics or industrial machine processes. For instance, in industrial processes, discrete values identify machine settings or quality outputs and continuous variables identify measurements of physical quantities. Graphical models have proven to be an effective tool for representation of conditional independences among variables. They have been extensively studied for separate set of continuous or discrete variables, but there have been limited efforts on joint statistical models, in simplified settings that constrain to specific model distributions or that approximate parts of the model otherwise computationally prohibitive. We propose a Bayesian Conditional Gaussian model able to deal with mixed data, for posterior parameter inference and for learning the graph-driven structure of dependences. We implement an MCMC algorithm and the method is compared with alternatives in the literature, showing favorable results.

## Title: Average Error Controlled Bayesian Sample Size Determination Methods

Speaker: *Sujit Ghosh*

IS

North Carolina State University

Sample size determination is one of most widely used methods in bio-pharmaceutical applications. The primary goal is to determine minimal sample size that would achieve a desired sampling precision for statistical estimation and/or maintain a pre-specified Type-I and Type-II error rates for hypotheses testing. However, Bayesian methods are not usually designed to achieve such desired precision or control error rates despite the need for such desired characteristics from regulatory perspectives. On the other hand, the success of classical sample size determination methods crucially depend on finding a pivotal quantity which becomes increasingly difficult for general composite null hypothesis (e.g., bio-equivalence and non-inferiority tests) involving nuisance parameters. Thus, a unified methodological framework is needed that not only provide theoretical guarantees for controlling desired level of errors but is also broadly applicable for composite null hypothesis. This talk will focus on presenting average error controlled Bayesian methodologies for sample size calculation primarily for hypotheses testing framework.

## Title: Immersion posterior: Empowering Bayesians to meet frequentist targets under structural restrictions

Speaker: *Subhashis Ghosal*

FL

North Carolina State University

Bayesian procedures can be attractive from the frequentist perspective, if carefully constructed. Classical results such as the complete class theorem and the Bernstein-von Mises theorem justify Bayesian estimators and credible intervals from a frequentist perspective in parametric problems. Posterior consistency and contraction rates theorems developed over the last sixty years gave conditions for frequentist validity of many Bayesian procedures in infinite-dimensional parameter spaces. In more recent years, finer properties like the frequentist coverage of credible regions for nonparametric problems have come into prominence, but most works have been done under smoothness regimes. In many contexts of statistical modeling, the shape of a function used in modeling plays a key role. Shape restrictions like monotonicity, convexity, log-concavity or unimodality may arise naturally. In a differential equation model, the regression function is implicitly given by a set of differential equations. These are examples of restricted parameter spaces, where the original parameter space may be viewed as a subset of larger parameter space where the model remains meaningful. In such situations, it is operationally and computationally convenient to put a prior on the bigger space, where classical tools like conjugacy may be available. To make an inference with a posterior distribution supported within the intended parameter space, we consider an immersion map that induces such a distribution from the posterior on the larger parameter space. A projection map is a natural type of immersion map, but more general immersion maps may be beneficial for some problems. Exploiting the explicit description of the immersion posterior distribution, we show that the corresponding “immersion posterior” leads to credible regions with frequentist validity for many problems including monotone regression, monotone density, multivariate monotone regression, multivariate monotone density, convex regression, ordinary and partial differential equation models, and so on. Interesting new phenomena such as over-coverage, recalibration, and exotic new limit distributions arise from the analysis. The immersion posterior approach also allows us to derive the optimal rate of contraction and obtain Bayesian tests for structural hypotheses with optimal asymptotic properties.

## **Title: Flexible spatial modeling of areal data: Introducing the Hausdorff-Gaussian Spatial Process**

**Speaker: Lucas Godoy**

CT

University of Connecticut

Accounting for spatial dependence when analyzing areal data is extremely important for efficient and valid statistical inferences. Most existing models for areal data employ adjacency matrices to quantify the spatial dependence structure, where spatial polygons of different shapes and sizes are treated the same way. Such methodologies impose some limitations. Remarkably, computing predictions in different maps may become impractical. We propose a flexible model for area data from a geostatistical perspective with a Gaussian process defined based on the Hausdorff distance instead of the Euclidean distance to circumvent these limitations. We present the benefits of the proposed method in Bayesian spatial modeling via a simulation study comparing its performance to popular spatial models. The model is used to fit the average income at the census tracts in a Brazilian city called Barbacena. Our methodology, has performed better (WAIC = 409) than a CAR model (WAIC = 433.1).

## **Title: Assessment of design operating characteristics in Bayesian adaptive trials**

**Speaker: Shirin Golchi**

IS

McGill University

Bayesian adaptive designs have gained popularity in all phases of clinical trials with numerous new developments in the past few decades. Designing Bayesian adaptive trials, however, requires extensive simulation studies that are generally considered challenging, particularly in time-sensitive settings such as a pandemic. In this talk, I lay out a set of methods for efficient estimation and uncertainty quantification for design operating characteristics of Bayesian adaptive trials. The proposed approach is to model the sampling distribution of Bayesian probability statements that are commonly used as the basis of decision making. To showcase the implementation and performance of the proposed approach, we use a clinical trial design with an ordinal disease-progression scale endpoint.

## **Title: Model-based clustering of high-dimensional data via Bayesian factor models**

*Speaker: Claire Gormley*

IS

University College Dublin, Ireland

Factor-analytic Gaussian mixture models are often employed as a model-based approach to clustering high-dimensional data. Typically, the numbers of clusters and latent factors must be specified in advance of model fitting, and remain fixed. The pair which optimises some model selection criterion is then chosen. For computational reasons, models in which the number of latent factors differ across clusters are rarely considered. Here the infinite mixture of infinite factor analysers (IMIFA) model is introduced. IMIFA employs a Pitman-Yor process prior to facilitate automatic inference of the number of clusters using the stick-breaking construction and a slice sampler. Furthermore, IMIFA employs multiplicative gamma process shrinkage priors to allow cluster-specific numbers of factors, automatically inferred via an adaptive Gibbs sampler. IMIFA is presented as the flagship of a family of factor-analytic mixture models, providing flexible approaches to clustering high-dimensional data. Applications to a benchmark data set, metabolomic spectral data, and a manifold learning handwritten digit example illustrate the IMIFA model and its advantageous features. These include obviating the need for model selection criteria, reducing the computational burden associated with the search of the model space, improving clustering performance by allowing cluster-specific numbers of factors, and quantifying uncertainty in the numbers of clusters and cluster-specific factors.

## **Title: Bayesian Multi-Study Non-Negative Matrix Factorization for Mutational Signatures**

*Speaker: Isabella Grabski*

IS

Harvard University

Mutational signatures reveal the processes underlying tumors and are typically identified from cancer genomes using non-negative matrix factorization (NMF). It is often of interest to compare signatures across conditions, such as treatments or cancer types, but existing techniques only decompose one dataset at a time. We introduce a Bayesian NMF method that jointly decomposes multiple datasets to estimate signatures and their sharing patterns across conditions. We propose two models: a completely unsupervised "discovery-only" model, and a "recovery-discovery" model building informative priors from previous findings to identify both known and novel signatures. We extend these models to estimate the effects of sample-level covariates, enforcing sparsity through a non-local spike-and-slab prior. Our approach is evaluated in extensive simulations and real data applications.

## Title: Spatio-temporal stick breaking

Speaker: Clara Grazian

IS

University of Sydney

Dirichlet processes and their extensions have reached a great popularity in Bayesian nonparametric statistics. They have also been introduced for spatial and spatio-temporal data, as a tool to analysis and predict surfaces. A popular approach to Dirichlet process in a spatial setting relies on a stick-breaking representation of the process, where the dependence over space is described in the definition of the stick-breaking probabilities. Extensions to include temporal dependence usually introduce a temporal dependence among the atoms of the Dirichlet process, however this approach does not allow to properly test and incorporate a possible interaction between space and time. In this talk, we propose a Dirichlet process where the stick-breaking probabilities are defined to incorporate both spatial and temporal dependence. An advantage of the method is that it offers a natural way to test for separability of the two components. The performance of this approach will be tested on simulations and a real-data example from meteorology.

## Title: Repulsive mixture models for high-dimensional data

Speaker: Alessandra Guglielmi

IS

Politecnico di Milano

Model-based clustering is customarily achieved in the Bayesian setting through finite or infinite mixture models, assuming that data are iid generated from homogeneous populations, represented by parametric densities. In this talk, we focus on the large- $p$  setting, i.e. for instance when  $p$  is in the order of hundreds or thousands, and possibly larger than the sample size. The poor performance of Bayesian mixtures in the large- $p$  setting is known. For instance, Chandra, Canale and Dunson (2020) show that Gaussian mixtures lead to inconsistent cluster estimates when  $p$  increases to infinity. They propose a latent factor model, where clustering is performed at the latent level, specifically on  $d$ -dimensional latent parameters, for  $d$  much smaller than  $p$ . In this talk we build on a class of mixtures of latent factor models, similar to the one in Chandra, Canale and Dunson (2020), mixing over the latent parameters. Our main contribution to the model is the assumption of a repulsive point process as mixing measure. The matrix of factor loadings drives the anisotropic behavior, so that separation is indeed induced between the high-dimensional centers of different clusters. In this talk (details in Lorenzo Ghilotti's talk), we propose a MCMC algorithm which extends the conditional algorithm introduced in Beraha et al. (2022) for repulsive mixture models. To sample from the full conditional of the factor loadings, we replace the standard Metropolis step by a Metropolis adjusted Langevin algorithm.

This is a joint work with Mario Beraha (Università di Bologna & Politecnico di Milano) and Lorenzo Ghilotti (Università di Milano Bicocca).

## **Title: Bayesian Generalized Sparse Symmetric Tensor-on-Vector Regression**

**Speaker: Sharmistha Guha**

IS

Texas A&M University

Motivated by brain connectome datasets acquired using diffusion weighted magnetic resonance imaging (dMRI), this article proposes a novel generalized Bayesian linear modeling framework with a symmetric tensor response and scalar predictors. The symmetric tensor coefficients corresponding to the scalar predictors are embedded with two features: low-rankness and group sparsity within the low-rank structure. Besides offering computational efficiency and parsimony, these two features enable identification of important “tensor nodes” and “tensor cells” significantly associated with the predictors, with characterization of uncertainty. The proposed framework is empirically investigated under various simulation settings and with a real brain connectome dataset. Theoretically, we establish that the posterior predictive density from the proposed model is “close” to the true data generating density.

## **Title: Bayesian data compression for spatially correlated data**

**Speaker: Rajarshi Guhaniyogi**

IS

Texas A&M University

We will introduce Bayesian data compression for spatial regression models to obviate computational challenges presented by large numbers of spatial locations. To address the challenges of analysing large spatial data, we compress spatially oriented data by a random linear transformation to achieve dimension reduction and conduct inference on the compressed data. Our approach distinguishes itself from several existing methods for analysing large spatial data in that it requires neither the development of new models or algorithms nor any specialised computational hardware while delivering fully model-based Bayesian inference. Well-established methods and algorithms for spatial regression models can be applied to the compressed data and compressed data protects privacy of the samples. We will discuss theoretical results on data compression idea and provide empirical understanding of the proposed approach.

## **Title: Bayesian approaches for clustering distributional features in neuroimaging experiments.**

**Speaker: Michele Guidani**

IS

University of California, Irvine

In the neurosciences, it is now widely established that brain processes are characterized by heterogeneity at several levels. For example, neuronal processes differ by external stimuli, and patterns of brain activations vary across subjects. In this presentation, we will discuss a few Bayesian strategies for characterizing heterogeneity in the neurosciences, where time-series data are assumed to be organized in different, but related, units (e.g., neurons and/or regions of interest), and some sharing of information is required to learn distinctive features of the units. First, we will briefly discuss models for multi-subject analysis that will identify population subgroups characterized by similar brain activity patterns, also by integrating available subject information. Then, we will look at how novel techniques in intracellular calcium signals may be used to analyze neural responses to external stimuli in awake animals. The performance of the models will be demonstrated by applications to data from fMRI, EEG, and fluorescence microscopy experiments.

## **Title: Handling Partial Identification: Sensitivity Analysis, Inference, or Stuck in the Middle with Bayes?**

**Speaker: Paul Gustafson**

IS

University of British Columbia

In many scientific application areas, realistic modeling of available data may lead to only partial statistical identification of parameters of interest. Intuitively put, with only a partially identified model, even an infinite amount of data would not reveal the exact value of a scalar target parameter. Though it could reveal lower and upper bounds for the target. This talk will review Bayesian inference in partially identified settings, describing how it works and when and why it might be useful. We will try to shed light on this perhaps murky realm, wedged as it is between what statisticians typically think of sensitivity analysis, and what they typically think of as inference. We discuss connections with non-Bayesian methods based on parameter bounds, and consider options for prior specification in such settings.

## **Title: Principled, practical, flexible, fast: a new approach to phylogenetic factor analysis**

**Speaker: Gabriel Hassler**

IS

Department of Computational Medicine, UCLA

Biological traits are products of complex evolutionary processes in which selective forces influence multiple biological characteristics in unknown ways. Phylogenetic factor analysis disentangles these relationships across the evolutionary history of a group of organisms. These relationships can include, for example, the influence of interacting viral mutations on transmissibility and clinical outcomes. Scientists seeking to employ this modeling framework confront numerous modeling and implementation decisions, the details of which pose computational and replicability challenges. General and impactful community employment requires a data scientific analysis plan that balances flexibility, speed and ease of use, while minimizing model and algorithm tuning. Even in the presence of non-trivial phylogenetic model constraints, we show that one may analytically address latent factor uncertainty in a way that (a) aids model flexibility, (b) accelerates computation (by as much as 500-fold) and (c) decreases required tuning. We further present practical guidance on inference and modeling decisions as well as diagnosing and solving common problems in these analyses. We codify this analysis plan in an automated pipeline that distills the potentially overwhelming array of modeling decisions into a small handful of (typically binary) choices.

## **Title: Shrinkage on the Simplex: Bayesian Inference with Sparse Generalized Dirichlet Distributions**

**Speaker: Matthew Heiner**

CT

Brigham Young University

The Dirichlet distribution is the default choice in many modeling settings including categorical data analysis, as a prior for mixture proportions, and for the population distribution of latent variables. Its popularity among modelers owes to: (i) its simple and interpretable structure, (ii) conjugacy to multinomial likelihoods, and (iii) shrinkage properties when the concentration parameter is close to zero. Property (iii) is often exploited to allow many elements of a probability vector to be close to zero; for example, favoring effective deletion of unnecessary components in probabilistic factorizations. However, the Dirichlet is actually very inflexible as a shrinkage prior. Motivated by this problem, we propose a new class of Sparse Generalized Dirichlet distributions that maintains (i)-(iii) while improving concentration around sparse vectors. We provide theoretical support for the new class of distributions and show improved performance in simulation settings, as well as an application to ecological data on species counts.

## **Title: Prior elicitation for variance parameters in Bayesian hierarchical models**

**Speaker: Ingeborg Gullikstad Hem**

IS

Norwegian University of Science and Technology

The prior distribution is a central component in Bayesian inference, and allow us to incorporate expert knowledge in the model and make the Bayesian inference robust. However, prior selection can be challenging even for statistical experts, and default choices, ad hoc procedures or suggestions from the literature are often used in practice. In this talk, I will present a new tool for formulating joint priors for variance parameters that eases the prior specification in hierarchical models. Using a hierarchical variance decomposition, we formulate a joint prior for variance parameters that takes the entire model structure into account. The user can utilize prior beliefs to proportion the total variance present in the model to the different random effect components. This allows for existing knowledge to be intuitively incorporated at the level it applies to. This framework is implemented in the user-friendly R-package 'makemyprior', with a graphical interface where the user can click through the model and select priors, making the process transparent and interactive.

## **Title: Bayesian calibration of a model of forced displacement**

**Speaker: Jason Hilton**

IS

University of Southampton

Recent years have seen a rapid increase in the number of people forcibly displaced from their homes, who often must make decisions about where to move to and what route to take quickly and with limited information. Changing policies of destination and transit countries together with access to new information frequently necessitates re-evaluation of these decisions. Information gathered by migrants may be shared with their social network, meaning that individual route choices depend upon the decisions of and interactions with others. Such migration systems are complex, in that it is difficult to predict the macro-level (aggregate) end state of such a system from knowledge of the micro-level behaviour. Understanding aggregate migration route dynamics is therefore difficult, but agent-based simulation provides a tool through which such systems can be studied. Agent-based simulation involves creating a computer model which allows software agents representing individuals to interact with each other and their environment according to decision rules governing individual behaviour. Agent-based models typically have free parameters which are difficult or impossible to measure accurately. Calibration involves finding the distribution of parameters such that the simulation output is close to empirical values, accounting for the various sources of uncertainty including parameter errors and stochasticity of the simulation itself. This paper reports on a simulation model of migration route dynamics for Syrian migrants to Europe. Using limited data on migrant presence at locations in North Africa, parameters governing migrant decision making and information transfer are calibrated so that the simulation can replicate empirical patterns.

## **Title: Bayesian approaches to unifying causal inference from mechanistic and statistical models**

**Speaker: Joseph Hogan**

IS

Brown University

Causal inference requires extrapolating unobserved potential outcomes from a model of an observed-data process. Causal inference from statistical models typically involves fitting a parametric or nonparametric model to observed data and using that model to generate predicted outcomes under different levels of an exposure of interest. Unless the observed data derive from a randomized trial, the inferences themselves depend on untestable assumptions about confounding (e.g., 'no unmeasured confounding'). Causal inference from mechanistic models, sometimes formulated as agent-based models, involves generating simulated values under different exposure levels using pre-specified models. For both types of models, a version of the g-formula can be applied to generate predicted or simulated outcomes.

This talk will describe a fully Bayesian formulation that casts causal inference as a posterior predictive inference problem, providing a unifying approach that applies to both statistical and mechanistic models. An appropriately formulated observed-data likelihood shows how information is used or, in the case of mechanistic models, integrated from multiple data sources; prior distributions are used to encode untestable assumptions and to assess sensitivity to violations of them. We draw connections to 'data fusion' methods and illustrate the fully Bayesian analyses using data and models related to COVID and HIV.

## **Title: A quantum parallel Markov chain Monte Carlo**

**Speaker: Andrew Holbrook**

CT

UCLA Biostatistics

We propose a novel quantum computing strategy for parallel MCMC algorithms that generate multiple proposals at each step. This strategy makes parallel MCMC amenable to quantum parallelization by using the Gumbel-max trick to turn the generalized accept-reject step into a discrete optimization problem. This allows us to embed target density evaluations within a well-known extension of Grover's quantum search algorithm. Letting  $P$  denote the number of proposals in a single MCMC iteration, the combined strategy reduces the number of target evaluations required from  $O(P)$  to  $O(\sqrt{P})$ .

## **Title: Using BART to Perform Pareto Optimization and Quantify its Uncertainties**

**Speaker: Akira Horiguchi**

IS

Duke University

Techniques to reduce the energy burden of an industrial ecosystem often require solving a multiobjective optimization problem. However, collecting experimental data can often be either expensive or time-consuming. In such cases, statistical methods can be helpful. This article proposes Pareto Front (PF) and Pareto Set (PS) estimation methods using Bayesian Additive Regression Trees (BART), which is a nonparametric model whose assumptions are typically less restrictive than popular alternatives, such as Gaussian Processes (GPs). These less restrictive assumptions allow BART to handle scenarios (e.g. high-dimensional input spaces, nonsmooth responses, large datasets) that GPs find difficult. The performance of our BART-based method is compared to a GP-based method using analytic test functions, demonstrating convincing advantages. Finally, our BART-based methodology is applied to a motivating engineering problem.

## **Title: The current state of undergraduate Bayesian education and recent developments**

**Speaker: Jingchen (Monika) Hu**

IS

Vassar College

With the advances in tools and the rise of popularity, Bayesian statistics is becoming more important for undergraduates. This session gathers three Bayesian educators who have made recent contribution to the teaching and learning of Bayesian statistics at the undergraduate level. In my discussant remark, I will describe a recent study on the current state of undergraduate Bayesian education in the United States, summarize the three talks, and give general recommendations for the future.

## **Title: Bayesian Nonparametric Estimation for Point Processes with Spatial Homogeneity: A Spatial Analysis of NBA Shot Locations**

**Speaker: Guanyu Hu**

IS

University of Missouri

Basketball shot location data provide valuable summary information regarding players to coaches, sports analysts, fans, statisticians, as well as players themselves. Represented by spatial points, such data are naturally analyzed with spatial point process models. We present a novel nonparametric Bayesian method for learning the underlying intensity surface built upon a combination of Dirichlet process and Markov random field. Our method has the advantage of effectively encouraging local spatial homogeneity when estimating a globally heterogeneous intensity surface. Posterior inferences are performed with an efficient Markov chain Monte Carlo (MCMC) algorithm. Simulation studies show that the inferences are accurate and the method is superior compared to a wide range of competing methods. Application to the shot location data of 20 representative NBA players in the 2017-2018 regular season offers interesting insights about the shooting patterns of these players. A comparison against the competing method shows that the proposed method can effectively incorporate spatial contiguity into the estimation of intensity surfaces.

## **Title: Calibrated Model Criticism Using Split Predictive Checks**

**Speaker: Jonathan Huggins**

CT

Boston University

Checking how well a fitted model explains the data is one of the most fundamental parts of a Bayesian data analysis. However, existing model checking methods suffer from trade-offs between being well-calibrated, automated, and computationally efficient. To overcome these limitations, we propose split predictive checks (SPCs), which combine the ease-of-use and speed of posterior predictive checks with the good calibration properties of predictive checks that rely on model-specific derivations or inference schemes. We develop an asymptotic theory for two types of SPCs: single SPCs and the divided SPC. Our results demonstrate that they offer complementary strengths, with single SPCs providing superior power in the small-data regime or when the misspecification is significant. We validate the finite-sample utility of SPCs through extensive synthetic experiments in exponential family and hierarchical models, and provide four real-data examples where SPCs offer complimentary insights to what is provided by posterior predictive checks.

## Title: Bayesian inference with models made of modules

Speaker: *Pierre Jacob*

SB

ESSEC Business School

In many cases, models employed in statistics are made of different "modules" that each relates to a specific part of the analysis or a specific source of data. For example, there might be a primary module with the parameters of interest, but also: a module for the imputation of missing data, a module for the estimation of propensity scores, a module that estimates unobserved explanatory variables using external data, a module that handles some pre-processing step, a module representing measurement error, etc. An appeal of the Bayesian approach is the ability to perform inference on all unknown quantities jointly, leading to a coherent treatment of uncertainty across all modules. However, misspecification of any module can motivate departures from the standard joint inference approach and has led practitioners to employ "modularized" strategies, designed to "cut" some unwanted feedback from some modules onto others, leading to "cut distributions" instead of joint posteriors. Modularization is sometimes perceived as unprincipled, and its implementation as either computationally convenient or computationally difficult depending on the context. This lecture will review (1) the arguments proposed to defend and to condemn modularization in Bayesian analysis, (2) describe some statistical properties of the "cut distributions" and (3) review a range of algorithms to implement them.

## Title: General Bayesian Loss Function Selection and the use of Improper Models

Speaker: *Jack Jewson*

IS

Universitat Pompeu Fara

Statisticians often face the choice between using probability models or a paradigm defined by minimising a loss function. Both approaches are useful and, if the loss can be re-cast into a proper probability model, there are many tools to decide which model or loss is more appropriate for the observed data, in the sense of explaining the data's nature. However, when the loss leads to an improper model, there are no principled ways to guide this choice. We address this task by combining the Hyvarinen score, which naturally targets infinitesimal relative probabilities, and general Bayesian updating, which provides a unifying framework for inference on losses and models. Specifically we propose the H-score, a general Bayesian selection criterion and prove that it consistently selects the (possibly improper) model closest to the data-generating truth in Fisher's divergence. We also prove that an associated H-posterior consistently learns optimal hyperparameters featuring in loss functions, including a challenging tempering parameter in generalised Bayesian inference. As salient examples, we consider robust regression and non-parametric density estimation where popular loss functions define improper models for the data and hence cannot be dealt with using standard model selection tools. These examples illustrate advantages in robustness-efficiency trade-offs and enable Bayesian inference for kernel density estimation, opening a new avenue for Bayesian non-parametrics.

## **Title: Variable Selection Consistency of Gaussian Process Regression**

**Speaker: Sheng Jiang**

IS

University of California, Santa Cruz

Bayesian nonparametric regression under a rescaled Gaussian process prior offers smoothness-adaptive function estimation with near minimax-optimal error rates. Hierarchical extensions of this approach, equipped with stochastic variable selection, are known to also adapt to the unknown intrinsic dimension of a sparse true regression function. But it remains unclear if such extensions offer variable selection consistency, i.e., if the true subset of important variables could be consistently learned from the data. In this talk, I will show that variable consistency may indeed be achieved with such models at least when the true regression function has finite smoothness to induce a polynomially larger penalty on inclusion of false positive predictors. Our result covers the high dimensional asymptotic setting where the predictor dimension is allowed to grow with the sample size.

## **Title: Scalable Gaussian Processes on Physically Constrained Domains**

**Speaker: Bora Jin**

IS

Duke University

Many applications in point-referenced geostatistics have measurements that are collected and meaningful only within a constrained domain. Typical spatial Gaussian process (GP) models ignore the unique geometry of the domain, which may lead to inappropriate smoothing over physical barriers. We focus on developing a scalable GP approach that incorporates the constrained domain. One way to construct a scalable GP is via sparsity-inducing directed acyclic graphs (DAGs) that limit neighbors and impose conditional independence. A main contribution of this paper is the development of the Barrier Overlap-Removal Acyclic directed graph GP (BORA-GP) that constructs neighbors conforming to barriers, enabling characterization of dependence in constrained domains. We analyze levels of sea surface salinity in the Arctic Ocean as a motivating application.

## **Title: Gaussian processes using network inputs with application to the microbiome**

**Speaker: Nathaniel Josephs**

IS

Yale University

While the study of a single network is well-established, technological advances now allow for the collection of multiple networks with relative ease. Increasingly, anywhere from several to thousands of networks can be created from brain imaging, gene co-expression data, or microbiome measurements. And these networks, in turn, are being looked to as potentially powerful features to be used in modeling. However, with networks being non-Euclidean in nature, how best to incorporate them into standard modeling tasks is not obvious. In this talk, we propose a Bayesian modeling framework that provides a unified approach to binary classification, anomaly detection, and survival analysis with network inputs. We encode the networks in the kernel of a Gaussian process prior via their pairwise differences and we discuss several choices of provably positive definite kernel that can be plugged into our models. Although our methods are widely applicable, we are motivated here in particular by microbiome research (where network analysis is emerging as the standard approach for capturing the interconnectedness of microbial taxa across both time and space) and its potential for reducing preterm delivery and improving personalization of prenatal care.

## **Title: Detecting and diagnosing prior and likelihood sensitivity with power-scaling**

**Speaker: Noa Kallioinen**

IS

Aalto University

Determining the sensitivity of a posterior to perturbations of the prior and likelihood is an important part of the Bayesian workflow. Here, we discuss a practical and computationally efficient sensitivity analysis approach using importance sampling to estimate properties of posteriors that would result from power-scaling the prior or likelihood. We suggest a diagnostic based on this that can indicate the presence of prior-data conflict or likelihood noninformativity and also discuss limitations to the power-scaling approach. The approach can be readily included in Bayesian workflows with minimal work by the model builder, and we present its implementation in our R package, priorsense. We further demonstrate the workflow on case studies of real data using models ranging from simple linear regression to Gaussian processes.

## **Title: Bayesian Spatially Varying Weight Neural Networks with the Soft-Thresholded Gaussian Process Prior**

**Speaker: Jian Kang**

IS

University of Michigan

Deep neural networks (DNN) have been adopted in the scalar-on-image regression which predicts the outcome variable using image predictors. However, training DNN often requires a large sample size to achieve a good prediction accuracy and the model fitting results can be difficult to interpret. In this work, we construct a novel single-layer Bayesian neural network (BNN) with spatially varying weights for the scalar-on-image regression. Our goal is to select interpretable image regions and to achieve high prediction accuracy with limited training samples. We assign the soft-thresholded Gaussian process (STGP) prior to the spatially varying weights and develop an efficient posterior computation algorithm based on stochastic gradient Langevin dynamics (SGLD). The BNN-STGP provides large prior support for sparse, piecewise-smooth, and continuous spatially varying weight functions, enabling efficient posterior inference on image region selection and automatically determining the network structures. We establish the posterior consistency of model parameters and selection consistency of image regions when the number of voxels/pixels grows much faster than the sample size. We compared our methods with state-of-the-art deep learning methods via analyses of multiple real data sets including the task fMRI data in the Adolescent Brain Cognitive Development (ABCD) study.

## **Title: Bayesian Dynamic Borrowing with Applications to Large-Scale Educational Assessments**

**Speaker: David Kaplan**

CT

University of Wisconsin - Madison

The purpose of this paper is to demonstrate and evaluate the use of Bayesian dynamic borrowing (Viele et al., 2014) as a means of systematically utilizing historical information with specific applications to large-scale educational assessments. Dynamic borrowing via Bayesian hierarchical models is a special case of a general framework of historical borrowing where the degree of borrowing depends on the heterogeneity among historical data and current data. A joint prior distribution over the historical and current data sets is specified with the degree of heterogeneity across the data sets controlled by the variance of the joint distribution. We apply Bayesian dynamic borrowing to both single-level and multilevel models and compare this approach to complete pooling and traditional power priors (Ibrahim & Chen, 2000). Two case studies using data from the PISA reveal the utility of Bayesian dynamic borrowing in terms of predictive accuracy. This is followed by two simulation studies that reveal the utility of Bayesian dynamic borrowing over simple pooling and power priors in cases where the historical data is heterogeneous compared to current data based on bias, mean squared error, total effective sample size, and predictive accuracy. In cases of homogeneous historical data, Bayesian dynamic borrowing performs similarly to data pooling and power priors. In contrast, for heterogeneous historical data, Bayesian dynamic borrowing performed at least as well, if not better, than other methods of borrowing with respect to MSE, percent bias, total effective sample size, and leave-one-out cross-validation.

## **Title: Dynamic factor models with common (drifting) stochastic trends**

**Speaker: Sylvia Kaufman**

IS

Study Center Gerzensee

High-dimensional macroeconomic and financial data are often composed of groups of series that display a common upward or downward drift, or a kind of long-run oscillating behavior. It is now common to capture these drifts by modelling a stochastic trend, i.e. a random walk with or without drift, respectively. When using a factor model to extract common components, common drifts traditionally are removed by mean-adjusting first differences of the level series. However, it may be of interest to extract factors that include the drift of the common stochastic trend component. We propose a dynamic factor model that captures the common drifting trend component as well as a remaining idiosyncratic stochastic trend component. We address identification issues and apply a MCMC sampler based on parameter expansion to obtain posterior inference.

## **Title: A Variational Inference Approach to Inverse Problems with Gamma Hyperpriors**

**Speaker: Hwanwoo Kim**

CT

University of Chicago

Hierarchical models with gamma hyperpriors provide a flexible, sparse-promoting framework to bridge L1 and L2 regularizations in Bayesian formulations to inverse problems. Despite the Bayesian motivation for these models, existing methodologies are limited to maximum a posteriori estimation. The potential to perform uncertainty quantification has not yet been realized. We introduce a variational iterative alternating scheme for hierarchical inverse problems with gamma hyperpriors. The proposed variational inference approach yields accurate reconstruction provides meaningful uncertainty quantification and is easy to implement. In addition, it lends itself naturally to conducting model selection for the choice of hyperparameters. We illustrate the performance of our methodology in several computed examples, including a deconvolution problem and sparse identification of dynamical systems from time series data.

## **Title: Bayesian inference when the model is wrong**

**Speaker: Alice Kirichenko**

IS

University of Warwick

Over the last years it became abundantly clear that Bayesian inference can behave rather poorly under misspecification. In this talk I will discuss theoretical and empirical approaches to using (modifications of) Bayesian methods when the model is wrong. For example, a possible remedy is to employ a generalized Bayesian method instead, i.e. to raise the likelihood in the Bayes equation to some power, which is called a learning rate. It turns out that under some conditions the posterior with a suitably chosen learning rate concentrates around the best approximation of the truth within the model, even when the model is misspecified. In particular, these conditions are satisfied for General linear models (GLMs). We also show that this method significantly outperforms other Bayesian estimation procedures on both, simulated and real data.

## **Title: Marginally calibrated response distributions for end-to-end learning in autonomous driving**

**Speaker: Nadja Klein**

CT

Humboldt-Universität zu Berlin

End-to-end learners for autonomous driving are deep neural networks that predict the instantaneous steering angle directly from images of the ahead-lying street. These learners must provide reliable uncertainty estimates for their predictions in order to meet safety requirements and initiate a switch to manual control in areas of high uncertainty. Yet end-to-end learners typically only deliver point predictions, since distributional predictions are associated with large increases in training time or additional computational resources during prediction. To address this shortcoming we investigate efficient and scalable approximate inference for the implicit copula neural linear model of Klein, Nott and Smith (2021) in order to quantify uncertainty for the predictions of end-to-end learners. The result are densities for the steering angle that are marginally calibrated, i.e. the average of the estimated densities equals the empirical distribution of steering angles. To ensure the scalability to large  $n$  regimes, we develop efficient estimation based on variational inference as a fast alternative to computationally intensive, exact inference via Hamiltonian Monte Carlo. We demonstrate the accuracy and speed of the variational approach in comparison to Hamiltonian Monte Carlo on two end-to-end learners trained for highway driving using the comma2k19 data set. The implicit copula neural linear model delivers accurate calibration, high-quality prediction intervals and allows to identify overconfident learners. Our approach also contributes to the explainability of black-box end-to-end learners, since predictive densities can be used to understand which steering actions the end-to-end learner sees as valid.

## Title: Detecting Stealthy Behavior within Big Data

Speaker: Michael Kouritzin

CT

University of Alberta

A real-time Bayesian approach to big data knowledge extraction and prediction that is based on probabilistic behavior modeling and counting is given. While its goal is to find, track and predict multiple behaviors within the data rather than noise removal, the approach still shares the nice analytic and implementation features of nonlinear filtering theory. In particular, data analysis, learning and prediction can be implemented by sequential Monte Carlo algorithms, especially model-rating particle algorithms, like the one that will be given, and conditional expectation estimates of behaviors are described in terms of a measure-valued stochastic equations. Finally, it is shown that the model rating particle algorithm given here converges to the given measure-valued process, establishing consistency for the algorithm.

## Title: Semiparametric Functional Factor Models with Bayesian Rank Selection

Speaker: Daniel Kowal

IS

Rice University

Functional data are frequently accompanied by parametric templates that describe the typical shapes of the functions. Although the templates incorporate critical domain knowledge, parametric functional data models can incur significant bias, which undermines the usefulness and interpretability of these models. To correct for model misspecification, we augment the parametric templates with an infinite-dimensional nonparametric functional basis. Crucially, the nonparametric factors are regularized with an ordered spike-and-slab prior, which implicitly provides consistent rank selection and satisfies several appealing theoretical properties. This prior is accompanied by a parameter-expansion scheme customized to boost MCMC efficiency, and is broadly applicable for Bayesian factor models. The nonparametric basis functions are learned from the data, yet constrained to be orthogonal to the parametric template in order to preserve distinctness between the parametric and nonparametric terms. The versatility of the proposed approach is illustrated through applications to synthetic data, human motor control data, and dynamic yield curve data. Relative to parametric alternatives, the proposed semiparametric functional factor model eliminates bias, reduces excessive posterior and predictive uncertainty, and provides reliable inference on the effective number of nonparametric terms—all with minimal additional computational costs.

## **Title: Bayesian Inversion with Hierarchical Random Field Priors: Computational Strategies**

*Speaker: Jonas Latz*

IS

Heriot-Watt University

Gaussian random fields are popular models for spatially varying uncertainties, arising, e.g., in geotechnical engineering, hydrology, or image processing. A Gaussian random field is fully characterised by its mean and covariance operator. In more complex models these can also be partially unknown. In this case we need to handle a family of Gaussian random fields indexed with hyperparameters. Sampling for a fixed configuration of hyperparameters is already very expensive due to the nonlocal nature of many classical covariance operators. Sampling from multiple configurations increases the total computational cost severely. In this talk we employ parameterised Karhunen-Loève expansions and adaptive cross approximations for sampling. To reduce the cost we construct a reduced basis surrogate built from snapshots of Karhunen-Loève eigenvectors in the first case. In the second case, we propose a parameterised version of the adaptive cross scheme.

In numerical experiments we consider Matérn-type covariance operators with unknown correlation length and standard deviation. Here, we study the approximation accuracy of reduced basis and cross approximation. As an application we consider Bayesian inversion with an elliptic partial differential equation where the logarithm of the diffusion coefficient is a parameterised Gaussian random field. Indeed, we employ Markov chain Monte Carlo on the reduced space to generate samples from the posterior measure.

## **Title: Wasserstein distance between Lévy measures with applications to Bayesian nonparametrics**

*Speaker: Hugo Lavenant*

IS

Bocconi University

Optimal transport is a powerful and thriving theory which enables to build couplings between random variables and measure distances between probability distributions. I will present an extension of optimal transport to the case of Lévy measures with potentially infinite mass. It will be applied to the study of Completely Random Measures (CRMs) and Completely Random Vectors, a central object to model in a Bayesian way the borrowing of information between similar, yet different sources of observations. In particular we are able to build a tractable index of dependence between CRMs which quantifies precisely the borrowing of information. I will also present some ongoing work about other applications of this new distance.

This is joint work with Marta Catalano, Antonio Lijoi and Igor Prünster.

## **Title: Concentration and robustness of discrepancy-based ABC through Rademacher complexity**

**Speaker: Sirio Legramanti**

IS

University of Bergamo

Approximate Bayesian Computation (ABC) typically employs summary statistics to measure the discrepancy among the observed data and synthetic data generated from each proposed parameter value. However, finding good summary statistics (that are close to sufficiency) is non-trivial for most of the models for which ABC is needed. This motivated summary-free versions of ABC based on discrepancies between the empirical distributions of observed and synthetic data. The studies on the properties of the corresponding ABC posteriors are quite uneven, ranging from empirical assessments to more theoretical investigations. Even when available, existing theory is often limited to a single discrepancy, or relies on hypotheses that are difficult to verify. We propose a unifying view through Rademacher complexity over a general class of discrepancies known as integral probability semimetrics, which include the maximum mean discrepancy and the total variation, Kolmogorov-Smirnov, and Wasserstein distances. For rejection ABC based on this class of semimetrics, we prove results on both posterior concentration and robustness. Such results connect the properties of the ABC posterior to the Rademacher complexity of the class of test functions that characterizes each integral probability semimetric. This provides a new understanding of why some discrepancies work well with ABC and others do not.

## **Title: Compound vectors of subordinators and their associated positive Lévy copulas**

**Speaker: Fabrizio Leisin**

IS

University of Nottingham

Lévy copulas are an important tool which can be used to build dependent Lévy processes. In a classical setting, they have been used to model financial applications. In a Bayesian framework they have been employed to introduce dependent nonparametric priors which allow to model heterogeneous data. This talk focuses on introducing a new class of Lévy copulas based on a class of subordinators recently appeared in the literature, called compound random measures. The well-known Clayton Lévy copula is a special case of this new class. Furthermore, we provide some novel results about the underlying vector of subordinators such as a series representation and relevant moments. This is a work in collaboration with Alan Riva-Palacio.

Reference: Alan Riva-Palacio, Fabrizio Leisen. Compound vectors of subordinators and their associated positive Lévy copulas, *Journal of Multivariate Analysis*, Volume 183, 2021, <https://doi.org/10.1016/j.jmva.2021.104728>.

## **Title: Bayesian Fixed-domain Asymptotics for Covariance Parameters in Spatial Gaussian Process Models**

**Speaker: Cheng Li**

IS

National University of Singapore

Gaussian process models typically contain finite dimensional parameters in the covariance functions that need to be estimated from the data. We present some recent development in Bayesian fixed-domain asymptotics for estimating the covariance parameters in the isotropic Matern covariance function. For the spatial Gaussian process regression model without nugget, we derive the limiting joint posterior distribution for the microergodic parameter and the range parameter. For the model with nugget, we explicitly derive the posterior contraction rates for both the microergodic parameter and the nugget parameter. We verify the new theory in simulation studies and real data applications.

## **Title: Principal Stratification: Uses, Bayesian Inference and Software**

**Speaker: Fan Li**

IS

Duke University

Post-treatment confounding is a common problem in causal inference, with special cases including non-compliance, truncation by death, surrogate endpoint and recruitment bias. Principal stratification is a general framework for post-treatment confounding. Despite its conceptual flexibility, principal stratification is not accessible to the vast majority of applied researchers because it inherently involves latent mixtures and thus requires complex modeling and inference tools that requires highly-customized programming. We fill this gap by developing a computing platform R package "PStrata" to automatize inference of principal stratification for the most common scenarios. PStrata supports both frequentist and Bayesian paradigms and accommodate different assumptions. For the Bayesian paradigm, the computing architecture combines R, C++, Stan, where R provides the user-interface, Stan automatizes posterior sampling, and C++ bridges the two owing to the object-oriented feature. For the frequentist paradigm, PStrata implements the weighting-based triply-robust estimator of Jiang et al (2020). PStrata can accommodate regular outcomes and time-to-event outcomes. We will illustrate PStrata using a high-profile clinical trial on aspirin use among heart disease patients.

## Title: Spatiotemporal Clustering with Neyman-Scott Processes via Connections to Bayesian Nonparametric Mixture Models

Speaker: *Scott Linderman*

IS

Stanford University

Neyman-Scott processes (NSPs) are point process models that generate clusters of points in time or space. They are natural models for a wide range of phenomena, ranging from neural spike trains to document streams. The clustering property is achieved via a doubly stochastic formulation: first, a set of latent events is drawn from a Poisson process; then, each latent event generates a set of observed data points according to another Poisson process. This construction is similar to Bayesian nonparametric mixture models like the Dirichlet process mixture model (DPMM) in that the number of latent events (i.e. clusters) is a random variable, but the point process formulation makes the NSP especially well suited to modeling spatiotemporal data. While many specialized algorithms have been developed for DPMMs, comparatively fewer works have focused on inference in NSPs. Here, we present novel connections between NSPs and DPMMs, with the key link being a third class of Bayesian mixture models called mixture of finite mixture models (MFMMs). Leveraging this connection, we adapt the standard collapsed Gibbs sampling algorithm for DPMMs to enable scalable Bayesian inference on NSP models. We demonstrate the potential of Neyman-Scott processes on a variety of applications including sequence detection in neural spike trains and event detection in document streams.

## Title: Mediation Analysis using Bayesian Tree Ensembles

Speaker: *Antonio Linero*

IS

University of Texas at Austin

We present a general framework for causal mediation analysis using nonparametric Bayesian methods in the potential outcomes framework. Our model, which we refer to as the Bayesian Causal Mediation Forests model, combines recent advances in Bayesian machine learning using decision tree ensembles, Bayesian nonparametric causal inference, and a Bayesian implementation of the  $g$ -formula for computing causal effects. Because of its strong performance on simulated data and because it greatly reduces researcher degrees of freedom, we argue that Bayesian Causal Mediation Forests are highly attractive as a default approach. Of independent interest, we also introduce a new sensitivity analysis technique for mediation analysis with continuous outcomes that is widely applicable. We demonstrate our approach on both simulated and real datasets, and show that our approach obtains low mean squared error and close to nominal coverage of 95% interval estimates, even in highly non-linear problems on which other methods fail.

## **Title: Discreteness and dependence: an effective interplay in Bayesian nonparametrics**

**Speaker: Antonio Lijoi**

FL

Bocconi University

Random discrete structures are increasingly popular as compelling tools for defining Bayesian nonparametric models for grouped data, affected by some source of heterogeneity. This talk will survey some recent proposals and will highlight the range of dependencies and the strength of the borrowing of information across data groups they induce. The main ingredients of the analysis are the random partitions associated to such models and their probability distributions. Special emphasis will be given to multivariate species sampling models, a new general class of random probabilities that encompasses as special cases most well known priors and serves as a unifying thread for illustrating the most relevant distributional properties.

## **Title: Handling Categorical Features with Many Levels Using a Product Partition Model**

**Speaker: Rossangela H. Loschi**

IS

Universidade Federal de Minas Gerais

A common difficulty in data analysis is how to handle categorical predictors with a large number of categories. Few proposals have been developed to tackle this important problem. We introduce a generative model that simultaneously carries out the model fitting and the aggregation of the categorical levels into larger groups. We represent the categorical predictor by a graph where the nodes are the categories and establish a probability distribution over meaningful partitions of this graph. Conditionally on the observed data, we obtain a posterior distribution for the levels aggregation, allowing the inference about the most probable clustering for the categories. Simultaneously, we extract inference about all the other regression model parameters. We compare our and state-of-art methods showing that it has equally good predictive performance and more interpretable results. Our approach balances out accuracy versus interpretability, a current important concern in statistics and machine learning. Financial Support: FAPEMIG, CAPES, CNPQ.

## Title: Identifying the latent space geometry of network formation models through analysis of curvature

Speaker: *Shane Lubold*

IS

Department of Statistics, University of Washington

Modeling networks, across numerous disciplines and contexts, is fundamentally challenging because of (often high-order) dependence between connections. A common approach assigns each person in the graph to a position on a low-dimensional manifold. Distance between individuals in this (latent) space is inversely proportional to the likelihood of forming a connection. The choice of the latent geometry (the manifold class, dimension, and curvature) has consequential impacts on the substantive conclusions of the model. More positive curvature in the manifold, for example, encourages more and tighter communities; negative curvature induces repulsion among nodes. Currently, however, the choice of the latent geometry is an a priori modeling assumption and there is limited guidance about how to make these choices in a data-driven way. In this work, we present a method to consistently estimate the manifold type, dimension, and curvature from an empirically relevant class of latent spaces: simply connected, complete Riemannian manifolds of constant curvature. Our core insight comes by representing the graph as a noisy distance matrix based on the ties between cliques. Leveraging results from statistical geometry, we develop hypothesis tests to determine whether the observed distances could plausibly be embedded isometrically in each of the candidate geometries. Our testing framework covers two empirically relevant cases: 1) when the graph is completely observed and 2) when only partial data about the network is available, such as aggregated relational data. We explore the accuracy of our approach with simulations and then apply our approach to data-sets from economics, sociology, and neuroscience.

## **Title: Balanced tree stick-breaking priors for covariate-dependent mixture models**

*Speaker: Li Ma*

IS

Duke University

Stick-breaking priors are often adopted in Bayesian nonparametric mixture models for generating the mixture weights. When covariates influence the sizes of the clusters, stick-breaking mixtures can leverage Pólya-gamma data augmentation to ease the demands of posterior computation. However, despite this computational convenience, we show that in the presence of covariates, existing stick-breaking processes in fact often induce artificial correlation among the class of random measures they model, and introduce excessive posterior uncertainty into inference on the covariates effects on all but the largest of the clusters, which propagates into posterior uncertainty on the cluster sizes and the number of clusters. We believe that these phenomena arise because existing stick-breaking models are constructed based on continually breaking a single remaining piece of a unit stick. This design results in an underlying one-sided unbalanced tree structure, which in turn leads to an inefficient representation of covariate effects that often confounds with the stochastic ordering of cluster sizes. Instead, we propose to generalize the stick-breaking priors to allow the continual breaking of all remaining pieces of a unit stick at each stage of the stick breaking. This corresponds to a balanced tree along which random weights are generated. When covariates are incorporated, this model decouples the stochastic ordering of the cluster sizes from that of the effect of covariates, thereby allowing covariate effects to be more sparsely represented and circumventing the challenges faced by existing stick breaking models. The computation under this new covariate-dependent balanced-tree stick-breaking mixture model, on the other hand, remains essentially identical to that of existing stick-breaking mixtures. We demonstrate through both simulated and a data set from flow cytometry that the new model induces more robust posterior inference for both the number and size of clusters as well as the covariate effects on the cluster sizes. This talk is based on joint work with Akira Horiguchi.

## **Title: Dependence structures for longitudinal Bayesian nonparametric models**

*Speaker: Steve MacEachern*

IS

The Ohio State University

Nonparametric Bayesian models have proven their worth in many settings. The bulk of work on the models considers a collection of conditionally independent observations, each of which is drawn from a distribution of arbitrary (hence nonparametrics) form. In this talk, with much of the work joint with Zhiguang Xu and Xinyi Xu, we consider models that capture both nonparametric distributions at a given time point and serial dependence within a collection of observations.

## **Title: Some computational strategies for richly parameterized multivariate Gaussian Markov random fields**

*Speaker: Ying MacNab*

IS

University of British Columbia

This presentation concerns with Bayesian hierarchical modeling of multivariate spatial data on lattice using richly parameterized multivariate Gaussian Markov random field (MGMRF) priors that are formulated via full latent conditionals and linear co-regionalization. I present computational methods and strategies, analytically and via case studies of Bayesian disease mapping and small area estimation, for posterior estimation and inference of MGMRFs with a  $p$  by  $p$  full matrix  $C$  of spatial dependence parameters and/or by a  $p$  by  $p$  full matrix of co-regionalization coefficients matrix  $A$ . These models allow for flexible characterizations of spatial and (a)symmetric cross-spatial dependence functions and/or cross-covariance functions. Bayesian estimation and inference of these models are challenged by complex requirements that the estimated MGMRF precision matrices be positive definite. I present and illustrate several solutions of placing hard or soft constraints, as well as no constraint, on the spatial parameterizations for posterior estimation of MGMRFs via Markov chain Monte Carlo simulations.

## **Title: Sparse linear mixed model selection via streamlined variational Bayes**

*Speaker: Luca Maestrini*

IS

The Australian National University

Variational approximations facilitate fast approximate Bayesian inference for the parameters of a variety of statistical models, including linear mixed models. However, for models with a high number of fixed or random effects, simple application of standard variational inference principles does not lead to fast approximate inference algorithms, due to the size of model design matrices and inefficient treatment of sparse matrix problems arising from the required approximating density parameters updates. We illustrate how recently developed streamlined variational inference procedures can be generalized to make fast and accurate inference for the parameters of linear mixed models with nested random effects and global-local priors for Bayesian fixed effects selection. Our variational inference algorithms achieve convergence to the same optima of their standard implementations, although with significantly lower computational effort, memory usage and time, especially for large numbers of random effects. Joint work with Emanuele Degani, Dorota Toczyłowska and Matt Wand.

## **Title: A graphical Gaussian process model for multi-fidelity emulation of expensive computer codes**

**Speaker: Simon Mak**

CT

Duke University

With advances in scientific computing, complex phenomena can now be reliably simulated via computer code. Due to the high computational cost of accurate (or high-fidelity) simulations, multi-fidelity emulation is often applied to leverage less accurate (or lower-fidelity) simulation data for better predictive performance of high-fidelity simulations. In many complex applications (e.g., heavy-ion physics), simulation data with different fidelities are connected via a directed acyclic graph (DAG), which cannot be integrated within existing multi-fidelity emulator models. We thus propose a new Graphical Multi-fidelity Gaussian process (GMGP) model, which embeds this underlying DAG structure within a Gaussian process predictive framework. We show that the GMGP has desirable modeling traits via two Markov properties, and admits a scalable recursive formulation for computing the posterior predictive distribution along sub-graphs. We also present a design framework for allocating experimental runs over the DAG given a computational budget and propose a nonlinear extension to GMGP. The advantages of the GMGP model over existing methods are then demonstrated via a suite of numerical experiments and an application to emulation of heavy-ion collisions, which can be used to study the conditions of matter in the Universe shortly after the Big Bang.

## **Title: Stochastically developed Langevin dynamics applied to generative adversarial networks**

**Speaker: Mariya Mamajiwala**

CT

University College London

Generative adversarial networks (GANs) are deep neural networks trained to transform standard Gaussian noise to the underlying multimodal distribution of a target dataset, for instance images or speeches. GANs comprise of two competing networks and hence their training is generally unstable, owing to the non-convexity of the losses, and the conflicting goals of the two networks. As a result, training GANs with stochastic gradient descent algorithms has been challenging, with models failing to fully capture all the modes of the underlying distributions. The update scheme we propose in the present work is based on evolving the parameters of the network via Langevin dynamics on a Riemannian manifold with a view to exploring the various modes efficiently. We show through several test examples, the benefits of the current approach versus the limitations of stochastic gradient as well as natural gradient based search algorithms.

## **Title: Statistical Approaches for Entity Resolution under Uncertainty**

**Speaker: Neil Marchant**

IS

University of Melbourne

When real-world entities are referenced in data, their identities are often obscured. This presents an obstacle for data cleaning and integration, as references to an entity may be scattered across multiple records or sources, without a means to identify and consolidate them. Entity resolution (ER) seeks to address this problem by linking references to the same entity based on imprecise information. ER has diverse applications: from construction of knowledge bases in the life sciences, to data sharing between government agencies, and integration of data silos in the enterprise. In this talk, I explore statistical approaches for managing uncertainty in ER applications. In the first part of the talk, I present work on improving the scalability and flexibility of Bayesian models for ER, which naturally allow for uncertainty quantification of ER predictions. This work was done in collaboration with the Australian Bureau of Statistics and US Census Bureau. In the second part of the talk, I present work on evaluating ER accuracy via an adaptive importance sampling framework. This allows practitioners to obtain more precise accuracy estimates, with a reduced ground truth label budget.

## **Title: Dynamic Bayesian predictive synthesis with large number of forecasts**

**Speaker: Kenchiro McAlinn**

IS

Temple University

We consider using a large set of models for predictive synthesis to deal with model misspecification and uncertainty. When the number of models to synthesize is large, many of the models produce redundant forecasts, providing little additional information. This produces a unique problem that is different from typical issues in model uncertainty (e.g. sparse modeling). To address this, we propose a novel method within the BPS framework that automatically clusters the forecasts into subgroups, utilizing reduced rank formulation, within dynamic synthesis. Our methodological development is motivated by real applications in macroeconomics.

## **Title: Bayesian remedies to identifiability issues arising in the estimation of infection fatality rates with incomplete data**

**Speaker: Gonzalo Mena**

CT

University of Oxford

Suppose we want to estimate stratified infection fatality rate (IFR) of a new disease. We propose a novel hierarchical bayesian methodology for the estimation of such rates under the assumption that stratified death information is reliable, but case (and infection) information is incomplete. Our estimates are based on a cascade of binomial models linking infection, cases, and deaths. To infer infections from cases we estimate a mapping from a reporting proxy (such as testing rates) to reporting rates. We discuss the identifiability issues that appear as a consequence of incomplete data and over-parameterization. We also discuss how model ensembles can lead to more robust estimates. We apply this model to understand COVID-10 related mortality in Santiago, Chile. From our model we find a strong socioeconomic gradient of IFRs in young age groups.

## **Title: Prior sample size extensions for assessing prior impact and prior-likelihood discordance**

**Speaker: Xialoi Meng**

IS

Harvard University

This paper (Reimherr, Meng, Nicolae, 2021, JRSSB, 413-437) outlines a framework for quantifying the prior's contribution to posterior inference in the presence of prior-likelihood discordance, a broader concept than the usual notion of prior-likelihood conflict. We achieve this dual purpose by extending the classic notion of prior sample size,  $M$ , in three directions: (I) estimating  $M$  beyond conjugate families; (II) formulating  $M$  as a relative notion, i.e., as a function of the likelihood sample size  $k$ ,  $M(k)$ , which also leads naturally to a graphical diagnosis; and (III) permitting negative  $M$ , as a measure of prior-likelihood conflict, i.e., harmful discordance. Our asymptotic regime permits the prior sample size to grow with the likelihood data size, hence making asymptotic arguments meaningful for investigating the impact of the prior relative to that of likelihood. It leads to a simple asymptotic formula for quantifying the impact of a proper prior that only involves computing a centrality and a spread measure of the prior and the posterior. We use simulated and real data to illustrate the potential of the proposed framework, including quantifying how weak is a "weakly informative" prior adopted in a study of lupus nephritis. Whereas we take a pragmatic perspective in assessing the impact of a prior on a given inference problem under a specific evaluative metric, we also touch upon conceptual and theoretical issues such as using improper priors and permitting priors with asymptotically non-vanishing influence.

## **Title: A Comparative Approach to Teaching Undergraduates**

**Speaker: Sierra Merkes**

IS

Virginia Tech

Since the inception of statistics, the frequentist/Bayesian divide has overwhelmed the forefront of most statistical conversations. While all geniuses in their time, Pearson, Fisher, and Lindley began a cultural divide steeped in mathematics. As a result, statisticians designed undergraduate statistics curriculums to develop courses either in an entirely Bayesian or classical perspective, thus, only propagating more historical debates. We take the position that both approaches have their merits, depending on the context of the problem, so a comparative approach would better serve students. With the increase in computational resources, we can now develop a comparative teaching approach through well-designed simulation studies that alleviate the need for students to have a strict mathematical background. Providing students with the computational "know-how" to compare various perspectives will develop their student's critical thinking skills, help them gain job market skills, and connect students to the material. In this talk, we discuss how to construct undergraduate classes that begin with Monte Carlo-based simulation studies to motivate mathematical theory, compare various assumptions of probability distributions/inferential branches of statistics, and apply their knowledge to real-world data analyses.

## **Title: Bayesian Optimal Experimental Design for Inferring Causal Structure**

**Speaker: Jeffrey Miller**

IS

Harvard TH Chan School of Public Health

Inferring the causal structure of a system typically requires interventional data, rather than just observational data. Since interventional experiments can be costly, it is preferable to select interventions that yield the maximum amount of information about a system. We propose a novel Bayesian method for optimal experimental design by sequentially selecting interventions that minimize the expected posterior entropy as rapidly as possible. A key feature is that the method can be implemented by computing simple summaries of the current posterior, avoiding the computationally burdensome task of repeatedly performing posterior inference on hypothetical future datasets drawn from the posterior predictive. After deriving the method in a general setting, we apply it to the problem of inferring causal networks. We present a series of simulation studies in which we find that the proposed method performs favorably compared to existing alternative methods. Finally, we apply the method to real and simulated data from a protein-signaling network.

## **Title: The Copula Opinion Pool: Modeling Expert Dependence**

**Speaker: James Mitchell**

IS

Federal Reserve Bank of Cleveland

This paper develops methods for combining density forecasts that accommodate stochastic dependence between different experts' predictions. Previous work combining density forecasts, using so-called "opinion pools", has tended to ignore dependence. The proposed basis for modeling the dependence among different experts' densities is a recalibration function, based on the probability integral transforms of the expert densities. We show that this reduces to a copula function in a special case and discuss the relationship with Bayesian predictive synthesis. We explore the properties of various approximations to the re-calibration function both via Monte Carlo simulations and in an application density forecasting U.K. inflation using the Bank of England's "fan" chart. We find that the copula opinion pool can deliver more accurate densities than traditional methods.

## **Title: A Two-Stage Adaptive Metropolis Algorithm for Bayesian Model Calibration**

**Speaker: Anirban Mondal**

CT

Case Western Reserve University

We propose a new sampling algorithm combining two quite powerful ideas in the Markov chain Monte Carlo literature – adaptive Metropolis sampler and two-stage Metropolis-Hastings sampler. The proposed sampling method is particularly very useful for high-dimensional posterior sampling in Bayesian model calibration which involves a computationally expensive forward model based likelihood. In the first stage of the proposed algorithm, an adaptive proposal is used based on the previously sampled states, and the corresponding acceptance probability is computed based on an approximated inexpensive posterior. The approximated posterior can include a surrogate model or a statistical emulator that runs much faster than the true complex forward model. The expensive target posterior involving the true forward model is evaluated while computing the second stage acceptance probability only if the proposal is accepted in the inexpensive first stage. The adaptive nature of the algorithm guarantees faster convergence of the chain and very good mixing properties. On the other hand, the two-stage approach helps in rejecting the bad proposals in the inexpensive first stage, making the algorithm computationally efficient. As the proposals are dependent on the previous states the chain loses its Markov property, but we prove that it retains the desired ergodicity property. The performance of the proposed algorithm is compared with the existing algorithms in a simulated and real data example in satellite remote sensing.

## **Title: $O(N)$ Approximate Bayesian Gaussian Process Regression**

**Speaker: Kelly Moran**

IS

Los Alamos National Laboratory

Gaussian processes (GPs) are common components in Bayesian non-parametric models having a rich methodological literature and strong theoretical grounding. The use of exact GPs in Bayesian models is limited to problems containing several thousand observations due to their prohibitive computational demands. We leverage our novel posterior sampling algorithm along with kernel approximation methods to develop  $O(n)$  fully Bayesian Gaussian process regression to problems having an input space of dimension up to 3. We show that this approximation's Kullback-Leibler divergence to the true posterior can be made arbitrarily small. This work represents an improvement on our previous scaling in inputs, in which  $d$ -dimensional surfaces were modeled as tensor products of univariate GPs. We illustrate the performance of this fast increased fidelity approximate GP using both simulated and real data sets.

## **Title: Bayesian Population Finding in a Randomized Clinical Trial**

**Speaker: Satoshi Morita**

IS

Kyoto University

We consider finding a sensitive subpopulation for a new treatment in a randomized clinical trial (RCT), and approach the problem as a Bayesian decision problem together with a flexible probability model, Bayesian additive regression trees (BART), to summarize observed data. We define a utility function that addresses the competing aims of the desired report so that the decision is constrained to be parsimonious and interpretable. We illustrate the approach with data from RCTs for advanced non-small-cell lung cancer and metastatic breast cancer.

## **Title: Single arm trials with a synthetic control arm built from RWD**

**Speaker: Peter Mueller**

IS

UT Austin

Randomized clinical trials (RCT) are the gold standard for approvals by regulatory agencies. However, RCT's are increasingly time consuming, expensive, and laborious with a multitude of bottlenecks involving volunteer recruitment, patient truancy, and adverse events. An alternative that fast tracks clinical trials without compromising quality of scientific results is desirable to more rapidly bring therapies to consumers. We propose a model-based approach using nonparametric Bayesian common atoms models for patient baseline covariates. This specific class of models has two critical advantages in this context: (i) The models have full prior support, i.e., allow to approximate arbitrary distributions without unreasonable restrictions or shrinkage in specific parametric families; (ii) inference naturally facilitates a reweighting scheme to achieve equivalent populations. We prove equivalence of the synthetic and other patient cohorts using an independent separate verification. Failure to classify a merged data set using a flexible statistical learning method such as random forests, support vector machines etc. proves equivalence. We implement the proposed approach in two motivating case studies.

## **Title: Preferential data subsampling in stochastic gradient MCMC**

**Speaker: Christopher Nemeth**

IS

Lancaster University

Stochastic gradient MCMC (SGMCMC) offers a scalable alternative to traditional Markov chain Monte Carlo by constructing an unbiased estimate of the gradient of the log-likelihood with a small, uniformly-weighted subsample of the data. While efficient to compute, the resulting gradient estimator may exhibit a high variance and impact the sampler's performance.

The problem of variance control has been traditionally addressed within the SGMCMC literature by constructing a better stochastic gradient estimator (e.g. using control variates). In this talk, we'll explore an alternative approach to reduce the gradient variance by preferentially subsampling data-points and re-weighting the stochastic gradient. Furthermore, as well as subsampling the data non-uniformly, we'll see how we can adaptively adjust the subsample size at each iteration of the SGMCMC algorithm to reduce the computational cost without increasing the Monte Carlo in the posterior expectations.

## Title: Ordinal Causal Discovery

*Speaker: Yang Ni*

IS

Texas A&M University

Causal discovery for purely observational, categorical data is a long-standing challenging problem. Unlike continuous data, the vast majority of existing methods for categorical data focus on inferring the Markov equivalence class only, which leaves the direction of some causal relationships undetermined. This paper proposes an identifiable ordinal causal discovery method that exploits the ordinal information contained in many real-world applications to uniquely identify the causal structure. The proposed method is applicable beyond ordinal data via data discretization. Through real-world and synthetic experiments, we demonstrate that the proposed ordinal causal discovery method combined with simple score-and-search algorithms has favorable and robust performance compared to state-of-the-art alternative methods in both ordinal categorical and non-categorical data. An accompanied R package OCD is freely available at [https://web.stat.tamu.edu/yni/files/OCD\\_0.1.0.tar.gz](https://web.stat.tamu.edu/yni/files/OCD_0.1.0.tar.gz).

## Title: Bayesian non-linear inverse problems in high-dimensions

*Speaker: Richard Nickl*

FL

University of Cambridge

In the last decade, Bayesian inference methods have propagated into various areas of applied mathematics and the physical sciences, in part because of an influential paper of Andrew Stuart (2010). Specifically, Bayes is now routinely used to solve statistical tasks in parameter identification with partial differential equations, which can be viewed as complex, non-linear and high-dimensional regression type. We explain why posterior inference has been such a success story in these areas of applied mathematics, and provide some theory about why it works in such settings. We also discuss how one can prove that Bayesian computation is *\*not\** NP hard in such problems, but actually feasible even in high dimensions and in absence of log-concavity of posterior measures, specifically when using gradient based methods. We will also discuss some of the mysteries of Bernstein-von Mises type phenomena in PDE settings.

## **Title: A Bayesian semiparametric Archimedean copula**

**Speaker: Luis Nieto-Barajas**

IS

ITAM, Mexico

An Archimedean copula is characterised by its generator. This is a real function whose inverse behaves as a survival function. We propose a semiparametric generator based on a quadratic spline. This is achieved by modelling the first derivative of a hazard rate function, in a survival analysis context, as a piecewise constant function. Convexity of our semiparametric generator is obtained by imposing some simple constraints. The induced semiparametric Archimedean copula produces Kendall's tau association measure that covers the whole range  $(-1, 1)$ . Inference on the model is done under a Bayesian approach and for some prior specifications we are able to perform an independence test. Properties of the model are illustrated with a simulation study as well as with a real dataset.

## **Title: Bayesian nonparametric modelling of covariance functions on the torus with an application to the analysis of wind speed data**

**Speaker: Bernardo Nipoti**

CT

University of Milano Bicocca

Motivated by the analysis of wind speed data, we propose a Bayesian nonparametric model for covariance functions on the torus. Wind speed data display directional and seasonal autocorrelation patterns: in order to account for both we consider a Gaussian process defined on the torus and aim at proposing a flexible model for its covariance function. To this end we define a prior for the covariance function by exploiting a convenient spectral representation of isotropic covariance functions on the torus. We study its properties and assess the performance of the model with an extensive simulation study. Finally, we analyse a real dataset consisting of hourly measurements of wind speed with accompanying wind mean direction.

## **Title: Hamiltonian zigzag sampler got more momentum than its Markovian counterpart: Equivalence of two zigzags under a momentum refreshment limit**

*Speaker: Akihiko Nishimura*

IS

Johns Hopkins University

Zigzag and other piecewise deterministic Markov process samplers have attracted significant interest for their non-reversibility and other appealing properties for Bayesian posterior computation. Hamiltonian Monte Carlo is another state-of-the-art sampler, exploiting fictitious momentum to guide Markov chains through complex target distributions. In this article, we uncover a remarkable connection between the zigzag sampler and a variant of Hamiltonian Monte Carlo based on Laplace-distributed momentum. The position and velocity component of the corresponding Hamiltonian dynamics travels along a zigzag path paralleling the Markovian zigzag process; however, the dynamics is non-Markovian in the position-velocity space as the momentum component encodes non-immediate pasts. This information is partially lost during a momentum refreshment step, in which we preserve its direction but re-sample magnitude. In the limit of increasingly frequent momentum refreshments, we prove that Hamiltonian zigzag converges strongly to its Markovian counterpart. This theoretical insight suggests that, when retaining full momentum information, Hamiltonian zigzag can better explore target distributions with highly correlated parameters by suppressing the diffusive behavior of Markovian zigzag. We corroborate this intuition by comparing performance of the two zigzag cousins on high-dimensional truncated multivariate Gaussians, including a 11,235-dimensional target arising from a Bayesian phylogenetic multivariate probit model applied to HIV virus data.

## **Title: Modelling the death rate among hospitalisations during the first wave of the Coronavirus pandemic: A causal mediation approach**

*Speaker: Widemberg Nobre*

CT

INSPER - Brazil

We propose a causal mediation analysis to handle time-varying observations. The motivating example involves time-varying observations that are affected by an intervention at a given point in time. The goal is to quantify the causal impact of an intervention when multiple observations are available before and after the intervention. The proposal suggests modelling mediator and outcome variables separately while accounting for the temporal evolution of the data. We illustrate the proposed method when modelling hospitalisation data due to respiratory diseases, during the first wave of the coronavirus pandemic, in the state of Rio de Janeiro, Brazil. The outcome of interest is the death rate, and the time point intervention is March 11th of 2020, the day at which the World Health Organisation (WHO) declared the Coronavirus a pandemic. While accounting for control variables affected by the intervention, our proposal allows us to describe the composition of a time-varying global (total) causal impact in terms of a mediator variable.

## **Title: Semiparametric Bayesian Estimation of Dynamic Discrete Choice Models**

**Speaker: Andriy Norets**

CT

Brown University

We propose a tractable semiparametric estimation method for dynamic discrete choice models. The distribution of additive utility shocks is modeled by location-scale mixtures of extreme value distributions with varying numbers of mixture components. Our approach exploits the analytical tractability of extreme value distributions and the flexibility of the location-scale mixtures. We implement the Bayesian approach to inference using Hamiltonian Monte Carlo and an approximately optimal reversible jump algorithm from Norets (2021). For binary dynamic choice model, our approach delivers estimation results that are consistent with the previous literature. We also apply the proposed method to multinomial choice models, for which previous literature does not provide tractable estimation methods in general settings without distributional assumptions on the utility shocks. We develop theoretical results on approximations by location-scale mixtures in an appropriate distance and posterior concentration of the set identified utility parameters and the distribution of shocks in the model.

## **Title: Bayesian Semiparametric Model for Sequential Decision Making in Continuous Time**

**Speaker: Arman Oganisian**

IS

Brown University

We present an analysis of an ongoing study of acute myeloid leukemia (AML) at the Children's Hospital of Philadelphia. Upon diagnosis, patients move through a sequence of four treatment courses where they are treated with either an anthracycline-based chemotherapy (ACT) agent or a non-anthracycline-based agent. While ACT is thought to more aggressively suppress AML, it is also cardiotoxic. Thus, treating overzealously with either may reduce survival. Since ACT is not randomized, its effect on survival is confounded over time. Our task is to estimate potential survival probabilities under hypothetical dynamic treatment strategies while adjusting for time-varying confounding. A key impediment is that the timing of each patient's treatment sequence varies depending on their unique disease progression, making the waiting times between courses a time-varying confounder of subsequent treatment and survival. We develop a generative Bayesian semiparametric model based on Gamma Process priors that capture subjects' transition to subsequent treatment, death, or censoring in continuous time. A g-computation procedure is used to compute a posterior over a causal survival probability under hypothetical strategies. We use the method to estimate the efficacy of various ejection fraction-based treatment rules that dynamically assign ACT at each course based on cardiotoxicity.

## **Title: Coupling and Parallelization in Statistical Inference**

**Speaker: John O'Leary**

IS

Tudor/Xantium, Portfolio Manager

This talk considers the design of Markov chain Monte Carlo (MCMC) estimators using couplings. We begin by describing a coupling-based construction that yields unbiased estimators based on MCMC algorithms. This yields a parallelizable framework in which consistency occurs in the number of replications rather than the number of time steps. The efficiency of the resulting estimators depends on the quality of the underlying couplings, so for the second part of the talk we turn to the question of coupling design. We describe a series of principles based on theory and numerical work. These results clarify what is possible for Markovian couplings and suggest a path toward efficient unbiased MCMC estimators.

## **Title: Latent Gauss-Markov models for spatial and spatiotemporal conditional extremes**

**Speaker: Thomas Opitz**

IS

INRAE

Structured additive regression for Gaussian or non-Gaussian response variables, with latent Gaussian processes included in the predictor, defines a flexible and widely-used class of statistical models, especially in spatial statistics. A Bayesian interpretation of these models is that we put a multivariate Gaussian prior on the vector of regression coefficients. By leveraging numerically convenient Gauss-Markov representations of the Matérn covariance function, an idea that is known as the Stochastic Partial Differential Equation (SPDE) approach, we can model complex and large spatial and spatiotemporal datasets. Recently, such modeling techniques have been adapted to extreme-value analysis. This talk will focus on the conditional extremes framework, which allows for event-based stochastic modeling of dependent extremes. In this framework, certain non-stationary Gaussian processes can be used as asymptotically-motivated models for the observed process conditioned on threshold exceedances at a fixed reference location and time. We explain how constraints on the spatial and spatio-temporal Gaussian processes, arising from the conditioning mechanism, can be implemented through the latent variable approach, such that full-likelihood-based inference becomes feasible even with thousands of observed locations. We implement a Bayesian inference approach using the integrated nested Laplace approximation (INLA), and we illustrate its flexibility with gridded Red Sea surface temperature data at over 6,000 observed locations. Posterior sampling is exploited to study the probability distribution of cluster functionals of spatial and spatio-temporal extreme episodes. This work is joint with Emma Simpson (University College London) and Jennifer Wadsworth (Lancaster University).

## **Title: Density Regression with Bayesian Additive Regression Trees**

**Speaker: Vittorio Orlandi**

IS

Duke University

Flexibly modeling how an entire density changes with covariates is an important but challenging generalization of mean and quantile regression. While existing methods for density regression primarily consist of covariate-dependent discrete mixture models, we consider a continuous latent variable model in general covariate spaces, which we call DR-BART. The prior mapping the latent variable to the observed data is constructed via a novel application of Bayesian Additive Regression Trees (BART). We prove that the posterior induced by our model concentrates quickly around true generative functions that are sufficiently smooth. We also analyze the performance of DR-BART on a set of challenging simulated examples, where it outperforms various other methods for Bayesian density regression. Lastly, we apply DR-BART to a U.S. census dataset to study returns to education. Our proposed sampler is efficient and allows one to take advantage of BART's flexibility in many applied settings where the entire distribution of the response is of primary interest. Furthermore, our scheme for splitting on latent variables within BART facilitates its future application to other classes of models that can be described via latent variables, such as those involving hierarchical or time series data.

## **Title: A hierarchical Hidden Markov Model for cancer detection**

**Speaker: Sally Paganin**

CT

Harvard University

There is growing interest in developing tools for cancer screening and monitoring based on the analysis of DNA sequencing data derived from non-invasive procedures such as blood samples. At early cancer stages, such samples contain DNA from a majority of normal cells and a low fraction of tumor cells. Cancer presence can be assessed measuring allelic imbalance: since a person inherits one allele from each parents, the allele proportion at heterozygous loci is close to 0.5 in normal cells, whereas significant deviations from 0.5 are indicative of the presence of cancer. To efficiently and sensitively detect such deviations, we model the allele proportions over the genome via a novel Bayesian hierarchical Hidden Markov Model. We leverage prior knowledge from population genome databases while borrowing information across multiple samples from the same subject. Hypothesis testing for cancer presence is embedded in the model via a spike and slab prior. We show the performance of our model at different levels of tumor fraction using in-silico mixed data.

## **Title: Clustering and Prediction with Variable Dimension Covariates**

**Speaker: Garritt Page**

IS

Brigham Young University

In many applied fields incomplete covariate vectors are commonly encountered. It is well known that this can be problematic when making inference on model parameters, but its impact on prediction performance is less understood. We develop a method based on covariate dependent partition models that seamlessly handles missing covariates while completely avoiding any type of imputation. The method we develop allows in-sample as well as out-of-sample predictions, even if the missing pattern in the new subjects' incomplete covariate vector was not seen in the training data. Any data type, including categorical or continuous covariates are permitted. In simulation studies the proposed method compares favorably. We illustrate the method in two application examples.

## **Title: Covariate-informed latent interaction models: Addressing geographic taxonomic bias in predicting bird-plant interactions**

**Speaker: Georgia Papadogeorgou**

IS

University of Florida

Climate change and reductions in natural habitats necessitate that we better understand species' interactivity and how biological communities respond to environmental changes. However, ecological studies of species' interactions are limited by their geographic and taxonomic focus which can lead to severe under-representation of certain species and distort our understanding of inter-species interactions. We illustrate that ignoring the studies' focus can result in poor performance. We develop a model for predicting species' interactions that (a) accounts for errors in the recorded interaction networks, (b) addresses the geographic and taxonomic biases of existing studies, (c) is based on latent factors to increase flexibility and borrow information across species, (d) incorporates covariates in a flexible manner to inform the latent factors, and (e) uses a meta-analysis data set from 166 individual studies. We focus on interactions among 242 birds and 511 plants in the Brazilian Atlantic Forest, and identify 5% of pairs of species with an unrecorded interaction, but posterior probability that the interaction is possible over 80%. Finally, we develop a permutation-based variable importance procedure for latent factor network models and identify that a bird's body mass and a plant's fruit diameter are most important in driving the presence and detection of species interactions, with a multiplicative relationship.

## **Title: Reliable variance matrix priors for Bayesian mixture models with Gaussian kernels for problems of moderately high-dimensionality**

**Speaker: Michael Papathomas**

CT

University of St Andrews

Bayesian mixture modelling is an increasingly popular approach for clustering and density estimation. We study the choice of prior for the variance or precision matrix when Gaussian kernels are adopted. Typically, mixture models are assessed by considering observations in a space of only a handful of dimensions. Instead, we are concerned with higher dimensionality problems, in a space of up to 20 dimensions, observing that the choice of prior becomes increasingly important as the dimensionality increases. After identifying certain undesirable properties of standard priors, we review and implement possible alternative priors. The most promising priors are identified, as well as factors that affect MCMC convergence. Results, using simulated and real data, show that the choice of prior and its implementation are critical for deriving reliable inferences. Although our work focusses on the Dirichlet Process Mixture Model, we also discuss its relevance to Bayesian Mixtures of Finite Mixture models.

## **Title: Network structure learning under uncertain interventions**

**Speaker: Stefano Peluso**

IS

Università degli Studi di Milano-Bicocca

Gaussian Directed Acyclic Graphs (DAGs) represent a powerful tool for learning the network of dependencies among variables, a task which is of primary interest in many fields and specifically in biology. Different DAGs may encode equivalent conditional independence structures, implying limited ability, with observational data, to identify causal relations. In many contexts however, measurements are collected under heterogeneous settings where variables are subject to exogenous interventions. Interventional data can improve the structure learning process whenever the targets of an intervention are known. However, these are often uncertain or completely unknown, as in the context of drug target discovery. We propose a Bayesian method for learning dependence structures and intervention targets from data subject to interventions on unknown variables of the system. Selected features of our approach include a DAG-Wishart prior on the DAG parameters, and the use of variable selection priors to express uncertainty on the targets. We provide theoretical results on the correct asymptotic identification of intervention targets and derive sufficient conditions for Bayes factor and posterior ratio consistency of the graph structure. Our method is applied in simulations and real-data world settings, to analyze perturbed protein data and assess antiepileptic drug therapies.

## Title: Bayesian Models for Scalable Machine Learning

Speaker: Valerio Perrone

IS

Amazon Web Services

In recent years we have witnessed an unprecedented increase in the volume and complexity of data, calling for scalable models that can leverage this wealth of information. For instance, high throughput methods such as DNA sequencing are becoming the norm but extracting meaningful representations from this data remains an open problem. In this talk, we introduce new Bayesian models to address some of the key challenges posed by large scale data, showing their benefits in a variety of real-world applications.

In the first part of the talk, we introduce a novel Bayesian nonparametric prior for time-evolving data. We apply the new model to a natural language processing scenario, giving insights into the evolution of research trends across thirty years of NeurIPS conference papers. We then show how Bayesian uncertainty estimates can be combined with the representations learned by deep neural networks to scale up to millions of datapoints and drive decision-making in industrial and population genetics applications. We demonstrate in experiments that the resulting approach accurately detects recombination hotspots from raw population genomic data, which can illuminate the biological mechanism that underlies recombination and help geneticists map the mutations causing genetic diseases.

## Title: Bayesian Fusion

Speaker: Murray Pollock

CT

Newcastle University

Combining several (sample approximations of) distributions, which we term sub-posteriors, into a single distribution proportional to their product, is a common challenge. Many existing approaches resort to approximating the individual sub-posteriors for practical necessity, then representing the resulting approximate posterior. The quality of the approximation in these approaches is poor when the sub-posteriors fall out-with a narrow range of distributional form. Recently, a Fusion approach has been proposed which finds a direct and exact Monte Carlo approximation of the posterior (as opposed to the sub-posteriors), circumventing the drawbacks of approximation. Unfortunately, existing Fusion approaches have a number of computational limitations, particularly with large number of sub-posteriors. In this work we introduce a practical Bayesian Fusion approach: extending the theory underpinning Fusion and, by embedding it within a sequential Monte Carlo algorithm, still recovering the correct target distribution, all while being computationally competitive with approximate schemes.

## Title: Accelerating Inference for Neural Stochastic Differential Equations with Checkpoints

Speaker: Dennis Prangle

CT

University of Bristol

Stochastic differential equation (SDE) models are popular in areas such as biology, finance, physics and medicine. Neural SDEs let the gradient term be the output of a neural network, allowing very flexible behaviour. They can be used as models in their own right, or to approximate a posterior distribution through variational inference.

However, current neural SDE training methods require sampling SDE paths for the entire time range of interest in each training iteration. This can become impractically costly for larger datasets. We propose a method where each training iteration only requires sampling a path over a randomly chosen short interval (analogous to mini-batch methods for IID data). Our method involves learning an approximation to the marginal SDE state at "checkpoints" - every possible interval start-point - using an amortised approach. Empirical studies show that the method achieves orders of magnitude speed-ups at only small reductions in accuracy.

## Title: Non-Separable Spatio-temporal Models via Transformed Multivariate Gaussian Markov Random Fields

Speaker: Marcos Prates

IS

Universidade Federal de Minas Gerais

Models that capture spatial and temporal dynamics are applicable in many scientific fields. Non-separable spatio-temporal models were introduced in the literature to capture these dynamics. However, these models are generally complicated in construction and interpretation. We introduce a class of non-separable Transformed multivariate Gaussian Markov random fields (TMGMRF) in which the dependence structure is flexible and facilitates simple interpretations concerning spatial, temporal and spatio-temporal parameters. Moreover, TMGMRF models have the advantage of allowing specialists to define any desired marginal distribution in model construction without suffering from spatio-temporal confounding. Consequently, the use of spatio-temporal models under the TMGMRF framework leads to a new class of general models, such as spatio-temporal Gamma random fields, that can be directly used to model Poisson intensity for space-time data. The proposed model was applied to identify important environmental characteristics that affect variation in the abundance of *Nenia tridens*, a dominant species of gastropod in a well-studied tropical ecosystem, and to characterize its spatial and temporal trends, which are particularly critical during the Anthropocene, an epoch of time characterized by human-induced environmental change associated with climate and land use.

## **Title: Efficient multifidelity likelihood-free Bayesian inference with adaptive resource allocation**

*Speaker: Thomas Prescott*

CT

The Alan Turing Institute

Likelihood-free Bayesian inference algorithms are popular approaches to calibrating the parameters of complex, stochastic models when the likelihood is intractable. These algorithms typically rely heavily on repeated model simulations. However, whenever the computational cost of model simulations is even moderately expensive, the significant burden incurred by these algorithms leaves them unviable in many practical applications. We present the multifidelity technique, which reduces the simulation burden of likelihood-free inference with no loss of accuracy through exploiting the information provided by simulating computationally cheap, approximate models in place of the model of interest. Analytical results on the optimal allocation of computational resources to simulations at different levels of fidelity can be practically implemented by learning the relationships between models at different fidelities and adapting resource allocation accordingly using gradient descent. We demonstrate that this adaptive multifidelity likelihood-free inference algorithm produces a Monte Carlo sample with near-optimal efficiency.

## **Title: Empirical Copulas and Bayesian Modeling**

*Speaker: Serge Provost*

CT

Western University

The use of copulas is becoming more prevalent in connection with the Bayesian modeling of data sets. Novel approaches for estimating the density functions associated with empirical copulas are proposed. One of them involves smoothing Deheuvel's joint empirical cumulative distribution function and another one provides a semi-parametric representation of the copula density estimates. A representation expressed in terms of Bernstein polynomials will also be discussed. Additionally, the case of a data set and its running maxima will be considered, and the resulting copula will be compared to that associated with a Brownian motion process and its running maxima. As well, an indirect technique for estimating joint density functions from copula densities will be discussed. Such density estimates and their graphical displays ought to contribute to facilitating the fitting of Bayesian models to multivariate observations. Such density estimates and their graphical displays ought to contribute to facilitating the fitting of Bayesian models to multivariate observations. Several illustrative examples involving stock market data will be presented.

## **Title: Differentially private synthetic data for disease mapping**

**Speaker: Harrison Quick**

IS

Drexel University

Motivated by the lack of formal privacy protections in place on CDC WONDER, recent work has proposed the creation of a Synthetic CDC WONDER based on a differentially private Poisson-gamma modeling framework inspired by the disease mapping literature. More specifically, the approach samples values from the posterior predictive distribution associated with modeling event count data with a Poisson-likelihood and assuming a gamma prior on the underlying event rate. The Poisson-gamma framework incorporates (and relies on) publicly available information such as estimates of the underlying population sizes and event rates to improve its utility and protects the sensitive data by increasing the informativeness of the prior distribution. The goal of this work is to present a comparison of the Poisson-gamma framework and the Laplace mechanism for the purpose of generating a synthetic dataset comprised of the 26,000 cancer-related deaths in Pennsylvania counties from 1980. We show that while the Poisson-gamma framework preserves inference on quantities such as urban/rural and black/white disparities in death rates, the Laplace mechanism – when forced to produce non-negative values – can fail to preserve both the magnitude and direction of such disparities.

## **Title: Bayesian Demography and Climate Change Assessment**

**Speaker: Adrian Raftery**

FL

University of Washington

Population forecasts are used by governments and the private sector for planning, with horizons up to about three generations (around 2100) for different purposes. The traditional methods are deterministic using scenarios, but probabilistic forecasts are desired to get an idea of accuracy, to assess changes, and to make decisions involving risks. In a major breakthrough, since 2015 the United Nations has issued probabilistic population forecasts for all countries using a Bayesian methodology that we review here. Assessment of the social cost of carbon relies on long-term forecasts of carbon emissions, which in turn rely on even longer-range population and economic forecasts, to 2300. We extend the UN method to very-long range population forecasts, by combining the statistical approach with expert review and elicitation. We find that, while world population is projected to grow for most of the rest of this century, it is likely to stabilize in the 22nd century, and to decline in the 23rd century. This is joint work with Hana Ševčíková.

## **Title: Graph-Aligned Random Partition Model**

**Speaker: Giovanni Rebaudo**

IS

The University of Texas at Austin

Bayesian nonparametric mixtures and random partition models are effective tools to perform probabilistic clustering. However, standard independent mixture models can be restrictive in some applications such as inference on cell-lineage due to the biological relations of the clusters. The increasing availability of large genomics data and studies require new statistical tools to perform model-based clustering and infer the relationship between the homogeneous subgroups of units. Motivated by single-cell RNA applications we develop a novel dependent mixture model to jointly perform cluster analysis and align the cluster on a graph. Our flexible graph-aligned random partition model exploits Gibbs-type priors as building blocks allowing us to derive analytical results on the probability mass function (pmf) of the graph-aligned random partition. From the pmf, we derive a generalization of the well-known Chinese restaurant process and a related neat MCMC algorithm to perform Bayesian inference. We perform posterior inference on real single-cell RNA data from mice stem cells. We further investigate the performance of our model in capturing underlying clustering structure as well as the underlying graph by means of a simulation study.

## **Title: Discussion of Presentations**

**Speaker: Jerome Reiter**

IS

Duke University

As the discussant in this session, I will present my thoughts on the papers.

## **Title: A Bayesian Gaussian Process for Estimating a Causal Exposure Response Curve and its Change Points**

**Speaker: Boyu Ren**

CT

McLean Hospital

Motivated by environmental policy questions, we address the challenges of estimation, change point detection, and uncertainty quantification of a causal exposure-response function (CERF). Under a potential outcome framework, the CERF describes the relationship between a continuously varying exposure (or treatment) and its causal effect on an outcome. We propose a new Bayesian approach that relies on a Gaussian process (GP) model to estimate the CERF non parametrically. To achieve the desired separation of design and analysis phases, we parametrize the covariance (kernel) function of the GP to mimic matching via a Generalized Propensity Score (GPS). The tuning parameters of the matching function are chosen to optimize covariate balance. Our approach achieves automatic uncertainty evaluation of the CERF with high computational efficiency and enables change point detection through inference on the one-sided derivatives of the CERF. We provide theoretical results showing the correspondence between our Bayesian GP framework and traditional approaches in causal inference for estimating causal effects of a continuous exposure. We apply the methods to 520,711 ZIP-code-level observations to estimate the causal effect of long-term exposures to PM2.5, ozone and NO2 on all-cause mortality, as well as critical levels above which drastic changes in their causal effects are observed, among Medicare enrollees in the United States.

## **Title: A Neural Network for Distributions**

**Speaker: Dan Richard**

CT

University of Alberta

Loss-based clustering methods, such as k-means and its variants, are standard tools for finding groups in data. However, the lack of quantification of uncertainty in the estimated clusters is a disadvantage. Model-based clustering based on mixture models provides an alternative, but such methods face computational problems and large sensitivity to the choice of kernel. This article proposes a generalized Bayes framework that bridges between these paradigms through the use of Gibbs posteriors. In conducting Bayesian updating, the log-likelihood is replaced by a loss function for clustering, leading to a rich family of clustering methods. The Gibbs posterior represents a coherent updating of Bayesian beliefs without needing to specify a likelihood for the data, and can be used for characterizing uncertainty in clustering. We consider losses based on Bregman divergence and pairwise similarities, and develop efficient deterministic algorithms for point estimation along with sampling algorithms for uncertainty quantification. Several existing clustering algorithms, including k-means, can be interpreted as generalized Bayes estimators under our framework, and hence we provide a method of uncertainty quantification for these approaches; for example, allowing calculation of the probability a data point is well clustered.

## **Title: A generalized Bayes framework for probabilistic clustering**

**Speaker: Tommaso Rigon**

CT

University of Milano-Bicocca

Particle filters are used to solve probability filtering problems by representing the posterior distribution over the state-space with a collection of particles, which are then analysed for decision making purposes. This analysis could include calculation of various statistics, like the mean, to guide a decision; however, this could be leaving out vital information required for making good decisions. How can a decision be made that encapsulates all the information contained in the particles? The answer is to provide the posterior distribution as an input to the newly developed “distributional neural network,” which differs from traditional neural networks that map real finite vectors to real finite vectors. The talk will provide a Universal Approximation result for neural network representations of arbitrary functions with measure valued or even topological space inputs. These approximations work well on a computer with empirical distributions and Bayes factor like those that arise naturally from particle filters.

## **Title: Frequentist asymptotics for parametric empirical Bayes**

**Speaker: Stefano Rizzelli**

CT

Università Cattolica, Milan

Empirical Bayes (EB) methods, via maximum marginal likelihood estimators (MMLE) of prior hyperparameters, provide popular alternatives to genuine Bayesian inference. For regular parametric models, we provide sufficient conditions for MMLE convergence to a set of oracle hyperparameter choices, asymptotically minimizing the Kullback-Leibler divergence between the true data generating density and the marginal likelihood. This complements a result previously established in Petrone et al. (2014). We also refine their comparison between Bayes and EB to higher-order asymptotics, showing that EB offers a closer approximation to posterior/predictive distributions obtained with oracle hyperparameters than with nonoracle choices. We illustrate applications to linear regression and finite mixture models. We further provide new results in a degenerate case where the EB prior can collapse to a point mass with nonvanishing frequentist probability, giving conditions under which the MMLE still asymptotically provides an oracle choice. Finally, we explore extensions to irregular, nonidentifiable statistical models.

## **Title: Testing for components in finite and infinite mixtures**

**Speaker: Christian Robert**

CT

Université Paris Dauphine

Testing for the number of components in a finite mixture model or against the fit of a finite mixture model for a given dataset has long been and still is an issue of much interest, albeit yet missing a fully satisfactory resolution. Using a Bayes factor to find the right number of components  $K$  in a finite mixture model is known to provide a consistent procedure. We furthermore establish the consistence of the Bayes factor when comparing a parametric family of finite mixtures against the nonparametric location Dirichlet Process Mixture (DPM) model. In practice, estimating the model evidence (1) is unfortunately a notoriously difficult task for finite and infinite mixture models and we reexamine here different Monte Carlo techniques advocated in the recent literature, as well as novel approaches based on Geyer's (1994) reverse logistic regression technique and Sequential Monte Carlo (SMC).

## **Title: Stereographic Markov chain Monte Carlo**

**Speaker: Gareth Roberts**

CT

University of Warwick

High dimensional distributions, especially those with heavy tails, are difficult for off-the-shelf MCMC samplers: the combination of unbounded state spaces, diminishing gradient information and local moves, results in empirically observed "stickiness" and poor theoretical mixing properties – lack of geometric ergodicity. This talk will introduce a new class of MCMC samplers that map the original high dimensional problem in Euclidean space onto a sphere and remedy these notorious mixing problems. In particular, we develop Random Walk Metropolis type algorithms as well as versions of Bouncy Particle Sampler that are uniformly ergodic for a large class of light- and heavy-tailed distributions and also empirically exhibit rapid convergence. This is joint work with Jun Yang (Oxford) and Krys Latuszynski (Warwick).

## **Title: A Bayesian Approach to Spherical Factor Analysis for Binary Data**

**Speaker: Abel Rodriguez**

IS

University of Washington

Factor models are widely used across diverse areas of application for purposes that include dimensionality reduction, covariance estimation, and feature engineering. Traditional factor models can be seen as an instance of linear embedding methods that project multivariate observations onto a lower dimensional Euclidean latent space. This paper discusses a new class of geometric embedding models for multivariate binary data in which the embedding space correspond to a spherical manifold, with potentially unknown dimension. The resulting models include traditional factor models as a special case, but provide additional flexibility. Furthermore, unlike other techniques for geometric embedding, the models are easy to interpret, and the uncertainty associated with the latent features can be properly quantified. These advantages are illustrated using both simulation studies and real data on voting records from the U.S. Senate. This is joint work with Xingchen Yu.

## **Title: Cauchy Markov random fields for Bayesian inversion**

**Speaker: Lassi Roininen**

IS

LUT University

We consider using Cauchy Markov random field priors for Bayesian inversion with applications in sawmill log X-ray imaging. Using Cauchy fields leads to posterior distributions which are non-Gaussian, high-dimensional, multimodal and heavy-tailed. Thus, we need sophisticated optimization and Markov chain Monte Carlo (MCMC) methods. We firstly propose a one-dimensional second order Cauchy difference prior, and construct new first and second order two-dimensional isotropic Cauchy difference priors. Another new Cauchy prior is based on the stochastic partial differential equation approach, derived from Matérn type Gaussian presentation. We consider maximum a posteriori and conditional mean estimation. We exploit state-of-the-art MCMC methodologies such as Metropolis-within-Gibbs, Repelling-Attracting Metropolis, and No-U-Turn sampler variant of Hamiltonian Monte Carlo. We demonstrate the applicability of the models and methods with two one and two-dimensional synthetic deconvolution problems, and with real data example arising from the sawmill X-ray imaging with extremely sparse measurement geometry requiring sophisticated prior constructions and efficient sampling methods.

## **Title: A Bayesian nonparametric approach for causal inference with multiple mediators**

**Speaker: Samrat Roy**

IS

University of Pennsylvania

Mediation analysis with contemporaneously observed multiple mediators is a significant area of causal inference. Recent methodologies dealing with multiple mediators are based on parametric models and thus may suffer from parametric misspecification. Also, the existing literature estimates the joint mediation effect as the sum of individual mediators effect, which, on many occasions, is not a reasonable assumption. In this paper, we propose a methodology that overcomes the two aforementioned drawbacks in the existing literature. Our method is based on a novel Bayesian nonparametric (BNP) approach, wherein, the joint distribution of the observed data (outcome, mediators, treatment, and confounders) is modeled flexibly using an enriched Dirichlet process with three levels: the first level characterizing the conditional distribution of the outcome given the mediators, treatment, and the confounders, the second level corresponding to the conditional distribution of each of the mediators given the treatment and the confounders, and the third level corresponding to the distribution of the treatment and the confounders. Using the joint distribution in the EDP, we use standardization (g-computation) to compute causal mediation effects. The efficacy of our proposed method is demonstrated with simulations. Also, we apply our proposed method to analyze data from a study of Ventilator-associated Pneumonia (VAP) co-infected patients, where the effect of the abundance of Pseudomonas on VAP infection is suspected to be mediated through various antibiotic exposures.

## **Title: An overview of Bayesian computational frameworks for teaching**

**Speaker: Colin Rundel**

IS

Duke University

Bayesian modeling has never been more accessible and widespread, this is due in large part to development of modern Bayesian computational frameworks like BUGS, JAGS, Stan, etc. While this wealth of choices is empowering, it can also be overwhelming to new users trying to decide where to start. This is particularly true for instructors who need to carefully weight the advantages and disadvantages of these frameworks in the context of their target audience and desired learning outcomes. In this talk we will discuss some of the more popular options (JAGS, Stan, PyMC3, and others) with a focus on integrating them into a Bayesian modeling courses at both the undergraduate and graduate level. Specific attention will be given to the larger ecosystem for each framework, as well as considerations around syntax, efficiency, and performance.

## **Title: Bayesian bi-clustering for temporally heterogeneous high-dimensional longitudinal data**

*Speaker: Massimiliano Russo*

CT

Harvard University

X-linked Dystonia-Parkinsonism (XDP) is a rare genetic form of dystonia found almost entirely among males of Filipino descent, characterized by highly heterogeneous symptoms and unknown progression patterns. Distinguishing sub-types of XDP is pivotal in advancing our understanding of the disease and in providing effective targeted treatment for the affected patients. However, analysis of existing data is complicated by the fact that (i) the patients are observed for only a short length of time at different stages of progression, and (ii) the disease symptoms are measured using a large number of interdependent scales. To overcome these challenges, we propose a novel Bayesian statistical model that simultaneously clusters subjects according to the trajectory of their progression and clusters variables into jointly relevant aspects of the disease. We apply the model to clinical XDP data and find that it reveals novel insights into the patterns of disease progression.

## **Title: ABC-ROM: Time-series forecasting with ABC prediction uncertainty and Reduced Order Models**

*Speaker: Robin Ryder*

CT

Université Paris-Dauphine

We consider high-dimensional Partial Differential Equations models with incomplete observations. Data arise from a stochastic process involving a large number of time-evolving compartments, but we only observe a small number of (collections of) these compartments. In such a setting, no analytical solution is available, so Numerical Analysts rely on Reduced Order Models instead, a simulation-based method which projects synthetic trajectories onto a lower-dimensional space. We propose a technique to combine Reduced Order Models with Approximate Bayesian Computation, and obtain an approximate, fully-Bayesian prediction interval. We apply this method to prediction uncertainty for detailed epidemiological models of the Covid pandemic.

## **Title: Bayesian Additive Multivariate Decision Trees for Spatial Nonparametric Function Estimation on Complex Domains**

**Speaker: Huiyan Sang**

IS

Texas A&M University

Spatial nonparametric regression on complex domains has been a challenging task. Many popular ensemble/boosting tree nonparametric regression models based on binary decision trees with univariate split rules are not designed to account for intrinsic geometries and domain boundaries. This article proposes a Bayesian additive multivariate decision tree regression model for spatial nonparametric regression on manifolds, with an emphasis on complex constrained domains or irregularly shaped spaces embedded in Euclidean spaces. Our model is built upon multivariate decision tree weak learners, where a node of the decision tree can split a multivariate spatial feature space while respecting spatial contiguity constraints as well as intrinsic geometries and domain boundary constraints. Equipped with a soft prediction scheme, our method is demonstrated to significantly outperform other competing methods such as BART and GP in simulation experiments and in an application to the chlorophyll data in Aral Sea, due to its strong local adaptivity to different levels of smoothness.

## **Title: Non-Gaussian geostatistical models using nearest neighbors processes**

**Speaker: Bruno Sanso**

CT

University of California, Santa Cruz

We present a framework for non-Gaussian spatial processes that encompasses large distribution families. Spatial dependence for a set of irregularly scattered locations is described with a mixture of pairwise kernels. Focusing on the nearest neighbors of a given location, within a reference set, we obtain a valid spatial process: the nearest neighbor mixture transition distribution process (NNMP). We develop conditions to construct general NNMP models with arbitrary pre-specified marginal distributions. Essentially, NNMPs are specified by a bi-variate distribution, with suitable marginals, used to specify the mixture transition kernels. Such distribution can be spatially varying, to capture non-homogeneous spatial features. The mixture structure of the model allows for efficient MCMC-based exploration of posterior distribution of the model parameters, even for relatively large number of locations. We illustrate the capabilities of NNMPs with observations corresponding to distributions with different non-Gaussian characteristics: Long tails; Compact support; skewness.

## **Title: Private Tabular Survey Data Products through Synthetic Microdata Generation**

**Speaker: Terrance Savitsky**

IS

U.S. Bureau of Labor Statistics

We propose two synthetic microdata approaches to generate private tabular survey data products for public release. We adapt a pseudo posterior mechanism that downweights by-record likelihood contributions with weights  $\in [0, 1]$  based on their identification disclosure risks to producing tabular products for survey data. Our method applied to an observed survey database achieves an asymptotic global probabilistic differential privacy guarantee. Our two approaches synthesize the observed sample distribution of the outcome and survey weights, jointly, such that both quantities together possess a privacy guarantee. The privacy-protected outcome and survey weights are used to construct tabular cell estimates (where the cell inclusion indicators are treated as known and public) and associated standard errors to correct for survey sampling bias. Through a real data application to the Survey of Doctorate Recipients public use file and simulation studies motivated by the application, we demonstrate that our two microdata synthesis approaches to construct tabular products provide superior utility preservation as compared to the additive-noise approach of the Laplace Mechanism. Moreover, our approaches allow the release of microdata to the public, enabling additional analyses at no extra privacy cost.

## **Title: Removing the mini-batching error in Bayesian inference using Adaptive Langevin dynamics**

**Speaker: Inass Sekkat**

IS

CERMICS- Ecole des Ponts ParisTech

The computational cost of usual Monte Carlo methods for sampling a posteriori laws in Bayesian inference scales linearly with the number of data points. One option to reduce it to a fraction of this cost is to resort to mini-batching in conjunction with unadjusted discretizations of Langevin dynamics, in which case only a random fraction of the data is used to estimate the gradient. However, this leads to an additional noise in the dynamics and hence a bias on the invariant measure which is sampled by the Markov chain. We advocate using the so-called Adaptive Langevin dynamics, which is a modification of standard inertial Langevin dynamics with a dynamical friction which automatically corrects for the increased noise arising from mini-batching. We investigate the practical relevance of the assumptions underpinning Adaptive Langevin (constant covariance for the estimation of the gradient), which are not satisfied in typical models of Bayesian inference, and quantify the bias induced by minibatching in this case. We also show how to extend AdL in order to systematically reduce the bias on the posterior distribution by considering a dynamical friction depending on the current value of the parameter to sample.

## **Title: Scalable Bayesian inference for time series via divide-and-conquer**

**Speaker: Deborshee Sen**

IS

University of Bath

Bayesian computational algorithms tend to scale poorly as data size increases. This had led to the development of divide-and-conquer-based approaches for scalable inference. These divide the data into subsets, perform inference for each subset in parallel, and then combine these inferences. While appealing theoretical properties and practical performance have been demonstrated for independent observations, scalable inference for dependent data remains challenging. In this work, we study the problem of Bayesian inference from very long time series. The literature in this area focuses mainly on approximate approaches that lack rigorous theoretical guarantees. We propose a simple and scalable divide-and-conquer method, and provide accuracy guarantees. Numerical simulations and real data applications demonstrate the effectiveness of our approach.

## **Title: Modeling First Arrival of Migratory Birds using a Hierarchical Max-infinitely Divisible Process**

**Speaker: Benjamin Shaby**

IS

Colorado State University

"Humans have recorded the arrival dates of migratory birds for millennia, searching for trends and patterns. As the first arrival among individuals in a species is the realized tail of the probability distribution of arrivals, the appropriate statistical framework with which to analyze such events is extreme value theory. Here, for the first time, we apply formal extreme value techniques to the dynamics of bird migrations. We study the annual first arrivals Magnolia Warblers using modern tools from the statistical field of extreme value analysis. Using observations from the eBird database, we model the spatial distribution of Magnolia Warbler arrivals as a max-infinitely divisible process, which allows us to spatially interpolate observed annual arrivals in a probabilistically coherent way, and to project arrival dynamics into the future by conditioning on climatic variables. "

## **Title: Bayesian analysis of matrix data with applications to professional basketball game analysis**

**Speaker: Weining Shen**

IS

University of California, Irvine

We propose a Bayesian nonparametric matrix clustering approach to analyze the latent heterogeneity structure in the shot selection data collected from professional basketball players in the National Basketball Association (NBA). The proposed method adopts a mixture of finite mixtures framework and fully utilizes the spatial information via a mixture of matrix normal distribution representation. We propose an efficient Markov chain Monte Carlo algorithm for posterior sampling that allows simultaneous inference on both the number of clusters and the cluster configurations. We also establish large-sample convergence properties for the posterior distribution. The compelling empirical performance of the proposed method is demonstrated via simulation studies and an application to shot chart data from selected players in the NBA's 2017–2018 regular season.

## **Title: Non-Stratified Chain Event Graphs for Modelling Asymmetric Processes**

**Speaker: Aditi Shenvi**

IS

University of Warwick

Probabilistic graphical models provide a powerful and intuitive framework for representing complex processes through the conditional independence relationships of their defining random variables or events. They have been successfully applied to wide range of domains. However, popular graphical models such as Bayesian networks (BNs) are unable to effectively represent processes with context-specific independencies (CSI) or structural asymmetries i.e. processes whose event spaces do not admit a natural product space structure. Chain event graphs (CEGs) are a recent family of graphical models that were proposed for modelling such processes. CEGs generalise the class of discrete state space BNs. The CEG literature, through the 'stratified' CEG class, has demonstrated its efficacy over BNs in modelling processes with CSI. However, despite structural asymmetries being common in many domains (occurring naturally e.g. a defendant vs prosecutor's version of events, or by design e.g. a non-factorial public health intervention), modelling such asymmetric processes with CEGs has largely been unexplored. In this talk we present the class of 'non-stratified' CEGs that can model such asymmetric processes. We demonstrate their suitability for asymmetric processes through applications in public health and policing.

## **Title: Sampling with Mirrored Stein Operators**

**Speaker: Jiaxin Shi**

CT

Microsoft Research New England

Accurately approximating an unnormalized distribution with a discrete sample is a fundamental challenge in machine learning and probabilistic inference. Particle evolution methods like Stein variational gradient descent tackle this challenge by applying deterministic updates to particles using operators based on Stein's method and reproducing kernels to sequentially minimize Kullback-Leibler divergence. However, these methods break down for constrained targets and fails to exploit informative non-Euclidean geometry. In this talk, I will introduce a new family of particle evolution samplers suitable for constrained domains and non-Euclidean geometries. These samplers are derived from a new class of Stein operators and have deep connections with Riemannian Langevin diffusion, mirror descent, and natural gradient descent. We demonstrate that these new samplers yield accurate approximations to distributions on the simplex, deliver valid confidence intervals in post-selection inference, and converge more rapidly than prior methods in large-scale unconstrained posterior inference. Finally, we establish the convergence of our new procedures under verifiable conditions on the target distribution.

## **Title: A Bayesian decision support system for counteracting activities of terrorist groups**

**Speaker: Jim Smith**

CT

University of Warwick & the Alan Turing Institute

Activities of terrorist groups present a serious threat to the security and well-being of the general public. Counterterrorism authorities aim to identify and frustrate the plans of terrorist groups before they are put into action. Whilst the activities of terrorist groups are likely to be hidden and disguised, its members need to communicate and coordinate to organise their activities. Such observable behaviour and communications data can be utilised by the authorities to estimate the threat posed by the group. We present a Bayesian integrating decision support system that can combine information relating to each member of a terrorist group as well as the combined activities of the group. Our model, thus, folds in the level of threat posed by each member of the group through a credible proxy measure (i.e. the observed activities of each member), which is crucial in such social networks where individuals cannot be considered exchangeable.

## **Title: Fast and Accurate Variational Inference for Models with Many Latent Variables**

**Speaker: Michael Smith**

IS

University of Melbourne

Models with a large number of latent variables are often used to utilize the information in big or complex data, but can be difficult to estimate. Variational inference methods provide an attractive solution. These methods use an approximation to the posterior density, yet for large latent variable models existing choices can be inaccurate or slow to calibrate. Here, we propose a family of tractable variational approximations that are more accurate and faster to calibrate for this case. It combines a parsimonious approximation for the parameter posterior with the exact conditional posterior of the latent variables. We derive a simplified expression for the re-parameterization gradient of the variational lower bound, which is the main ingredient of optimization algorithms used for calibration. Implementation only requires exact or approximate generation from the conditional posterior of the latent variables, rather than computation of their density. In effect, our method provides a new way to employ Markov chain Monte Carlo (MCMC) within variational inference. We illustrate with random coefficients tobit model applied to two million sales by 20,000 individuals in a consumer panel. Last, we show how to implement data sub-sampling in variational inference for our approximation, further reducing computation time.

## **Title: Federated Learning with Hamiltonian Monte Carlo**

**Speaker: Qifan Song**

IS

Purdue University

Federated learning (FL) becomes an emerging learning paradigm. It is a distributed machine learning architecture that trains a centralized global model using decentralized data sets. This work proposed a Bayesian federated learning algorithm, namely, the federated averaging stochastic Hamiltonian Monte Carlo (FA-HMC), for parameter estimation and uncertainty quantification. We provide rigorous convergence guarantees of FA-HMC for non-iid distributed data sets, under the strongly convexity and Hessian smoothness assumptions. Our analysis reveals how the dimension of parameter space, the frequency of communication between the center node and worker nodes, and the heterogeneity level among local data sets affect the convergence of FA-HMC. The results are supported by our simulations.

## **Title: Fast Approximate BayesBag Model Selection via Taylor Expansions**

*Speaker: Neil Spencer*

CT

Harvard University

In recent years, BayesBag has emerged as an effective remedy for the brittleness of Bayesian model selection under model misspecification. However, computing BayesBag can be prohibitively expensive for large datasets. In this talk, I propose a fast approximation of BayesBag model selection based on Taylor approximations of the log marginal likelihood, which can achieve results comparable to BayesBag in a fraction of the computation time. I provide concrete bounds for the approximation error and establish conditions under which it converges to zero as the dataset grows.

## **Title: Divide-and-Conquer Bayesian Inference in Hidden Markov Models**

*Speaker: Sanvesh Srivastava*

IS

University of Iowa

Divide-and-conquer Bayesian methods consist of three steps: dividing the data into smaller computationally manageable subsets, running a sampling algorithm in parallel on all the subsets, and combining parameter draws from all the subsets. The combined parameter draws are used for efficient posterior inference in massive data settings. Several innovative methods have been developed over the years, but a major restriction common to all is that their first two steps assume that the observations are independent. We address this problem by developing a divide-and-conquer method for Bayesian inference in parametric hidden Markov models, where the state space is known and finite. First, we show that the dependence can be preserved on the subsets by appropriately modifying the subset likelihoods. Second, if the number of subsets is chosen appropriately depending on the mixing properties of the hidden Markov chain, then we show that the subset posterior distributions defined using the modified likelihood are asymptotically normal as the subset sample size tends to infinity. Finally, we present numerical results to justify the empirical validity of the theoretical results.

## **Title: Uncertainty calibration and exemplar identification for heterogeneous treatment effects with Individualized Bayesian Causal Forests (iBCF)**

**Speaker: Jennifer Starling**

CT

Mathematica Policy Research

Bayesian Causal Forests (BCF) has proven to be an effective framework for estimating heterogeneous treatment effects, particularly where standard methods yield biased estimates due to residual confounding. BCF parameterizes BART models to allow for separate regularization of treatment and prognostic effects, making it possible to shrink towards homogeneity. However, BCF's credible intervals are known to under-cover, and treatment effect estimates are purely a function of measured covariates, without the ability to identify "exemplar" observations that respond to treatment more strongly than covariates predict. We introduce iBCF, a new version of the BCF prior that incorporates observation-level random effects to allow for robust estimation of individual causal impacts. We demonstrate the utility of this approach in calibrating uncertainty and identifying exemplar observations. We showcase the benefits of this approach in a simulation based on real data from a Medicare primary care program.

## **Title: Measuring the robustness of Gaussian processes to kernel choice**

**Speaker: Will Stephenson**

IS

MIT

Gaussian processes (GPs) are used to make medical and scientific decisions, including in cardiac care and monitoring of atmospheric carbon dioxide levels. Notably, the choice of GP kernel is often somewhat arbitrary. In particular, uncountably many kernels typically align with qualitative prior knowledge (e.g. function smoothness or stationarity). But in practice, data analysts choose among a handful of convenient standard kernels (e.g. squared exponential). In the present work, we ask: Would decisions made with a GP differ under other, qualitatively interchangeable kernels? We show how to answer this question by solving a constrained optimization problem over a finite-dimensional space. We can then use standard optimizers to identify substantive changes in relevant decisions made with a GP. We demonstrate in both synthetic and real-world examples that decisions made with a GP can exhibit non-robustness to kernel choice, even when prior draws are qualitatively interchangeable to a user.

## **Title: Bayesian Covariate-Dependent Gaussian Graphical Models with Varying Structure**

**Speaker: Francesco Stingo**

IS

University of Florence

We introduce Bayesian Gaussian graphical models with covariates (GGMx), a class of multivariate Gaussian distributions with covariate-dependent sparse precision matrix. We propose a general construction of a functional mapping from the covariate space to the cone of sparse positive definite matrices, that encompasses many existing graphical models for heterogeneous settings. Our methodology is based on a novel mixture prior for precision matrices with a non-local component that admits attractive theoretical and empirical properties. The flexible formulation of GGMx allows both the strength and the sparsity pattern of the precision matrix (hence the graph structure) change with the covariates. Posterior inference is carried out with a carefully designed Markov chain Monte Carlo algorithm which ensures the positive definiteness of sparse precision matrices at any given covariates' values. Extensive simulations and a case study in cancer genomics demonstrate the utility of the proposed model. Joint work with Yang Ni and Veerabhadran Baladandayuthapani.

## **Title: Bayesian Spatio-Temporal Models for EPV in Hockey: A Simulation Approach**

**Speaker: Tyrel Stokes**

IS

McGill/Zelus Analytics

Building off previous work in Basketball, Soccer, and Hockey exploring expected possession value we consider offensive play sequences from event-level hockey data as realizations from stochastic processes. We construct more than 40 spatio-temporal submodels fit in INLA which allow us to forward simulate entire play sequences at the resolution of the data from the posterior predictive distribution. We use time since zone entry as a proxy for unobserved offensive and defensive structure. This allows us to effectively propagate and quantify uncertainty in outcomes of interest, namely expected value of the sequences and added value of player actions. We discuss insights from the model on OHL data and extensions to full-ice data.

## **Title: Collapsed blocked Gibbs sampling for improved convergence in latent variable models**

*Speaker: David Swanson*

CT

University of Oslo

Marginalizing over nuisance parameters can improve convergence of MCMC, as can blocking correlated parameters. This is straightforward to visualize and think about in contexts like low-dimensional multivariate normal distributions, but for latent variable models it is difficult because of the non-smooth way in which label assignment and cluster-specific parameters relate to one another. In general, however, cluster assignment of observations close to one another will tend to correlate since their weight pulls a cluster's mean to a center held in common. One can explore with respect to this correlation by sampling cluster assignments for a random block of observations, not updating cluster-specific parameters for several iterations, then performing an accept-reject step for each proposed block move. One can cater blocking to the context to push acceptance probabilities close to 1 and assure ergodicity by only using a weighted coin. The approach also yields computational advantages.

## **Title: Non-reversible parallel tempering on optimized paths**

*Speaker: Saifuddin Syed*

IS

University of Oxford

MCMC methods are a popular tool in computation science used to evaluate expectations with respect to complex probability distributions over general state spaces. They work by averaging over the trajectory of a Markov chain stationary with respect to the target distribution. In theory, the MCMC algorithms converge asymptotically, but, in practice, for challenging problems where the target distributions are high-dimensional with well-separated modes, MCMC algorithms can get trapped exploring local regions of high probability and suffer from poor mixing.

Physicists and statisticians independently introduced parallel tempering (PT) algorithms to tackle this issue. PT delegates the task of exploration to additional annealed chains running in parallel with better mixing properties. They then communicate with the target chain of interest and help discover new unexplored regions of the sample space. Since their introduction in the 90s, PT algorithms are still extensively used to improve mixing in challenging sampling problems arising in statistics, physics, computational chemistry, phylogenetics, and machine learning.

The classical approach to designing PT algorithms was developed using a reversible paradigm that is difficult to tune and deteriorates in performance when too many parallel chains are introduced. This talk will introduce a new non-reversible paradigm for PT that dominates its reversible counterpart while avoiding the performance collapse endemic to reversible methods. We will then establish near-optimal tuning guidelines and efficient black-box methodology scalable to GPUs. Our work outperforms state-of-the-art PT methods and has been used at scale by researchers to study the evolutionary structure of cancer and discover magnetic polarization in the photograph of the supermassive black hole M87.

## **Title: Recent developments in Gibbs posterior inference**

**Speaker: Nicholas Syring**

IS

Iowa State University

Bayesian inference has certain advantages, but a fully (and generally correctly) specified statistical model is needed to realize those advantages. What if the quantity of interest is not naturally described as a "model parameter"? Then there is no sense in which a specified statistical model could be "correct" and, hence, there's risk of model misspecification bias. To avoid this bias, one can construct a so-called Gibbs posterior that directly targets the quantity of interest, compared to a Bayesian posterior that does so only indirectly through a (possibly misspecified) statistical model and marginalization. In this talk, first I'll motivate and define Gibbs posteriors. Second, I'll present asymptotic properties of Gibbs posteriors with a focus on specific examples. Finally, I'll discuss the need to properly scale the Gibbs posterior so that inferences (or predictions) are valid, and present an algorithm that achieves this.

## **Title: Decision-guided Bayesian model uncertainty analysis: Betting on better models**

**Speaker: Emily Tallman**

IS

Duke University

I discuss background and examples of Bayesian predictive decision synthesis (BPDS) in the evaluation, comparison, and combination of candidate models. BPDS extends the foundations and practical advances of Bayesian predictive synthesis (BPS) by formally involving decision-analytic outcomes in assessing and weighting candidate models. A subjective Bayesian perspective admits the integration of context-specific utility functions that emphasize the intended uses of models when evaluating predictions. BPDS theory builds on methods of entropic tilting to enable implementation. Specific examples come from applied contexts including optimal design in regression prediction and sequential time series forecasting for financial portfolio decisions.

## Title: Guided sequential ABC schemes for intractable Bayesian models

Speaker: *Massimiliano Tamborrino*

IS

University of Warwick

Sequential algorithms such as sequential importance sampling (SIS) and sequential Monte Carlo (SMC) have shown to be fundamental in Bayesian inference for models not admitting a readily available likelihood function. For approximate Bayesian computation (ABC), sequential Monte Carlo ABC is the state-of-art sampler. However, since the ABC paradigm is intrinsically computationally wasteful, sequential ABC schemes can benefit from well-targeted proposal samplers that efficiently avoid improbable parameter regions. We contribute to the ABC modeller's toolbox with novel proposal samplers that are conditional to summary statistics of the data. In a sense, the proposed parameters are "guided" to rapidly reach regions of the posterior surface that are compatible with the observed data. This speeds up the convergence of these sequential samplers, thus reducing the computational effort, while preserving the accuracy in the inference. We provide a variety of guided (Gaussian or copula-based) samplers for both SIS-ABC and SMC-ABC and consider several case-studies, including a simulation study of cell movements with high-dimensional summary statistics. This is a joint work with Umberto Picchini (Chalmers University and the University of Göteborg).

## Title: Partial Membership Models for Functional Data

Speaker: *Donatello Telesca*

IS

University of California, Los Angeles

Partial membership models, or mixed membership models, are a flexible unsupervised clustering method that allows observations to belong to multiple clusters at the same time. In this paper, we propose a Bayesian partial membership model for functional data. By using the multivariate Karhunen-Loeve theorem, we are able to derive a scalable representation that maintains data-driven covariance structures and establish conditional posterior consistency. Compared to previous work on partial membership models, our proposal allows for increased flexibility, with the benefit of direct interpretation of the mean and covariance functions. Our work is motivated by studies in functional brain imaging of children with Autism Spectrum Disorder (ASD).

## **Title: Pathwise Conditioning of Gaussian Processes**

**Speaker: Alexander Terenin**

CT

University of Cambridge

In Gaussian processes, conditioning and computation of posterior distributions is usually done in a distributional fashion by working with finite-dimensional marginals. Recently, an alternative way of thinking about conditioning, based on actual random functions rather than their probability distributions, has gained prominence within the machine learning community. This perspective is particularly helpful in decision-theoretic settings such as Bayesian optimization, where it enables efficient computation of a wider class of acquisition functions than otherwise possible. In this talk, we describe these recent advances in a language accessible to the statistics community, and discuss their broader implications to Bayesian nonparametrics.

## **Title: Black Box Probabilistic Numerics**

**Speaker: Onur Teymur**

CT

University of Kent

Probabilistic numerics casts numerical tasks, such the solution of differential equations, as inference problems to be solved. One approach is to model the unknown quantity of interest as a random variable, and then to constrain it using information generated during the course of a traditional numerical method. However, this information may be nonlinearly related to the quantity of interest, rendering the proper conditioning of random variables difficult or impossible. Instead, we construct probabilistic numerical methods based only on the final output from a traditional method. A convergent sequence of approximations to the quantity of interest constitute a dataset, from which the limit can be extrapolated, in a probabilistic analogue of Richardson's extrapolation. This approach massively expands the range of tasks to which probabilistic numerics can be applied, inheriting the features and performance of state-of-the-art numerical methods, as well as enabling provably higher orders of convergence to be achieved.

## Title: Functional Priors for Bayesian Deep Learning

Speaker: *Ba-Hien Tran*

CT

EURECOM

The Bayesian treatment of neural networks dictates that a prior distribution is specified over their weight and bias parameters. This poses a challenge because modern neural networks are characterized by a large number of parameters and non-linearities. The choice of these priors has an unpredictable effect on the distribution of the functional output which could represent a hugely limiting aspect of Bayesian deep learning models. Differently, Gaussian processes offer a non-parametric rigorous framework to define prior distributions over the space of functions. With this contributed talk we aim to introduce a novel and robust framework to impose such functional priors on modern neural networks through minimizing the Wasserstein distance between samples of stochastic processes. We provide extensive experimental evidence that coupling these priors with scalable Markov chain Monte Carlo sampling offers systematically large performance improvements over alternative choices of priors and state-of-the-art approximate Bayesian deep learning approaches.

## Title: Information geometry in Variational Bayes

Speaker: *Minh-Ngoc Tran*

IS

University of Sydney

Variational Bayes (VB) has become a widely-used tool for Bayesian inference in statistics and machine learning. Nonetheless, the development of the existing VB algorithms is so far generally restricted to the case where the variational parameter space is Euclidean, which hinders the potential broad application of VB methods. We extend the scope of VB to the case where the variational parameter space is a Riemannian manifold. We develop an efficient manifold-based VB algorithm that exploits both the geometric structure of the constraint parameter space and the information geometry of the manifold of VB approximating probability distributions. Our algorithm is provably convergent and achieves a good convergence rate. We demonstrate through numerical experiments that the proposed algorithms are stable, less sensitive to initialization and compares favourably to existing VB methods.

## **Title: Dealing with large data sets in spatio-temporal disease mapping**

*Speaker: Maria Dolores Ugarte*

IS

Universidad Pública de Navarra

In this talk we propose a general procedure to analyse large data sets in spatio-temporal disease mapping. Model fitting is carried out over a partition of the spatio-temporal domain. This is a simple idea that works very well in this context as the models are defined to borrow strength locally in space and time, providing reliable risk estimates. Parallel and distributed strategies are proposed to speed up computations in a setting where Bayesian model fitting is generally prohibitively time-consuming and even unfeasible. We evaluate the whole procedure in a simulation study with a twofold objective: to estimate risks accurately and to detect extreme risk areas while avoiding false positives/negatives. We show that our method outperforms classical global models. A real data analysis will be also presented."

## **Title: The G-Wishart Weighted Proposal Algorithm: Efficient Posterior Computation for Gaussian Graphical Models**

*Speaker: Willem van den Boom*

IS

National University of Singapore

Gaussian graphical models can capture complex dependency structures amongst variables. For such models, Bayesian inference is attractive as it provides principled ways to incorporate prior information and to quantify uncertainty through the posterior. However, posterior computation under the conjugate G-Wishart prior on the precision matrix is expensive for general non-decomposable graphs. We therefore propose a Markov chain Monte Carlo (MCMC) method named the G-Wishart weighted proposal algorithm (WWA). WWA's distinctive features include delayed acceptance MCMC, Gibbs updates for the precision matrix and an informed proposal on the graph space that enables embarrassingly parallel computations. Compared to existing approaches, WWA reduces the frequency of the relatively expensive sampling from the G-Wishart distribution. This results in faster MCMC convergence, improved MCMC mixing and reduced computation time. Numerical studies show that WWA provides a more efficient tool for posterior inference than competing state-of-the-art MCMC algorithms.

## **Title: A Flexible Predictive Density Combination Model for Large Financial Data Sets in Regular and Crisis Periods**

**Speaker: Herman van Dijk**

IS

Tinbergen Institute, Erasmus University and Norges Bank

A flexible predictive density combination model is introduced for large financial data sets which allows for dynamic weight learning and model set incompleteness. Dimension reduction procedures allocate the large sets of predictive densities and combination weights to relatively small sets. Given the representation of the probability model in extended nonlinear state-space form, efficient simulation-based Bayesian inference is proposed using parallel sequential clustering as well as nonlinear filtering, implemented on graphics processing units. The approach is applied to combine predictive densities based on a large number of individual stock returns of daily observations over a period that includes the Covid-19 crisis period. Evidence on the quantification of predictive accuracy, uncertainty and risk, in particular, in the tails, may provide useful information for investment fund management. Information on dynamic cluster composition, weight patterns and model set incompleteness give also valuable signals for improved modelling and policy specification.

## **Title: Likelihood-free Sequential Transport Monte Carlo**

**Speaker: Cecilia Viscardi**

IS

University of Florence

Approximate Bayesian computation (ABC) is a class of methods for drawing inferences when the likelihood function is unavailable or computationally demanding to evaluate. Importance sampling and other algorithms using sequential importance sampling steps are state-of-art methods in ABC. Most of them get samples from tempered approximate posterior distributions defined by considering a decreasing sequence of ABC tolerance thresholds. Their efficiency is sensitive to the choice of an adequate proposal distribution and/or forward kernel function. We present a novel ABC method addressing this problem by combining importance sampling steps and optimization procedures. We resort to Normalising Flows (NFs) to optimize proposal distributions over a family of densities to transport particles drawn at each step towards the next tempered target. Therefore, the combination of sampling and optimization steps allows tempered distributions to get efficiently closer to the target posterior. Finally, we show the performance of our method on examples that are a common benchmark for likelihood-free inference.

## Title: Dependence between Bayesian neural network units

*Speaker: Mariia Vladimirova*

CT

Inria Grenoble

The connection between Bayesian neural networks and Gaussian processes gained a lot of attention in the last few years, with the flagship result that hidden units converge to a Gaussian process limit when the layers' width tends to infinity. Underpinning this result is the fact that hidden units become independent in the infinite-width limit. Our aim is to shed some light on hidden units dependence properties in practical finite-width Bayesian neural networks. In addition to theoretical results, we assess empirically the depth and width impacts on hidden units dependence properties.

## Title: Active Bayesian Causal Discovery

*Speaker: Julius von Kügelgen*

IS

Max Planck Institute for Intelligent Systems, Tübingen, Germany & University of Cambridge, Cambridge, United Kingdom

Observational causal discovery typically requires large amounts of data and can only recover the Markov equivalence class of graphs consistent with observation. On the other hand, experimental data facilitates more efficient learning and uniquely recovering the causal graph (subject to certain assumptions). In low-data regimes and safety-critical applications, a Bayesian approach is particularly attractive as it allows for uncertainty (over possible graphs and the parameters of the associated causal models) to be taken into account when making downstream predictions, e.g., about the effect of interventions on the rest of the system. I will present recent work in which a Bayesian agent (e.g., a robot or lab scientist) actively interacts with its environment through causal interventions on observed variables to learn a probabilistic causal model as efficiently as possible. In particular, we consider the class of nonlinear, additive Gaussian noise structural causal models (using Gaussian process priors over structural equations), employ a recently proposed differentiable graph parametrisation to scale our method to tens of nodes, and use Bayesian optimisation to select not only intervention targets but also intervention values by efficiently approximately maximising an estimate of expected information gain.

## **Title: Colombian Women's Life Patterns: A Multivariate Density Regression Approach**

**Speaker: Sara Wade**

IS

University of Edinburgh

Women in Colombia face difficulties related to the patriarchal traits of their societies and well-known conflict afflicting the country since 1948. In this critical context, our aim is to study the relationship between baseline socio-demographic factors and variables associated to fertility, partnership patterns, and work activity. To best exploit the explanatory structure, we propose a Bayesian multivariate density regression model, which can accommodate mixed responses with censored, constrained, and binary traits. The flexible nature of the models allows for nonlinear regression functions and non-standard features in the errors, such as asymmetry or multi-modality. The model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of categorical and continuous covariates. Computational difficulties for inference are overcome through an adaptive truncation algorithm combining adaptive Metropolis-Hastings and sequential Monte Carlo to create a sequence of automatically truncated posterior mixtures. For our study on Colombian women's life patterns, a variety of quantities are visualised and described, and in particular, our findings highlight the detrimental impact of family violence on women's choices and behaviors.

## **Title: Bayesian consistency with various metrics; including the Wasserstein**

**Speaker: Stephen Walker**

IS

University of Texas at Austin

Traditional approaches to Bayesian consistency rely on sieves, measures of entropy and the suitable allocation of prior mass. Consistency is then typically obtained for the weakest of the strong metrics; i.e. the Hellinger distance. This talk will investigate an alternative strategy which is based on inequalities between metrics, including the Wasserstein, the Kolmogorov and the sup metrics. The smoothing of density functions also plays a role, using, for example, the Fourier integral theorem. The aim is to explicitly use weak convergence to establish stronger forms of convergence using the inequalities.

## **Title: Bayesian Model Assessment for Jointly Modeling Multidimensional Response Data With Application to Computerized Testing**

*Speaker: Xiaojing Wang*

IS

University of Connecticut

Computerized assessment provides rich multidimensional data including trial-by-trial accuracy and response time (RT) measures. A key question in modeling this type of data is how to incorporate RT data, for example, in aid of ability estimation in item response theory (IRT) models. To address this, we propose a joint model consisting of a two-parameter IRT model for the dichotomous item response data, a log-normal model for the continuous RT data, and a normal model for corresponding paper-and-pencil scores. Then, we reformulate and reparameterize the model to capture the relationship between the model parameters, to facilitate the prior specification, and to make the Bayesian computation more efficient. Further, we propose several new model assessment criteria based on the decomposition of deviance information criterion (DIC) the logarithm of the pseudo-marginal likelihood (LPML). The proposed criteria can quantify the improvement in the fit of one part of the multidimensional data given the other parts. Finally, we have conducted several simulation studies to examine the empirical performance of the proposed model assessment criteria and have illustrated the application of these criteria using a real dataset from a computerized educational assessment program.

## **Title: Leveraging Historical Data via the Marginal Likelihood Criterion**

*Speaker: Yu-Bo Wang*

IS

School of Mathematical and Statistical Sciences, Clemson University, Clemson

One of desirable features of Bayesian analysis is that an end user can incorporate the available historical data into the prior setting to attain more precise analysis. Among many historical data-driven priors, the power prior that allows for different amount of information borrowing is attractive and has become increasingly popular in many applications. The determination of how much information can be leveraged from the historical data relies on comparability between the current and historical data. The marginal likelihood that measures the overall model fit to the data has been used to determine the choice of the discounting parameter in the use of the power prior under a simple model such as the normal linear regression model. However, the utilization of the marginal likelihood for a complex model has been woefully hindered by the analytical intractability of the complicated posterior kernel. To circumvent this computational challenge, we apply the recent developments of the partition weighted kernel (PWK) estimation and its variations to the power prior framework. With this powerful computation technique, even separately leveraging multiple historical data sets via the marginal likelihood becomes computation scalable. We illustrate how our proposed strategies are useful in determining the power priors in a benchmark dose toxicology study and a clinical trial for the development of anti-depression drugs.

## **Title: Data Augmentation for Bayesian Deep Learning**

**Speaker: Yuexi Wang**

CT

University of Chicago

Deep Learning (DL) methods have emerged as one of the most powerful tools for functional approximation and prediction. While the representation properties of DL have been well studied, uncertainty quantification remains challenging and largely unexplored. Data augmentation techniques are a natural approach to provide uncertainty quantification and to integrate MCMC stochastic search with stochastic gradient descent methods currently used in DL. Our paper shows that training DL architectures with data augmentation can lead to efficiency gains. To demonstrate our methodology, we develop data augmentation algorithms for a variety of commonly used activation functions: logit, ReLU and SVM. We compare our methods with traditional stochastic gradient descent with back-propagation. Our optimization procedure leads to a version of iteratively re-weighted least squares and can be implemented at scale with accelerated linear algebra methods providing substantial performance improvement. We illustrate our methodology on a number of standard datasets. Finally, we conclude with directions for future research.

## **Title: Sound the alarm: modeling behavioral changes in response to epidemic intensity**

**Speaker: Caitlin Ward**

CT

University of Calgary

For many infectious disease outbreaks, the at-risk population changes their behavior in response to the outbreak severity, changing the transmission dynamics to change in real-time. Various approaches to behavioral change modeling have been proposed, but work assessing the statistical properties of these models is limited. 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 Montreal, Canada.

## **Title: Bayesian projection of international migration flows**

**Speaker: Nathan Welch**

IS

University of Washington

We propose a method for forecasting global human migration flows. A Bayesian hierarchical model is used to make probabilistic projections of the 39,800 bilateral migration flows among the 200 largest countries from 2020 through 2045. The model is fit to estimates of quinquennial bilateral migration flows from 1990 through 2020. We find that the model produces well calibrated out-of-sample forecasts of bilateral flows, as well as total country-level inflows, outflows, and net flows. The mean absolute error decreased by 61% using our method compared to a leading model of international migration. Out-of-sample analysis indicates that simple models of migration flows offer accurate projections of bilateral migration flows in the near term. Our method matches or improves on the out-of-sample performance using these simple deterministic alternatives, while also accurately assessing uncertainty. Finally, we adapt a fully probabilistic population projection model to generate bilateral migration flow forecasts by age and sex for the 200 most populous countries.

## **Title: Subjectivist Model Uncertainty Analysis for Prediction and Decisions**

**Speaker: Mike West**

BF

Duke University

Bruno de Finetti (in his 1951 paper at the 2nd Berkeley Symposium on Mathematical Statistics and Probability) had the view that "... from the subjective standpoint, no assertion is possible without a priori opinion, but the variety of possible opinions makes problems depending on different opinions interesting."

A light adaptation to current interests emphasises the variety of possible "inferences and decisions" that makes problems depending on different "models" interesting.

Questions of model assessment, calibration, comparison and combination define continuing conceptual and practical challenges to all areas of quantitative analysis. Recent developments in Bayesian thinking have advanced methodology in areas including forecast distribution combination with defined predictive goals, while also highlighting foundational questions on the scope of Bayesian model uncertainty analysis as it is traditionally understood. A renewed focus on foundational questions is emphasising both the inferential "Yin" and the decision analytic "Yang" of Bayesian analysis, with resulting methodological implications. Linking to such perspectives underlying de Finetti's development and promotion of subjective Bayesian reasoning, the broader conversation explicitly addresses both predictive and decision analytic goals in the model uncertainty, evaluation and synthesis enterprise.

## Title: Myths and Reality in Bayesian Deep Learning

Speaker: Andrew Wilson

IS

New York University

Bayesian inference makes more sense for modern neural networks than virtually every other model class, because these models can represent many compelling and complementary explanations for data, corresponding to different settings of their parameters. However, a number of myths have emerged about modern Bayesian deep learning. In this talk we will evaluate the following questions: (1) is Bayesian deep learning practical? (2) are standard (e.g. Gaussian) priors arbitrary and poor? (3) is "deep ensembles" a non-Bayesian competitor to standard approximate inference approaches? (4) does the common practice of posterior tempering, leading to "cold posteriors", mean the Bayesian posterior is poor? I will also discuss the success stories, future opportunities, and challenges in Bayesian deep learning.

Key references: [1] Bayesian Deep Learning and a Probabilistic Perspective of Generalization (<https://arxiv.org/abs/2002.08791>) [2] What are Bayesian Neural Network Posteriors Really Like? (<https://arxiv.org/abs/2104.14421>) [3] Dangers of Bayesian Model Averaging under Covariate Shift (<https://arxiv.org/abs/2106.11905>) [4] On Uncertainty, Tempering, and Data Augmentation in Bayesian Classification (<https://arxiv.org/abs/2203.16481>) [5] Bayesian Model Selection, the Marginal Likelihood, and Generalization (<https://arxiv.org/abs/2202.11678>)

## Title: An Equivalence between Bayesian Priors and Penalties in Variational Inference

Speaker: Pierre Wolinski

CT

Inria

In machine learning, it is common to optimize the parameters of a probabilistic model, modulated by an ad hoc regularization term that penalizes some values of the parameters. Regularization terms appear naturally in Variational Inference (VI), a tractable way to approximate Bayesian posteriors: the loss to optimize contains a Kullback–Leibler divergence term between the approximate posterior and a Bayesian prior. We fully characterize which regularizers can arise this way, and provide a systematic way to compute the corresponding prior. This viewpoint also provides a prediction for useful values of the regularization factor in neural networks. We apply this framework to regularizers such as L2, L1 or group-Lasso.

## **Title: Properties of the bridge sampler with a focus on splitting the MCMC sample**

**Speaker: Jackie Wong**

CT

University of Essex

Computation of normalizing constants is a fundamental problem in various disciplines, particularly in Bayesian model selection. A sampling-based technique known as bridge sampling (Meng and Wong 1996) has been found to produce accurate estimates with good asymptotic properties. For small to moderate sample, however, we demonstrate that the (optimal) bridge sampler produces biased estimates. Specifically, when one density is constructed to be close to the target density, our simulation-based results indicate that the correlation-induced bias is non-negligible and grows with dimensionality. We show how the simple adaptive approach we termed "splitting" alleviates the bias at the expense of higher standard errors, irrespective of dimensionality. We then modify the approach by Wang et al. (2019) to address the higher standard error due to splitting, which is later generalized. We conclude the talk by applying a combination of these adaptive methods to improve bridge sampling estimates in a practical example, and offer recommendations for Bayesian applications.

## **Title: Graph Laplacian based Gaussian processes on restricted domains**

**Speaker: Nan Wu**

IS

Department of Mathematics, Duke University

In nonparametric regression, it is common for the inputs to fall in a restricted subset of Euclidean space. Typical kernel-based methods that do not take into account the intrinsic geometry of the domain across which observations are collected may produce sub-optimal results. In this talk, we focus on solving this problem in the context of Gaussian process models, proposing a new class of graph Laplacian based Gaussian processes, which learn a covariance that respects the geometry of the input domain. The graph Laplacian is constructed from a kernel depending only on the Euclidean coordinates of the inputs. The covariance is defined through finitely many eigenpairs of the graph Laplacian. When the subset is a submanifold of Euclidean space, we show that the covariance incorporates the intrinsic geometry of the manifold by relating it to the heat kernel. The graph Laplacian based Gaussian process is computationally efficient, as we can benefit from the full knowledge about the kernel to extend the covariance structure to newly arriving samples by a Nystrom type extension. We show the performance of our method in various applications. This talk is based on the joint work with David Dunson and Hau-Tieng Wu.

## Title: Tree-Informed Bayesian Multi-Source Domain Adaptation

Speaker: Zhenke Wu

IS

University of Michigan, Ann Arbor

Determining causes of deaths (COD) occurred outside of civil registration and vital statistics systems is challenging. A technique called verbal autopsy (VA) is widely adopted to gather information on deaths in practice. A VA consists of interviewing relatives of a deceased person about symptoms of the deceased in the period leading to the death, often resulting in multivariate binary responses. While statistical methods have been devised for estimating the cause-specific mortality fractions (CSMFs) for a study population, continued expansion of VA to new populations (or “domains”) necessitates approaches that recognize between-domain differences while capitalizing on potential similarities. In this talk, we propose such a domain-adaptive method that integrates external between-domain similarity information encoded by a pre-specified rooted weighted tree. Given a cause, we use latent class models to characterize the conditional distributions of the responses that may vary by domain. We specify a logistic stick-breaking Gaussian diffusion process prior along the tree for class mixing weights with node-specific spike-and-slab priors to pool information between the domains in a data-driven way. Posterior inference is conducted via a scalable variational Bayes algorithm. Simulation studies show that the domain adaptation enabled by the proposed method improves CSMF estimation and individual COD assignment. We also illustrate and evaluate the method using a validation data set. The talk concludes with a discussion on limitations and future directions.

## Title: Can we do better than Spike-and-Slab? — New theory about L1-ball priors in variable selection

Speaker: Maoran Xu

CT

University of Florida

Spike-and-slab priors have been viewed as the "gold standard" for variable selection in high dimensional regression. Although some optimality has been shown for signal recovery when both  $n$  and  $p$  diverge, in practice under a large but finite  $p$  (with  $p \gg n$ ), we found that the shrinkage effect of spike-and-slab is often overly-aggressive, causing elimination of true signals in estimates. In this study, we conduct theoretical analysis on the alternative, "L1-ball prior", whose sparsity is induced by projecting a continuous precursor random variable onto the surface of L1-ball. We show that when the precursor is allowed to be distributed anisotropically, the induced L1-ball prior gains substantial adaptiveness to different scales of signals. This improvement leads to a much superior signal detection limit, compared to common spike-and-slab priors using iid Bernoulli inclusions. We demonstrate practical advantages of these L1-ball priors in both simulations and data applications, where we obtain a much-improved balance between controlling false discovery rate and maintaining high statistical power.

## **Title: Nonparametric Bayesian Estimation of Heterogeneous Causal Effects Using Real-World Data**

*Speaker: Xinyi Xu*

IS

Ohio State University

Inferring a causal relationship is an important task in social science and health research. In a large population, different subgroups might respond differently to certain treatments. Identifying and estimating the heterogeneous effects can help researchers improve treatments or better allocate resources to meet the needs. In studies with real world data, propensity score is often used as a dimension reduction tool to aid the inference under the ignorable treatment assignment assumption. In this work, we develop a semiparametric Bayesian approach for efficient heterogeneous causal effect estimation in observational studies. Our model incorporate propensity scores and observable factors into potential outcome models via flexible Gaussian process regressions. We show that our model produces estimators outperform several popular causal methods with improved efficiency and better identification of heterogeneous effects. Furthermore, we apply our method to investigate the impact of college attendance on women fertility, which is known to suffer from the potential heterogeneous effects.

## **Title: Bayesian Sparse Mixture Models in High Dimensions**

*Speaker: Yanxun Xu*

IS

Johns Hopkins University

We propose a Bayesian method to estimate high-dimensional Gaussian mixture models whose component centers exhibit sparsity using a continuous spike-and-slab shrinkage prior. The number of components is not required to specify a priori and can be adaptively estimated from posterior inference. We establish the minimax risk for the parameter estimation in sparse Gaussian mixture models and show that the posterior contraction rate of the proposed Bayesian model is minimax optimal. We also obtain a contraction rate for the misclustering error as a by-product by using tools from matrix perturbation theory. Computationally, posterior inference can be implemented via a standard Gibbs sampler with data augmentation, circumventing the challenging frequentist nonconvex optimization-based algorithms. The validity and usefulness of the proposed approach is demonstrated through simulation studies and the analysis of a real-world single-cell sequencing data.

## **Title: Towards the Next Generation of Artificial Intelligence with its Applications in Practice**

**Speaker: Hongxia Yang**

IS

Alibaba DAMO Academy

Artificial intelligence has reached or surpassed human standards in the perceptual intelligence fields such as “listening, speaking, and seeing”, but it is still in its infancy in the field of cognitive intelligence that requires external knowledge, logical reasoning, or domain migration. After long-term exploration and verification of large-scale online businesses such as Taobao and Alipay, we have built a “super brain” through the extremely large scale pre-training model M6, and built “flexible limbs” through the edge-cloud collaboration platform Gemini, to fully present the comprehensive picture of the next generation of AI. The 10 trillion M6 is currently the world’s largest pre-training model, achieving the industry’s ultimate low-carbon efficiency. Compared to GPT-3, M6 uses only 1% of the energy consumption which greatly promotes the development of ubiquitous AI. This year, M6 supported the world’s largest shopping festival Double 11 for the first time. Relying on its multi-modal understanding ability, M6 greatly improved search and recommendation accuracy of Taobao and Alipay; with its smooth writing ability, M6 created scripts and copywriting for Tmall virtual streamers; depending on its generated high-definition images, M6 has been on duty in Rhino Intelligent Manufacturing and increased the efficiency of designers by 5 times. We developed Gemini, the industry’s first edge-cloud collaborative learning framework with the definition of five paradigms: cloud, edge, cloud-centric, edge-centric, and edge-cloud collaboration learning modes. It not only protects data privacy, but also has outstanding performances in personalized recommendations. This talk we will include details of state-of-the-art recommendation system in practice, M6 and Gemini.

## **Title: Variational inference for cutting feedback in misspecified models**

**Speaker: Xuejun Yu**

IS

National University of Singapore

Bayesian analyses combine information represented by different terms in a joint Bayesian model. When one or more of the terms is misspecified, it can be helpful to restrict the use of information from suspect model components to modify posterior inference. This is called “cutting feedback”, and both the specification and computation of the posterior for such “cut models” is challenging. In this paper, we define cut posterior distributions as solutions to constrained optimization problems, and propose optimization-based variational methods for their computation. These methods are faster than existing Markov chain Monte Carlo (MCMC) approaches for computing cut posterior distributions by an order of magnitude. It is also shown that variational methods allow for the evaluation of computationally intensive conflict checks that can be used to decide whether or not feedback should be cut. Our methods are illustrated in a number of simulated and real examples, including an application where recent methodological advances that combine variational inference and MCMC within the variational optimization are used.

## **Title: Robust leave-one-out cross-validation for high-dimensional Bayesian models**

**Speaker: Giacomo Zanella**

CT

Bocconi University

Leave-one-out cross-validation (LOO-CV) is a popular method for estimating out-of-sample predictive accuracy. However, computing LOO-CV criteria can be computationally expensive due to the need to fit the model multiple times. In the Bayesian context, importance sampling provides a possible solution but classical approaches can easily produce estimators whose variance is infinite, making them potentially unreliable. Here we propose and analyze a novel mixture estimator to compute Bayesian LOO-CV criteria. Our method retains the simplicity and computational convenience of classical approaches, while guaranteeing finite variance of the resulting estimators. Both theoretical and numerical results are provided to illustrate the improved robustness and efficiency. The computational benefits are particularly significant in high-dimensional problems, allowing to perform Bayesian LOO-CV for a broader range of models.

## **Title: Spatial scale-aware tail dependence modeling for high-dimensional spatial extremes**

**Speaker: Likun Zhang**

IS

Lawrence Berkley National Laboratory

Extreme events over large spatial domains like the contiguous United States may exhibit highly heterogeneous tail dependence characteristics, yet most existing spatial extremes models yield only one dependence class over the entire spatial domain. To accurately characterize 'storm dependence' in analysis of extreme events, we propose a mixture component model that achieves flexible dependence properties and allows truly high-dimensional inference for extremes of spatial processes. We modify the popular random scale construction that multiplies a Gaussian random field by a single radial variable; that is, we add non-stationarity to the Gaussian process while allowing the radial variable to vary smoothly across space. As the level of extremeness increases, this single model exhibits both long-range asymptotic independence and short-range weakening dependence strength that leads to either asymptotic dependence or independence. Under the assumption of local stationarity, we make inference on the model parameters using local Bayesian hierarchical models, and run adaptive Metropolis algorithms concurrently via parallelization. Then, after conducting posterior inference locally, the mixture component representation of the model coherently ties the local posteriors together to obtain a globally nonstationary model.

## Title: Bayesian Design Uncertainty in Two-Stage Propensity Score Analysis

Speaker: Corwin Zigler

IS

University of Texas at Austin

Propensity score methods are frequently framed in terms of a “design stage” where propensity scores (PSs) are estimated and implemented to approximate a randomized experiment and an “analysis stage” where treatment effects are estimated conditional on the design. Here we consider a Bayesian perspective on how uncertainty associated with the design stage impacts estimation of causal effects in the analysis stage. Such design uncertainty can derive from the fact that the PS itself is an estimated quantity, but also from other features of the design stage tied to choice of PS implementation. We offer a procedure for obtaining the posterior distribution of causal effects after marginalizing over a distribution of design-stage outputs, the details of which will depend on both how the PS is implemented in the design stage and how the treatment effect is estimated in the analysis stage. Instances where the analysis does not depend on a parametric likelihood (e.g., matching or weighted estimators) are particularly complicated for Bayesian inference. We consider a sample of commonly used PS implementations to assess Bayesian design uncertainty for a variety of design-related issues such as confounder selection, propensity score pruning for overlap, matching uncertainty, and attribution of treatment effect uncertainty to notions of between- and within-design variability. Methodology is illustrated in investigations of the association between levels of ambient air pollution and elevated exposure to emissions from power plants.