Composite Likelihood Methods
April 22-27, 2012

MEALS

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

MEETING ROOMS

All lectures will be held in the new lecture theater in the TransCanada Pipelines Pavilion (TCPL). LCD projector and blackboards are available for presentations.

SCHEDULE

Sunday	**April 22**
16:00	Check-in begins (Front Desk — Professional Development Centre — open 24 hours) Breakout rooms in the TransCanada Pipelines Pavilion are available from 16.00-22.00; note that the main lecture room (201) remains locked until Monday morning
17:30–19:30	Buffet Dinner, Sally Borden Building
20:00	Informal gathering in 2nd floor lounge, Corbett Hall (if desired) Beverages and a small assortment of snacks are available on a cash honor system.

For presentations, please allow 5 minutes for questions at the end.

Monday **April 23**

7:00–8:45 Breakfast

8:45–9:00 Introduction and Welcome by BIRS Station Manager, TCPL

9:00-10:15 **Opening session**
Chair: David Firth, University of Warwick

9:00 Cristiano Varin, Ca' Foscari University
Overview of composite likelihood methods

9:40 Alessandra Salvan and Nicola Sartori, University of Padova
Adjusting pseudo likelihood ratios for a parameter of interest

10:15–10:45 Coffee Break, TCPL

10:45–11:30 **Session 2**
Chair: Nancy Reid, University of Toronto

10:45 Bruce Lindsay, Pennsylvania State University
Sensible construction of composite likelihoods, with two case studies

11:30–13:00 Lunch

13:00–14:00 Guided Tour of The Banff Centre; meet in the 2nd floor lounge, Corbett Hall

14:00–15:00 **Session 3: Discussions to promote interactions**
Chair: Peter Song, University of Michigan
Handling high-dimensional data analysis and presenting background of application areas. What are some challenges for statistical modelling/inference in the themed areas such as spatial statistics, genomics, psychometrics, latent variable modelling.
Brief presentations welcomed from those looking for research collaborations and other expertise.

15:00-15:30 Coffee Break, TCPL

15:30-17:15 **Session 4: Spatial statistics**
Chair: Richard Davis, Columbia University

15:30 Jian Kang, Emory University
Accounting for spatial temporal correlation in fMRI data analysis using a composite likelihood approach

16:05 Yun Bai, University of Michigan
Joint composite estimating functions in spatio-temporal models

16:40 Marc Genton, Texas A&M University
On the likelihood function of Gaussian max-stable processes

17:30–19:30 Dinner

Tuesday **April 24**

7:00–9:00 Breakfast

9:00-10:10 **Session 5**

Chair: Kanti Mardia, University of Leeds

9:00 Xin Gao, York University

The composite likelihood estimation of high-dimensional Gaussian graphical models with symmetry

9:35 Subhash Lele, University of Alberta

Data cloning based estimability diagnostics and model selection for composite likelihood methods: Theory with application in modeling ecological communities

10:10–10:40 Coffee Break, TCPL

10:40-12:00 **Session 6: Psychometrics**

Chair: Vassilis Vasdekis, Athen University of Economics and Business

10:40 Irini Moustaki, London School of Economics

Composite likelihood estimation in models with latent variables and random effects

11:15 Albert Maydeu-Olivares, University of Barcelona

Limited information methods in psychometrics

11:50 Vassilis Vasdekis

Discussion

12:00–13:30 Lunch

13:30-14:40 **Session 7**

Chair: Cristiano Varin, Ca' Foscari University

13:30 Chi Tim Ng, Chinese University of Hong Kong

Information criteria and model comparisons with composite likelihood

14:05 Helen Jordan, University of Warwick

Robustness properties of marginal composite likelihood estimators

14:40–15:10 Coffee Break, TCPL

15:10 Group Photo; meet in the foyer of TCPL

15:30-17:30 **Session 8: Spatial extremes**

Chair: Anthony Davison, École Polytechnique Fédérale de Lausanne

15:30 Mathieu Ribatet, Université Montpellier II

Bayesian inference from composite likelihoods, with an application to spatial extremes

16:05 Simone Padoan, University of Padova

Extreme dependence models based on event magnitude

16:40 Raphael Huser, École Polytechnique Fédérale de Lausanne

Modelling of extreme rainfall in space and time

17:15 Anthony Davison

Discussion

17:30–19:30 Dinner

Wednesday **April 25**

7:00–9:00 Breakfast

9:00-12:00 **Session 9. Genetics/genomics**

Chair: Kung-Yee Liang, Johns Hopkins University

9:00 Fabrice Larribe, Université du Québec à Montréal

Composite likelihood in population genetics

9:45 Yun-Hee Choi, University of Western Ontario, and Laurent Briollais, Mount Sinai Hopsital

An EM composite likelihood approach for family studies with incomplete genetic data

10:30–11:00 Coffee Break, TCPL

11:00 Peter Xuekun Song, University of Michigan

Composite likelihood EM algorithm with application to gene network construction

11:45 Laurent Briollais

Discussion

12:00–13:30 Lunch

Free Afternoon

17:30–19:30 Dinner

Thursday **April 26**

7:00–9:00 Breakfast

9:00-10:10 **Session 10**

Chair: Annamaria Guolo, University of Verona

9:00 Geert Molenberghs, Hasselt University

Doubly robust pseudo-likelihood estimation for incomplete data

9:35 Paolo Vidoni, University of Udine

Prediction in a composite likelihood framework

10:10-10:40 Coffee Break, TCPL

10:40-11:50 **Session 11**

Chair: Alex de Leon, University of Calgary

10:40 Chandra Bhat, University of Texas

The estimation of non-normally mixed multinomial probit models using the Maximum Approximate Composite Marginal Likelihood (MACML) approach

11:15 Dimitris Karlis, Athens University of Economics and Business

Composite likelihood methods for discrete valued time series

12:00–13:30 Lunch

13:30-14:40 **Session 12**

Chair: Don Fraser, University of Toronto

13:30 Grace Yun Yi, University of Waterloo

A composite likelihood approach for analysis of survey data with sampling weights incorporated under two-level models

14:05 Ruggero Bellio, University of Udine

Some experiments with the maximization by parts algorithm

14:40–15:10 Coffee Break, TCPL

15:10-17:30 **Session 13. Computing issues**

Chair: Harry Joe, University of British Columbia

15:10 Computer Software Demos

Demonstration of software (e.g., R packages) with composite likelihood and other estimation methods for complex models with spatial dependence, random effects, latent variables etc.

16:15 Computing Discussion

The purpose of this session is to learn from each other about computational implementations and numerical methods. Additional computational details for previous presentations can be made here.

17:30–19:30 Dinner

Friday	**April 27**
7:00–9:00	Breakfast
9:00-11:00	**Session 14. Conclusion**
	Chair: Nancy Reid, University of Toronto
9:00	*Summarization of the week, discussion of important research directions and open research problems*
	Comments are welcomed from researchers in theme areas, such as spatial statistics, genomics, latent variable modelling, computing.
11:30–13:30	Lunch
Checkout by	**12 noon.**

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

Composite Likelihood Methods
April 22-27, 2012

ABSTRACTS
(in alphabetic order by speaker surname)

Speaker: **Yun Bai** (University of Michigan)
Title: *Joint composite estimating functions in spatio-temporal model*
Abstract: Modeling of spatio-temporal processes has received considerable attention in recent statistical research. However, due to the high dimensionality of the data, the joint modeling of spatial and temporal processes presents a great computational challenge, in both likelihood-based and Bayesian approaches. We propose a joint composite estimating function (JCEF) approach to estimating spatio-temporal covariance structures. This substantially reduces the computational complexity and is more efficient than existing composite likelihood methods. The novelty of the proposed JCEF is rooted in the construction of three sets of estimating functions from spatial, temporal and spatio-temporal cross pairs, respectively, which results in over-identified estimating functions. Thus, we form a joint inference function in a spirit similar to Hansen's generalized method of moments. We show that under practical scenarios the proposed estimator is consistent and asymptotically normal. Simulation studies prove that our method performs well in finite samples. Finally, we illustrate the JCEF method by estimating the spatio-temporal dependence structure of airborne particulates (PM10) in the northeastern United States over a 32-month period.
Coauthors: Peter X.-K. Song and T.E. Raghunathan

Speaker: **Ruggero Bellio** (University of Udine)
Title: *Some experiments with the maximization by parts algorithm*
Abstract: Maximization by parts (MBP) algorithm is a numerical procedure for obtaining the maximum likelihood estimator (MLE) without using the second order derivative of the log-likelihood function. The main idea of the MBP algorithm is to take advantage of a working log-likelihood providing a consistent, albeit inefficient, estimator of the parameter.
We argue that the MBP algorithm has a prominent role in the set of available tools for MLE computation in complex models. This is true for the original formulation of the algorithm, and it can be strengthened by considering some suitable variants. In particular, in this talk we explore some modifications of the basic MBP algorithm formulation, that may include step-length determination and upgrade of the working Hessian matrix.
These ideas are illustrated by means of some experiments with models commonly used in empirical applications. In particular, we consider generalized linear models with crossed random effects, and exponential random graph models for network data.

Speaker: **Chandra Bhat** (University of Texas at Austin)
Title: *The estimation of non-normally mixed multinomial probit models using the Maximum Approximate Composite Marginal Likelihood (MACML) approach*
Abstract: The multivariate skew-normal distribution function is used to accommodate non-normal mixing in cross-sectional and panel multinomial probit (MNP) models. The combination of skew-normal mixing and the MNP kernel lends itself nicely to estimation using Bhat's (2011) maximum approximate composite marginal likelihood (MACML) approach. Simulation results for the cross-sectional case show that our proposed approach does well in recovering the underlying parameters, and also highlights the pitfalls of ignoring non-normality of the continuous mixing distribution when such non-normality is present. At the same time, the proposed model obviates the need to assume a pre-specified parametric distribution for the mixing, and allows the estimation of a very flexible, but still parsimonious, mixing distribution form.

Speaker: **Yun-Hee Choi and Laurent Briollais** (University of Western Ontario, Mount Sinai Hospital)
Title: *An EM composite likelihood approach for family studies with incomplete genetic data*
Abstract: Composite likelihood is an important inference method derived by multiplying a collection of individual component likelihoods (Varin et al., 2011). Recently, we developed an EM composite likelihood approach (Choi and Briollais, 2011) for multistage sampling of family data in genetic epidemiology. Family data can be considered as arising from a retrospective cohort study and analyzed using various likelihood methods. We are interested in modelling a time-to-event response among carriers and non-carriers of a specific gene mutation (i.e., the penetrance). We use a weighted composite likelihood for time to event data where the weights are related to the inverse sampling probability of the families, and the components of the composite likelihood are for completely observed units. Missing genotype data for a major gene are incorporated via the EM algorithm. Conditional on the major gene, the observations are assumed independent and a robust variance estimator is used to account for possible model misspecification. In our context, we showed that the resulting estimating equation obtained from the derivative of composite log-likelihood is an unbiased estimating equation. In addition, it has been shown that the EM retains its key properties when applied to composite likelihood (Gao and Song, 2011). In genetic epidemiology, the interest lies not only in the estimation of the major gene characteristics (allele frequency, penetrance) but also of residual familial correlation, an unobserved random effect (i.e., frailty), which could indicate the presence of additional genetic factors besides the major gene or shared environment within families. To make inference about the frailty parameter, we propose to extend our EM composite likelihood framework using a pairwise likelihood formulation to account for the residual correlation within pairs of relatives. The bivariate survival function for each pair is modelled through a gamma frailty model and the family contribution to the composite likelihood becomes a weighted product over pairs of relatives. We discuss optimal weights for this new composite likelihood approach and show its interest through an application to a familial study of early-onset breast cancer.

Speaker: **Xin Gao** (York University)
Title: *The composite likelihood estimation of high-dimensional Gaussian graphical models with symmetry*
Abstract: We discuss the composite likelihood estimation of high-dimensional Gaussian graphical models. When there is symmetry constraints on the concentration matrix or partial correlation matrix, the estimation can be computational intensive. The composite likelihood formulation offers an alternative formulation of the objective function to maximize and yield consistent estimators. When a sparse model is considered, the penalized composite likelihood analysis can yield estimates satisfying the symmetry and sparsity constraints and also possess the ORACLE property.
Coauthor: Helene Massam

Speaker: **Marc G. Genton** (Texas A&M University)
Title: *On the likelihood function of Gaussian max-stable processes*
Abstract: We derive a closed form expression for the likelihood function of a Gaussian max-stable process indexed by R^d at $p \leq d+1$ sites, $d \geq 1$. We demonstrate the gain in efficiency in the maximum composite likelihood estimators of the covariance matrix from $p = 2$ to $p = 3$ sites in R^2 by means of a Monte Carlo simulation study. We further investigate the use of a weighted composite likelihood approach for inference in spatio-temporal max-stable processes based on pairs, triples, and quadruples of observations. Specifically, we propose a weighting strategy based on the estimated covariance matrix of the parameter estimates of the max-stable process, which substantially reduces the computational cost and improves the efficiency of the composite likelihood estimators.
Coauthors: Yanyuan Ma and Huiyan Sang

Speaker: **Raphael Huser** (École Polytechnique Fédérale de Lausanne)
Title: *Modelling of extreme rainfall in space and time*
Abstract: Extreme values for complex settings, for example spatial or space-time data, have become increasingly important, and over the past few years there has been a burst of activity in statistical modelling

for them. The structure of multivariate extremal models implies that composite likelihood must be used for inference in most realistic cases, and there is a growing literature on its application to spatial extremes using block maxima, with data from different blocks treated as independent. Threshold modelling is more efficient, enables more detailed analysis of extremes and is widely used in applications, but requires a more sophisticated treatment, because there are no natural independent blocks [Coles 2001]. The work described in this talk makes three contributions to this area.
First, we develop composite likelihood inference for extremes of multivariate time series using a censored likelihood approach, and show that under suitable mixing conditions the resulting estimators are consistent and asymptotically normally distributed.
A crucial practical element is the selection of suitable pairs of observations, and a second contribution is to investigate the efficiencies of different choices of pairs for conventional time series, to give insight into the extreme context. We also look into the use of triplets in constructing the composite likelihood, and obtain results that complement those of Genton et al. 2011, for a more common and realistic class of models.
The third contribution is to apply the methods to the extremes of space-time rainfall in Western Switzerland, using max-stable processes [deHaan 1984, Schlather 2002]. We use the composite likelihood approach to fit a three-dimensional max-stable random set model for extreme rainfall. Although the model is rather simple, it requires a major computational effort to fit it, but appears to fit the data surprisingly well.

Speaker: **Helen Jordan** (University of Warwick)
Title: *Robustness properties of marginal composite likelihood estimators*
Abstract: To form a marginal composite likelihood, we do not need to know the joint distribution of all the components of the full observation vector. Instead, we only need to specify the marginal distributions of some subsets of those components. For example, the independence likelihood only requires specification of the marginal distribution of each component, and the pairwise likelihood only uses the distribution of each pair of components. If these marginal distributions are correctly specified, and the parameter of interest is identifiable from them, then the corresponding marginal composite likelihood estimator will be consistent, even under incorrect specification of the dependence between subsets. The maximum likelihood estimator need not be consistent under such a model misspecification.
It is tempting to conclude from this that a marginal composite likelihood estimator will always be at least as robust to model misspecification as the maximum likelihood estimator.
However, in some cases the misspecification may be such that it affects the marginal distributions as well as the dependence between those margins. I will discuss one such situation, in which various marginal composite likelihoods give inconsistent estimators, even though the full maximum likelihood estimator is consistent.

Speaker: **Jian Kang** (Emory University)
Title: *Spatial-temporal fMRI analysis using composite likelihood*
Abstract: Functional magnetic resonance imaging (fMRI) studies that collect a huge number of time series over a three-dimensional brain. It is very important to estimate the spatial-temporal correlation of fMRI signals, which helps researchers to get better understanding of human brain function and structure. Computationally, however, it is challenging to fit a spatial-temporal model based on a likelihood approach, since it requires a full specification of the probabilistic model and involves inversion of high-dimensional matrices. In this talk, we propose an efficient and computationally feasible approach to spatial-temporal fMRI analysis using conditional composite likelihood. Also, we develop a fast parameter estimation algorithm. We illustrate our method on simulation studies and a real data example.
Coauthors: Yun Bai and Peter X.K. Song

Speaker: **Dimitris Karlis** (Athens University of Economics and Business)
Title: *Composite likelihood methods for discrete valued time series*
Abstract: When working with discrete valued time series it is often difficult to maximize the likelihood of the data, and hence switching to some composite likelihood is challenging from both the theoretical and

numerical perspective. Two specific families of time series models for count data will be considered. The first family consists of parameter-driven models based on a latent Gaussian autoregressive process. This is a class of models for which the likelihood consists of a high-dimensional integral and hence composite likelihood can be considered as a mean to avoid such an integral. The second family of models consists of multivariate integer autoregressive models. In such models the conditional likelihood implies the usage of a multidimensional discrete distribution. Typically such a distribution is too complicated to be numerically tractable (e.g., case where the model is defined through copulas). Thus, we consider a composite likelihood approach to allow for estimation in such models. Simulation evidence on the efficiency loss will be discussed.

Speaker: **Fabrice Larribe** (Université du Québec à Montréal)
Title: *Composite likelihood in population genetics*
Abstract: Due to the dimension and the dependency structure of genetic data, composite likelihood methods have found their natural place in the statistical methodology involving such data. After a brief description of the type of data one encounters in population genetic studies, questions of interest concerning the main genetic parameters in population genetics will be presented, followed by an up-to-date review on how composite likelihoods have been used to estimate these parameters.

Speaker: **Subhash Lele** (University of Alberta)
Title: *Data cloning based estimability diagnostics and model selection for composite likelihood methods: Theory with application in modeling ecological communities*
Abstract: Data cloning is a simple computational approach for likelihood inference in general hierarchical models. One of the attractive features of data cloning is that it provides a simple diagnostic test for parameter estimability. We extend the use of data cloning to composite likelihood methods. Composite likelihood methods are used when usual likelihood function is difficult to compute. However, use of lower-dimensional marginal and conditional distributions to form the composite likelihood can lead to lack of parameter estimability. It also leads to difficulties in using information based model selection criteria. We show how data cloning can be used to diagnose possible lack of estimability when using composite likelihood methods. We also develop an approach to compute information-based model selection criteria. We illustrate our methods to conduct inference for ecological community models.
Coauthor: Peter Solymos

Speaker: **Bruce Lindsay** (Pennsylvania State University)
Title: *Sensible construction of composite likelihoods, with two case studies*
Abstract: The construction of a statistical model is as much art as science. As discussed in Lindsay et al. (2011), there is also at least as much art involved in the construction of a composite likelihood. In particular, as viewed from the perspective of a "true" model, designing efficient composite likelihoods is rather difficult. In this talk I will focus on two constructions that seem sensible.
In the first model, there is a two-way mixture model structure that creates computational headaches. We pair an "independence likelihood" for the columns with an "independence likelihood" for the rows to build a lower cost method that provides very nice results.
The second model is a nonparametric segmentation problem familiar in genomics. Viewed nonparametrically, it is clear how the selection of the composite likelihood determines which parameters are identifiable. A set of Markov Chain likelihoods is considered; it results in a hierarchy of composite likelihoods that depend on the order of the Markov Chain. There is also a nested method of estimation, so that one can adapt the method to a given limit on computational time.
Coauthors: Jianping Sun and Prabhani Kuruppumullage Don

Speaker: **Albert Maydeu-Olivares** (University of Barcelona)
Title: *Limited information methods in psychometrics*
Abstract: We provide an overview of limited information methods in Psychometrics and relate it to bivariate composite likelihood (BCL) estimation. Limited information methods are popular in psychometrics.

For continuous data, they are used to estimate systems of equations possibly with latent variables when normality assumptions are not made. Estimation proceeds by minimizing a quadratic form in residual means and covariances. They are also used to estimate latent variable models for discrete data (aka item response theory (IRT) models). The most popular method assumes discretized multivariate normality and proceeds in stages. First, thresholds and polychoric correlations are estimated separately. Model parameters are then estimated by minimizing a quadratic form in thresholds and polychoric correlations. The method compares well to full information maximum likelihood (FIML) when data are not too small and variables are not skewed. But it can only handle MCAR missing data and one model. BCL methods can be used to estimate models with a larger number of latent variables than FIML can, provided an underlying continuous distribution is assumed. We compare BCL, FIML and BCL to estimate IRT models.
Coauthor: Carl Falk

Speaker: **Geert Molenberghs** (Universiteit Hasselt, Diepenbeek, Belgium)
Title: *Doubly robust pseudo-likelihood estimation for incomplete data*
Abstract: In applied statistical practice, incomplete measurement sequences are the rule rather than the exception. Fortunately, in a large variety of settings, the stochastic mechanism governing the incompleteness can be ignored without hampering inferences about the measurement process. While ignorability only requires the relatively general missing at random assumption for likelihood and Bayesian inferences, this result cannot be invoked when non-likelihood methods are used. A direct consequence of this is that a popular non-likelihood-based method, such as generalized estimating equations, needs to be adapted towards a weighted version or doubly-robust version, when a missing at random process operates. So far, no such modification has been devised for pseudo-likelihood based strategies. We propose a suite of corrections to the standard form of pseudo-likelihood, to ensure its validity under missingness at random. Our corrections follow both single and double robustness ideas, and is relatively simple to apply. When missingness is in the form of dropout in longitudinal data or incomplete clusters, such a structure can be exploited towards further corrections. The proposed method is applied to data from a clinical trial in onychomycosis and a developmental toxicity study.
Reference: Molenberghs, G., Kenward, M.G., Verbeke, G., and Teshome Ayele, B. (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica*, 21, 187–206.

Speaker: **Irini Moustaki** (London School of Economics)
Title: *Composite likelihood estimation in models with latent variables and random effects*
Abstract: We will review various extensions and applications of composite likelihood estimation for models with latent variables and random effects. Full information maximum likelihood (FIML) can quickly become infeasible for models with many latent variables or random effects and in some model specifications even when the number of observed variables is bigger than five. Composite likelihood estimation methods and more specifically pairwise likelihood can in some cases offer reduction in estimation complexity and provide a feasible alternative to FIML methods. We have applied various alternatives of composite likelihood estimation in latent variable models for longitudinal ordinal variables, structural equation models for continuous and categorical variables and growth curve models. Simulated and real examples will be presented to illustrate how well the pairwise likelihood performs on its own and in comparison to existing methods such as FIML, WLS and Robust maximum likelihood. Emphasis will be given on parameter estimation, standard errors and goodness-of-fit tests and measures.

Speaker: **Chi Tim Ng** (Chinese University of Hong Kong)
Title: *Information criteria and model comparisons with composite likelihood*
Abstract: Information criteria for model selection are discussed under the framework of maximum composite likelihood. Based on (i) the scale invariance of weights and (ii) the probabilities of selecting the correct model, we compare Varin-Vidoni's information criterion that generalizes the AIC and Gao-Song's information criterion that generalizes the BIC. Several new information criteria are introduced. Comparisons are

made for the amount of agreement of the Varin-Vidoni composite likelihood information criterion and AIC when choosing a simpler model nested within a bigger model, under a sequence of local alternatives to the simpler model. The behaviors of information criteria are illustrated via theory and simulation examples of the Gaussian mixed-effect model.
Coauthor: Harry Joe

Speaker: **Simone A. Padoan** (University of Padova)
Title: *Extreme dependence models based on event magnitude*
Abstract: We propose a new process model suitable for modelling the dependence of spatial extremes. By considering pointwise maxima of independent stationary random processes with dependent Cauchy marginals, we obtain a new limiting process whose families of bivariate distributions have Fréchet marginals, interpolating between complete independence and complete dependence. The dependence structure of the process that emerges is characterized in being dependent on the magnitude of the events and in particular it decreases with intensity increases. This is in contrast with one of the main feature of max-stable processes, namely the dependence is assumed constant for all the levels of a real process and even for events more extreme than those that have already occurred. In all practical situations where such a restrictive assumption is not verified, our new model is therefore useful to describe the dependence of extreme values. Of the proposed model we describe some properties of the dependence structure and we illustrate its utility in assessing the dependence. Combining marginal likelihoods we are able to estimate, by composite likelihood approach, the spatial dependence of the process. We show the good performance of the model by analysing the surge levels recorded along the cost of United Kingdom.

Speaker: **Mathieu Ribatet** (Université Montpellier II)
Title: *Bayesian Inference from Composite Likelihoods, with an Application to Spatial Extremes*
Abstract: Composite likelihoods are increasingly used in applications where the full likelihood is analytically unknown or computationally prohibitive. Although some frequentist properties of the maximum composite likelihood estimator are akin to those of the maximum likelihood estimator, Bayesian inference based on composite likelihoods is in its early stages. This talk discusses inference when one uses composite likelihood in Bayes' formula. We establish that using a composite likelihood results in a proper posterior density, though it can differ considerably from that stemming from the full likelihood. Building on previous work on composite likelihood ratio tests, we use asymptotic theory for misspecified models to propose two adjustments to the composite likelihood to obtain appropriate inference. We also investigate use of the Metropolis-Hastings algorithm and two implementations of the Gibbs sampler for obtaining draws from the composite posterior. We test the methods on simulated data and apply them to a spatial extreme rainfall dataset. For the simulated data, we find that posterior credible intervals yield appropriate empirical coverage rates. For the extreme precipitation data, we are able to both effectively model marginal behavior throughout the study region and obtain appropriate measures of spatial dependence.
Coauthors: Daniel Cooley and Anthony C. Davison

Speaker: **Alessandra Salvan and Nicola Sartori** (University of Padova)
Title: *Adjusting pseudo likelihood ratios for a parameter of interest*
Abstract: For inference about a parameter of interest in the presence of nuisance parameters, we consider a pseudo likelihood obtained from a genuine or composite likelihood by replacing the nuisance component with an estimate based on a generic estimating equation. Suitable adjustments are developed for the resulting pseudo likelihood ratio statistic, taking into account both nuisance estimation procedure and possible misspecification.

Speaker: **Peter Xuekun Song** (University of Michigan)
Title: *Composite likelihood EM algorithm with application to gene network construction*
Abstract: The method of composite likelihood provides an estimation procedure useful in high-dimensional parametric models that involve complicated dependency or hierarchical structures. In such cases, the full

likelihood approach is often computationally intractable. This talk will focus on an extension of the EM algorithm in the framework of composite likelihood estimation in the presence of missing data or latent variables. Several properties of the proposed algorithm will be discussed. Both analytic and empirical performances of the proposed algorithm are illustrated through simulation studies and real world data examples. In particular, a comparison of the new algorithm to the full likelihood EM algorithm. As a primary motivating example, the proposed algorithm will be applied to construct gene networks through multivariate hidden Markov models for time-course mRNA expressions.
Coauthor: Xin Gao

Speaker: **Cristiano Varin** (Ca' Foscari University)
Title: *Overview of composite likelihood methods*
Abstract: In recent years composite likelihood methods have received considerable interest in modelling data with complex forms of dependence. In this talk I will survey recent developments in the theory and application of composite likelihood methods. Special emphasis will be given to open questions and possible limitations of composite likelihood inference.
The talk is partially based on a recent review paper jointly written with Nancy Reid and David Firth.

Speaker: **Paolo Vidoni** (University of Udine)
Title: *Prediction in a composite likelihood framework*
Abstract: The notion of predictive likelihood relies on the likelihood principle for prediction and it corresponds to any function of the future observation, obtained by eliminating the model parameter from the joint likelihood. Usually, predictive likelihoods are evaluated by studying how well they generate point predictors, predictive densities and prediction intervals. Thus, from the frequentist viewpoint, this involves the evaluation of mean predictive errors of predictors, the specification of expected loss of predictive densities and the computation of coverage probabilities of prediction intervals and prediction limits.
Several problems, with the use of predictive likelihood, may arise when the joint distribution of the data, and consequently the joint likelihood function, is not available in a closed form and analytical approximations or computational intensive numerical procedures are not feasible. In this context, a viable solution could be to consider a suitable joint pseudo-likelihood as a starting point in order to specify a useful pseudo-predictive likelihood. In this talk we aim at specifying new prediction procedures based on pseudo-likelihood, which can be considered when the genuine notion of likelihood cannot be applied, due to computational or modelling problems. Among various notions of pseudo-likelihoods, we focus on composite likelihood, which in fact defines a rich class of simplified likelihoods, obtained as a weighted product of likelihood factors related to marginal or conditional events associated to the observed sample. More precisely, we concentrate on the evaluation of the frequentist properties of alternative pseudo-predictive likelihoods by considering mean predictive errors of predictors, expected loss of predictive densities and coverage probabilities of prediction limits.

Speaker: **Grace Y. Yi** (University of Waterloo)
Title: *A composite likelihood approach for analysis of survey data with sampling weights incorporated under two-level models*
Abstract: Multi-level models provide a convenient framework to handle survey data arising from complex sampling designs. Relative to well established inferential procedures for single-level models, however, there are limited methods that are valid for general multi-level models with design issues accommodated. We proposed a unified inference method to analyze survey data with multi-stage sampling probabilities. Our method has a broad applicability and is straightforward to implement. Extensive empirical studies demonstrate its satisfactory performance.
Coauthors: J.N.K. Rao and Haocheng Li

PARTICIPANTS

(in alphabetic order by surname)

Participant	**Affiliation**	**E-mail**
Bai, Yun	University of Michigan	yunbai@umich.edu
Bellio, Ruggero	Università di Udine	ruggero.bellio@uniud.it
Bhat, Chandra	University of Texas	bhat@mail.utexas.edu
Briollais, Laurent	Samuel Lunenfeld Research Institute	laurent@lunenfeld.ca
Carroll, Raymond	Texas A&M University	carroll@stat.tamu.edu
Choi, Yun-Hee	University of Western Ontario	ychoi97@uwo.ca
Davis, Richard	Columbia University	rdavis@stat.columbia.edu
Davison, Anthony	École Polytechnique Fédérale de Lausanne (EPFL)	Anthony.Davison@epfl.ch
de Leon, Alex	University of Calgary	adeleon@math.ucalgary.ca
Firth, David	University of Warwick	d.firth@warwick.ac.uk
Fraser, Don	University of Toronto	dfraser@utstat.utoronto.ca
Gao, Xin	York University	xingao@mathstat.yorku.ca
Genton, Marc	Texas A&M University	genton@stat.tamu.edu
Guolo, Annamaria	Università di Verona	annamaria.guolo@univr.it
Huser, Raphal	École Polytechnique Fédérale de Lausanne	raphael.huser@epfl.ch
Joe, Harry	University of British Columbia	Harry.Joe@ubc.ca
Jordan, Helen	University of Warwick	H.E.Jordan@warwick.ac.uk
Kang, Jian	Emory University	jian.kang@emory.edu
Karlis, Dimitris	Athens University of Economics and Business	karlis@aueb.gr
Larribe, Fabrice	Université du Québec à Montréal	larribe.fabrice@uqam.ca
Lele, Subhash	University of Alberta	slele@ualberta.ca
Liang, Kung-Yee	Johns Hopkins University	kyliang@jhsph.edu
Lindsay, Bruce	Pennsylvania State University	bgl@psu.edu
Liu, Juxin	University of Saskatchewan	liu@math.usask.ca
Ma, Yanyuan	Texas A&M University	ma@stat.tamu.edu
Mardia, Kanti	University of Leeds	K.V.Mardia@leeds.ac.uk
Maydeu-Olivares, Albert	Universitat de Barcelona	amaydeu@ub.edu
Molenberghs, Geert	Universiteit Hasselt	geert.molenberghs@uhasselt.be
Moustaki, Irini	London School of Economics	I.Moustaki@lse.ac.uk
Ng, Chi Tim	Chinese University of Hong Kong	easterlyng@gmail.com
Padoan, Simone	Università di Padova	simone.padoan@stat.unipd.it
Reid, Nancy	University of Toronto	reid@utstat.utoronto.ca
Ribatet, Mathieu	Université Montpellier II	mathieu.ribatet@math.univ-montp2
Salvan, Alessandra	Università di Padova	salvan@stat.unipd.it
Sartori, Nicola	Università di Padova	sartori@stat.unipd.it
Song, Peter Xuekun	University of Michigan	pxsong@umich.edu
Varin, Cristiano	Università Ca' Foscari	sammy@unive.it
Vasdekis, Vassilis	Athens University of Economics and Business	vasdekis@aueb.gr
Vidoni, Paolo	Università di Udine	paolo.vidoni@uniud.it
Wang, Naisyin	University of Michigan	nwangaa@umich.edu
Xu, Ximing	University of Toronto	ximing@utstat.toronto.edu
Yi, Grace Yun	University of Waterloo	yyi@uwaterloo.ca