Poster Session II

JUSTIN ANGEVAARE, University of Guelph
The Utility of Catch Per Unit Effort Variance

Population models play a central role in fisheries management. Commercial harvest data, especially catch per unit effort (CPUE), are used to predict population parameters and abundance. A survey of recent literature reveals that CPUE variability is typically ignored. The common practice is to aggregate harvest and effort by year. We present a simulation study to explore traditional models and those that have been modified to include CPUE variability. Daily fish populations and harvest events are simulated over 60 years. Catch and effort are aggregated over several temporal scales. Population level parameter estimates are compared across models and temporal aggregation scales.

SHAHEENA BASHIR, University Health Network, Toronto
Joint Tests of Effect Size in Multivariate Meta-Analysis

Multivariate meta-analysis combines estimates of several related parameters over several studies. This work investigates the feasibility of obtaining an F-statistic that can test for differences among more than two groups in a meta-analysis context. The question of obtaining a general F-statistic arose in a meta-analysis of gene expression microarray data with 4 treatment groups and 6 possible pairwise comparisons across 2 studies. We evaluate the performance of our test statistic in comparison to standard random effect meta-analysis approaches in simulations. The simulations show that the quadratic form of the pairwise t statistic from random effect model is distributed as F.

VICTORIA BORG DEBONO, McMaster University
The Quality of Reporting of Randomized Controlled Trials (RCTs) in Postoperative Pain Management

Good reporting of RCTs is important because it is a proxy for researchers to appraise RCT methodological rigor. The aims of this study were to assess the quality of reporting (QOR) in postoperative pain management RCTs – using the CONSORT criteria – and to determine factors associated with high QOR. Overall, the QOR of anesthesia trials was found to be poor to moderate, results which are consistent with observations from other clinical disciplines. The information gained from this can be used by journals to register the urgency for enhancing the reporting to make RCT findings accessible and comparable.

ZHENGFEI CHEN, University of Toronto
Riemann Manifold HMCMC for Log-Gaussian Cox Processes

Log Gaussian Cox Point Processes are a rich class of models for clustered point patterns, and are useful and popular for location data which is distributed inhomogeneously. Riemann Manifold Hamiltonian Monte Carlo method produce highly efficient sampling even in very high dimensions where different scaling may be required for the transient and stationary phases of the Markov chain. Some nice properties LGCP with Matern covariance using HMCMC will be explored, along with an application of RMHMC in LGCP with Matern covariance matrix on a high dimensional data set.

JING DING, Acadia University
Population Assessments by Using Age-at-harvest Data and Catch Effort

Age-at-harvest data and catch effort are readily available quantities in wildlife management. They provide essential information for modelling populations so as to reveal the situation of animals under exploitation. Three practicable methods are introduced. Downing (1980)'s reconstruction is a commonly used technique which uses backward addition through cohorts to reconstruct populations over time. A linear regression model incorporating catch effort, developed by Paloheimo (1980), is another applicable method. Stochastic catch effort method (Dupont, 1983) is the most advanced method since it requires weaker assumptions about the birth-death-capture process. Hazard models are assumed, and ordinary least-squares can be used for parameter estimation.

JAYDE EUSTACE, Memorial University of Newfoundland
A Gumbel State Space Model for Extreme-value Data

Extreme value theory has many applications in areas such as meteorology, finance and engineering. Recent years have seen a revival of interest in the analysis of series of extreme-value data. For this presentation, some properties of the mixtures of Gumbel and α-stable random variables will be used to construct a state-space model with Gumbel marginals. One feature of α-stable random variables is that their density cannot be expressed in closed form and their expectation does not exist. This poses some inference challenges. The methods proposed will be compared through simulation to currently used models using real data.

REMYA M. JOHN, McMaster University
Statistical Analysis of RNA-Seq Data: A Comparison of Three R Packages

Next generation sequencing of RNA (RNA-Seq) is a new tool which can be used for analysis of differential expression of genes. A number of R/Bioconductor packages have been developed in recent years to deal with such data. In this project we examine the performance of three widely used packages (DESeq, edgeR and DEGseq) on two datasets with biological replication and show that a Poisson model as used in DEGseq is not appropriate in the presence of biological replicates and that there is a lack of consistency between the other two methods even though their overall performance is similar.

XIAOQING NIU, University of Alberta
Testing Homogeneity in Normal Scale Mixture Models

Normal scale mixture models are a special type of the mixture models and have many applications. We propose the retooled EM-test for testing homogeneity in them. Two approaches are carried out. In one way, the estimation of the common mean is initialized by the sample mean. In the other way, to initialize, we maximize the penalized log-likelihood function. We show that the EM-test statistics asymptotically follow the simple distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. Simulation studies show that the EM-test has accurate type I errors, and the second approach enjoys higher power than the first. Two real data examples are analyzed.

MATTHEW RUEFFER, University of Guelph
Spatial and Temporal Modelling of Change in Marine Harvests

From 1979 until 2010, catch data on whitefish (Coregonus clupeaformis) has been collected for the Canadian portion of Lake Huron. This data, obtained from native fishers and the Ontario Ministry of Natural Resources, contains the geographic location, harvest, relative effort and mesh size. This poster will examine catch per unit effort (CPUE = the quotient of harvest and effort) across regions and years, using exploratory techniques, temporal and spatial models.

KATHLEEN RYAN, University of Guelph
Depth Distribution of Larval Lake Whitefish (Coregonus clupeaformis) in a Great Lakes Embayment

This study investigated the relationship between larval lake whitefish (Coregonus clupeaformis) density and water depth in Stokes Bay, Lake Huron. Nearshore waters (1-3 m) of the Great Lakes have been hypothesized to provide favourable habitat for larval whitefish, leading investigators to focus survey work there. The distribution of larvae in Stokes Bay was investigated by surface tows of 500 micron nets between April and May 2011. A novel stratified random sampling design assigned sample sites over depths ranging from 1-11m (n=71). Statistical analyses suggest depth was not significant in predicting larval fish density.

OLGA SHESTAKOVSKA, Population Health Research Institute, McMaster University and Hamilton Health Sciences
Large Hazard Ratios Do Not Always Mean High Predictive Accuracy: A Simulation Study

In survival analysis, hazard ratio is a measure of relative risk frequently used to assess the effect of demographics, clinical characteristics, and biomarkers on event rates. We explore the relationship between hazard ratio and measures of predictive accuracy such as Harrell's c-statistic and time-dependent ROC curves in a simulation study. Motivation came from the analysis of a risk score calculated to predict stroke in patients with atrial fibrillation. We observed a strong association between the score and event rates (HR=2.45; 95% CI=1.66-3.75) yet the Harrell's c-statistic was 0.567.

MONICA SIRSKI, University of Manitoba
On the Statistical Analysis of Functional Data Arising from Designed Experiments

We investigate various methods for testing whether two groups of curves are statistically significantly different, with the motivation to apply the techniques to the analysis of data arising from designed experiments. We propose a set of tests based on pairwise differences between individual curves. Our objective is to compare the power and robustness of a variety of tests, including a collection of permutation tests. We illustrate the application of these tests in the context of a designed 2^4 factorial experiment with a case study using data provided by NASA.

GARY SNEDDON, Mount Saint Vincent University
Panel Data Models and Electoral Success of Ex-Cabinet Ministers

Panel data models are of interest in a variety of disciplines where data are collected over time on cross-sections of individuals. We examine these models and estimation methods in studying data on the success of former Canadian federal cabinet ministers in subsequent elections. This will include an examination of some statistical approaches more commonly used in the social science literature.

MARY THIESS, Fisheries and Oceans Canada
Time Series Analysis: Assessing the "pink salmon" effect

Fraser River sockeye salmon have received a lot of press recently due to highly variable adult returns in recent years amid a general overall decline. One of many hypotheses to account for some of this variability suggests the potential for competition between sockeye salmon and pink salmon. Isolating the relationship between their respective time series requires the assessment (and quantification) of auto-correlation and non-monotonic trend over time. This poster will outline the challenges of assessing the "pink salmon" effect and highlight some approaches to analyzing these data.

MONICA H.T. WONG, University of Guelph
Cluster Analyses to Assess Weight Loss Maintenance: An Application of Clustering in Nutrigenomics

Within nutrigenomics, data generated by microarray gene expression profiles can be clustered to help identify sub-populations of subjects that respond differently to a dietary intervention. The use of clustering analyses in obesity-related research is burgeoning as personalized nutrition gains popularity. This study focuses on clustering a human adipose tissue gene expression dataset obtained during a low-calorie diet intervention to identify differential responders. The aims of the study were to identify the best performing clustering method for samples, pick out differential responders to the low-calorie diet, and help find the biological pathways affected during the low-calorie diet by the differential responders.

HAIYAN YANG, Memorial University of Newfoundland
Joint Modelling of Genetic Linkage and Association

A joint modelling of genetic linkage and association within families having a remote common ancestor is presented. This joint modelling uses a likelihood approach that allows the inclusion of other covariates into the model for quantitative traits; for binary traits, a score test is proposed. The simulation study show that the joint test is more powerful than linkage or association test alone where both sources of variation are present. Furthermore, the proposed method also allows to test against specific alternatives. For example, against the significance of linkage given that there is no association and so on.

LI HUA YUE, University of Western Ontario
Logistic Prediction on Prostate Cancer Status Using Imaging Data

New imaging techniques, such as CT and MRI, have become commonly available for diagnosis and management of prostate cancer. Accurately identifying prostate cancer is essential to guide the treatment therapy. With the pathological cancer status available in our study, we can build prediction models using imaging characteristics as predictors to predict prostate cancer status, and the model is expected to be used for future diagnosis of cancer using imaging tools. We show some preliminary results on the prediction of prostate cancer status using CT, T2WMR, and DCEMR features. The prediction accuracy is evaluated using AUC with a 10-fold cross-validation.

WEI YUHONG, University of Guelph
Mixture-model Averaging

Criteria commonly used for choosing the number of components in a finite mixture model include the Akaike information criterion, Bayesian information criterion, approximate weight of evidence, and the Integrated complete likelihood. We argue that it is not reasonable when the difference between the biggest value of such a criterion and some other values is small. It is, therefore, worth considering a model-averaging approach. We consider an averaging of the top M mixture models with applications in clustering and classification.