# ISM Symposium on Environmental Statistics 2014

Date & Time: Feb. 5(Wed), 2014 , 10:30~16:45
Venue: The Institute of Statistical Mathematics,
10-3 Midori-cho Tachikawa, Tokyo 190-8562, Japan
(2F Auditorium)

Speakers:
・**Alan Welsh(The Australian National University)**
・**Richard Huggins (The University of Melbourne)**
・**Anne Chao (National Tsing Hua University)**
・**Shuangzhe Liu(University of Canberra)**
・**Kunio Shimizu (Keio University)**
・**Hideyasu Shimadzu(University of St Andrews)**
・**Keiichi Fukaya (The Institute of Statistical Mathematics)**

Aim of Symposium:
In order to enhance the understanding of the global environment, statistical science is extremely important. Centered around the topic of directional statistics, we are holding a symposium in order to better develop research on statistical theory which can be applied to solve specific issues in the fields of environmental and ecological data.

Host Organization:The Institute of Statistical Mathematics

Organizing Committee:
Koji Kanefuji (ISM), Alan Welsh (ANU), Kenichiro Shimatani (ISM),
Atsushi Yoshimoto (ISM), Kunio Shimizu (Keio University)

ISM Symposium on Environmental Statistics 2014
 (Second Report)


Date & Time
Feb. 5 (Wed), 2014 10:00～17:30


Venue
2rd Floor Auditorium
 The Institute of Statistical Mathematics,
 10-3 Midori-cho Tachikawa, Tokyo 190-8562, Japan


Topics
In order to enhance the understanding of the global environment, statistical science is extremely important. Centered around the topic of directional statistics, we are holding a symposium in order to better develop research on statistical theory which can be applied to solve specific issues in the fields of environmental and ecological data.


Joint Hosting
Ministry of Education, Culture, Sports, Science and Technology (MEXT)


Organizers
Koji Kanefuji (ISM)
Alan Welsh (ANU)
Atsushi Yoshimoto (ISM)
Kunio Shimizu (Keio University)
Kenichiro Shimatani (ISM)

Program(Tentative)

10:30-10:40 Opening Address:

10:40-11:20 The prediction of random effects in hurdle models
        E. Cantoni*1, J. Mills Flemming*2 and A.H. Welsh*3
        (*1:The University of Geneva, *2:Dalhousie University, *3:The
        Australian National University)

11:20-12:00 Coverage-based rarefaction and extrapolation: standardizing
        samples by completeness rather than size
        Anne Chao (Institute of Statistics, National Tsing Hua
        University)

Lunch

13:00-13:40 Semivarying coefficient models for capture-recapture data
        :Colony size estimation for the little penguin
        Richard Huggins (Department of Mathematics and Statistics,
        The University of Melbourne)
13:40-14:20 Estimation of Markov transition probabilities for ecological
        communities using dynamic site occupancy models
        Keiichi Fukaya (The Institute of Statistical Mathematics)
14:20-15:00 Investigating species interactions in a fish community
        Hideyasu Shimadzu, Maria Dornelas, Peter A. Henderson, Anne
        E. Magurran(School of Biology, University of St Andrews)

Coffee Break

15:10-15:50   A new probability model for cylindrical data
        Kunio Shimizu(Department of Mathematics, Keio University)
15:50-16:30   Influence diagnostics in regression and time series models of
        circular data.
        Shuangzhe Liu(Department of Mathematics and Statistics,
        University of Canberra)

16:30-16:35 Closing address

[1]The prediction of random effects in hurdle models

E. Cantoni*1, J. Mills Flemming*2 and A.H. Welsh*3
*1:The University of Geneva, *2:Dalhousie University, *3:The Australian National University

For many endangered marine species, such as sharks, the only available data on their abundance are counts of when they are caught unintentionally in a fishery (that is, as bycatch). These data typically involve a larger number of zero counts (indicating that none were caught as bycatch in a particular haul, for example) and very few positive counts (obtained if one or more are caught as bycatch in a haul) and are clustered because hauls are clustered within trips which may also be clustered within vessels. We are therefore interested in fitting models to the data which accommodate the zeros and the clustering, and then using these models to properly answer important scientific questions such as predicting the probability of bycatch (and other cluster specific targets).

We will discuss the prediction of random effects and functions of random effects in the context of fitting hurdle models (also referred to as two-part, zero-altered or separated models) with random effects for modelling clustered count data with excess zeros. We implement empirical best predictors of the random effects and other cluster specific targets. We discuss estimating their prediction mean squared errors using a fast bootstrap approach. The methodology is validated through simulation and demonstrated using real data on critically endangered hammerhead sharks where the prediction of cluster specific targets is essential for informing conservation and management decisions.

[2]Coverage-based rarefaction and extrapolation: standardizing samples by completeness rather than size

ANNE CHAO
(Institute of Statistics, National Tsing Hua University)

In this talk, I will review a recent proposal of an integrated sampling, rarefaction, and extrapolation methodology to compare species richness of a set of communities based on samples of equal completeness (as measured by sample coverage) instead of equal size. The concept of "sample coverage" (or simply "coverage") was originally developed by the founder of modern computer science, Alan Turing, and I. J. Good. It is a measure of sample completeness, giving (in our context) the proportion of the total number of individuals in a community that belong to the species represented in the sample. Traditional rarefaction or extrapolation to equal-sized samples can misrepresent the relationships between the richnesses of the communities being compared, because a sample of a given size may be sufficient to fully characterize the lower-diversity community but insufficient to characterize the richer community. Thus the traditional method systematically biases the degree of differences between community richnesses. A new analytic method for seamless coverage-based rarefaction and extrapolation was recently developed by Chao and Jost (2012). I will use examples to show that this method yields less-biased comparisons of richness between communities, and manages this with less total sampling effort. Several hypothetical and real examples demonstrate these advantages. I will also briefly mention the extension of this new rarefaction/extrapolation method to other measures of biodiversity, including Shannon diversity (the exponential of entropy) and Simpson diversity (the inverse Simpson concentration).
(This is a joint work with Lou Jost)

[3]Semivarying coefficient models for capture-recapture data: Colony size
estimation for the little penguin
Richard Huggins
(Department of Mathematics and Statistics, The University of Melbourne)

To accommodate seasonal effects that change from year to year into models
for the size of an open population we consider a time-varying coefficient
model. We fit this model to a capture-recapture data set collected on the little
penguin in south-eastern Australia over a 25 year period, using Jolly-Seber
type estimators and nonparametric P-spline techniques. The time-varying
coefficient model identified strong changes in the seasonal pattern across the
years which we further examine using functional data analysis techniques.
(Joint work with Jakub Stoklosa of The University of New South Wales and
Peter Dann from the Phillip Island Nature Parks)

[4]Estimation of Markov transition probabilities for ecological communities using dynamic site occupancy models

Keiichi Fukaya

(The Institute of Statistical Mathematics)

Dynamics of ecological communities, especially of sessile organisms, have often been described by Markov models where changes in relative abundance of species are summarized by transition probability matrices. Transition probabilities can be estimated from time series data of species occupancy states collected at fixed sampling points, although a naive estimation of transition probabilities may not be robust when resampling error exists. Here I present a model based approach for the estimation of transition probabilities of the communities in which transition probabilities are estimated in the framework of multistate dynamic site occupancy model. The model explicitly describes both transitions among occupancy states and observation processes and hence takes the effect of resampling error in the estimation of transition probabilities into account. I also show that the resampling error rate can be estimated without any additional data other than the annual (or seasonal) data of occupancy states, suggesting the usefulness of this framework to estimate transition probabilities with limited field data.

[5]Investigating species interactions in a fish community

Hideyasu Shimadzu, Maria Dornelas, Peter A. Henderson, Anne E. Magurran
(School of Biology, University of St Andrews)

Understanding the interactions between species within an ecological community is a key challenge in biodiversity research. Here, we focus on monthly time series records of an exceptionally well-documented estuarine fish assemblage in the Bristol Channel. Given the multi-species time series data, we have developed a model for a multivariate feedback system in which the outputs can be the inputs and vice versa. The model assumes linear interactions between species as a tractable approximation. To examine the extent to which the abundance of a given species is driven by the other species, we have analysed the model in the spectrum domain and calculated the contribution ratio of each species at each frequency. The result suggests that our modelling approach dealing with an ecological community as a multivariate feedback system provides new insights into species interactions. We demonstrate how it enables further analysis into ecologically relevant groups of species that underpin the functioning of the system as a whole.

[6]A new probability model for cylindrical data

Kunio Shimizu
(Department of Mathematics, Keio University)

Environmental data such as wind speed and direction, ozone concentration and wind direction, and earthquakes magnitudes and successive turning angles of the epicenter are considered data on the cylinder, i.e. the dada consist of a combination of linear and circular observations. In the literature, probability distributions are introduced by Mardia and Sutton (1978, JRSS), Johnson and Wehrly (1978, JASA) and Kato and Shimizu (2008, JSPI) as models for cylindrical data. In this talk, we give a new probability distribution which is derived as a conditional distribution of a mixture of trivariate normal distributions. The distribution should be called the Pearson Type VII distribution on the cylinder, and includes the Mardia-Sutton and Kato-Shimizu cylindrical distributions as special cases. We study some distributional properties such as marginal and conditional distributions, modality, circular-linear correlation, regression, and so on.
(This is joint work with Shonosuke Sugasawa, The University of Tokyo, and Shogo Kato, The Institute of Statistical Mathematics.)

[7]Influence diagnostics in regression and time series models of circular data.

Shuangzhe Liu
(Department of Mathematics and Statistics, University of Canberra)

Distributional studies, regression and time series models have played important roles in statistical analysis of circular data. In this paper, we consider a likelihood approach to identify possible influential observations in circular data for these models. We use the maximum likelihood estimation and influence diagnostics methods. The observed information matrices and normal curvatures are derived. Simulated and real data examples are then provided to illustrate our approach and results to be useful.