

NOTES FOR LOUISE RYAN RE KEIO “DATA SCIENCE WORKSHOP”.

Workshop Title: ***Australia-Japan Workshop on Data Science***
24, 25 and 27th March, 2009
Yagami Camps, Keio University, Japan.

The workshop is partly supported by the Australia-Japan Foundation.

Presenters at the workshop include scientists and academics from CMIS, CMAR, Keio University, Institute of Statistical Mathematics Japan (IS), RIKEN, Tokyo Institute of Marine Science and Technology, Geoscience Australia, SRA Wellington (NZ), and Hitotsubashi University (Program below).

Links with between CMIS and Keio University

Professor Shibata visited Ted Hannan at ANU and met Murray Cameron. (not sure of year)

From 2002, Keio University succeeded to get a funding for “Centre of Excellence Programme” by Japanese Ministry of Education for five years at which Prof Shibata played a key role as a programme leader and organised subsequent workshops called “Cherry Bud Workshop”. From CMIS a number of scientists were invited to the workshop every year, Glen Stone attended first “Cherry Bud Workshop” at Keio in 2003. Other CMISers to attend

subsequent “Cherry Bud Workshops” include Ian Saunders, Pavel Shevchenko, Richard Jarrett, Charis Burrige, John Donnelly, Harry Kivili,

Dr Peter Thomson (SRA, NZ) who was also a student of Ted Hannan and senior of Murray, was a member of advisory board for the COE programme at Keio (Peter is also a good friend of Prof Shibata since 1980 or 1984), suggested to invite CMISer to the Cherry Bud Workshop when the plan of the workshop was proposed.

Dr Hideyasu Shimadzu, a student of Prof Shibata, now with GeoScience Australia is a collaborator in the CERF Hub project, a large marine species biodiversity project (WfO flagship) lead by Nick Bax, CMAR Hobart.

About Prof Shibata

Prof Shibata is a charming man☺

Prof Shibata in his early career provided a theoretical justification of the Akaike Information Criterion, routinely used for model selection in many statistical analyses.

As a member of the S software development team based at Bell Laboratories, he pioneered the use and development of S in Japan. The S language now forms the basis for R, a statistical computing environment widely used by research statisticians worldwide, including CMIS.

He was awarded the Japan Statistical Society Medal in 2006. In 2004 he helped win a national COE (Center of Excellence) 5 year research programme for Keio University.

About AJF Project

Under the AJF Grant theme of Education and Science, the Division was successful in obtaining funds for bilateral collaboration for the development of innovative tools in ecological management.

The AJF Grant of \$44,000 is supported by DFAT.

The project sought to provide the knowledge-base to inform future strategies in conservation and sustainable use of ecological and environmental assets, particularly in marine zones around Australia,

The Grant allows collaboration on several methodological fronts, publication of the analyses of existing datasets and to establish future collaboration between future PhD students and CMIS. The areas of expertise of the Keio University group are of interest to several CMIS program areas.

Whilst this initial contractual agreement has a focus on ecological and environmental assets (as relevant to the CMIS Environmental Informatics Research Program) – this definition has been broadened to enable collaborative effort to fund aspects relevant to the CMIS Bio Research Program. Dr Ian Saunders, CMIS Adelaide, has co-supervised Keio PhD

student Yuki Sugaya investigating the analysis of genomics data (and links with RIKEN).

July 2008

Dr Ross Darnell, Statistician CMIS and Mr Mick Haywood Biologist CMAR, visited Professor Ritei Shibata's department at Keio University from the 30th June to 4th July to collaborate on a project aimed at measuring the effect of trawling on marine benthic species. During this time, data issues were discussed and project goals were established.

August/Sept 2008

Prof Shibata and PhD student Yuki Sugaya visited Cleveland to present their progress on the "effect of trawling" project, working with CMIS and CMAR staff to progress their studies.

Prof Shibata and Prof Kamatani (RIKEN) also visited Adelaide to further the statistical genomics study.

Which brings us to the current workshop

A third visit to Keio by CMIS staff in March 2009 was planned to present the completed projects and establish links with PhD studentships with Professor

Shibata's current PhD and Masters students. This work will be presented during the workshop.

Further Collaboration between CMIS and Keio

Recently, a proposal to the Japanese and Australian governments was submitted by Prof Shibata and Dr Darnell under the scheme titled "Strategic Japanese-Australian Cooperative on Marine Science", a joint funding of Japanese and Australian research cooperation. This submission aims to develop scientifically based, statistical models of marine abundance and diversity from marine survey data, which form the basis of reliable and long run management systems for marine ecology. The space-time models obtained by this project should be useful for prediction as well as for explaining fish or seabed fauna ecology.

Australia-Japan Workshop on Data Science 2009

Provisional Programme

Tuesday 24 March

- 09:00-10:00 Registration (Coffee/Tea provided)
10:00-11:00 Louise Ryan (Opening by video conference)
11:00-12:30 Ross Darnell and Louise Marquart
Clustering species according to their relationship with their physical environment
Hideyasu Shimadzu and Ross Darnell
Investigating the Issues of Sampling in Marine Surveys
12:30-13:30 Lunch
13:30-15:00 Ritei Shibata and Yuki Sugaya
Modeling Counts in Trawling Data
Mayumi Naka and Ritei Shibata
Weight Distribution in Trawling Data
15:00-15:30 Coffee Break
15:30-17:50 Mick Haywood
TBA
Mihoko Minami
Statistical Challenges for Modeling Data with Many Zeros
Sara Lennox
Understanding water quality measurements
18:00-20:00 Welcome Party

Wednesday 25 March

- 09:00-10:30 Scott Foster
RAD Biodiversity: Modelling many species' counts together.
Toshihide Kitakado
Statistical genetic approaches for estimating population structure with application to fisheries populations
10:30-11:00 Coffee Break
11:00-12:30 Naoyuki Kamatani
A short history of genome-wide association study (GWAS) and statistical challenges
Natsuhiko Kumasaka
Textile plot: a graphical representation for multiple SNP genotype data
12:30-13:30 Lunch
13:30-15:00 Yuki Sugaya and Ritei Shibata
Likelihood-based method for estimating penetrance and disease susceptibility allele frequency
Bill Wilson
TBA
15:00-15:30 Coffee Break
15:30-17:00 Kunio Shimizu
Circular-circular regression, functional relationship and measurement error models
Murray Cameron

Collaboration, statistics and major research challenges: avoiding boring people

18:30-20:30 Workshop Dinner

Thursday 26 March

05:00-12:00 Visit to Tokyo Fish Market
Excursion to Ueno Park, Lunch at Tofu Restaurant

Friday 27 March

09:00-10:30 Masaaki Sibuya
Bell polynomials in discrete probability distributions
Peter Thomson

Mixed methods for fitting GEV distribution

10:30-11:00 Coffee Break

11:00-12:30 Arnaud Doucet

Particle Markov chain Monte Carlo

Daisuke Yokouchi and Ryozi Miura

What can we do for hedge fund return data under the DandD environment?

12:30-13:30 Lunch

13:30-15:00 Pavel Shevchenko

Quantitative modelling of financial risks

Ritei Shibata and Yuka Tanizawa

Smile Curve and Local Volatility

15:00- Farewell (with Japanese cake)