ISM Symposium on Environmental Statistics 2015

Date & Time: Feb. 24(Tue.), 2015

Venue: The Institute of Statistical Mathematics,

10-3 Midori-cho Tachikawa, Tokyo 190-8562, Japan

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

Joint Hosting:
The Graduate University for Advanced Studies 【SOKENDAI】
NPO Japan Institute of Environmental Statistics

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

Program

9:50-10:00 Opening Address: Yoshinori Kawasaki(ISM)

[Section 1]Chair:Shuhei Mano(ISM)

10:00-10:45 Spatio-temporal point patterns, Periodicities and earthquakes
Pierre Dutilleul (McGill University)

10:45-11:15 Modeling of the typhoon trajectory pattern and its variations on annual and longer time scales
Shinya Nakano (The Institute of Statistical Mathematics)

11:15-12:00 Statistical modelling of wet and dry spell frequencies over Langat river basin, Malaysia
Chua Kuan Chin (Universiti Tunku Abdul Rahman)

Lunch

[Section 2]Chair:Shinto Eguchi(ISM)

13:00-13:45 Exploring co-occurrence of closely-related guild members in a fragmented landscape subject to rapid transformation
David B. Lindenmayer, Alan Welsh, Wade Blanchard, Philip Tennant and Christine Donnelly (The Australian National University)

13:45-14:15 Ecological statistics beginning with intensive field observations
Kenichiro Shimatani (The Institute of Statistical Mathematics)

14:15-15:00 Study of rain induced landslides prediction and casualty prevention in Malaysia
Swee Peng Koay, Habibah Lateh and Pei Shan Fam (Universiti Sains Malaysia)

Coffee Break

[Section 3]Chair:Satoshi Kuriki(ISM)

15:15-16:00 Circular regression and animal movement
Louis-Paul Rivest (Université Laval)

16:00-16:30 Atractable, parsimonious and highly flexible model for cylindrical data
Toshihiro Abe (Tokyo University of Science)

16:30-17:00 Deforestation modeling based on statistical and machine learning approaches
Ryuei Nishii (Institute of Mathematics for Industry, Kyushu University)

17:00-17:10 Closing Address:Kunio Shimizu(ISM)

[1] Spatio-temporal point patterns, Periodicities and earthquakes
Pierre Dutilleul(McGill University, Canada)

Earthquakes represent, at the same time, a life-threatening phenomenon and a fascinating subject of investigation for many scientists, including seismologists and statisticians. While bearing in mind the seismological perspective, the statistical approach will prevail in this talk, and with more or less adjustment, the methodological aspects that will be discussed could be applied to other environmental phenomena such as volcanic eruptions and extreme climatic events (e.g. droughts, flooding).

The location of an earthquake in space and time, through the latitude, longitude and depth of its hypocenter in 3D and the date and clock time of rupture, respectively, makes it a "point" in a spatio-temporal point pattern, observed over a given territory and decades/years/months; the association of a magnitude with each earthquake will mark the point pattern. The frontier between point pattern analysis and time series analysis is not hermetic (Brillinger, 1994). Therefore, I shall move from one framework to the other and *vice versa* in my talk, (i) searching for periodicities in earthquake occurrence and (ii) attempting to understand the effects of earthquake catalog declustering algorithms based on ETAS models (Ogata, 1998; Zhuang et al., 2002, 2011) on the characteristics of the resulting time series of earthquake numbers (e.g. on a monthly basis); declustering is aimed at separating dependent earthquakes (aftershocks, foreshocks) from independent background event occurrences.

Formally, I shall show how the declustering of an earthquake spatial point pattern affects the mean and autocovariance functions of the time series of earthquake numbers. Empirically, I shall use central California earthquake datasets to illustrate differences in results of periodicity analysis with complete vs. declustered catalogs depending on the magnitude range. Two methods of periodicity analysis will be compared: the "Schuster spectrum", which belongs more to the point-pattern approach (Schuster, 1897; Ader and Avouac, 2013), and the "multi-frequential periodogram analysis" (MFPA; Dutilleul, 2001), for which this is a first documented application to seismological time series.

The periodicity analysis of central California earthquakes is joint work with Roland Bürgmann and Christopher Johnson from University of

California in Berkeley.

References

Ader, T. J., and J. P. Avouac (2013), Detecting periodicities and declustering in earthquake catalogs using the Schuster spectrum, application to Himalayan seismicity, *Earth and Planetary Science Letters*, 383, 26-36.

Brillinger, D. R. (1994), Time series, point processes, and hybrids, *Canadian Journal of Statistic*s, 22, 177-206.

Dutilleul, P. (2001), Multi-frequential periodogram analysis and the detection of periodic components in time series, *Communications in Statistics - Theory and Methods*, 30, 1063-1098.

Ogata, Y. (1998), Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, 50, 379-402.

Schuster, A. (1897), On lunar and solar periodicities of earthquakes, *Proceedings of the Royal Society of London*, 61, 455-465.

Zhuang, J., Y. Ogata, and D. Vere-Jones (2002), Stochastic declustering of space-time earthquake occurrences, *Journal of the American Statistical Association*, 97, 369-382.

Zhuang, J., M. J. Werner, S. Hainzl, D. Harte, and S. Zhou (2011), Basic models of seismicity: Spatiotemporal models, Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-07487583. Available at http://www.corssa.org.

[2] Modeling of the typhoon trajectory pattern and its variations on annual and longer time scales

Shin'ya Nakano(The Institute of Statistical Mathematics)

A typhoon is an intense tropical cyclone forming in the western north Pacific region. Since a typhoon can cause a serious disaster in East Asia, it is crucial to evaluate risks of typhoon hazards. In principle, It would be possible to evaluate such risks by compiling a large number of the simulation results for various possible scenarios. However, numerical fluid dynamics simulations that could satisfactorily reproduce a realistic typhoon trajectory would require prohibitive computational cost for each run. In order to enable us to assess the risks due to typhoons with low computational cost, we aim at developing a new probabilistic model which can generate a variety of surrogate typhoon trajectories with high reality. Since the trajectory pattern of typhoons has variations on annual and longer time scales, the model should be designed so as to consider these long-term variations to allow detailed assessment of the risks due to typhoons. In order to reproduce such features of the typhoon trajectories, we employ a Gaussian process regression and obtain the typical typhoon velocity as a function of latitude, longitude, day of year, and year. The characteristics of the long term change of the typical typhoon velocity and their implications are discussed on the basis of the results.

[3] Statistical modelling of wet and dry spell frequencies over Langat river basin, Malaysia

Chua Kuan Chin(Universiti Tunku Abdul Rahman, Malaysia)

Recent water crisis that occurred in Selangor state of Malaysia has raised the importance of understanding the rainfall characteristic. By having able to model a rainfall processes, runoff of water and natural disaster due to heavy rainfall are able to be supervised. A frequency analysis is the most commonly applied method in hydrology data, while Normal, Log-normal, Gamma, Gumbel and Weibull distribution are among the important probability distributions commonly used in hydrology analysis. In this study, the analysis of distributions of wet and dry spells based on daily rainfall is investigated and it is fitted to the daily rainfall data for the period 1995-2005. The candidate distributions are Hurwitz-Lerch Zeta distribution, Eggemberger-Polya distribution, Logarithmic distribution, Truncated Poisson distribution and Geometric distribution. The parameters of distributions are estimated by Maximum Likelihood Estimation using simulated annealing optimization method. The distributions are then plotted and compared to the histogram of dry and wet spell of the daily rainfall. Model selection technique such as AIC is employed to study the fittings of those distributions.

[4] Exploring co-occurrence of closely-related guild members in a fragmented landscape subject to rapid transformation

David B. Lindenmayer, Alan Welsh, Wade Blanchard, Philip Tennant, Christine Donnelly (The Australian National University, Australia)

We use data from a 15-year experimental study, to explore intra-guild co-occurrence of six closely-related and functionally-similar sets of birds within 55 woodland fragments. Areas surrounding these remnants are undergoing transformation from grazed paddocks to Pinus radiata plantations, leading to a novel assemblage of forest and woodland birds. We therefore sought to determine if the occurrence of a given species in a guild influenced the occurrence of other closely-related species in that guild, and through this relationship whether there was evidence of co-occurrence between species.

After controlling for environmental variables which can affect species occurrence like time since commencement of landscape transformation, patch size and vegetation type, we found the occurrence of a given species was influenced by the occurrence of a closely-related species in the same guild. Kinds of resulting co-occurrence varied among bird guilds and included: (1) positive co-occurrence in which occurrence of one species within woodland patches positive affected the occurrence of another closely-related species in the same guild (e.g. Eastern and Crimson Rosellas); and (2) negative co-occurrence in which the occurrence of one species was negatively associated with the occurrence of another within the same guild (e.g. Willie Wagtail and Grey Fantail).

We also identified interactions between patch size and abundance within members of two bird guilds. For example, results of our modelling for mean conditional abundance showed that the Eastern Rosella increased with increasing numbers of the Crimson Rosella in large patches, but decreased with increasing numbers of the Crimson Rosella in small patches.

Our results provide empirical evidence of co-occurrence among closely related and functionally similar species in a rapidly transforming landscape. Our findings underscore the complexity of biotic responses to landscape transformation

[5] Ecological statistics beginning with intensive field observations
Ichiro ken Shimatani（ISM）

　In this presentation, I am talking examples of ecological studies that heavily relied on intensive field observations and statistical modeling. In one case about long-term forest dynamics, I was one of the intensive field workers, and the study also utilized old historical records. The statistical analyses were elementary applications of Bayesian modeling. In one case about long-term population dynamics of large ocean mammals (seal), the study began with field surveys by volunteer students and two miracle (!) field zoologists 40 years ago. The up-to-date Bayesian modeling (integrated population model, IPM) was applied to 40-years census data, and detail records during a few months uninhabited island stays by the miracle zoologists were also used in the Bayesian formulation. We discuss long-term trends of the population and applications to marine ecosystem management.

[6] Study of rain induced landslides prediction and casualty prevention in Malaysia

Swee Peng Koay(Universiti Sains Malaysia), Habibah Lateh(Universiti Sains Malaysia), Pei Shan Fam, (Universiti Sains Malaysia)

In Malaysia, landslides occur more often than before due to climate change. Unlike Japan, most of the landslides are rain induced slope failures. The Malaysian Government has to allocate millions of Malaysian Ringgit for slope monitoring in the budget every year.

However, there are thousands of slopes which are classified as high risk slopes in Malaysia. Installing site monitoring devices, such as extensometer, piezometer and inclinometer in these slopes to monitor the movement of the soil in the slopes, are too costly and almost impossible.

In our study, we propose Accumulated Rainfall vs. Rainfall Intensity prediction method to predict the slope failure by referring to the predicted rain data from radar and the rain volume from rain gauges. The critical line which determines if the slope is in danger, is generated by simulator with well-surveyed the soil property in the slope and compared with historical data. By establishing such predicting system, the slope failure warning information can be obtained and disseminated to the surroundings via IT systems and siren.

Besides establishing the early warning dissemination system, educating school children and the community by giving knowledge on landslides, such as landslide's definition, how and why does the slope failure happen and when will it fail, to raise the risk awareness on landslides will reduce landslides casualty, especially in rural area.

[7] Circular regression and animal movement
Louis-Paul Rivest (Université Laval, Canada)

In ecology, data on animal's movement is easily collected using GPS tracking devices. This information is merged with data on an animal's environment, available in satellite photos for instance, using a GIS (Geographic Information System). To investigate how an animal's environment influences its motion, one constructs predictive models for y_t, the angle of an animal's movement at time t. The explanatory variables are y_{t-1}, the direction of displacement at the previous time step, the directions at time t, x_t, of targets of interest, such as a lair or a patch rich in nutriment, and the distances z_t to these targets. In ecology, these models are known as biased correlated random walks, see Benhamou (2014). As the dependent variable is circular, standard regression models do not apply and multivariate generalizations of the circular-circular regression model of Rivest (1997) and of the circular-linear model of Presnell et al. (1998) are introduced. Likelihood inference methods for these models are developed. Generalizations of the models for the movement angle y_t that include hidden states to account for an animal changes in behavior (Langrock et al., 2012) are also introduced. The statistical challenges encountered when fitting them to movement data will be discussed. This is joint work with A. Nicosia, T. Duchesne and D. Fortin of Université Laval.

References

Benhamou, S. (2014) Of scales and stationarity in animal movements. Ecology Letters, 17, 261–272.

Langrock, R., King, R., Matthiopulous, J., Thomas, L., Fortin, D., and Morales, J. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. Ecology, 93, 2336-2342

Presnell, B., Morrison, S.P. & Littell, R.C. (1998) Projected multivariate linear models for directional data. Journal of the American Statistical Association, 93, 1068–1077.

Rivest, L.P. (1997) A decentred predictor for circular-circular regression. Biometrika, 84, 717–726.

[8] A tractable, parsimonious and highly flexible model for cylindrical data
Toshihiro Abe
(Tokyo University of Science)

Cylindrical data are observations that consist of a directional part (a set of angles), which is often of a circular nature (a single angle), and a linear part (mostly a positive real number). This explains the alternative terminology of directional-linear or circular-linear data. Such data occur frequently in natural sciences; typical examples are wind direction and another climatological variable such as wind speed or air temperature, the direction an animal moves and the distance moved, or wave direction and wave height. In this talk, we consider cylindrical distributions obtained by combining the sine-skewed von Mises distribution (circular part) with the Weibull distribution (linear part). This new model, the Weibull-sine-skewed von Mises (WeiSSVM) distribution, enjoys numerous advantages: simple normalizing constant and hence very tractable density, parameter-parsimony and interpretability, maximum entropy characterization, good circular-linear dependence structure, easy random number generation thanks to known marginal/conditional distributions, and flexibility illustrated via excellent fitting abilities. We also introduce other new circular-linear models, based on the same idea by making use of the Gamma and Generalized Gamma distributions. Our flexible models are applied in analyses of two cylindrical data sets. We conclude the talk by showing a straightforward extension to the directional-linear cylindrical distributions.

[9] Deforestation modeling based on statistical and machine learning approaches

Ryuei Nishii (Institute of Mathematics for Industry, Kyushu University, Japan)

Deforestation is caused by various factors. In the literature, the impact of human activities as well as geographic circumstances on forests has been extensively discussed. We have studied statistical models for prediction of forest area ratio by covariates: human population density and relief energy observed in a grid-cell system. Parametric non-linear regression functions of the covariates were used for predicting forest coverage ratio and cubic spline functions were also used for detection of small fluctuation of regression functions. Furthermore, zero-one inflated distributions were proposed for classification of each site into one of three categories: completely-deforested, fully-forest-covered or partly-deforested areas. These methods took the spatial dependency into the modeling, which is not an easy task.

Our aim here is to substitute the previous statistical approach for machine learning approach based on SVM (support vector machine) and SVR (support vector regression). SVM will be used for classification of each site into one of the above-mentioned categories, and SVR for prediction of the forest coverage ratio. The proposed approach implements a neighbors' effect into the modeling easily. By our numerical study, it will be shown that the performance of the machine learning methods is comparable or superior to that of the statistical methods.

ISM Symposium on Environmental Statistics 2015

Feb. 24, 2015. The Institute of Statistical Mathematics