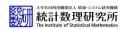# LMBP-ISM joint workshop

Report of Contributions

Contribution ID: **10**   Type: **not specified**

Generalised hyperbolic state space models with application to spatio-temporal heat wave prediction

Tuesday, 24 October 2023 10:00 (50 minutes)

As global warming progresses, it is increasingly important to monitor and analyse spatio-temporal patterns of heat waves and other extreme climate-related events that impact urban areas. In this work, we present a novel dynamic spatio-temporal model by combining a state space model (SSM) and a generalised hyperbolic distribution to flexibly describe a spatial-temporal profile of the tail behaviour, skewness and kurtosis of the local urban temperature distribution of the greater Tokyo metropolitan area. Such a model can be used to study local dynamics of temperature effects, specifically those that characterise extreme heat or cold. The focus of the application in this paper will be heat wave events in the greater Tokyo metropolitan area which is known to be prone to some of the most severe heat wave events that have one of the largest population exposures due to high density living in Tokyo city.

Presenter: SEPTIER, François (Université Bretagne Sud (UBS))

Session Classification: Session: Statistical Modeling

Contribution ID: **11**   Type: **not specified**

Sub-model aggregation for scalable spatially varying coefficient modeling

Tuesday, 24 October 2023 11:00 (45 minutes)

This study aggregates/combines global and local sub-models to build a fast and flexible spatially varying coefficient model. An approach inspired by the generalized product-of-experts method is used to aggregate the sub-models. The aggregated model has the following properties: (i) computationally efficient; (ii) the marginal likelihood is available in closed-form; (iii) each sub-model can be estimated independently to maximize the likelihood. Owing to (ii) and (iii), the proposed model describes complex spatial patterns flexibly and computationally efficiently. The accuracy and computational efficiency of the proposed method are compared with alternatives through Monte Carlo simulation experiments. Then, the method is applied to a regression analysis of residential land prices in Japan.

Presenter: Prof. MURAKAMI, Daisuke (The Institute of Statistical Mathematics (ISM))

Session Classification: Session: Statistical Modeling

Contribution ID: **12**   Type: **not specified**

Minimum information dependence modeling:

Tuesday, 24 October 2023 11:45 (45 minutes)

We propose a method to construct a joint statistical model for mixed-domain data to analyze their dependence. The model is characterized by two orthogonal parameters: the dependence parameter and the marginal parameter. To estimate the dependence parameter, a conditional inference together with a sampling procedure is proposed and is shown to provide a consistent estimator. Illustrative examples of data analyses involving penguins and earthquakes are presented.

Presenter: Prof. YANO, Keisuke (The Institute of Statistical Mathematics (ISM))

Session Classification: Session: Statistical Modeling

Contribution ID: **13**   Type: **not specified**

Statistical modeling and machine learning studies on the forecast of earthquakes and fault slip

Wednesday, 25 October 2023 10:00 (20 minutes)

Earthquakes refer to sudden and spontaneous rupture, or slip, of a geologic fault. Although the physics of fault slips are not fully understood, the governing equations of fault slips are used along with data assimilation methods to make forecast of future fault slip behavior. On the other hand, by focusing on the location and timing of occurrences, earthquakes have also been regarded and modeled as a point process. In both kinds of studies –that focuses on the fault slip evolution and that focuses on the earthquake occurrence patterns –statistical modeling and machine learning studies are changing the research landscape. Recently, science findings have been increasingly used in society, in particular for providing alert/advisory information related to future earthquake occurrence. It means that methodological developments having a larger predictive ability may change how society prepares for future earthquakes. In the talk, I will review the recent advances of statistical modeling and machine learning studies in earthquake science after explaining the fundamental backgrounds, and subsequently on how the research advances can contribute to society for earthquake disaster mitigation.

Presenter: FUKUSHIMA, Yo (Tohuku University - IRIDeS)

Session Classification: Session: Earthquake Hazard

Contribution ID: **14**   Type: **not specified**

Estimation of the fault slip of short-term slow slip events from GNSS data using deep learning

Wednesday, 25 October 2023 11:00 (30 minutes)

Since dense geodetic and seismic networks reveal the presence of slow earthquakes and the close relationship between regular and slow earthquakes, many studies have focused on the detection of slow earthquakes and their source characterization. Global Navigation Satellite System (GNSS) continuously monitors ground deformation and is one of the most common tools used to detect slow slip events (SSEs), a kind of slow earthquake. However, GNSS data sometimes needs manual preprocessing due to its low signal-to-noise ratio. Furthermore, automated analysis methods are becoming increasingly important in today's world of huge data volumes. Deep-learning approaches, especially convolutional neural networks (CNN), have largely contributed automation process to deal with big data. These brand-new technologies have brought significant breakthroughs into many fields including seismology (e.g., Yano et al., 2021) and geodesy (e.g., Rouet-Leduc et al., 2021).

In this study, we aim to develop a deep-learning method to monitor spatio-temporal evolutions of short-term SSEs based on a dense GNSS network. We theoretically create two types of horizontal deformation data including synthetic noise by assuming 272 subfaults in western Shikoku, southwest Japan; 16 subfaults along the strike multiplied by 17 subfaults along the dip. One is deformations at 113 GNSS stations, and the second is those at 900 virtual stations which are regularly located over the target area. We tailor two supervised-learning Convolutional Neural Network (CNN) models to estimate the slip area and the slip amount of SSE by learning those deformation images as input data. Nakagawa et al. (2021, Fall Meeting in Geodetic Society of Japan) showed that the model trained with GNSS stations estimated SSEs with 91.8% variance reduction (VR) while the other model achieved 98.3% VR. We concluded that this difference in estimation accuracy is contributed to the dissimilarity between input deformation images. Therefore, we newly implement Model-supervised Interpolation (MSI) approach to overcome this problem. MSI successfully reproduces the deformations at 900 virtual stations only from the deformations at GNSS stations with 97.4% VR although nearly half of the target area is located on the offshore region. It shows this deep-learning approach is effective to estimate SSEs from GNSS data in this region.

Presenter: NAKAGAWA, Ryo (Tohuku University - IRIDeS)

Session Classification: Session: Earthquake Hazard

Contribution ID: **15** Type: **not specified**

COVID-19 Case Prediction Using Emotion Trends via Twitter Emoji Analysis

Wednesday, 25 October 2023 16:35 (20 minutes)

The worldwide COVID-19 pandemic, which began in December 2019 and has lasted for almost 3 years now, has undergone many changes and has changed public perceptions and attitudes. Various systems for predicting the progression of the pandemic have been developed to help assess the risk of COVID-19 spreading. In a case study in Japan, we attempt to determine whether the trend of emotions toward COVID-19 expressed on social media, specifically Twitter, can be used to enhance COVID-19 case prediction system performance. We use emoji as a proxy to shallowly capture the trend in emotion expression on Twitter. Two aspects of emoji are studied: the surface trend in emoji usage by using the tweet count and the structural interaction of emoji by using an anomalous score. Our experimental results show that utilizing emoji improved system performance in the majority of evaluations.

Presenter: Prof. MATSUI, Tomoko (The Institute of Statistical Mathematics (ISM))

Session Classification: Session: Statistics and AI Applications

Contribution ID: **16**
Type: **not specified**

Goal Oriented random forest

Tuesday, 24 October 2023 15:15 (35 minutes)

TBA

Presenter: Prof. MAUME-DESCHAMPS, Véronique (Institut Camille Jordan, Université Lyon 1)

Session Classification: Session : Machine Learning

Contribution ID: **17**   Type: **not specified**

Unlinked or shuffled monotone regression

Wednesday, 25 October 2023 16:00 (30 minutes)

In standard regression models, pairs of covariates and response variables are observed. In the more complex case of shuffled regression (on anonymized data), we only observe a sample of covariates on the one hand, and a sample of responses on the other, but we don't know which response corresponds to each covariate. In the even more complex case where responses and covariates are not necessarily measured on the same individuals, both samples of covariates and responses are still observed, but there is not necessarily a link between them. The data are unlinked. This raises the question of whether the link between the two samples in the shuffled case provides any real information compared with the unlinked case, i.e. whether or not the optimal rates of convergence of estimators are identical in the two models. We provide some answers to this question.

Presenter: Prof. DUROT, Cécile (Université Paris Nanterre)

Session Classification: Session: Statistics and AI Applications

Contribution ID: 18   Type: **not specified**

Limit theorems for neural networks modeling

Tuesday, 24 October 2023 14:30 (35 minutes)

TBA

Presenter: Prof. GUILLIN, Arnaud (Laboratoire de Mathémariques Blaise Pascal (LMBP))

Session Classification: Session : Machine Learning

Contribution ID: **21**   Type: **not specified**

Creep characteristics along the central part of the Philippines Fault on Leyte Island from ALOS-2 time-series analysis

Wednesday, 25 October 2023 11:30 (20 minutes)

Primary authors: OKUR, Yagizalp (Laboratoire de Mathémariques Blaise Pascal (LMBP) and IRIDeS); FUKUSHIMA, Yo (Tohuku University - IRIDeS)

Presenter: OKUR, Yagizalp (Laboratoire de Mathémariques Blaise Pascal (LMBP) and IRIDeS)

Session Classification: Session: Earthquake Hazard

Contribution ID: **23**   Type: **not specified**

Assessing Portfolio Diversification via Two-Sample Graph Kernel Inference.

Tuesday, 24 October 2023 16:15 (30 minutes)

In this work we seek to enhance the frameworks practitioners in asset management and wealth management may adopt to assess how different screening rules may influence the diversification benefits of portfolios. The problem arises naturally in the area of Environmental, Social, and Governance (ESG) based investing practices as practitioners need to select subsets of the total available assets using screening rules of ESG ratings and to compare the subsequent risk and return profile of the portfolios created from different selective portfolios. We propose a novel method to compare the diversification relationships of assets in different portfolios based on a machine learning hypothesis testing framework called the kernel two-sample test. The objective of the test is to determine whether two samples come from the same underlying probability distribution. In the case of asset management, the samples are sequences of graph-valued data points that represent a dynamic portfolio obtained by a certain ESG screening rule and certain portfolio optimization criteria such as the global minimum variance or max Sharpe. The fact that the sample data points are graphs means that one needs graph testing frameworks to compare diversification benefits. The problem is natural for kernel two-sample testing as one can use so-called graph kernels to work with samples of graphs. The objective is then to determine if the two dynamic portfolios have the same generating mechanism. A failure to reject the null hypothesis would indicate that ESG screening does not affect diversification while rejection would indicate that ESG screening does have an effect. The article describes the graph kernel two-sample testing framework, further, it provides a brief overview of different graph kernels. We then demonstrate the power of the graph two-sample testing framework under different realistic scenarios. We finally apply the framework to demonstrate the workflow one can use in asset management to test for structural differences in diversification of portfolios under different ESG screening rules.

Primary authors: LEVI GUDMUNDARSON, Ragnar (Herriot WATT); PETERS, Gareth W. (University of California Santa Barbara)

Presenters: LEVI GUDMUNDARSON, Ragnar (Herriot WATT); PETERS, Gareth W. (University of California Santa Barbara)

Session Classification: Session: Insurance and Risk

Contribution ID: **24**   Type: **not specified**

Data-Driven Framework for Uncovering Hidden Control Strategies in Evolutionary Analysis

Wednesday, 25 October 2023 15:00 (30 minutes)

Natural risks are characterized primarily by their uncontrollability, or at least, their difficulty in being controlled. In this instance, the goal is to try to understand the mechanism that causes these hazards as well as the variables that affect the processes' evolutionary behavior. In this talk, we provide a data-driven approach for identifying an evolutionary system's hidden control mechanisms. This approach enables deciphering of the hidden processes that leads to these kinds of hazard circumstances. Model predictive control is extended by using new techniques to determine the best control together with the parameters for evolution in general dynamical systems. This is a major divergence from traditional control approaches, which call for an understanding of the system, the capacity to influence its course, and the controller's approach or parameters.

Presenter: AZZAOUI, Nourddine (Laboratoire de Mathématique Blaise Pascal - UCA)

Session Classification: Session: Volcano Hazard

Contribution ID: **25**   Type: **not specified**

Statistical analysis of the runout distance of pyroclastic density currents and implications for emplacement mechanisms

Wednesday, 25 October 2023 14:15 (30 minutes)

Pyroclastic density currents (PDCs) are hot mixtures of gas and particles generated by volcanic eruptions. They propagate on the ground at high velocity and can travel distances that are commonly of several tens of kilometers. Understanding the factors that control the long runout distance of PDCs is important for hazard assessment. In this context, we collected data on PDCs in more than 200 publications to create a database. Through statistical analysis of the data, we show that the runout distance of PDCs correlates with the mass eruption rate. A model selection procedure further shows that it is possible to determine two well-defined power law relationships, respectively for fully dilute turbulent currents and for currents with a concentrated base. For dilute currents, the runout distance scales with the ratio of the eruption rate over the particles settling velocity to the power 0.5, in agreement with theory. Runout distances of some concentrated currents are longer than 300 km and outside predictions intervals, and in this regard we argue that these extreme travel distances were reached because the currents were confined in large paleovalleys. We conclude that statistical analysis can help to better understand the mechanisms of complex natural phenomena such as pyroclastic density currents.

Presenter: ROCHE, Olivier (Laboratoire Magma et Volcans (LMV) - UCA)

Session Classification: Session: Volcano Hazard

Contribution ID: **26**   Type: **not specified**

Opening by the LMBP director/ Prof. Julien Bichon

Tuesday, 24 October 2023 09:45 (10 minutes)

Presenter: Prof. BICHON, Julien (Laboratoire de Mathémariques Blaise Pascal (LMBP))

Session Classification: Opening Remarks and Reception