Omi Statistical Seminar

New Frontiers in Statistical Inference on Geometry, Topology and Time Series

Date: August 11~13, 2026

Venue: Meeting Room 207, Piazza-Omi, Shiga

(Access map: https://otsu-mice.jp/facilities/piazza-omi/)

Organizers: Masanobu Taniguchi (Waseda University)

Hernando Ombao (KAUST)

Yan Liu (Waseda University)

Supported by

- JSPS Grant-in-Aid for Scientific Research (C) 26K14735 (M. Taniguchi, Waseda Univ.)

- JSPS Grant-in-Aid for Scientific Research (C) 26K14748 (Y. Liu, Waseda Univ.)

Program

August 11

09:00 – 09:30 Rinka Sagawa (Waseda Univ.)
On the Asymptotic Normality of Exponential Smoothing Predictors

09:45 – 10:15 Paolo Redondo (KAUST)
Extremal Vulnerability

10:30 – 11:00 Ruoqian Huang (Waseda Univ.)
Cross-validation for Locally Stationary Spatial Data

11:15 – 11:45 Fletcher Christensen (Univ. New Mexico)
Efficient Marginalization of the Deviance Information Criterion for Generalized Linear Mixed Models

12:00 – General Discussion

August 12

09:00 – 12:00 Small Group Discussion (I)

13:00 – 16:00 Small Group Discussion (II)

18:00 – 18:30 Angel Lopez (KAUST)
Time series clustering: forecasting and amortized inference

18:45 – 19:15 Giovanni Motta (Columbia Univ.)
Frequency-domain Regularization for Dynamic Factor Models: Forecasting and Sparsity

19:30 – 20:00 Cathy W-S Chen (Feng Chia Univ.)
Advances in Spatial Integer-Valued Time Series Modeling

20:15 – 20:45 Masanobu Taniguchi (Waseda Univ.)
Second order generalization of Hajek’s convolution theorem and its applications

August 13

09:00 – 09:30 Johannes Lederer (Hamburg Univ)
How can mathematics contribute to AI?

09:45 – 10:15 Ziling Ma (KAUST)
Forecasting Multivariate Time Series under Predictive Heterogeneity: A Validation-Driven Clustering Framework

10:30 – 11:00 Yan Liu (Waseda Univ.)
A Distance-Preserving Bootstrap for Topological Features

11:15 – 11:45 Anass El Yaagoubi (KAUST)
Topological Causal Modeling of Information Flow in Brain Networks

11:45 – 12:00 Final Remarks
By Masanobu Taniguchi, Hernando Ombao and Yan Liu

12:00 – General Discussion

Abstract

Chen, Cathy W.S.

Title: Advances in Spatial Integer-Valued Time Series Modeling

Abstract: This talk reviews recent advances in spatial integer-valued time series modeling, focusing on spatial hurdle integer-valued GARCH and spatial zero-inflated generalized Poisson (ZIGP) integer-valued GARCH models for count data exhibiting spatiotemporal dependence and excess zeros. Bayesian inference is employed throughout, including an empirical Bayes hurdle formulation that introduces a data-adaptive prior to improve model parsimony and estimation stability. To enhance epidemiological relevance, seasonal effects are incorporated into the log-intensity equations using two approaches: Fourier-based harmonic terms and meteorological covariates. Applications to dengue fever data illustrate the importance of accounting for spatial interaction, temporal dependence, seasonality, and excess zeros. The results further demonstrate that Fourier-based seasonal components provide an effective alternative when environmental data are limited or unavailable.

Christensen, Fletcher

Title: Efficient Marginalization of the Deviance Information Criterion for Generalized Linear Mixed Models

Abstract: Spiegelhalter et al. (2002) and Celeux et al. (2006) developed various forms of the Deviance Information Criterion (DIC), which are commonly used for Bayesian model selection. The differing forms involve different choices of penalty term for overfitting, and differing choices about whether or not to regard latent variables as parameters. We focus on the Deviance Information Criterion based on the marginal probability model for the data—what Celeux et al. describe as the "observed DIC"—having integrated out any latent variables. Obtaining the marginalized pdf for the data generally involves numerical integration of a joint probability model including the latents, all conditional on model parameters. We present an analytical approximation to the marginal pdf of the data, given the model parameters, and and show that its use gives accurate numerical approximations to the corresponding marginalized DIC.

El Yaagoubi, Anass

Title: Topological Causal Modeling of Information Flow in Brain Networks

Abstract: A graph-based framework for quantifying self-similarity in chromatin folding from Hi-C contact data is introduced. Genomic loci are represented as nodes and Hi-C interactions as weighted edges, enabling the estimation of fractal dimension using correlation and sandbox methods. Validation on synthetic networks demonstrates that the approach captures hierarchical multiscale organization while distinguishing non-fractal clustered structures. Applied to IMR90 Hi-C data, the framework reveals approximate scaling behavior consistent with fractal-like chromatin organization, with larger chromosomes tending toward more compact, space-filling structures and smaller chromosomes exhibiting lower-dimensional folding patterns. These results suggest that graph-based fractal dimension provides an interpretable quantitative summary of genome organization across scales.

Huang, Ruoqian

Title: Cross-validation for Locally Stationary Spatial Data

Abstract: We study bandwidth selection for nonparametric estimation in spatially varying processes, where classical global models often fail to capture complex dependencies that evolve smoothly across a geographic space. We formulate a spatially varying autoregressive (svAR(1)) model and propose a rigorous spatial cross-validation framework to determine the optimal smoothing bandwidth for local M-estimation. We bridge two-dimensional spatial grid frameworks with established cross-validation theories for locally stationary time series, employing a dual-kernel local least squares estimator that decouples structural lag approximation from local spatial smoothing. The theoretical properties of this data-driven bandwidth selector are rigorously established, including a formal derivation of its asymptotic bias-variance decomposition and mathematical proofs of its consistency. Finite-sample simulation studies evaluate the performance of our proposed methods, and an empirical application modeling Tokyo's population density illustrates its practical utility in real-world scenarios. Overall, the proposed framework provides a principled, theoretically justified, and practically reliable approach to spatial cross-validation in locally stationary spatial modeling.

Lederer, Johannes

Title: How can mathematics contribute to AI?

Abstract: Artificial intelligence is arguably the hottest topic in science. Computer science and engineering currently set the agenda in this field, sidelining mathematics to a large extent. This talk, however, will highlight that mathematics has a lot to offer. We will introduce mathematical guarantees that provide deep insights into the inner workings of AI, and we will show how statistical principles can make AI more efficient. More generally, we will discuss the role of mathematics, especially mathematical statistics, in AI and data science.

Liu, Yan

Title: A Distance-Preserving Bootstrap for Topological Features

Abstract: We propose a distance-preserving bootstrap method for topological features. Statistical aspects of topological data analysis have recently attracted increasing attention, and bootstrap methods for invariants such as Betti numbers have begun to be explored. However, statistical methods based on analysis over topological spaces often involve intrinsic difficulties, and it is well known that they may suffer from slow convergence rates. To address these issues, we introduce a metric structure into the framework of topological data analysis, thereby enriching the analysis from a purely topological perspective. By incorporating the notion of distance, we develop a bootstrap procedure that leverages metric information while preserving topological characteristics. Our numerical simulations demonstrate that the proposed method achieves more stable statistical performance compared to conventional approaches, in accordance with the theoretical considerations. This is a joint work with Yuto Hamazaki.

Lopez, Angel

Title: Time series clustering: forecasting and amortized inference

Abstract: Time series clustering is a central task in statistics, with wide-ranging applications in demography, finance, biology, medicine, and many other fields. Its objective is to partition a collection of unlabeled time series into homogeneous groups so

that similar series are assigned to the same cluster, while dissimilar series are placed in different clusters. Traditionally, time series clustering has been used primarily for pattern recognition, that is, for detecting and summarizing patterns in collections of time series. More recently, however, it has been employed for tasks beyond pattern recognition. In this context, this talk presents a clustering framework designed to enhance forecasting by grouping time series according to predictive accuracy, yielding improved results across benchmark datasets. We also discuss how amortized inference can address major challenges in time series clustering, including algorithm selection, optimization limitations, and determining the number of clusters, offering a promising unified solution to these issues.

Ma, Ziling

Title: Forecasting Multivariate Time Series under Predictive Heterogeneity: A Validation-Driven Clustering Framework

Abstract: We study adaptive pooling under predictive heterogeneity in high-dimensional multivariate time series forecasting, where global models improve statistical efficiency but may fail to capture heterogeneous predictive structure, while naive specialization can induce negative transfer. We formulate adaptive pooling as a statistical decision problem and propose a validation-driven framework that determines when and how specialization should be applied. Rather than grouping series based on representation similarity, we define partitions through out-of-sample predictive performance, thereby aligning data organization with predictive risk, defined as expected out-of-sample loss and approximated via validation error. Cluster assignments are iteratively updated using validation losses for both point (Huber) and probabilistic (pinball) forecasting, improving robustness to heavy-tailed errors and local anomalies. To ensure reliability, we introduce a leakage-free fallback mechanism that reverts to a global model whenever specialization fails to improve validation performance, providing a safeguard against performance degradation under a strict training-validation-test protocol. Experiments on large-scale traffic datasets demonstrate consistent improvements over strong baselines while avoiding degradation when heterogeneity is weak. Overall, the proposed framework provides a principled and practically reliable approach to adaptive pooling in high-dimensional forecasting problems.

Motta, Giovanni

Title: Frequency-domain Regularization for Dynamic Factor Models: Forecasting and Sparsity

Abstract: The generalized dynamic factor model (GDFM) of Forni and Lippi (2001) has become very popular for the theoretical and empirical analysis of large panels of time series data. The corresponding estimators introduced by Forni et al. (2000) rely on Brillinger's dynamic principal components and thus involve two-sided filters, which are unsuitable for forecasting. Forni et al. (2015, 2017) derive a semi-parametric estimator of common components and common shocks based on one-sided filters. However, this newer approach relies on the additional assumptions that the common components have a rational spectral density. Moreover, their estimation procedure involves several time-domain and frequency-domain steps.

To overcome these limitations, we have recently introduced a novel approach that estimates the common components and the common shocks directly in the frequency domain. Our model does not require the assumption of a rational spectral density, and our estimation method is computationally simpler and faster.

In this companion work on frequency-domain estimation of the GDFM, we extend our methodology in two important directions: forecasting and sparsity. We present a one-sided regularized estimator specifically designed for forecasting. Additionally, we introduce regularization techniques directly in the frequency domain to induce sparsity in the transfer-function (or frequency-domain loading) matrix. We establish asymptotic properties under increasing cross-section and time dimensions, and demonstrate that our novel method outperforms existing one-sided estimators in both forecasting accuracy and interpretability.

Redondo, Paolo

Title: Extremal Vulnerability

Abstract: In many complex systems, identifying the most vulnerable components is essential for effective prevention, intervention, and risk management. Extreme events

rarely occur in isolation but instead arise jointly across interconnected components. A method is developed to identify the nodes that are most vulnerable to such systemic shocks. These joint extreme occurrences provide insight into patterns of tail dependence and help identify the components that are systematically exposed to extreme shocks propagating through the network. The notion of extremal vulnerability is introduced, defined as the long-run tendency of a component to be affected by extreme events propagating from other components. A novel framework is developed that builds on the tail dependence matrix (TDM), which captures directed extremal dependence, and leads to the Extremal Vulnerability Rank (XVRank), a PageRank-inspired algorithm for quantifying extremal vulnerability. Theoretical properties of the resulting estimators, including consistency and asymptotic normality, are established, and their performance is validated through Monte Carlo simulations. As an empirical application, the XVRank method is applied to the Magnificent Seven (MAG7) equities to identify assets most exposed to severe market downturns, and to 30 industry portfolios to characterize extremal vulnerability across sectors of the US economy.

Sagawa, Rinka

Title: On the Asymptotic Normality of Exponential Smoothing Predictors

Abstract: Forecasting is one of the fundamental problems in time series analysis, and various forecasting methods have been proposed in the literature. Among them, exponential smoothing is a widely used forecasting method based on exponentially weighted moving averages, where the weights decrease exponentially as observations become older. Conventionally, the weights are selected by minimizing empirical criteria such as the mean squared error (MSE), relative absolute error (RAE), and mean absolute percentage error (MAPE). In this talk, we investigate the weights under which the exponential smoothing predictor satisfies asymptotic normality. This property enables the construction of asymptotic confidence intervals and prediction intervals for the predictor. Through simulation studies, we illustrate the asymptotic normality of the proposed predictors. In data analysis, we compare the prediction accuracies of exponential smoothing methods with our proposed weights using monthly data on foreign visitors to Japan from four countries. This is a joint work with Kazuki Koizumi and Yan Liu.

Taniguchi, Masanobu

Title: Second order generalization of Hajek's convolution theorem and its applications

Abstract: For a class of regular estimators, Hajek showed that the contiguous asymptotic distribution of regular estimator converges to a convolution of distribution of efficient estimator and the residual distribution, called as the Hajek convolution theorem. This provides the foundation of asymptotic efficiency etc of regular estimators. In this talk, introducing a class of second-order regular estimators by the valid Edgeworth expansion, we derive its second-order contiguous distribution, which is a convolution of the second-order efficient distribution and its second-order residual distribution, which is a second-order generalization of Hajek's convolution theorem. Then, we discuss the second-order efficiency etc. Also we introduce a second-order robustness for the second-order regular estimators. For a class of general minimum contrast estimators in time series models, the second-order robustness will be evaluated. Cowork with J.Hirukawa & M.Hallin.