Workshop on causal inference in complex marine ecosystems

Date: 11-12 June 2019
Venue: Institute of Marine Research, Bergen
11 June, Tuesday: SG196 Nordishavet
12 June, Wednesday: Pynten

This workshop is supported by:
Japan Scoiety for the Promotion of Science Kiban (S) No.18 H05290 (Prof.Taniguchi, M),
Waseda Research Institute for Science & Engineering, Institute for Mathematical Science,
and The Institute of Marine Research

Program

11 June SG196 Nordishavet

10:00 – 10:10 Opening Solvang, Hiroko, IMR
10:10 – 10:25 Planque, Benjamin, IMR
Evaluating the performance of causal inference for marine ecological systems
10:25 – 10:40 Subbey, Sam, IMR
Challenges of Scale and Granularity in Quantifying Causality in Marine Ecosystems

Chair Michoel, Tom

10:40 – 11:40 Hannisdal, Bjarte, University of Bergen, Norway
Earth system interactions from geological records: a transfer operator approach

Lunch break

Chair Parviainen, Pekka

13:00 – 13:40 Akashi, Fumiya, The University of Tokyo, Japan
Hybrid GEL test for rotational symmetry on spheres

13:40 – 14:20 Yan Liu, Kyoto University, Japan
Sphericity test for high-dimensional time series

14:20 – 15:00 Dou, Xiaoling, Waseda University, Japan
Testing for Granger Causality by Use of Box-Cox Transformations

Coffee break

Chair Subbey, Sam

15:30 – 16:30 Sullivan, Patrick, Cornel University, USA
Is Knowing Causality Necessary, Sufficient or Even Useful? Reflections on Alternative Modelling Philosophies

16:30 – 17:30 Liudas Giraitis, Queen Mary University of London, UK
Standard testing procedures for white noise and heteroskedasticity

18:30 – Dinner

12 June Pynten

Chair Vansteelandt, Stijn

10:00 – 11:00 Taniguchi, Masanobu, Waseda University, Japan
Joint circular distribution in view of higher order spectra

11:00 – 12:00 Tjøstheim, Dag, University of Bergen, Norway
Local Gaussian partial correlation, conditional independence and causality

Lunch break

Chair Sullivan, Patrick

13:00 – 13:40 Pierre, Maud, National Research Institute of Science and Technology for Environment and Agriculture, France
EDM (Empirical Dynamical Modelling) methods applied in aquatic ecology: different case studies

14:00 – 15:00 Vansteelandt, Stijn, Ghent University, Belgium
How to estimate the effect of a time-varying exposure?

Coffee break

Chair Sullivan, Patrick

15:30 – 16:30 Parviainen, Pekka, University of Bergen, Norway
Bayesian networks and causality

16:30 – 17:30 Michoel, Tom, University of Bergen, Norway
Causal inference in systems genetics

17:30 – 17:40 Closing Solvang, Hiroko, IMR

18:30 – Dinner

Abstract

Planque, Benjamin Institute of Marine Research, Norway

Title: Evaluating the performance of causal inference for marine ecological systems

Abstract: Developing ecological predictive capacity requires understanding, description and quantification of causal relationships in ecological systems. Numerical methods exist to explore causal relationships and the field of causality research is rapidly evolving. In this presentation, I will focus on the application of causality research to the case of sea fisheries and large marine ecosystems. These systems are often characterised by complicated structures, multiple interactions and limited observations. This makes them particularly difficult to analyse and model. I will discuss the challenges ahead for causality research in marine ecological systems and the importance of evaluation tools to measure the performance of causality methods for the specific case of marine ecological systems.

Subbey, Sam Institute of Marine Research, Norway

Title: Challenges of Scale and Granularity in Quantifying Causality in Marine Ecosystems

Abstract: Observation data on causal phenomena in marine ecosystems are often collected on different spatial and temporal scales, and granularity. Quantitative methods for quantifying causal flows use the data either as processed observations (e.g. acoustic data) or information derived from computational models (e.g., population sizes).

This talk will highlight potential computational challenges that need to be addressed in quantifying causality for marine ecosystems, given data limitations in scale and granularity. Illustrative examples will be presented using data from the Barents Sea ecosystem.

Hannsidal, Bjarte University of Bergen, Norway

(co-author: Haaga, Kristian A. and Diego**,** David)

Title: Earth system interactions from geological records: a transfer operator approach

Abstract: Global environmental projections raise the spectre of climate tipping points, sea level rise, and ecological disruption on a scale that is unprecedented in the history of human civilization. Researchers therefore turn to geological archives for insights into causes, consequences, and time scales of global change in the past. Preserved geological time series typically provide indirect access to only a very small subset of the vast complexity of interacting Earth system components. Hence, a dynamical systems reconstruction approach to causal inference is particularly attractive.

Here we present a method for computing the transfer entropy between time series using Ulam's approximation of the Perron-Frobenius (transfer) operator associated with the map generating the dynamics. Unlike standard transfer entropy estimators, the invariant measure is estimated not directly from the data points but from the invariant distribution of the transfer operator approximated from the data points. Our method shows promise for detecting dynamical coupling in relatively sparse time series, exemplified by a case study of paleoenvironmental proxy records.

Next, we present a causal asymmetry index based on computing transfer entropy over a range of prediction lags, both negative (non-causal) and positive (causal). If an underlying directional coupling exists, then the transfer entropy corresponding to positive lags will exceed that corresponding to negative lags. This asymmetry criterion provides a suprisingly robust estimate of directional

influence in a wide range of dynamical systems, with as little as a few hundred observations or less. Thus, in conjunction with surrogate testing, the causal asymmetry index may be a valuable tool for detecting directional influences from observed time series.

We briefly showcase the implementation of these methods in our open-source CausalityTools.jl Julia package. Finally, we highlight the tantalizing prospect of using the transfer operator to improve the statistical characterization of an unknown system from observed time series, and discuss possible future extensions.

Akashi, Fumiya **The Universiy of Tokyo, Japan**

Title: Hybrid GEL test for rotational symmetry on spheres

Abstract: This talk considers a nonparametric test for rotational symmetry of directional data. Most of classical density functions on the unit spheres share the common important feature called rotational symmetry. However, recently some authors found real data which do not satisfy this condition, and Ley and Verdebout (J. Multivariate Anal. 2017, 159:67-81) proposed a family of skew-rotationally-symmetric distributions on spheres. On the other hand, it is often severe to assume certain parametric family for real data. To overcome such hurdle, this talk employees a measure of skewness proposed by Mardia (Biometrika 1970, 57(3):519-530), and uniformity test proposed by Watson (1983, Wiley) for projected observations. This talk integrates these two methods into a hybrid-type generalized empirical likelihood statistic for the rotational symmetry test. Unlike the classical research, the proposed method do not require any parametric assumption for the underlying model. Some simulation experiments also illustrate finite sample performance of the proposed method.

Liu, Yan **Kyoto Universiy, Japan**

(co-author: Tamura, Yurie and Taniguchi, Masanobu)

Title: Sphericity test for high-dimensional time series

Abstract: We consider the testing problem for the sphericity hypothesis regarding the covariance matrix of high-dimensional time series. Recently, test statistics for sphericity of independent and identically distributed high-dimensional random variables have been studied under the condition that both the sample size n and the dimension p diverge to infinity. A test statistic for sphericity has been proved to be well-behaved even when p is larger than n. We investigate the test statistic under the situations of high-dimensional time series. The asymptotic null distribution of the test statistic is shown to be standard normal when the observations come from Gaussian stationary processes. In the simulation study, we illustrate the properties of the test statistic for several high-dimensional time series models. In our empirical study, we apply the test to the portfolio selection problem.

Dou, Xiaoling **WasedaUniversiy, Japan**

(co-author: Koike, Ryunosuke and Taniguchi, Masanobu)

Title: Testing for Granger Causality by Use of Box-Cox Transformations

Abstract: Granger causality test is a statistical hypothesis test for asserting the predictive causality between two time series. Granger causality can be tested by using an F-test assuming a VAR model. However, since in financial economics, many financial variables are not normally distributed, an F-test based on the raw data may produce misleading results. When the data are not normal distributed, we propose to transform the distribution of data into nearly normal and then use the F-test. In our paper,

we use Box-Cox transformation and the Yeo-Johnson power transformation. We briefly introduce the simultaneous estimation of the parameters of transformations and AR models, and apply the methods to two real data sets. We test the Granger causalities of 5 areas of Australia with the number of cases of flu and the Granger causalities of the number of tourists visiting Japan and appreciation of the yen as an application in tourism statistics.

Sullivan, Patrick J Cornel University, USA

Title: Is Knowing Causality Necessary, Sufficient or Even Useful? Reflections on Alternative Modelling Philosophies

Abstract: As George E. P. Box famously put it, "All models are wrong, but some are useful". With the growth of computational power, the continual evolvement of modeling methods and the increasing volume and availability of data, one might ask to what degree one needs to know cause. Obviously, for fully understanding mechanisms, processes, relationships and flows, understanding underlying causes and effects are centrally important. For prediction, risk assessment and decision making, perhaps demonstrating cause isn't as important. Then there is also the question of whether cause should be inferred from a model; especially if "all models are wrong". Einstein was able to show instances where Newtonian models will fail. Still, those models are useful (e.g. building bridges, dams, airplanes), but could those models have been used to predict the Einstein's discoveries on relativity? Knowing causalities is useful, inferring causality is yet another matter. Should we worry about such things?

Giraitis, Liudas Queen Mary University, UK

Title: Standard testing procedures for white noise and heteroskedasticity

Abstract: Commonly used tests to assess evidence for the absence of autocorrelation or serial correlation between time series in applied work rely on procedures whose validity holds for i.i.d. data. When the series are not i.i.d., the size of correlogram and cumulative Ljung-Box tests can be significantly distorted. This paper adapts standard correlogram tests to accommodate hidden dependence and non-stationarities involving heteroskedasticity, thereby uncoupling these tests from limiting assumptions that reduce their applicability in empirical work. To enhance the Ljung-Box test for non i.i.d. data a new cumulative test is introduced. Asymptotic size of these tests is unaffected by hidden dependence and heteroskedasticity in the series. An extensive Monte Carlo study confirms good performance in both size and power for the new tests. Applications to real data reveal that standard tests frequently produce spurious evidence of serial correlation.

Taniguchi, Masanobu Waseda University, Japan

(co-author: Kato, Shogo, Ogata, Hiroaki and Pewsey, Arthur)

Title: Joint circular distributions in view of higher order spectra of time series

Abstract: Circular data analysis is emerging as an important component of statistics. For this half century, various circular distributions have been proposed, e.g., von Mises distribution, wrapped Cauchy distribution, among other things.

Also, regarding the joint distribution, Wehrly and Johnson (1980) proposed a bivariate circular distribution which is related to a family of Markov processes on the circle.

Because the sample space is on a circle, various new statistical methods have been developed. In this talk we provide a new look at circular distributions in view of spectral distributions of time series because the typical circular distributions correspond to spectral densities of time series models. For

example, autoregressive AR(1) spectral density corresponds to wrapped Cauchy distribution, and von Mises distribution corresponds to exponential spectral density (Bloomfield(1973)), etc.
Furthermore, we introduce a class of joint circular distributions from the higher order spectra of time series, which can describe very general joint circular distributions. Hence, we can develop the statistical inference for dependent observations on the circle. We present a family of distributions on the circle derived from the ARMA spectral density. It is seen that the proposed family includes some existing circular families as special cases. For these special cases, the normalizing constant and trigonometric moments are shown to have simple and closed form. We develop the asymptotic optimal inference theory based on the local asymptotic normality (LAN) on the circle. Because the observations are permitted to be dependent, the theory opens a new paradigm in the estimation for joint circular distributions.

Tjøstheim, Dag B University of Bergen, Norway

Title: Gaussian partial correlation: conditional independence and causality

Abstract: A very important tool in learning the structure of a network or graph is the concept of conditional independence. For Gaussian variables the partial correlation function can be used to test for conditional independence, and one has conditional independence if and only if the appropriate partial correlation is zero. Unfortunately, the partial correlation function as a measure of conditional dependence may fail spectacularly for non-Gaussian and nonlinear relationships. One reason for this is that it is essentially based on moments. To handle nonlinear cases distribution theory is needed. We use a distributional approach, where an arbitrary distribution is replaced by a family of Gaussian distributions, and where each Gaussian member approximates the original distribution in a local region of that distribution. This makes it possible to locally exploit the uniquely useful properties of the multivariate and conditional Gaussian. In particular, we introduce a local Gaussian partial correlation function (LGPC). A number of properties of the LGPC is given, and it is shown how the LGPC can measure and visualize conditional dependence between random variables, and how it can be used to test for conditional independence. Applications will be given to the testing and detecting of Granger causality. Further, we will indicate potential applications to causality theory on graphs and networks.

Pierre, Maud National Research Institute of Science and Technology for Environment and Agriculture, France

Title: EDM (Empirical Dynamical Modelling) methods applied in aquatic ecology: different case studies

Abstract: Studying causal relationships has been a long standing issue in ecology. Ecological time series can be considered as the result of complex dynamic systems (i.e. nonlinear), which reflect the evolution of dynamic behaviour over time. Therefore, linear statistical approaches are inappropriate for this kind of data. Recent methods - flexible, non-linear and non-parametric - based on dynamics encoded in time series (empirical dynamic modelling - EDM) have been developed to allow the investigation of causal links. The use of these promising methods will be addressed through different case studies.

Vansteelandt, Stijn Ghent University, Belgium

Title: How to estimate the effect of a time-varying exposure?

Abstract: When estimating the effect of a time-varying treatment or exposure on an outcome of interest measured at a later time, standard regression-based methods have been shown to give biased

results. While they need to adjust for so-called time-varying confounders (i.e., common causes of exposure and outcome, such as the history of previous exposures and outcomes), they are unable to make a valid adjustment when those confounders are themselves affected by earlier exposures. This is often the case in practice; it occurs for instance whenever there are dynamic feedback relations with exposure and outcome mutually influencing each other over time. In this talk, I will develop insight into this problem, and will review 3 solutions proposed by James Robins and colleagues, which have become very popular in epidemiology and beyond. They include the g-computation formula, inverse probability weighted estimation of marginal structural models and g-estimation of structural nested models.

Parviainen, Pekka University of Bergen, Norway

Title: Bayesian networks and causality

Abstract: Bayesian networks are compact representations of multivariate probability distributions. Under certain assumptions, they can represent causal relationships between variables. In this talk, I give a short introduction to Bayesian networks and how they can be used in causal analysis.

Michoel, Tom University of Bergen, Norway

Title: Causal inference in systems genetics

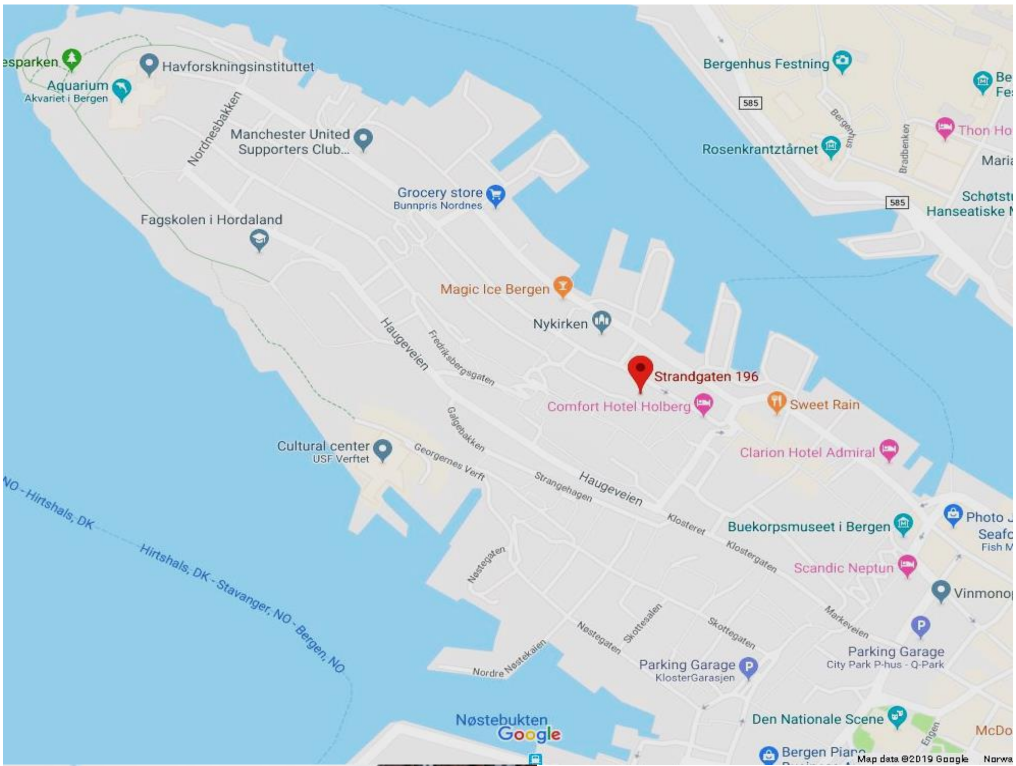
Abstract: Abstract: Correlation does not imply causation, so what does imply causation? To transform patterns of co-expression among transcripts, proteins, metabolites and phenotypes into predictive models of biological systems, causal inference is crucial. In genetics, the random segregation of alleles allows to conduct massively parallel randomized experiments, where the direction of causality between co-expressed genes can be inferred from their joint genetic linkage to cis-regulatory DNA sequences. While the general principle of causal inference in this field of "systems genetics" is well-established, numerous challenges remain. To deal with the vast amount of data that can now be generated for ever more samples, advances in software engineering are needed. To account for multiple levels of known and unknown confounding factors and noise in the data, development of more sensitive statistical models is required. I will present ongoing work in my group on the development of new methods and software for causal inference from systems genetics data and demonstrate how they have contributed to an improved understanding of the tissue-specific effects of genetic risk factors for cardiovascular and metabolic disease.

Speakers List

Akashi, Fumiya	The University of Tokyo, Japan	akashi@e.u-tokyo.ac.jp
Dou, Xiaoling	Waseda University, Japan	xiaoling@aoni.waseda.jp
Giraitis, Liudas	Queen Mary University, UK	l.giraitis@qmul.ac.uk
Hannisdal, Bjarte	University of Bergen, Norway	bjarte.hannisdal@uib.no
Liu, Yan	Kyoto University, Japan	yan.liu@i.kyoto-u.ac.jp
Michoel, Tom	University of Bergen, Norway	tom.michoel@uib.no
Parviainen, Pekka	University of Bergen, Norway	Pekka.Parviainen@uib.no
Planque, Benjamin	IMR, Norway	benjamin.planque@hi.no
Pierre, Maud	National Research Institute of Science and Technology for Environment and Agriculture, France	maud.pierre@irstea.fr
Solvang, Hiroko,	IMR, Norway	hirokos@hi.no
Subbey, Sam	IMR, Norway	samuel.subbey@hi.no
Sullivan, Patrick	Cornel University, USA	pjs31@cornell.edu
Taniguchi, Masanobu	Waseda University, Japan	taniguchi@waseda.jp
Tjøstheim, Dag B	University of Bergen, Norway	Dag.Tjostheim@math.uib.no
Vansteelandt, Stijn	Ghent University, Belgium	stijn.vansteelandt@ugent.be

Meeting room

(11 June) Nordishavet address: Strandgaten 196



(12 June) Pynten address: Nordnesparken