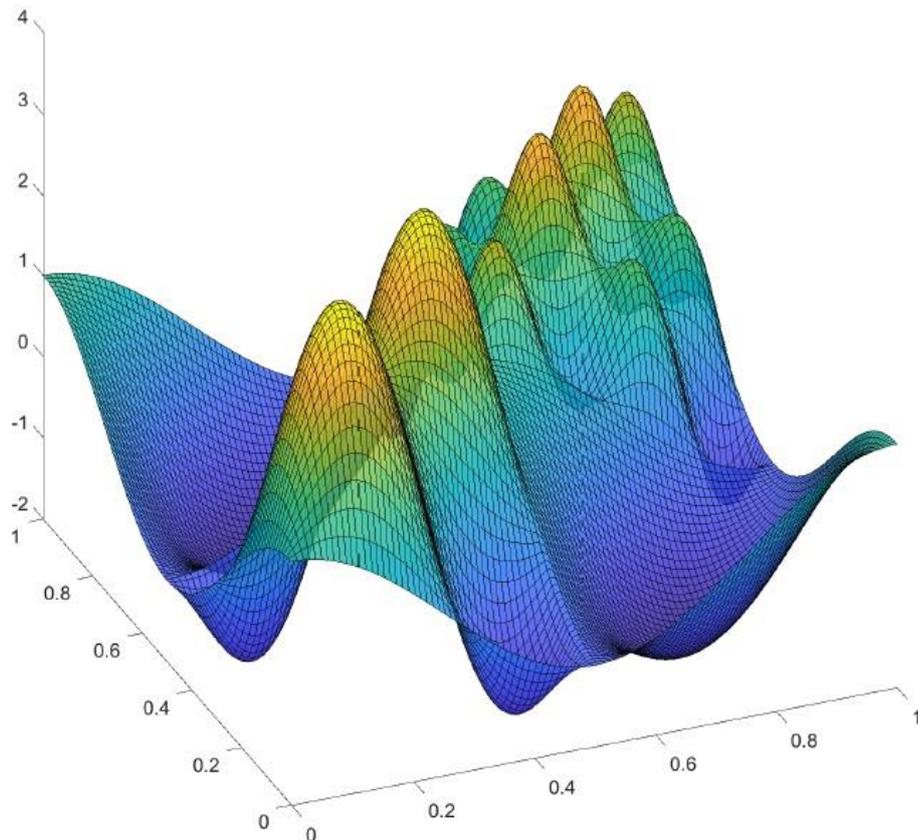# Functional Data Analysis and Beyond Conference

Programme and Abstracts

MATRIX/ACEMS, Melbourne
8 – 9 December 2018

ACEMS/MATRIX Conference on Functional Data Analysis

Saturday, 8 December

FDA and Regression	
9 :00 am – 9 :45 am	*Predicting One-day-ahead Wind Power Capacity Factor via Functional Inverse Regression* **C.-R. Jiang (**Academia Sinica)
9 :45 am – 10 :30 am	*Distributed Estimation of Functional Linear Regression* **H. Lian** (University of Hong Kong)
10 :30 am – 11 :00 am	Coffee break
FDA and Dynamics	
11 :00 am – 11 :45 am	*From Brain to Hand to Statistics with Dynamic Smoothing* **J. Ramsay** (University of Ottawa)
11 :45 am – 12 :30 am	*Estimation of temperature-dependent growth profiles of fly larvae for criminology* **F. Ferraty** (Toulouse University)
12 :30 am – 2 :00 pm	Lunch break
Applications of FDA	
2 :00 pm – 2 :45 pm	*Differential Privacy in Functional Data Analysis* **M. Reimherr** (Pennsylvania State University)
2 :45 pm – 3 :30 pm	*A new knn-classifier for functional data with applications* **J.-T. Zhang** (National University of Singapore)
3 :30 pm – 4 :00 pm	Coffee break
Modeling FDA	
4 :00 pm – 4 :45 pm	*A two-stage procedure for detecting multiple change points in functional data sequence* **J.-M. Chiou** (Academia Sinica)
4 :45 pm – 5 :30 pm	*Reconstructing partially observed functional data with (non-) systematically missing part* **D. Liebl** (Bonn University)

MATRIX **ACEMS** — AUSTRALIAN RESEARCH COUNCIL CENTRE OF EXCELLENCE FOR MATHEMATICAL AND STATISTICAL FRONTIERS

ACEMS/MATRIX Conference on Functional Data Analysis

Sunday, 9 December

FDA and Spatial DATA	
9 :00 am – 9 :45 am	*Functional linear spatial autoregressive modeling* **S. Dabo-Niang** (University of Lille 3)
9 :45 am – 10 :30 am	*Dynamic modeling with Data2PDE* **M. Carey** (University College Dublin)
10 :30 am – 11 :00 am	Coffee break
FDA and measurement errors	
11 :00 am – 11 :45 am	*Functional Regression on Manifold with Contamination* **F. Yao**
11 :45 am – 12 :15 am	*Instrumental Variable Approach to Estimating the Scalar-on-Function Regression Model with Measurement Error with Application to Energy Expenditure Assessment in Childhood Obesity* **C. Tekwe** (Texas A&M University)
12 :15 am – 1 :45 pm	Lunch break
FDA and testing	
1 :45 pm – 2 :30 pm	*A New Test for Functional One-Way ANOVA with Application to Ischemic Heart Screening* **M-Y. Cheng**
2 :30 pm – 3 :15 pm	*Depth-based nonparametric tests for homogeneity of functional data* **G. Geenens**
3 :15 pm – 3 :45 pm	Coffee break
FDA and Forecasting	
3 :45 pm – 4 :30 pm	*Testing for stationarity of functional time series in the freqency domain* **A. Aue** (University of California, Davis)
4 :30 pm – 5 :15 pm	*Seasonal functional autoregressive models* **Rob Hyndman** (Monash University)
5 :15 pm – 5 :45 pm	*Simultaneous estimation of seasonal long memory parameters of functional time series* **A. Olenko** (La Trobe University)

INVITED TALKS

Speaker: Alexander Aue (University of California at Davis)

Title: Testing for stationarity of functional time series in the frequency domain

Abstract: Interest in functional time series has spiked in the recent past with papers covering both methodology and applications being published at a much-increased pace. This talk contributes to the research in this area by proposing stationarity tests for functional time series based on frequency domain methods. Setting up the tests requires a delicate understanding of periodogram- and spectral density operators that are the functional counterparts of periodogram- and spectral density matrices in the multivariate world. Two sets of statistics are proposed. One is based on the eigendecomposition of the spectral density operator, the other on a fixed projection basis. Their properties are derived both under the null hypothesis of stationary functional time series and under the smooth alternative of locally stationary functional time series. The methodology is theoretically justified through asymptotic results. Evidence from simulation studies and an application to annual temperature curves suggests that the tests work well in finite samples. The talk is based on joint work with Anne van Delft (Ruhr-Universität Bochum)

Speaker: Michelle Carey (University College Dublin)

Title: Dynamic modeling with Data2PDE

Abstract: Geo spatial data are observations of a process that are collected in conjunction with reference to their geographical location. This type of data is abundant in many scientific fields, some examples include: population census, social and demographic (health, justice, education), economic (business surveys, trade, transport, tourism, agriculture, etc.) and environmental (atmospheric and oceanographic) data. They are often distributed over irregularly shaped spatial domains with complex boundaries and interior holes. Modeling approaches must account for the spatial dependence over these irregular domains as well as describing their temporal evolution.

Dynamic systems modeling has a huge potential in statistics, as evidenced by the amount of activity in functional data analysis. Many seemingly complex forms of functional variation can be more simply represented as a set of differential equations, either ordinary or partial.

In this talk, I will present a class of semi parametric regression models with differential regularization in the form of PDEs. This methodology will be called Data2PDE "Data to Partial Differential Equations". Data2PDE characterizes spatial processes that evolve over complex geometries in the presence of uncertain, incomplete and often noisy observations and prior knowledge regarding the physical principles of the process characterized by a PDE.

Speaker: Ming-Yen Cheng (Hong Kong Baptist University)
Title: A New Test for Functional One-Way ANOVA with Application to Ischemic Heart Screening

Abstract: Motivated by an ischemic heart screening problem, a new global test for one-way ANOVA in functional data analysis is studied. The test statistic is taken as the maximum of the pointwise F-test statistic over the interval the functional responses are observed. Nonparametric bootstrap, which is applicable in more general situations and easier to implement than parametric bootstrap, is employed to approximate the null distribution and to obtain an approximate critical value. Under mild conditions, asymptotically our test has the correct level and is root-n consistent in detecting local alternatives. Simulation studies show that the proposed test outperforms several existing tests in terms of both size control and power when the correlation between observations at any two different points is high or moderate, and it is comparable with the considered competitors otherwise. Application to an ischemic heart data set suggests that resting electrocardiogram signals may contain enough information for ischemic heart screening at outpatient clinics, without the help of stress tests required by the current standard procedure.

Speaker: Jeng-Min Chiou (Academia Sinica)

Title: A two-stage procedure for detecting multiple change points in functional data sequence

Abstract: We introduce a two-stage procedure for identifying multiple changes in the mean functions for a functional data sequence. The method first searches for change point candidates by the dynamic segmentation that recursively adjusts the endpoints of the subsequences by an optimality criterion. Then, it iteratively removes unlikely change point candidates until all the selected change points are statistically significant. We show the consistency property of the algorithm, illustrate the method by a traffic data application, and examine its practical performance through a simulation study.

Speaker: Sophie Dabo-Niang (University of Lille 3)

Title: Functional linear spatial autoregressive modeling

Abstract: Complex issues arise in spatial statistics and econometrics (statistical techniques to address economic modeling), many of which are neither clearly defined nor completely resolved but form the basis for current research. Among the practical considerations that influence the available methods used in spatial data modeling, particularly in econometrics, is data dependency. In fact, spatial data are often dependent, and a spatial model must be able to account for this characteristic. Linear spatial models, which are common in geostatistical modeling, generally impose a dependency structure model based on linear covariance relationships between spatial locations. However, under many circumstances, the spatial index does not vary continuously and may be of the lattice type, the baseline of this current talk.
This is, for instance, the case in a number of problems. In images analysis, remote sensing form satellites, agriculture and so one, data are often received as regular lattice and identified as the centroids of square pixels, whereas a mapping forms often an irregular lattice. Basically, statistical

models for lattice data are linked to nearest neighbors to express the fact that data are nearby. We are concerned here about spatial functional models for lattice data. We consider a spatial functional linear model with a random functional covariate and a real-valued response using spatial autoregression on the response based on a weight matrix.

We investigate parameter identification and asymptotic properties of the quasi-maximum likelihood estimator of the functional parameter using the so-called increasing domain asymptotics. We provide identification conditions combining that in the classical spatial autoregressive model (SAR) model and in the functional linear model. Monte Carlo experiments illustrate the performance of the QML estimation.

Invited Speaker: Frederic Ferraty

Title: Estimation of temperature-dependent growth profiles of fly larvae with application to criminology

Abstract: It is not unusual in cases where a body is discovered that it is necessary to determine a time of death or more formally a post mortem interval (PMI). Forensic entomology can be used to estimate this PMI by examining evidence obtained from the body from insect larvae growth. In this talk, we propose a method to estimate the hatching time of larvae (or maggots) based on their lengths, the temperature profile at the crime scene and experimental data on larval development. This requires the estimation of a time-dependent growth curve from experiments where larvae have been exposed to a relatively small number of constant temperature profiles. Since the temperature influences the developmental speed, a crucial step is the time alignment of the curves at different temperatures. We then propose a model for time varying temperature profiles based on the local growth rate estimated from the experimental data. This allows us to find out the most likely hatching time for a sample of larvae from the crime scene. We explore via simulations the robustness of the method to errors in the estimated temperature profile and apply it to the data from two criminal cases from the United Kingdom. Asymptotic properties are also provided for the estimators of the growth curves and the hatching time.

(Joint work with D. Pigoli, J.A.D. Aston, A. Mazumder, C. Richards and M.J.R. Hall)

Speaker: Gery Geenens (University of New South Wales)

Title: Depth-based nonparametric tests for homogeneity of functional data

Abstract: In this work we study some tests for the homogeneity between two independent samples of functional data. The null hypothesis of "homogeneity" here means that the latent stochastic processes which generated the two samples have the same distribution. Most instances of functional data are so complex that it seems natural to opt for nonparametric procedures in this setting. Making use of recent developments on functional depths, we adapt some Kolmogorov-Smirnov-and-Cramer-von-Mises-type of criteria to the functional context. Exact p-values for the test can be obtained via permutations, or, in case of too large samples, a bootstrap algorithm is easily implemented. Some real data examples are analyzed.

Speaker: Rob Hyndman (Monash University)

Title: Seasonal functional autoregressive models

Abstract: Functional autoregressive models have been widely used in functional time series analysis, but no attention has been given to handling seasonality within this framework. I will discuss a proposed seasonal functional autoregressive model and explore some of its statistical properties including stationarity conditions and limiting behaviour. I will also look at methods for estimation and prediction of seasonal functional autoregressive time series of order one. The ideas will be illustrated using simulation studies and real data.

Speaker: Ci-Ren Jiang (Academia Sinica)

Title: Predicting One-day-ahead Wind Power Capacity Factor via Functional Inverse Regression

Abstract: Inverse regression is an appealing dimension reduction method for regression models with multivariate covariates. Recently, it has been extended to the cases with functional or longitudinal covariates. However, the extensions simply focus on one single functional or longitudinal covariate. Motivated by a real application, we extend functional inverse regression to the cases with multiple functional covariates, whose domains could be different. The asymptotical properties of the proposed estimators are investigated. The computational issues are taken care with data binning, the fast Fourier transformation and random projections on a multi-core computation platform. In addition to simulation studies, the proposed approach is applied to predict the one-day-ahead wind power capacity factors in Germany from 2016 to 2017. Both demonstrate the good performance of our method.

Speaker: Heng Lian

Title: Distributed Estimation of Functional Linear Regression

Abstract: We consider distributed computation in fitting functional linear regression, with a very large sample size so that a centralized estimation may be infeasible. The covariates are functional and both scalar and functional responses are considered. We show that if the tuning parameter is chosen based on the size of the entire data, the aggregated estimator has the same (optimal) convergence rate as the estimator based on the entire data. Some numerical illustrations are presented to demonstrate the performance of the distributed estimators.

Speaker: Dominik Liebl (University of Bonn)

Title: Reconstructing partially observed functional data with (non-)systematically missing parts

Abstract: The first part of the talk considers the case of partially observed functional data with non-systematically missing parts. A new reconstruction operator is proposed which aims to recover the missing parts of a function given the observed parts. This new operator belongs to a

new, very large class of functional operators which includes the classical regression operators as a special case. The optimality of our reconstruction operator is shown, and it is demonstrated that the usually considered regression operators generally cannot be optimal reconstruction operators. The estimation theory allows for autocorrelated functional data and considers the practically relevant situation in which each of the n functions is observed at m_i, $i=1,?,n$, discretization points plus noise. Rates of consistency are derived for the nonparametric estimation procedures using a double asymptotic. The second part of the talk proposes new estimators for the mean and the covariance function for partially observed functional data using a detour via the fundamental theorem of calculus. These new estimators allow for a consistent estimation of the mean and covariance function under specific violations of the missing-completely-at-random assumption. A simulation study compares the new estimators with the classical estimators from the literature in different missing data scenarios.

Speaker: Jim Ramsay (McGill University)

Title: From Brain to Hand to Statistics with Dynamic Smoothing

Abstract: Discrete observations of curves are often smoothed by attaching a penalty to the error sum of squares, and the most popular penalty is the integrated squared second derivative of the function that fits the data. But it has been known since the earliest days of smoothing splines that, if the linear differential operator D^2 is replaced by a more general differential operator L that annihilates most of the variation in the observed curves, then the resulting smooth has less bias and greatly reduced mean squared error.

This talk will show how we can use the data to estimate such a linear differential operator for a system of one or more variables. The differential equations estimated in this way represents the dynamics of the processes being estimated. This idea can be used to estimate a forcing function that defines the output of a linear system and apply this to handwriting data to show that both the static and dynamic aspects of handwriting are well represented by a surprisingly simple second order differential equation.

Speaker: Fang Yao (University of Toronto/Peking University)

Title: Functional Regression on Manifold with Contamination (by Zhenhua Lin and **Fang Yao**)

Abstract: We propose a new perspective on functional regression with a predictor process via the concept of manifold that is intrinsically finite-dimensional and embedded in an infinite-dimensional functional space, where the predictor is contaminated with discrete/noisy measurements. By a novel method of functional local linear manifold smoothing, we achieve a polynomial rate of convergence that adapts to the intrinsic manifold dimension and the level of sampling/noise contamination with a phase transition phenomenon depending on their interplay. This contrasts with the logarithmic convergence rate in the literature of functional nonparametric regression. We demonstrate that the proposed method enjoys favorable finite sample performance relative to commonly used methods via simulated and real data examples.

Speaker: Jin-Ting Zhang (National University of Singapore)

Title: A new knn-classifier for functional data with applications

Abstract: In this talk, we discuss a new *knn* (k-nearest neighbors) classifier for functional data. For supervised classification of functional data, several classifiers have been proposed in the literature, including the well-known classic *knn* classifier. The classic knn classifier selects k nearest neighbors around a new observation and determines its class-membership according to a majority vote. A difficulty arises when there are two classes having the same largest number of votes. To overcome this difficulty, we propose a new *knn* classifier which selects *k* nearest neighbors around a new observation from each class. The class-membership of the new observation is determined by the minimum average distance or semi-distance between the k nearest neighbors and the new observation. Good performance of the new *knn* classifier is demonstrated by simulation studies and real data examples.

CONTRIBUTED TALKS

Speaker: Carmen Tekwe (Texas A&M University)

Title: Instrumental Variable Approach to Estimating the Scalar-on-Function Regression Model with Measurement Error with Application to Energy Expenditure Assessment in Childhood Obesity

Abstract: Wearable device technology allows continuous monitoring of biological markers and thereby enables study of time-dependent relationships. For example, in this paper, we are interested in the impact of daily energy expenditure over a period of time on subsequent progression toward obesity among children. Data from these devices appear as either sparsely or densely observed functional data and methods of functional regression are often used for their statistical analyses. We study the scalar-on-function regression model with imprecisely measured values of the predictor function. In this setting, we have a scalar-valued response and a function-valued covariate that are both collected at a single time period. We propose a generalized method of moments-based approach for estimation while an instrumental variable belonging in the same time space as the imprecisely measured covariate is used for model identification. Additionally, no distributional assumptions regarding the measurement errors are assumed, while complex covariance structures are allowed for the measurement errors in the implementation of our proposed methods. We demonstrate that our proposed estimator is L2 consistent and enjoys the optimal rate of convergence for univariate nonparametric functions. In a simulation study, we illustrate that ignoring measurement error leads to biased estimations of the functional coefficient. The simulation studies also confirm our ability to consistently estimate the function-valued coefficient when compared to approaches that ignore potential measurement errors. Our proposed methods are applied to our motivating example to assess the impact of baseline levels of energy expenditure on BMI among elementary school-aged children.

Speaker: Andriy Olenko (La Trobe University)

Title: Simultaneous estimation of seasonal long memory parameters of functional time series

Abstract: We study functional time series with spectral singularities outside the origin. These time series can be used to model seasonal long-memory behaviour. For a semiparametric statistical model new simultaneous estimates for singularity location and long-memory parameters are proposed. The approach is based on general filter transforms that include wavelet transformations as a particular case. It is proved that the estimates are almost surely convergent to the true values of parameters. Solutions of the estimation equations are studied, and adjusted statistics are proposed. Monte-Carlo study results are presented to confirm the theoretical findings.