# Abstracts

September 16, 2022

1. **Bobrowski, Omer**, Technion Haifa, Israel.
*Joint authors:*

**Title:** Persistent homology, time series, and universality

**Abstract:** Topological data analysis (TDA) promotes the use of qualitative structural information in data and network analysis. The main workhorse in TDA, called Persistent Homology, is a tool that provides multi-scale information about various types of "holes" that may appear in the data. Persistent homology has been successfully applied in a wide variety of data analytic problems such as neuroscience, genomics, material sciences, cosmology, and more.

In the context of time series, we can use time-delay embeddings to transform sequences into high-dimensional point-clouds. We can then use persistent homology to analyze the structure of the underlying dynamical system. This method has been proved useful in detecting periodic and quasi-periodic patterns, bifurcations, and chaotic behavior. It also serves as a powerful feature extraction tool, for various signal classification tasks. In the first part of the talk, we will provide a quick introduction to persistent homology and its use in analyzing time series.

One of the most elusive challenges in TDA is understanding the distribution of noise in persistent homology. Despite much effort, this is still largely an open problem. In the second part of the talk we will present new conjectures regarding the noise distribution in persistent homology. Our main claim is that, viewed in the right way, persistence values obey a universal probability law, that does not depend on the underlying system. We will present this phenomenon in simulated data, as well as in real data, including recorded time-series. We will also discuss the powerful potential of this newly discovered behavior in applications.

2. **Bolin, David**, KAUST, Saudi Arabia.
*Joint authors:*

**Title:** Gaussian Whittle-Matérn fields on metric graphs

**Abstract:** We define a new class of Gaussian processes on compact metric graphs such as street or river networks. The proposed models, the Whittle-Matérn fields, are defined via

a fractional stochastic partial differential equation on the compact metric graph and are a natural extension of Gaussian fields with Matérn covariance functions on Euclidean domains to the non-Euclidean metric graph setting. Existence of the processes, as well as their sample path regularity properties are derived. The model class in particular contains differentiable Gaussian processes. To the best of our knowledge, this is the first construction of a valid differentiable Gaussian field on general compact metric graphs. We then focus on a model subclass which we show contains processes with Markov properties. For this case, we show how to evaluate finite dimensional distributions of the process exactly and computationally efficiently. This facilitates using the proposed models for statistical inference in big data settings without the need for any approximations.

3. **Caponera, Alessia**, EPFL, Switzerland.
*Joint authors:* J. Fageot, M. Simeoni, V. M. Panaretos

**Title:** Functional Estimation of Anisotropic Covariance and Autocovariance Operators on the Sphere

**Abstract:** We propose nonparametric estimators for the second-order central moments of possibly anisotropic spherical random fields, within a functional data analysis context. We consider a measurement framework where each random field among an identically distributed collection of spherical random fields is sampled at a few random directions, possibly subject to measurement error. The collection of random fields could be i.i.d. or serially dependent. Though similar setups have already been explored for random functions defined on the unit interval, the nonparametric estimators proposed in the literature often rely on local polynomials, which do not readily extend to the (product) spherical setting. We therefore formulate our estimation procedure as a variational problem involving a generalized Tikhonov regularization term. The latter favours smooth covariance/autocovariance functions, where the smoothness is specified by means of suitable Sobolev-like pseudo-differential operators. Using the machinery of reproducing kernel Hilbert spaces, we establish representer theorems that fully characterize the form of our estimators. We determine their uniform rates of convergence as the number of random fields diverges, both for the *dense* (increasing number of spatial samples) and *sparse* (bounded number of spatial samples) regimes. We moreover demonstrate the computational feasibility and practical merits of our estimation procedure in a simulation setting, assuming a fixed number of samples per random field. Our numerical estimation procedure leverages the sparsity and second-order Kronecker structure of our setup to reduce the computational and memory requirements by approximately three orders of magnitude compared to a naive implementation would require.

4. **Chong, Carsten**, Columbia University, USA.
*Joint authors:* Marc Hoffmann, Yanghui Liu, Mathieu Rosenbaum and Grégoire Szymanski

**Title:** Estimating roughness in volatility

**Abstract:** In this talk, we discuss how to obtain a feasible central limit theorem for a nonparametric estimator of roughness in volatility from high-frequency price observations. In contrast to previous works in the literature, we do not assume any parametric rough volatility model or any a priori relationship between volatility and volatility proxies. The rates of our estimators are optimal in a minimax sense and, surprisingly, are better if volatility is rougher.

5. **Cirkovic, Daniel**, Texas A&M University, USA.
*Joint authors:*

**Title:** Likelihood-based Changepoint Detection in Preferential Attachment Networks

**Abstract:** Generative, temporal network models play an important role in analyzing the dependence structure and evolution patterns of complex networks. Due to the complicated nature of real network data, it is often naive to assume that the underlying data-generative mechanism itself is invariant with time. Such observation leads to the study of changepoints or sudden shifts in the distributional structure of the evolving network. In this paper, we propose a likelihood-based methodology to detect changepoints in undirected, affine preferential attachment networks, and establish a hypothesis testing framework to detect a single changepoint, together with a consistent estimator for the changepoint. The methodology is then extended to the multiple changepoint setting via both a sliding window method and a more computationally efficient score statistic. We also compare the proposed methodology with previously developed non-parametric estimators of the changepoint via simulation, and the methods developed herein are applied to modeling the popularity of a topic in a Twitter network over time.

6. **Constantin, Alexandre**, Université Grenoble Alpes, France.
*Joint authors:* Mathieu Fauvel et Stéphane Girard

**Title:** Mixture of Multivariate Gaussian Processes for Classification of irregularly sampled time-Series: Application to Massive Satellite Image

**Abstract:** This work takes place in the context of the processing of the Satellite Image Time-Series (SITS) data from Sentinel-2 mission, launched in 2015 by the European Space Agency. It focuses on the classification task, i.e. land use or land cover maps that can be produced using spectro-temporal aspect of the Sentinel-2 SITS. The main difficulty is the acquisition noise and irregular revisiting of the satellite resulting in irregular and unevenly sampled time-series. Conventional approaches process time-series to a set of fixed and known time stamps, then they use machine learning techniques to classify vectors which scales with a large number of SITS. The main disadvantage of this two-step processing approach is that

it increases the number of operations applied to the SITS, implying a more difficult transition to massive amount of data. To a lower extent, the temporal re-sampling step may slightly alter the characteristics of the data.

To tackle this concern, we introduce a novel model-based approach with the ability to classify irregularly sampled time-series based on a mixture of multivariate Gaussian processes using two dimensions from SITS - spectral and temporal - thanks to a Kronecker structure of the covariance operator of the Gaussian process. The model allows jointly the imputation of missing values and is scalable to large-scale data-sets. It is evaluated numerically on simulated data-sets to illustrate the importance of taking into account spectral correlation and on Sentinel-2 SITS in terms of classification and imputation accuracy and is compared with conventional approaches. Analyses of the results illustrate the relevance of the model and the benefit of using interpretable parametric models.

7. **Dahlhaus, Rainer**, Heidelberg University, Germany.
*Joint authors:* Stefan Richter

**Title:** Adaptive estimation for locally stationary processes

**Abstract:** In this talk we discuss adaptive estimation of parameter curves for locally stationary processes. We investigate different cross validation approaches and a method based on contrast minimization and derive asymptotic properties of both methods. The results are applicable for different statistics under a general setting of local stationarity including nonlinear processes. At the same time we deepen the general framework for local stationarity based on stationary approximations. For example a general Bernstein inequality is derived for such processes. The properties of the bandwidth selection methods are investigated in several simulation studies.

8. **van Delft, Anne**, Columbia University, USA.
*Joint authors:*

**Title:** A general framework to quantify deviations from structural assumptions in the analysis of nonstationary function-valued processes

**Abstract:** We present a general theory to quantify the uncertainty from imposing structural assumptions on the second-order structure of nonstationary Hilbert space-valued processes, which can be measured via functionals of time-dependent spectral density operators. The second-order dynamics are well-known to be elements of the space of trace-class operators, the latter is a Banach space of type 1 and of cotype 2, which makes the development of statistical inference tools more challenging. A part of our contribution is to obtain a weak invariance principle as well as concentration inequalities for (functionals of) the sequential time-varying spectral density operator. In addition, we derive estimators of the deviation measures in the

nonstationary context that are asymptotically pivotal. We then apply this framework and propose statistical methodology to investigate the validity of structural assumptions for nonstationary response surface data, such as low-rank assumptions in the context of time-varying dynamic fPCA and principle separable component analysis, deviations from stationarity with respect to the square root distance, and deviations from zero func- tional canonical coherency.

9. **Dudek, Anna**, AGH University of Science and Technology, Poland.
*Joint authors:* B. Majewski, A. Napolitano, H. Ombao

**Title:** Modeling ECG signals using OACS processes

**Abstract:** One of the main interests of modern statistical analysis are nonstationary signals. The reason for that is very natural. Most of the observed phenomena cannot be described using stationary processes. Analysis of nonstationary processes originates in 20th century and one could think that similarly to stationary case, it is quite well investigated. However, most of phenomena are very complex and existing nonstationary models usually allow to run analysis under many strong simplifying assumptions. An example that illustrates this problem are signals modeled under assumption of periodicity. Although, theory of periodic processes and their modifications is well developed, this approach is not always suitable to analyze 'approximately periodic' data.
In this talk, we show how the Electrocardiogram (ECG) signal can be modeled using oscillatory almost-cyclostationary processes, which are generalization of oscillatory processes. We also discuss how the Extension of the Moving Block Bootstrap approach can be applied to this problem and how the optimal bootstrap block length can be chosen.

10. **Düker, Marie-Christine**, Cornell University, USA.
*Joint authors:* Robert Lund and Vladas Pipiras

**Title:** High-dimensional latent Gaussian count time series

**Abstract:** The focus of this talk are stationary vector count time series models defined via deterministic functions of a latent stationary vector Gaussian series. The construction is very general and ensures a pre-specified marginal distribution for the counts in each dimension, depending on unknown parameters that can be marginally estimated. The vector Gaussian series injects flexibility in the model's temporal and cross-sectional dependencies, perhaps through a parametric model akin to a vector autoregression. This talk discusses how the latent Gaussian model can be estimated by relating the covariances of the observed counts and the latent Gaussian series. In a possibly high-dimensional setting, concentration bounds are established for the differences between the estimated and true latent Gaussian autocovariances, in terms of those for the observed count series and the estimated marginal parameters. An application of the result is given to the case when the latent Gaussian series

follows a VAR model, and its parameters are estimated sparsely through a LASSO-type procedure.

11. **Freyermuth, Jean-Marc**, Université Aix-Marseille, France.
*Joint authors:* Dehay, D. and Dudek, A.

**Title:** Some contributions to harmonizable time series analysis

**Abstract:** Harmonizable time series are natural extensions of stationary time series with a spectral decomposition whose components are correlated. Thus, the covariance function of a harmonizable time series is bivariate and admits a two-dimensional Fourier decomposition (Loève spectrum). They form a broad class of nonstationary processes that has been subject of investigation for a long time, starting with Loève (1948-1963), Rozanov (1959) and Cramèr (1961). In this talk, we will introduce harmonizable VARMA time series and a way to generate them from given Loève spectra. Then, we will provide a quick look on tools for nonparametric spectral analysis based on replicated realizations of spatiotemporal processes that are locally time-harmonizable. Real data illustration inspired from the analysis of EEG data will be provided.

12. **Goude, Yannig**, EDF Paris, France.
*Joint authors:* Anestis Antoniadis and Solenne Gaucher

**Title:** Hierarchical transfer learning with applications for electricity load forecasting.

**Abstract:** The recent abundance of data on electricity consumption at different scales opens new challenges and highlights the need for new techniques to leverage information present at finer scales in order to improve forecasts at wider scales. In this work, we take advantage of the similarity between this hierarchical prediction problem and multi-scale transfer learning. We develop two methods for hierarchical transfer learning, based respectively on the stacking of generalized additive models and random forests, and on the use of aggregation of experts. We apply these methods to two problems of electricity load forecasting at national scale, using smart meter data in the first case, and regional data in the second case. For these two usecases, we compare the performances of our methods to that of benchmark algorithms, and we investigate their behaviour using variable importance analysis. Our results demonstrate the interest of both methods, which lead to a significant improvement of the predictions.

13. **Hirsch, Christian Pascal**, Aarhus University, Denmark.
*Joint authors:* A. Cipriani and M. Vittorietti.

**Title:** Topology-based goodness-of-fit tests for sliced spatial data

**Abstract:** In materials science and many other application domains, 3D information can often only be extrapolated by taking 2D slices. In topological data analysis, persistence vineyards have emerged as a powerful tool to take into account topological features stretching over several slices. In this talk, I will illustrate how persistence vineyards can be used to design rigorous statistical hypothesis tests for 3D microstructure models based on data from 2D slices. More precisely, by establishing the asymptotic normality of suitable longitudinal and cross-sectional summary statistics, goodness-of-fit tests will be devised that become asymptotically exact in large sampling windows. Finally, I will illustrate the testing methodology through a detailed simulation study and provide a prototypical example from materials science.

14. **Huckermann, Stephan**, Universität Göttingen, Germany.
*Joint authors:* Benjamin Eltzner, Kanti V. Mardia and Henrik Wiechers

**Title:** PCA and Clustering on Manifolds

**Abstract:** Many clustering methods rely on PCA and several extensions of PCA to manifolds have been proposed. Most of them rely on geodesics, the analogs of Euclidean straight lines. While geodesics pose little problems on spheres, shape spaces and hyperbolic spaces, say, they are, in contrast, problematic on a torus, as they can be dense, approximating every data set perfectly, leaving no room for statistics based on residuals.

This work has been motivated by the challenge of clustering for RNA structure prediction, where, among others, structure at atomic level is described by dihedral angles, leading to data on a torus. We propose torus PCA by mapping a torus to a stratified sphere and performing a variant of principal nested spheres (PNS) taylored to a stratified sphere.

As for a training data set the number of clusters is unknown and we anticipate varying densities, we pre cluster via iterative hierarchical clustering and decompose pre clusters via circular mode hunting on the main one-dimensional component.

On a mesoscopic (molecular) level, RNA structure can be described by shape spaces and it turns out that the interdependence between atomic and molecular level can be used for structure prediction, this will be expanded in Henrik Wiechers talk with application to SARS-CoV-2.

Notably shape spaces are quotients of spheres which also qualify for PNS. In fact this extends to every manifold, as it can be better approximated by a tangent sphere than by a tangent space.

15. **Kengne, William C.**, Université de Cergy, France.
*Joint authors:* Lamine Diop

**Title:** Statistical learning for $\psi$-weakly dependent processes

**Abstract:** We consider the statistical learning question for $\psi$-weakly dependent processes, that unifies a large class of weak dependence conditions such as mixing, association,$\cdots$ The consistency of the empirical risk minimization algorithm is established. We derive the generalization bounds and provide the learning rate, which, on some Hölder class of hypothesis, is close to the usual $O(n^{-1/2})$ obtained in the i.i.d. case. Application to time series prediction is carried out, with the example of causal models with exogenous covariates.

16. **Killick, Rebecca**, Lancaster University, UK.
*Joint authors:*

**Title:** Detecting changes in mixed-sampling rate data sequences

**Abstract:** Different environmental variables are often monitored using different sampling rates; examples include half-hourly weather station measurements, daily $CO_2$ data and six-day satellite data. When researchers want to combine the data into a single analysis this often requires data aggregation or down-scaling. Further, when one is seeking to identify changes within multivariate data, the aggregation and/or down-scaling processes obscure the changes we seek. In this talk, we propose a novel changepoint detection algorithm which can analyse multiple time series for co-occurring changepoints with potentially different sampling rates, without requiring preprocessing to a standard sampling scale. We demonstrate the algorithm on synthetic data before providing an example identifying simultaneous changes in multiple variables at a location on the Greenland ice sheet using synthetic apature radar (SAR) and weather station data.

17. **Kirch, Claudia**, University Magdeburg, Germany.
*Joint authors:* Marco Meyer and Philipp Klein.

**Title:** Scan statistics for the detection of anomalies in random fields

**Abstract:** Anomaly detection in random fields is an important problem in many applications including the detection of cancerous cells in medicine, obstacles in autonomous driving and cracks in the construction material of buildings. Scan statistics have the potential to detect local structure in such data sets by enhancing relevant features. Frequently, such anomalies are visible as areas with different expected values compared to the background noise where the geometric properties of these areas may depend on the type of anomaly. Such geometric properties can be taken into account by combinations and contrasts of sample means over differently-shaped local windows. For example, in 2D image data of concrete both cracks, which we aim to detect, as well as integral parts of the material (such as air bubbles or gravel) constitute areas with different expected values in the image. Nevertheless, due to their different geometric properties we can define scan statistics that enhance cracks and at the same time discard the integral parts of the given concrete. Cracks can then be detected using a

suitable threshold for appropriate scan statistics.

In order to derive such thresholds, we prove weak convergence of the scan statistics towards a functional of a Gaussian process under the null hypothesis of no anomalies. The result allows for arbitrary (but fixed) dimension, makes relatively weak assumptions on the underlying noise, the shape of the local windows and the combination of finitely-many of such windows. These theoretical findings are accompanied by some simulations as well as applications to semi-artifical 2D-images of concrete.

18. **Kley, Tobias**, University of Goettingen, Germany.
*Joint authors:* Yuhan Philip Liu, Hongyuan Cao and Wei Biao Wu.

**Title:** Change point analysis with irregular signals: when did the COVID-19 pandemic start?

**Abstract:** The problem of testing and estimation of change point is considered where signals after the change point can be highly irregular, which departs from the existing literature that assumes signals after the change point to be piece-wise constant or vary smoothly. A two-step approach is proposed to effectively estimate the location of the change point. The first step consists of a preliminary estimation of the change point that allows us to obtain unknown parameters in the second step. In a second step we use a new procedure to determine the position of the change point. We show that the optimal $O_P(1)$ rate of convergence of the estimated change point to the true change point can be obtained based on the proposed method. We apply our method to analyze the Baidu search index of COVID-19 related symptoms and find 7 December 2019 to be the starting date of the COVID-19 pandemic.

19. **Krampe, Jonas**, University of Mannheim, Germany.
*Joint authors:*

**Title:** Frequency Domain Statistical Inference for High-Dimensional Time Series

**Abstract:** Analyzing time series in the frequency domain enables the development of powerful tools for investigating the second-order characteristics of multivariate stochastic processes. Parameters like the spectral density matrix and its inverse, the coherence or the partial coherence, encode comprehensively the complex linear relations between the component processes of the multivariate system. In this paper, we develop inference procedures for such parameters in a high-dimensional, time series setup. In particular, we first focus on the derivation of consistent estimators of the coherence and, more importantly, of the partial coherence which possess manageable limiting distributions that are suitable for testing purposes. Statistical tests of the hypothesis that the maximum over frequencies of the coherence, respectively, of the partial coherence, do not exceed a prespecified threshold value are developed. Our approach allows for testing hypotheses for individual coherences and/or partial coherences as well as for multiple testing of large sets of such parameters. In the latter case, a consistent

procedure to control the false discovery rate is developed.

20. **Lee, Jeongyin**, University of Namur, Belgium.
*Joint authors:* Dan Cooley

**Title:** Partial Tail Correlation for Extremes

**Abstract:** We develop a method for investigating conditional extremal relationships between variables. We consider an inner product space constructed from transformed-linear combinations of independent regularly varying random variables. By developing the projection theorem for the inner product space, we derive the concept of partial tail correlation via projection theorem. We show that the partial tail correlation can be understood as the inner product of the prediction errors associated with transformed- linear prediction. Similar to Gaussian cases, we connect partial tail correlation to the inverse of the inner product matrix and show that a zero in this inverse implies a partial tail correlation of zero. We develop a hypothesis test for the partial tail correlation of zero and demonstrate the performance in a simulation study as well as in two applications: extreme river discharges in the upper Danube basin and high nitrogen dioxide levels in Washington DC.

21. **Leonte, Dan**, Imperial College London, UK.
*Joint authors:*

**Title:** Simulation and forecasting method for trawl processes and ambit fields

**Abstract:** Trawl processes are continuous-time, stationary and infinitely divisible processes which can describe a wide range of possible serial correlation patterns in data. This talk introduces a new algorithm for the efficient and exact simulation of monotonic trawl processes. The algorithm accommodates any monotonic trawl shape and any infinitely divisible distribution described via the Lévy seed, requiring only access to samples from the distribution of the Lévy seed. Furthermore, we describe how Monte Carlo methods can generalise the above method to a simulation scheme for ambit fields. Finally, we discuss deterministic and probabilistic forecasting of trawl processes and ambit fields.

22. **Leucht, Anne**, University of Bamberg, Germany.
*Joint authors:* Paul Doukhan and Michael Neumann

**Title:** Mixing properties of non-stationary INGARCH(1,1) processes

**Abstract:** In this talk, I will present mixing properties for a broad class of Poisson count time series satisfying a certain contraction condition. Using specific coupling techniques, we obtain absolute regularity at a geometric rate not only for stationary Poisson-GARCH pro-

cesses but also for models with an explosive trend. Easily verifiable sufficient conditions for absolute regularity can be deduced from our general results for a variety of models including classical (log-)linear models.

23. **Li, Yuan**, Imperial College, UK.
*Joint authors:* Mikko S. Pakkanen and Almut E.D. Veraart

**Title:** Limit theorems for the realised semicovariances of multivariate Brownian semistationary processes

**Abstract:** In this project, we will introduce the realised semicovariance for Brownian semistationary (BSS) processes, which is obtained from the decomposition of the realised covariance matrix into components based on the signs of the returns and study its in-fill asymptotic properties. More precisely, weak convergence in the space of càdlàg functions endowed with the Skorohod topology for the realised semicovariance of a general Gaussian process with stationary increments is proved first. The proof is based on the Breuer-Major theorem and on a moment bound for sums of products of non-linearly transformed Gaussian vectors. Furthermore, we establish a corresponding stable convergence. Finally, a central limit theorem for the realised semicovariance of multivariate BSS processes is established. These results extend the limit theorems for the realised covariation to a result for non-linear functionals.

24. **Masak, Tomas**, Ecole polytechnique fédérale de Lausanne, Switzerland.
*Joint authors:* Soham Sarkar and Victor M. Panaretos

**Title:** Separable Expansions for Covariance Estimation via the Partial Inner Product

**Abstract:** The non-parametric estimation of covariance lies at the heart of functional data analysis, whether for curve or surface-valued data. The case of a two-dimensional domain poses both statistical and computational challenges, which are typically alleviated by assuming separability. However, separability is often questionable, sometimes even demonstrably inadequate.

We propose a framework for the analysis of covariance operators of random surfaces that generalises separability, while retaining its major advantages. Our approach is based on the expansion of the covariance into a series of separable terms. The expansion is valid for any covariance over a two-dimensional domain. Leveraging the key notion of the partial inner product, we generalise the power iteration method to general Hilbert spaces and show how the aforementioned expansion can be efficiently constructed in practice at the level of the surface observations. Truncation of the expansion and retention of the leading terms automatically induces a non-parametric estimator of the covariance, whose parsimony is dictated by the truncation level. The resulting estimator can be calculated, stored and manipulated with little computational overhead relative to separability.

Consistency and rates of convergence are derived under mild regularity assumptions, illustrating the trade-off between bias and variance regulated by the truncation level. Finally the empirical performance and computational feasibility of our methods is demonstrated in an extensive simulation study and on a real data set.

25. **Mhalla, Linda**, EPFL, Switzerland.
*Joint authors:*

**Title:** Extremal connectedness of hedge funds

**Abstract:** We propose a dynamic measure of extremal connectedness tailored to the short reporting period and unbalanced nature of hedge funds data. Using multivariate extreme value regression techniques, we estimate this measure conditional on factors reflecting the economic uncertainty and the state of the financial markets, and derive risk indicators reflecting the likelihood of extreme spillovers. Empirically, we study the dynamics of tail dependencies between hedge funds grouped per investment strategies, as well as with the banking sector. We show that during crisis periods, some pairs of strategies display an increase in their extremal connectedness, revealing a higher likelihood of simultaneous extreme losses. We also find a sizable tail dependence between hedge funds and banks, indicating that banks are more likely to suffer extreme losses when the hedge fund sector does. Our results highlight that a proactive regulatory framework should account for the dynamic nature of the tail dependence and its link with financial stress.

26. **Ombao, Hernando**, KAUST, Saudi Arabia.
*Joint authors:*

**Title:** Exploring General Dependence in a Brain Network

**Abstract:** Brain physiological and cognitive process over the entire network is complex. A full understanding of brain activity requires careful study of its multi-scale spatial-temporal organization (from small volume neurons to communities of regions of interest; and from transient events to long-term temporal dynamics). Motivated by these challenges, we will explore some characterizations of dependence between components of a multivariate time series and then apply these to the study of brain functional connectivity. There is no single measure of dependence that can capture all facets of brain connectivity. Here, we shall explore dependence between band-specific oscillations as well as scale-specific subprocesses of a locally stationary wavelet process. We develop a method that explores potential interactions between oscillations in multivariate time using these subprocesses. This is potentially interesting for brain scientists because functional brain networks are associated with cognitive function, and mental and neurological diseases. This method is used to study differences in functional brain connectivity between the healthy and the children diagnosed with attention deficit hyperac-

tivity disorder.

27. **Opitz, Thomas**, Biostatistics and Spatial Processes, INRAE, 84914, Avignon, France.
*Joint authors:*

**Title:** Spatiotemporal modeling of extreme events: a state of the art

**Abstract:** In this talk, I will attempt to outline the state of the art of spatiotemporal extreme-value modeling and give some examples from my own work. Due to the very rich body of work already available in this area, I will present a certainly subjective selection of topics that is intended to encourage discussion about "what could/should be done" in the near future. I plan to allude to the following aspects (among others):

- What are typical goals of modeling spatiotemporal extremes?
- What are the specifics of modeling extremes in the spatiotemporal setting as opposed to the purely spatial or time series setting?
- What types of models are currently available, and what are their strengths and weaknesses?
- Are asymptotic models such as max-stable processes or Pareto processes always a good choice, or should we consider more flexible subasymptotic modeling in certain contexts?
- What are numerical challenges of inference and simulation, e.g. with respect to scalability to large datasets?

Finally, I will give some examples of areas for future research.

28. **Panaretos, Victor M.**, Institute of Mathematics, EPFL, Switzerland.
*Joint authors:* K.G. Waghmare

**Title:** The Extrapolation of Correlation

**Abstract:** We discuss the problem of positive-semidefinite extension: extending a partially specified covariance kernel from a subdomain $\Omega$ of a rectangular domain $I \times I$ to a covariance kernel on the entire domain $I \times I$. For a broad class of domains $\Omega$ called *serrated domains*, we present a complete theory. Namely, we demonstrate that a canonical completion always exists and can be explicitly constructed. We characterise all possible completions as suitable perturbations of the canonical completion, and determine necessary and sufficient conditions for a unique completion to exist. We interpret the canonical completion via the graphical model structure it induces on the associated Gaussian process. Furthermore, we show how the determination of the canonical completion reduces to the solution of a system of linear inverse problems in the space of Hilbert-Schmidt operators, and derive rates of convergence when the kernel is to be empirically estimated. We conclude by providing extensions of our

theory to more general forms of domains, and by demonstrating how our results can be used in statistical inverse problems associated with stochastic processes.

29. **Pasche, Olivier**, Université de Genève, Switzerland.
*Joint authors:* Sebastian Engelke

**Title:** Neural Networks for Extreme Quantile Regression with an Application to Forecasting of Flood Risk

**Abstract:** Risk assessment for extreme events requires accurate estimation of high quantiles that go beyond the range of historical observations. When the risk depends on the values of observed predictors, regression techniques are used to interpolate in the predictor space. We propose the EQRN model that combines tools from neural networks and extreme value theory into a method capable of extrapolation in the presence of complex predictor dependence. Neural networks can naturally incorporate additional structure in the data. We develop a recurrent version of EQRN that is able to capture complex sequential dependence in time series. We apply this method to forecasting of flood risk in the Swiss Aare catchment. It exploits information from multiple covariates in space and time to provide one-day-ahead predictions of return levels and exceedances probabilities. This output complements the static return level from a traditional extreme value analysis and the predictions are able to adapt to distributional shifts as experienced in a changing climate. Our model can help authorities to manage flooding more effectively and to minimize their disastrous impacts through early warning systems.

Link: https://arxiv.org/abs/2208.07590

30. **Pfitzner, Leo**, Institute of Mathematics, EPFL, Switzerland.
*Joint authors:* Eric Adjakossa, Olivier Wintenberger and Olivier Mestre

**Title:** Temperature forecasting with expert aggregation

**Abstract:** A lot of Numerical Weather Prediction (NWP) models and their associated Model Output Statistics (MOS) are available. Expert aggregation has a lot of advantages to deal with all these models, like being online, adaptive to model changes and having theoretical guarantees. With a new expert aggregation algorithm - FSBOA - a combination of BOA (Wintenberger 2017) and FS (Herbster and Warmuth 1998), and the use of a sliding window, in collaboration with Olivier Wintenberger, Eric Adjakossa and Olivier Mestre, we improved the temperature prediction on average without loosing too much reactivity of the expert weights. We also tested several aggregation strategies in order to improve the prediction of extrem temperature events like cold and heat waves. To do so, we added some biased experts of the Météo-France 35-member ensemble forecast (PEARP) to the set of models. We also tried out the SMH (Mourtada et al. 2017) algorithm which fits the sleeping experts framework.

31. **Philippe, Anne**, Université de Nantes, France.
*Joint authors:* M. Ould Haye

**Title:** Detection of non-stationarity in dependent data

**Abstract:** Distinguishing long memory behaviour from nonstationarity can be very difficult as in both cases the sample autocovariance function decays very slowly. Available stationarity tests either do not include long memory or fare poorly in terms of empirical size, especially near the boundary between long memory and nonstationarity. We propose a parameter-free decision rule, that is based on evaluating periodograms at different epochs. We establish some asymptotic theorems in order to validate the method. Limiting distributions are easily tractable as sum of weighted independent chi-square random variables. Moreover, numerical studies are provided to show that the proposed approach outperforms existing methods. We also apply our method to a well-known empirical data, often cited as an example of confusion between long memory and nonstationarity.

32. **Rademacher, Daniel**, University Heidelberg, Germany.
*Joint authors:*

**Title:** Inference for the Spectral Mean of Functional Time Series in Hilbert Space

**Abstract:** A variety of statistics for functional time series allows for a representation as weighted average of corresponding periodogram operators over the frequency domain. We study the consistency and asymptotic distribution of such spectral mean estimates under mild assumptions. We show that the weak convergence of spectral mean estimates can be reduced to the (joint) weak convergence of the sample autocovariance operators, and that the latter condition holds for a large class of weakly dependent functional time series, which admit expansions as Bernoulli shifts. The weak dependency is quantified by the condition of $L^4 - m$-approximability and a (functional version) of the classical mixing condition on the 4th order cumulant operators.

33. **Robinson, Peter M.**, LSE, United Kingdom.
*Joint authors:*

**Title:** Issues in Spatial Processes with Long Range Dependence

**Abstract:** We discuss issues of statistical inference for spatial processes with 'long range dependence'. Long range dependence, or strong dependence, in time series is a topic that has been quite extensively studied in recent years. After a number of probabilistic contributions and empirical studies, serious treatment of issues of statistical inference could be said to

have begun in the mid-1980s, with activity then increasing through the 1990s and this century. Predominately this literature has focussed on observations that are regularly-spaced over time, and the bulk of the theoretical development has been in terms of asymptotic statistical theory, with the number of observations, $n$, regarded as diverging, finite-sample theory proving mathematically intractable, even under the precise distributional assumptions that are typically not required in a large-sample treatment.

Parametric, semiparametric and nonpararmetric models have all featured. The major characteristic feature of a long range dependent covariance stationary time series process is that its autocovariance function decays so slowly with increase in lag length as not to be summable, or, nearly equivalently, its spectral density diverges, typically at zero frequenccy, while some nonstationary processes, such as ones with a unit root can, a fortiori, be regarded as having even longer memory. By contrast, 'short range dependent' time series typically have autocovariances that are summable and spectral density that is more or less smooth (for example a stationary autoregressive moving average (ARMA) has exponentuially decaying autocovariance and analytic spectral density), though for some relevant purposes a short range dependent process is sometimes defined as merely having spectral density that is positive and finite at zero frequency.

Spatial data have long attracted the attention of statisticians, and the configurations of some such data, especially some arising in such fields as meteorology, cosmology and agriculture, can be viewed as generalisations of the typical regularly-spaced time series one mentioned above. In particular they constitute observations in $M \geq 2$ dimensions, (where $M = 1$ in the tme series case), such that there are $n_m$ observations across dimension $m$, for $m = 1, ..., M$, so $n = \Pi_{m=1}^{M} n_m$ . We call this a '(rectangular) lattice', and data recorded on it belong to the class of 'random fields'. There is equal spacing across each dimension, but the spacing can vary across dimensions.

This brings to mind regularly spaced agricultural plots on a field. In fact, 'spatial long range dependence' goes back at least to the agriculturally-motivated paper of Smith (1938), which is also a very early reference relative to the literature on time series long range depedence. It is interesting that Smith (1938) thought of a power law decay: this might seem natural to one coming to the subject unschooled in time series modellng which, after World War II, stressed exponential decay, as in ARMA modelling.

Since then, many papers on 'spatial long range dependence' have appeared, but the topic has not been developed as systematically or comprehensively as 'l long range dependent time series'. Some distinctive issues arising in the 'spatial' case, all of which have been studied far more under short range dependence than long range dependence, are:

1. Is there isotropy or not? If not, we might model each dimension separately, or with some interaction, and possibly with a different memory parameter for each dimension.

2. Is sampling regularly or irregularly spaced? Whereas in time series regular spacing has been predominately studied, irregular spacing is perhaps more likely to be found with spatial data.

3. Should modelling be unilateral or multilateral? For time series unilateral modelling, reflecting one-sided transition from past to future, is usually natural, but this is not the case with spatial data, where, for example, the dimensions might be latitude and longitude.

4. The edge effect. In estimating lagged quantities there is loss of information at the boundary of the observation region, which has negligible effect when $M = 1$, but increasing, and damaging, effect as $M$ increases unless corrected for.

We discuss the following topics:

1. Inference on location and regression with long range dependent errors.

2. Inference on second-order properties of long range dependent stationary processes.

3. Miscellaneous topics: nonstationary processes, irregular spacing, adaptive estimation, nonparametric regression.

We do not consider 'spatial autoregressive'-type ('SAR') models (which depend on a user-chosen weight matrix of geographic or economic inverse distances); these do not fit into our framework but typically possess a kind of short range dependence.

34. **Roueff, François**, Institut Polytechnique de Paris, France.
*Joint authors:*

**Title:** Fractionally Integrated Autoregressive Moving Average processes valued in a separable Hilbert space

**Abstract:** Fractionally integrated autoregressive moving average (FIARMA) processes have been widely and successfully used to model univariate time series exhibiting long range dependence. Vector and functional extensions of these processes have also been considered more recently. Here we extend this class of models to processes valued in a separable Hilbert space. To this end we will briefly rcall the spectral theory for weakly stationary processes valued in a separable Hilbert space. In such a framework, the usual univariate long memory parameter d is replaced by a long memory operator D acting on the observation space. This operator is used as the exponent of the fractional integration filter expressed in the spectral domain. In the case where D is normal we provide a sufficent and necessary condition for the corresponding FIARMA process to be well defined as a weakly stationary process. Finally, we will address the question of parametric statistical inference in this context.

35. **Santoro, Leonardo**, Institute of Mathematics, EPFL, Switzerland.
*Joint authors:* Neda Mohammadi and Victor M. Panaretos

**Title:** Nonparametric Estimation for SDE with Sparsely Sampled Paths: an FDA Perspective

**Abstract:** We consider the problem of nonparametric estimation of the drift and diffusion coefficients of a Stochastic Differential Equation (SDE), based on $n$ independent replicates $\{X_i(t) : t \in [0,1]\}_{1\leq i\leq n}$, observed sparsely and irregularly on the unit interval, and subject to additive noise corruption. By *sparse* we intend to mean that the number of measurements per path can be arbitrary (as small as two), and can remain constant with respect to $n$. We focus on time-inhomogeneous SDE of the form $dX_t = \mu(t)X_t^\alpha dt + \sigma(t)X_t^\beta dB_t$, where $\alpha \in \{0,1\}$ and $\beta \in \{0,1/2,1\}$, which includes prominent examples such as Brownian motion, the Ornstein-Uhlenbeck process, geometric Brownian motion, and the Brownian bridge. Our estimators are constructed by relating the local (drift/diffusion) parameters of the SDE to its global parameters (mean/covariance, and their derivatives) by means of an apparently novel Partial Differential Equation (PDE). This allows us to use methods inspired by functional data analysis, and pool information across the sparsely measured paths. The methodology we develop is fully non-parametric and avoids any functional form specification on the time-dependency of either the drift function or the diffusion function. We establish almost sure uniform asymptotic convergence rates of the proposed estimators as the number of observed paths $n$ grows to infinity. Our rates are non-asymptotic in the number of measurements per path, explicitly reflecting how different sampling frequency might affect the speed of convergence. Our framework suggests possible further fruitful interactions between FDA and SDE methods in problems with replication.

36. **Sauri, Orimar**, Aalborg University, Denmark.
*Joint authors:* A. Veraart

**Title:** Nonparametric estimation of trawl processes: Theory and Applications

**Abstract:** Trawl processes belong to the class of continuous-time, strictly stationary, infinitely divisible processes; they are defined as Lévy bases evaluated over deterministic trawl sets. In this talk, we present the first nonparametric estimator of the trawl function characterising the trawl set and the serial correlation of such a class of processes. Under an in-fill and a long-span sampling scheme, we establish a detailed asymptotic theory for the proposed estimator, including a law of large numbers and a central limit theorem. With the aim of providing a feasible asymptotic theory, we study the estimation problem of non-linear functionals of the trawl function. In an empirical illustration, we apply our new methodology to modelling and forecasting high-frequency financial spread data from a limit order book.

37. **Segers, Johan**, Université Catholique de Louvain, Belgium.
*Joint authors:* Hu, Shuang and Peng, Zuoxiang

**Title:** Modelling multivariate extreme value distributions via Markov trees

**Abstract:** Multivariate extreme value distributions are a common choice for modelling multivariate extremes. In high dimensions, however, the construction of flexible and parsimonious models is challenging. We propose to combine bivariate extreme value distributions into a Markov random field with respect to a tree. Although in general not an extreme value distribution itself, this Markov tree is attracted by a multivariate extreme value distribution. The latter serves as a tree-based approximation to an unknown extreme value distribution with the given bivariate distributions as margins. Given data, we learn an appropriate tree structure by Prim's algorithm with estimated pairwise upper tail dependence coefficients or Kendall's tau values as edge weights. The distributions of pairs of connected variables can be fitted in various ways. The resulting tree-structured extreme value distribution allows for inference on rare event probabilities, as illustrated on river discharge data from the upper Danube basin.

38. **Strokorb, Kristin**, Cardiff University, UK.
*Joint authors:* Sebastian Engelke and Jevgenijs Ivanovs

**Title:** Conditional Independence in Extremes

**Abstract:** Statistical modelling of complex dependencies in extreme events requires meaningful sparsity structures in multivariate extremes. In this context two perspectives on conditional independence and graphical models have recently emerged: One that focuses on threshold exceedances and multivariate pareto distributions, and another that focuses on max-linear models and directed acyclic graphs. What connects these notions is the exponent measure that lies at the heart of each approach. In this work we develop a notion of conditional independence defined directly on the exponent measure (and even more generally on measures that explode at the origin) that extends recent work of Engelke and Hitz (2019), who had been confined to homogeneous measures with density. We prove easier checkable equivalent conditions to verify this new conditional independence in terms of a reduction to simple test classes, probability kernels and density factorizations. This provides a pathsway to graphical modelling among general multivariate (max-)infinitely distributions. Structural max-linear models turn out to form a Bayesian network with respect to our new form of conditional independence.

39. **Truquet, Lionel**, ENSAI, France.
*Joint authors:*

**Title:** Asymptotics for estimating the Taylor's law parameters in Ecology

**Abstract:** Taylor's power law in Ecology is a power relation between the variance and the mean of the number of a species per unit area. In a spatio-temporal context, parameters of this law are estimated from arrays of data representing some counts of a species across different locations and units of time. This law, which has been verified empirically for various

species, has an important interpretation in term of aggregation properties of a species in time or in space. The parameters of this law are typically estimated by regressing the logarithm of the empirical variances of the counts on their (log) empirical means. In this talk, we will first discuss which type of assumptions it is necessary to make on the population mean/variance for being compatible with these empirical findings. We will then study consistency and weak convergence of the corresponding estimates. We will complete the talk with some numerical experiments as well as illustrations with real data sets.

40. **Wiechers, Henrik**, University of Göttingen, Germany.
*Joint authors:* Benjamin Eltzner, Kanti V. Mardia and Stephan F. Huckemann

**Title:** Learning torus PCA based classification for multiscale RNA correction with application to SARS-CoV-2

**Abstract:** Reconstructions of structure of biomolecules, for instance via X-ray crystallography or cryo-EM frequently contain clashes of atomic centers which are not chemically permissible. Methods to correct these clashes are usually based on simulations approximating biophysical chemistry, making them computationally expensive and often not correcting all clashes. Using RNA data, we propose fast, data-driven multiscale learning reconstructions from clash free RNA benchmark data. Multiscale here is based on two levels of shape analysis for RNA geometry. The shape of RNA at microscopic scale (suites) can be described by dihedral angles of the backbone leading to appropriate landmarks at another scale, the mesoscopic scale which is described by landmarks obtained as centers of so-called sugar rings. Based on our analysis that concentrated neighborhoods at mesoscopic scale closely relate to clusters at microscopic scale, we correct within-suite-backbone-to-backbone clashes exploiting Fréchet means at the two scales; one uses angular shape (microscopic), the other uses size-and-shape (mesoscopic). We validate our reconstructions by showing that the classes learned are in high correspondence with clusters from existing literature and that the reconstructions proposed are well within physical resolution limits. While this method is general for RNA we illustrate its power by the cutting-edge RNA example of SARS-CoV-2.

41. **Yang, Junho**, Academia Sinica.
*Joint authors:* Akihiko Inoue

**Title:** Baxter-type convergence results for a block Toeplitz system under long memory with an application to linear prediction problem

**Abstract:** The Wiener-Hopf (WH) equation is a semi-infinite Toeplitz system of equations that has diverse applications. In practice, truncation of the system is inevitable to calculate the solution of the WH equation in a finite time. Baxter's inequality provides an $L_1$-bound for the approximation error between the WH solution and its finite-section approximation.

However, this inequality is only valid for a symbol with short memory. In this talk, we derive the Baxter-type convergence results for a block Toeplitz system when the corresponding matrix-valued symbol has a long memory. A key ingredient is using a series expansion of the inverse of a finite-order Toeplitz matrix. Based on these results, we show the Baxter-type convergence for linear prediction problems for multivariate long memory stationary processes.

42. **Yu Yi**, University of Warwick, UK.
*Joint authors:* Haotian Xu, Daren Wang and Zifeng Zhao

**Title:** Change point inference in high-dimensional regression models under temporal dependence

**Abstract:** This paper concerns about the limiting distributions of change point estimators, in a high-dimensional linear regression time series context, where a regression object $(y_t, X_t) \in \mathbb{R} \times \mathbb{R}^p$ is observed at every time point $t \in \{1, \ldots, n\}$. At unknown time points, called change points, the regression coefficients change, with the jump sizes measured in $\ell_2$-norm. We provide limiting distributions of the change point estimators in the regimes where the minimal jump size vanishes and where it remains a constant. We allow for both the covariate and noise sequences to be temporally dependent, in the functional dependence framework, which is the first time seen in the change point inference literature. We show that a block-type long-run variance estimator is consistent under the functional dependence, which facilitates the practical implementation of our derived limiting distributions. We also present a few important byproducts of their own interest, including a novel variant of the dynamic programming algorithm to boost the computational efficiency, consistent change point localisation rates under functional dependence and a new Bernstein inequality for data possessing functional dependence. Extensive numerical results are provided to support our theoretical results.

The paper is available at https://arxiv.org/abs/2207.12453