MATRIX Research Program:

Multivariate Dependence Modelling: Theory and Applications

Titles and Abstracts

22 JULY – 2 AUGUST 2024

CRESWICK, AUSTRALIA

List of Abstracts

MONDAY, JULY 22

Dietrich von Rosen, Swedish University of Agricultural Sciences

Safety belt regression, a mixture of classical regression and ridge regression

Penalized estimation methods will be considered (Ridge estimation). If there is a model including parameters and an estimation function, e.g., the least squares function or likelihood function, so called penalized estimators can be obtained. This means that the estimation function is modified by adding some "fitting" term. However, in this presentation of the subject, restrictions are put on the parameters instead of the estimation function which in turn also will lead to a penalized estimation function. Indeed the two different approaches are similar but the estimators differ. In some way, from a likelihood point of view, it is more natural to put restrictions on the parameters in a model than on the estimation function. The approach is based on convex optimization theory.

Lin Deng, University of Melbourne

Large Skew-t Copula Models and Asymmetric Dependence in Intraday Equity Returns

Skew-t copula models are attractive for the modeling of financial data because they allow for asymmetric and extreme tail dependence. We show that the copula implicit in the skew-t distribution of Azzalini and Capitanio (2003) allows for a higher level of pairwise asymmetric dependence than two popular alternative skew-t copulas. Estimation of this copula in high dimensions is challenging, and we propose a fast and accurate Bayesian variational inference (VI) approach to do so. The method uses a generative representation of the skew-t distribution to define an augmented posterior that can be approximated accurately. A stochastic gradient ascent algorithm is used to solve the variational optimization. The methodology is used to estimate skew-t factor copula models with up to 15 factors for intraday returns from 2017 to 2021 on 93 U.S. equities. The copula captures substantial heterogeneity in asymmetric dependence over equity pairs, in addition to the variability in pairwise correlations. In a moving window study we show that the asymmetric dependencies also vary over time, and that intraday predictive densities from the skew-t copula are more accurate than those from benchmark copula models. Portfolio selection strategies based on the estimated pairwise asymmetric dependencies improve performance relative to the index.

Gery Geenens, University of New South Wales

Universal Copulas

The early 21st century has seen the advent of copulas as primary statistical tools when it comes to model dependence between numerical random variables. A copula is classically understood as a cumulative distribution function on the unit hypercube with standard uniform margins – we refer to such distributions as "Sklar's copulas", owing to their central role in the decomposition of multivariate distributions established by the celebrated Sklar's theorem. The argument habitually put forward for outlining the appeal of copula models is that they allow pulling apart the dependence structure of a bivariate vector, characterised by its copula, from the individual behaviour of its marginal components. Though, this interpretation can only be justified in the continuous framework, as copulas lose their "margin-free" nature outside of it, making Sklar's copula models unfit for modelling dependence between non-continuous variables. In this work, we argue that the very notion of copula should not be imprisoned into Sklar's theorem, and we propose an alternative definition of copulas which follows from approaching their role and meaning more broadly. This definition coincides with Sklar's copulas in the continuous framework, but leads to different concepts in other settings. We call this construction "universal copulas" and show that these maintain all the pleasant properties (in particular: 'margin-freeness') which make Sklar's copulas sound and effective in the continuous case. We illustrate our findings with some examples of "universal copula modelling" between two discrete variables, and between one continuous variable and one binary (Bernoulli) variable.

Michael Smith, University of Melbourne

Cutting Feedback in Misspecified Copula Models

In copula models the marginal distributions and copula function are specified separately. We treat these as two modules in a modular Bayesian inference framework, and propose conducting modified Bayesian inference by "cutting feedback". Cutting feedback limits the influence of potentially misspecified modules in posterior inference. We consider two types of cuts. The first limits the influence of a misspecified copula on inference for the marginals, which is a Bayesian analogue of the popular Inference for Margins (IFM) estimator. The second limits the influence of misspecified marginals on inference for the copula parameters by using a rank likelihood to define the cut model. We establish that if only one of the modules is misspecified, then the appropriate cut posterior gives accurate uncertainty quantification asymptotically for the parameters in the other module. Computation of the cut posteriors is difficult, and new variational inference methods to do so are proposed. The efficacy of the new methodology is demonstrated using both simulated data and a substantive multivariate time series application from macroeconomic forecasting. In the latter, cutting feedback from misspecified marginals to a 1096 dimension copula improves posterior inference and predictive accuracy greatly, compared to conventional Bayesian inference.

TUESDAY, JULY 23

Elif Acar, University of Manitoba

Conditional copula survival models for incomplete data

Incomplete data is a common feature in large biomedical databases, including among others those of phenotyping experiments and genetic studies. A fundamental challenge in these settings is to develop flexible modelling strategies for clustered censored survival data. To this end, the effectiveness of copula survival models has been studied extensively for right censored data. However, interval-censored and semi-competing risk settings are considerably less explored. In this talk, I will discuss and demonstrate the use of conditional copula models for clustered survival data under different censoring mechanisms, focusing on the cases with interval-censoring and semi-competing risks.

Karim Oualkacha, University of Quebec in Montreal

Multiple-Output Excpectile Regression via Copulas

Expectiles summarize a distribution in a manner similar to quantiles. Expectile regression has recently gained great popularity, in part due to its attractive statistical and computational properties. Despite the renewed interest, it unfortunately remains limited to single-output problems. To enhance on this and to gain insight into multivariate data, we build on a class of multivariate expectile loss functions to develop a unified and flexible copula-based multivariate expectile regression framework. Our approach provides a new class of multiple-output expectile regression estimators, which are unique solutions to convex risk minimization problems. We model the joint distribution of the multiple-output and the regressors using a copula model, which separates modelling the dependence and the marginal distributions. Then, we rewrite the multivariate expectile regression loss function in terms of the copula and the marginal distributions. We prove the asymptotic properties of our estimators (weak convergence and i.i.d. representation). We demonstrate the effectiveness of our approach through simulation studies and by analyzing the Fourth Dutch Growth data.

Mohamed Chaouch, Qatar University

Semiparametric copula-based multivariate M-estimators

This paper introduces copula-based estimation of a class of conditional multivariate M-estimators for stationary and ergodic processes. This class encompasses, as special cases, multivariate regression, spatial median, geometric quantiles and multivariate expectiles. We investigate several asymptotic properties of the estimator. That includes uniform almost sure convergence as well as asymptotic distribution. Application to multivariate data outliers detection as well as multistep forecasting are also discussed.

Bahareh Ghanbari, RMIT University

A Non-parametric estimator for high-dimensional factor copula models with the use of proxy variables

Factor copula models have emerged as parsimonious tools for capturing dependence in observed variables based on one or multiple unobserved common factors. These models represent canonical vine copulas, integrating both observed and latent variables, thereby facilitating the representation of tail, asymmetric, and non-linear dependence structures. Most of the literature on factor copulas focuses on parametric modeling, but most of the time parametric assumptions are made solely for practical convenience. However, relying on parametric assumptions in many cases results in a misspecified model, leading to a loss of consistency in estimators. This issue is particularly critical when one or more linking copulas deviate from popular parametric copula families. Although nonparametric approaches offer a potential solution, they typically involve more intricate and computationally intensive methods. In light of this, we propose a novel nonparametric kernel estimator for such models and compute its rate of convergence. The efficacy of our estimation method is demonstrated through an extensive simulation study.

MATRIX

WEDNESDAY, JULY 24

Guilherme Ludwig, University of Campinas

A sparse non-stationary spatio-temporal covariance model with advection effects for rainfall data

We propose a sparse, non-stationary model constructed using a mixture of spatio-temporal covariance models with advection (Gupta and Waymire, 1987; Cox and Isham, 1988); namely, models that have larger covariance values along an orientation vector in the spatio-temporal index set, that simulate wind direction and cloud movement. We show that a mixture of such models can allow for wind direction change in data, and produce seasonal variation. We construct a MCMC procedure for Bayesian estimation, and illustrate the problem with rainfall data from the southeastern coast of Brazil. This is a joint work with Pedro Nasevicius Ramos (UNICAMP).

Tingjin Chu, University of Melbourne

A geostatistical analysis of metallicity variations in galaxies

The metallicity of diffuse ionised gas (DIG) cannot be determined using strong emission line diagnostics, which are calibrated to calculate the metallicity of Hii regions. Because of this, resolved metallicity maps from integral field spectroscopy (IFS) data remain largely incomplete. In this work, we introduce the geostatistical technique of universal kriging, which allows the complete 2D metallicity distribution of a galaxy to be reconstructed from metallicities measured at Hii regions, accounting for spatial correlations between nearby data points. We apply this method to construct metallicity maps of the local spiral galaxy NGC 5236 using data from the TYPHOON/PrISM survey.

Yan Wang, RMIT University

Joint Species Distribution Modelling

Understanding species interactions and their responses to environmental variables is fundamental to ecology. Joint Species Distribution Modelling (JSDM) has emerged as a powerful tool for investigating the complex relationships between species in ecological communities. By considering multiple species simultaneously, JSDM provides valuable insights into community assembly, biodiversity conservation, and ecosystem functioning. These modelling approaches offer a comprehensive framework for ecologists to disentangle the intricacies of species interactions and to make more accurate predictions about community responses to changing environments. In this presentation, I will introduce the main approaches of the JSDM and their applications, highlighting their utilities in unravelling the co-occurrence patterns, community assembly, and the impacts of environmental factors on biodiversity.

THURSDAY, JULY 25

Hans Manner, University of Graz

Testing the equality of changepoints

Testing for the presence of changepoints and determining their location is a common problem in time series analysis. Applying changepoint procedures to multivariate data results in higher power and more precise location estimates, both in online and offline detection. However, this requires that all changepoints occur at the same time. We study the problem of testing the equality of changepoint locations. One approach is to treat common breaks as a common feature and test, whether an appropriate linear combination of the data can cancel the breaks. We propose how to determine such a linear combination and derive the asymptotic distribution resulting CUSUM and MOSUM statistics. We also study the power of the test under local alternatives and provide simulation results of its finite sample performance. Finally, we suggest a clustering algorithm to group variables into clusters that are co-breaking.

Archer Yang, McGill University

Regularized Estimating Equation: Some New Perspectives

We make some observations about the equivalences between regularized estimating equations, fixed-point problems and variational inequalities: (a) A regularized estimating equation is equivalent to a fixed-point problem, specified via the proximal operator of the corresponding penalty. (b) A regularized estimating equation is equivalent to a (generalized) variational inequality. Both equivalences extend to any estimating equations with convex penalty functions. To solve large-scale regularized estimating equations, it is worth pursuing computation by exploiting these connections. While fast computational algorithms are less developed for regularized estimating equation, there are many efficient solvers for fixed-point problems and variational inequalities. In this regard, we apply some efficient and scalable solvers which can deliver hundred-fold speed improvement. These connections can lead to further research in both computational and theoretical aspects of the regularized estimating equations. The applications in Multivariate Regularized Estimating Equations will be discussed.

Guoqi Qian, University of Melbourne

Mapping Australia's precipitation – harnessing the synergies of multi-satellite remote sensing and gauge network data

We develop Precipitation Profiler-Observation Fusion and Estimation (PPrOFusE), a tool to deliver high-quality gauge and multi-satellite fused precipitation data. PPrOFusE consists of two steps. Step 1, the relationship among the multiple sources of data is modelled by multiple linear regression at each rain gauge location, returning the least squares estimates for the associated regression coefficient vector. Step 2, such regression coefficient vector estimates for all rain gauge locations are fitted by a spatial autoregression model, whereafter the multiple linear regression coefficient vectors for those locations void of rain gauges are predicted by spatial interpolation. PPrOFusE is applied to the precipitation system in Australia over years 2000 to 2022, and has generated interesting results.

Jaehong Jeong, Hanyang University

Wind vector modeling using space-time cross-covariance models

Wind energy has significant potential for future energy portfolios without negative environmental impacts. We analyze wind vectors in East Asia from the fifth-generation ECMWF atmospheric reanalysis. To model the wind vectors, we consider non-Gaussian processes based on the Tukey g-and-h transformation, along with multivariate covariance functions. The proposed model can account for non-Gaussian features and some nonstationary dependence structures of the wind vectors. We show that the proposed method with a covariance structure reflecting the nonstationarity due to the latitude and the land-ocean differences, leads to better predictions of wind speed and wind potential, which is crucial for wind power generation planning.

MATRIX

FRIDAY, JULY 26

Arnab Hazra, Indian Institute of Technology Kanpur

Spatial wildfire risk modeling using a tree-based multivariate generalized Pareto mixture model

Wildfires pose a severe threat to the ecosystem and economy, and risk assessment is typically based on fire danger indices such as the McArthur Forest Fire Danger Index (FFDI) used in Australia. Studying the joint tail dependence structure of high-resolution spatial FFDI data is thus crucial for estimating current and future extreme wildfire risk. However, existing likelihood-based inference approaches are computationally prohibitive in high dimensions due to the need to censor observations in the bulk of the distribution. To address this, we construct models for spatial FFDI extremes by leveraging the sparse conditional independence structure of Hüsler–Reiss-type generalized Pareto processes defined on trees. These models allow for a simplified likelihood function that is computationally efficient. Our framework involves a mixture of tree-based multivariate generalized Pareto distributions with randomly generated tree structures, resulting in a flexible model that can capture nonstationary spatial dependence structures. We fit the model to summer FFDI data from different spatial clusters in Mainland Australia and 14 decadal windows between 1999–2022 to study local spatiotemporal variability with respect to the magnitude and extent of extreme wildfires. Our proposed method fits the margins and spatial tail dependence structure adequately and is helpful in providing extreme wildfire risk measures. Our results identify a significant increment in spatially aggregated fire risk across a substantially large portion of Mainland Australia, which indicates serious climatic concerns.

Kate Saunders, Monash University

An extreme value post-mortem of the 2022 Australian extreme rainfall

Extreme rainfall caused flooding in multiple Australia states in 2022; Queensland (Feb – Mar 2022), New South Wales (Mar, Jul, Oct 2022), Victoria (Oct 2022), Tasmania (Oct 2022) and the year rounded out with an ex-tropical cyclone in Western Australia (Dec – Jan 2023). In addition to the extreme and widespread nature of the flooding, some communities experienced back-to-back events. One community was flooded five times in 18 months.

In this talk, I'll undertake a post-mortem of the extreme rainfall events that impacted Australia in 2022. The goal will be to determine where we can reliably assign a probability to the recent events, and where we need concessions to make the research questions scientifically tractable. This will involve going beyond just the methods and considering the practical limitations imposed by observational records. Of particular concern in this post-mortem, is understanding whether the third consecutive La Nina contributed to the likelihood of spatially compounding and temporally compounding events in Australia, and what recommendations we could reasonably make to policy makers concerned about future extreme rainfall events.

Jenny Wadsworth, University of Lancaster

Statistical inference for multivariate extremes via a geometric approach

A geometric representation for multivariate extremes, based on the shapes of scaled sample clouds in light-tailed margins and their so-called limit sets, has recently been shown to connect several existing extremal dependence concepts. However, these results are purely probabilistic, and the geometric approach itself has not been fully exploited for statistical inference. We outline a method for parametric estimation of the limit set shape, which includes a useful non/semi-parametric estimate as a pre-processing step. More fundamentally, our approach provides a new class of asymptotically-motivated statistical models for the tails of multivariate distributions, and such models can accommodate any combination of simultaneous or non-simultaneous extremes through appropriate parametric forms for the limit set shape. Extrapolation further into the tail of the distribution is possible via simulation from the fitted model. A simulation study confirms that our methodology is very competitive with existing approaches, and can successfully allow estimation of small probabilities in regions where other methods struggle. This talk will review the basic ideas of this new framework for multivariate extremes, and illustrate recent updates to the methodology.

Pavel Krupskiy, University of Melbourne

Skewed multivariate distributions for spatial data and their extreme-value limits

Skewed multivariate distributions, such as skew-normal or skew-t distributions, are flexible parametric models that are suitable for modeling data sets with complex dependence structures. The respective limiting extreme-value distributions can capture both symmetric and asymmetric tail dependence structures thus providing greater flexibility when modeling multivariate extremes data, including spatial extremes. In this talk, we consider different methods of constructing multivariate extreme-value distributions for spatial data based on skewed multivariate distributions and study their properties.

MATRIX

MONDAY, JULY 29

Bruno Rémillard, HEC Montréal

Copula models for time series

In this presentation, I will present fundamental results for using copula-based models in time series. In particular, I will cover modelling, estimation, goodness-of-fit and tests of independence/randomness.

Bouchra Nasri, University of Montreal

On factor copula-based mixed regression models

In this talk, a copula-based method for mixed regression models is proposed, where the conditional distribution of the response variable, given covariates, is modelled by a parametric family of continuous or discrete distributions, and the effect of a common latent variable pertaining to a cluster is modelled with a factor copula. We show how to estimate the parameters of the copula and the parameters of the margins and find the asymptotic behaviour of the estimation errors. Numerical experiments are performed to assess the precision of the estimators for finite samples. An example of an application is given using COVID-19 vaccination hesitancy from several countries. This is a joint work with Pavel Krupskii and Bruno Remillard.

Harry Joe, University of British Columbia

Vine copulas with latent variables

Consider the case of a large number of variables where latent variables can explain the dependence in different groups of variables. Theory and algorithms are given for estimating the latent variables via proxies, and using the proxy variables in vine copula estimation.

Luciana Dalla Valle, University of Plymouth

Bayesian Multivariate Nonlinear State Space Copula Models for Air Pollution

A novel flexible class of multivariate nonlinear non-Gaussian state space models, based on copulas, is proposed. Specifically, it is assumed that the observation equation and the state equation are defined by copula families that are not necessarily equal. Inference is performed within the Bayesian framework, using the Hamiltonian Monte Carlo method. Simulation studies show that the proposed copula-based approach is extremely flexible, since it is able to describe a wide range of dependence structures and, at the same time, allows us to deal with missing data. The application to atmospheric pollutant measurement data in specific geographical regions shows that the approach is suitable for accurate modelling and prediction of data dynamics in the presence of missing values. Comparison to traditional approaches shows the superior performance of the proposed model with respect to predictive accuracy.

TUESDAY, JULY 30

Sévérien Nkurunziza, University of Windsor

On a class of tensor shrinkage estimators and recent identities

In this talk, we present a class of tensor shrinkage estimators as well as some challenges related to the risk analysis of such estimators. We also present some recent identities which are useful in establishing the risk dominance of tensor shrinkage estimators.

Haibo Li, University of Melbourne

Subspace projection regularization for high-dimensional Bayesian linear inverse problems

The Bayesian statistical framework provides a systematic approach to enhance the regularization model by incorporating prior information about the desired solution. For the linear Bayesian inference inverse problems with Gaussian prior, we propose a new iterative regularization algorithm based on subspace projection regularization (SPR). The SPR projects the high-dimensional problem to a series of lower-dimensional problems by iteratively constructing solution subspaces that incorporate information about the prior distribution of the underlying solution. With the new designed early stopping rules, this iterative algorithm can obtain a regularized solution with a satisfied accuracy. Several theoretical results about the algorithm are established to reveal the regularization properties of it. Both low- and high-dimensional Bayesian inference inverse problems are used to test the proposed algorithm and demonstrate its robustness and efficiency. The most computationally intensive operations in the proposed algorithm only involve matrix-vector products, making it highly efficient for high-dimensional problems.

Grace Y. Yi, Western Ontario University

Estimation and variable selection under the function-on-scalar linear model with covariate measurement error

Function-on-scalar linear regression has been widely used to model the relationship between a functional response and multiple scalar covariates. Its utility is, however, challenged by the presence of measurement error, a ubiquitous feature in applications. Naively applying the usual function-on-scalar linear regression to error-contaminated data often yields biased inference results. Further, the estimation of model parameters is complicated by the presence of inactive variables, especially when handling data with a large dimension. Building parsimonious and interpretable function-on-scalar linear regression models is in urgent demand to handle error-contaminated functional data. In this talk, I will discuss this problem and investigate measurement error effects. I will talk about procedures to simultaneously estimate functional coefficients and select salient predictors. Numerical studies will be presented to illustrate the performance of the proposed method.

Wenqing He, Western Ontario University

Feature Screening with Large Scale and High Dimensional Censored Data

Data with a huge size present great challenges in modeling, inferences, and computation. We will present a screening method for large-sized survival data, where the sample size is large and the dimension of covariates is of non-polynomial order of the sample size. We rigorously establish theoretical results and conduct numerical studies to assess the performance of the proposed method. The method capitalizes on the connections among useful regression settings and offers a computationally efficient screening procedure.

WEDNESDAY, JULY 31

Jarryd Chapman, Monash University

Modelling the Temporal Clustering of Eastern Australian Rainfall Extremes

Extreme rainfall events can cause considerable social and economic costs to the impacted communities. These costs are heightened when extreme events cluster together in time, creating a compounding hazard that may result in a natural disaster such as flooding. The East Coast of Australia was subject to repeated rainfall extremes and flooding events during the 2020-2022 “triple dip” La Nina, with the February 2022 event resulting in $4.8b in insurance claims and 27 deaths. This research uses a Non-Homogeneous Poisson Point Process to model the temporal clustering in the extreme arrival process and to analyse the spatial distribution of extremes clustering on Australia’s East Coast, finding evidence that extreme rainfall events across the East Coast exhibit temporally clustered behaviour and that this clustering is spatially dependent. The research has applications for policy purposes, risk/disaster management and the insurance industry.

Thong Nguyen-Huy, University of Southern Queensland

Applications of Copulas for Multivariate Dependence Modelling: Extreme Events and Financial Risk Transfer

Compound weather and climate events, often resulting from multiple climate drivers and hazards, have significant economic impacts. Even if these factors are not individually extreme, their combined effect can have severe consequences for society and the environment. As such, it is crucial to evaluate the potential impact of compound events. Unfortunately, existing risk assessments concentrate on single factors, neglecting scenarios where multiple factors are at play. This is problematic as climate-related catastrophes frequently result from compound events. There is still much to be discovered regarding understanding, estimating, and predicting these events.
This talk presents an analytical and modelling framework based on copulas to model the joint dependency of extreme events and subsequently develop financial strategies to transfer their risks. A copula is a mathematical function that expresses the joint cumulative probability distribution of multiple variables, allowing for accurate risk assessment of compound events in present and future climates, considering associated uncertainties.

Xueming Li, University of Melbourne

Blending Method for Various Precipitation Data Sources

Precipitation analysis is an important component of meteorological and hydrological applications. Quantifying precipitation is essential for understanding rainfall variations over time and space, as well as the impact of climate change and the society. There are two primary sources of precipitation data: rain gauge station observations, known for accuracy but sparse coverage, and remote sensing data, offering broader spatio-temporal coverage but are less accurate. The first part of this study aims to improve the accuracy of an existing fusion method called PPrOFusE by incorporating time lag and seasonal factors. This approach allows us to effectively blend the strengths of both rain gauge and satellite data, leveraging the accuracy of the former and the broader coverage of the latter. In the second part of this study, we address scenarios where the rain gauge station data is absent. We propose a new method, Triple Collocation Spatial Autoregressive under Dirichlet distribution, which fuses satellite precipitation datasets exclusively. The resulting fused precipitation dataset aims to provide a comprehensive and reliable representation of precipitation patterns particularly in regions without gauge stations.

Rachael Quill, State Electricity Commission, Victoria/Department of Energy, Environment and Climate Action

Modelling Dependence in the Renewable Energy Sector

To meet our renewable energy targets in Australia, multi-modelling approaches are required to understand the interdependence between large-scale systems, such as optimising the production of energy from wind and solar generation along-side grid-scale battery storage. From modelling the spatial dynamics of wind fields to predicting the temporal dynamics of the energy market, a multitude of techniques are required to address problems across the sector. This talk will provide an overview of some of these, many of which require hybrid techniques, blending physical approaches with statistical modelling and machine learning.

THURSDAY, AUGUST 1

Xiaoting Li, University of British Columbia

A Bayesian factor-vine copula model for extreme flood insurance losses

Statistical inference on the dependence of multivariate extremes poses notable challenges, particularly in contexts characterized by large dimensions and sparse extreme observations. While multivariate copula models, notably the vine copula and the factor copula, provide flexible parametric methods for dependence modeling, caution is warranted when using them for extremal dependence inference or tail extrapolation.
In this study, we introduce a novel class of factor-vine copula models. It is designed for modeling the dependence of extreme insurance losses within the context of the National Flood Insurance Program (NFIP), with broader applicability to spatio-temporal dependence modeling in multivariate time series data featuring a clustering structure. The factor-vine copula model is a variant of vine copula with a latent factor structure. It integrates the advantages of both vine and factor copulas by allowing for great flexibility in tail dependence modeling while maintaining interpretability through a parsimonious latent structure. Within the factor-vine model, we demonstrate how the incorporation of univariate extreme value margins and tail-weighted dependence measures can address the existing challenges associated with using parametric copulas for extreme inference. Furthermore, we discuss the practical applications of the proposed model in the context of the NFIP, focusing on its efficacy in evaluating the risks associated with extreme weather events.

Jialing Han, University of York

ARCH and GARCH-type Processes Replicated by Vt-s-vine models

Parametric and semi-parametric method for modelling and forecasting the time series based on the vt-s-vine copula approach for financial time series developed in Bladt and McNeil (2021) is applied in this paper. By combining the v-transform to describe the stochastic volatility and s-vine copulas to capture the serial dependence with non-Gaussian margins and copulas, the models offer an improvement of the replication and prediction of ARCH and GARCH-type processes.

Reza Nosratpour, RMIT University

Machine Learning approaches to classify and estimate precipitation

Precipitation as a key hydrological component has a significant impact on climate, agriculture, and public safety. Hence, precise estimation of precipitation is important for hydrological and climate studies. In recent decades many projects such as the Global Precipitation Measurement Mission (GPM), Tropical Rainfall Measurement Mission (TRMM), and Climate Prediction Center MORPHing technique (CMORPH) have been developed for estimation and forecasting of precipitation with a global and quasi-global coverage using Remotely sensed data. However, these satellite datasets have systematic and random errors, they provide useful information for improving our understanding of precipitation regimes. Moreover, Machine Learning approaches have shown significant performance in modelling complex interactions of precipitation and other atmospheric variables such as surface temperature and Precipitable Water Vapor (PWV). Multi-Layer Perceptron, Convolution Neural Networks, and Rand Forest are some of the most popular ML techniques used in different studies. By harnessing the power of machine learning (ML), precipitation models were developed that significantly improved the accuracy of satellite-based precipitation estimates. In addition, Due to the extensive spatial coverage, satellite precipitation estimates have provided researchers with an excellent opportunity to investigate various precipitation regimes. To this end, Unsupervised ML algorithms such as K-means++ have shown capabilities to classify and analysis of spatial regimes of precipitation. The results of these studies can be used as inputs for flood management models and Extreme Precipitation Early Warning systems. Ultimately, these researches have significant value for water resource management, climate studies, and improvement of satellite rainfall estimates.

Huiyan Sang, Texas A&M University

Graph-Split-Based Decision Trees and Their Ensembles for Spatial and Network Clustering, Regression and Classification

Decision trees and their ensembles such as CART, BART, XGBoost, and RF have become extremely popular in data science due to their exceptional performance in machine-learning tasks such as clustering, regression, and classification. We present a novel ensemble decision tree method on graphs to handle spatial and network data with complex geometries. The new method utilizes a highly flexible graph split-based decision rule to relax the axis-parallel split rule assumption found in most existing ensemble decision tree methods. We describe a scalable recursive algorithm on graphs for model estimations. The superior performance of the proposed method is demonstrated through various clustering, regression, and classification tasks on spatial and network data.