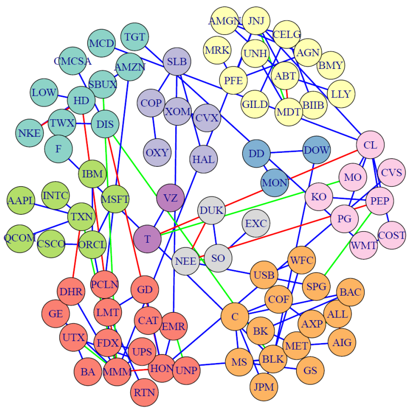
Book of Abstracts

Vine Copulas and their Applications

Workshop at the Technical University of Munich

July 8–9, 2019



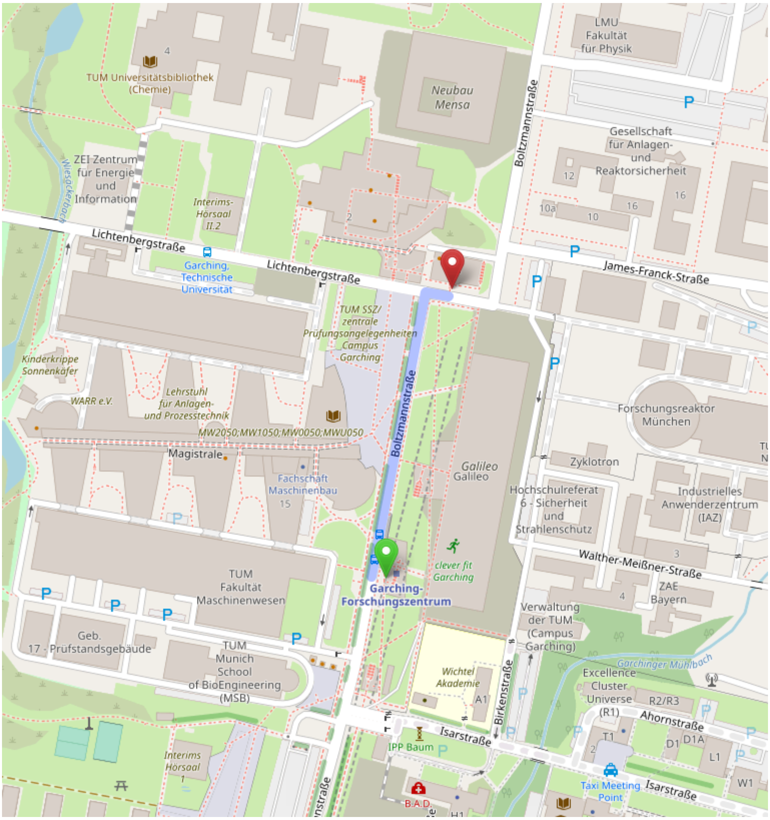
Contents

		Monday		Tuesday
08:45–09:00		Opening: F. Bornemann		
09:00–09:45		**K. Aas**		**H. Joe**
09:45–11:00	FINANCE C. Czado	**S. Petitjean**	MIXED DATA A. Kreuzer	**S. H. Kadhem**
		A. Stephan		**A. K. Nikoloulopoulos**
		coffee break		**W. Hao**
11:00–12:40	COMPUTING T. Nagler	**A. Panagiotelis**	TECHNOLOGY T. Nagler	*coffee break*
		J. Kang		**K. Doubleday**
		T. Vatter		**D. Carrera**
		A. Kreuzer		**A. Müller**
12:40–14:00		*lunch break*		*lunch break*
14:00–14:45		**C. Bernard**		**D. Kurowicka**
14:45–16:00	INSURANCE N. Barthel	**L. Yang**	REGRESSION N. Barthel	**E. Torre**
		J. Y. Ahn		**A. Möller**
		P. Shi		
16:00–17:40		*coffee break*		*coffee break*
	THEORY A. Min	**G. Puccetti**		**R. M. Cooke**
		P. X.K. Song		
		O. Okhrin		
19:00–		*conference dinner*		

Institute for Advanced Study (IAS)
Technical University of Munich
Lichtenbergstraße 2a 85748 Garching

The venue is close to the subway station *U6 Garching-Forschungszentrum*. For more details and other options to get there, see

https://www.ias.tum.de/directions/



The dinner takes place on Monday, July 8, 7 PM, at

Garchinger Augustiner
Freisinger Landstraße 4
85748 Garching b. München

The venue is close to the subway station *U6 Garching*.

- **gate-kitchen** (Thai)
  Lichtenbergstr. 8
- **Food trucks** (Kebab, Grill, Asia, Sandwiches)
  Boltzmannstr. 11
- **Herr Lichtenberg** (Menu changes daily)
  Lichtenbergstr. 6
- **FMI-Bistro** (Menu changes daily)
  Boltzmannstr. 3 (Mathematics building)

Explaining predictions from machine learning methods when features are dependent

Kjersti Aas

Machine Learning has been growing at an incredible pace during the last few years. In many applications, complex hard-to-interpret machine learning models like deep neural networks, random forests and gradient boosting machines, are currently outperforming the traditional linear and logistic regression models. Interpretability is crucial when a complex machine learning model is to be applied in areas such as medicine, fraud detection or credit scoring. Often there is a clear trade-off between model complexity and model interpretability, meaning that it might be hard to understand why these sophisticated models perform so well. In response, a new line of research has emerged that focuses on helping users to interpret the predictions from advanced machine learning methods.

In my talk I will describe one such explanation method, which is denoted *Shapley values*. This method was originally invented for assigning payouts to players in a game depending on their contribution towards the total payout. In the explanation setting, the features are the players and the prediction is the total payout. Kernel SHAP is a computationally efficient approximation to Shapley values in higher dimensions. Like several other existing explanation methods, this approach assumes independent features. In observational studies and machine learning problems, it is very rare that the features are statistically independent. If there is high degree of dependence among some or all the features, the resulting explanations might be very wrong. In our paper, we have extended the Kernel SHAP method to handle dependent features. We propose three different approaches; assuming a Gaussian distribution, assuming a Gaussian copula and using a non-parametric method. We provide several examples of linear and non-linear models with linear and non-linear feature dependence, where our method gives more accurate approximations to the true Shapley values than the original Kernel SHAP approach.

Option implied dependence

Carole Bernard

We propose a novel model-free approach to infer a joint risk-neutral dependence among several assets. The dependence can be estimated when traded options are available on individual assets as well as on their index. In the empirical application, we implement our approach using options on the S&P 500 index and its nine sectors. We find that option-implied dependence is highly non-normal and time-varying. Using the estimated dependence, we then study the correlation risk conditional on the market going down or up. We find that the risk premium for the down correlation is strongly negative, whereas it is positive for the up correlation. These findings are consistent with the economic intuition that the investors are particularly concerned with the loss of diversification when financial markets fall. As a result, they are willing to pay a considerable premium to hedge against increases in correlation during turbulent times. However, the investors actually prefer high correlation when markets rally.

Algorithms for truncated vine structures

Harry Joe

Methods and algorithms for determining some parsimonious truncated vine structures will be presented. These include truncated vine structures with latent variables that are implicit in factor copulas based on vines. One new approach uses a Monte Carlo tree search algorithm. Our methods use edge weights $-\log(1-\rho_e^2)$ on edge e of any tree of a vine, where ρ_e is a correlation or partial correlation from the van der Waarden correlation matrix. With these edge weights, fit statistics are used to compare truncated vines with and without latent variables. Examples will be presented to illustrate the methods.

Common sampling orders of regular vines with application to model selection

Dorota Kurowicka

Vine copulas become recently very popular in modelling continuous distributions with complicated dependence structures. In this model a joint density of random vector $X = (X_1, \ldots, X_d)$ is specified by the product of marginal distributions and $(d-1)d/2$ (un)conditional bivariate copulas. There are exponentially many decompositions of a density into these bivariate building blocks, and theoretically all these vine structures are equivalent. In practice, however, when copulas are fitted to data sequentially, level by level of the vine structure, and conditional copulas are assumed not to depend directly on the conditioning variables some vine structures constitute a better model of the data than the others. In [1] a heuristic search of the 'best' vine structure has been briefly introduced. It has been observed that for two vine structures, common sampling orders consistent with these vine structures, give an indication of how similar (how many repeated bivariate copulas they have in the decomposition) these vines are. In this talk we present a thorough evaluation of the heuristic based on sampling order proximity. We present an algorithm to find all vine structures with the given number of sampling orders in common and show results of an extensive simulation study of a heuristic search based on sampling order proximity.

[1] Cooke, R.M., Kurowicka, D. and Wilson, K. (2015). *Sampling, conditionalizing, counting, merging, searching regular vines.* Journal of Multivariate Analysis, 138:4-18.

Dependence modeling in expert elicitation on ice sheet dynamics

Roger M. Cooke

We report the findings of a structured expert judgement study using new techniques for modeling correlations between inter- and intra- ice sheet processes and, more crucially, their tail dependences. We find that, since the AR5, expert uncertainty has grown, in particular due to uncertain ice dynamic effects. For a +2°C temperature scenario consistent with the Paris Agreement, we obtain a median estimate of 26cm SLR contribution by 2100, with a 95th percentile value of 81cm. For a +5°C temperature scenario more consistent with unchecked emissions growth, the corresponding values are 51 and 178cm, respectively. Inclusion of thermal expansion and glacier contributions, results in a global total SLR estimate that exceeds 2m at the 95th percentile. Our findings support the use of scenarios of 21st century global total SLR exceeding 2m for planning purposes. Beyond 2100, uncertainty and projected SLR increase rapidly. The 95th percentile ice sheet contribution by 2200, for the +5°C scenario, is 7.5m as a result of instabilities coming into play in both West and East Antarctica. Introducing process correlations and tail dependences increase this value by roughly 15%.

Financial markets dependence modeling using vine copulae

Simon Petitjean

The first part of the talk will focus on selection methods for vine structures with non-monotone dependence measures with a focus on the Euclidean distance which presents several advantages.

In the second part of the talk, I will develop a benchmark construction methodology to replicate a financial portfolio considering its underlying financial characteristics such as the underlying risks, the currency hedging structure, the pricing policy, the number of observations, portfolio validation period, the management fees etc.

In last part of the talk I present an application with more than 500 asset manager strategies. This large panel is used to underline the numerical and statistical limitations of the methodology.

Portfolio optimization based on forecasting models using vine copulas: An empirical assessment for the financial crisis

Maziar Sahamkhadam and Andreas Stephan

We employ and examine vine copulas in modeling the symmetric and asymmetric dependency structure and forecasting of financial returns. Asset allocation is performed during the 2007-2010 financial crisis and different portfolio strategies are tested including maximum reward-to-risk ratio, minimum variance and minimum conditional Value-at-Risk. Regular, drawable and canonical vine copulas are specified including Clayton, Frank, Joe and mixed copula. Both in-sample and out-of-sample analyses of portfolio performances are conducted. The out-of-sample portfolio back-testing shows that vine copulas reduce portfolio risk more than simple copulas. Considering portfolio out-of-sample CVaR, Frank and mixed vine copulas result in lower downside risk. The results of the VaR back-testing shows improvement in forecasting of the downside risk for all portfolio strategies obtained from using simple Clayton and mixed copula families, implying time-varying tail dependence of stock market returns. Copula families which capture no tail dependence (Frank) and upper tail dependence (Joe) lead to higher terminal values of portfolios over the financial crisis.

Vine copulas for Approximate Bayesian Computation

Anastasios Panagiotelis, Robert Kohn, and Scott Sisson

Approximate Bayesian Computation (ABC) has become an increasingly popular technique particularly when the data generating process is easy to simulate from but when computation of the likelihood is difficult or impossible. A shortcoming of ABC is poor performance in high-dimensional settings since low acceptance rates arise due to the curse of dimensionality in matching a large number of summary statistics. In light of this methods that propose approximations based on Gaussian copulas have been proposed by Li et al (2017). To more accurately approximate posteriors with asymmetry and tail dependence, we propose ABC methods based on using a vine copula as an approximation.

[1] Li, J., Nott, DJ, Fan, Y., Sisson SA (2017). Extending approximate Bayesian computation methods to high dimensions via a Gaussian copula. *Computational Statistics and Data Analysis, 106(1): 77–89.*

Joint analysis of multimodal imaging data via nested vine copulas

Jian Kang and *Peter X.K. Song*

In cognitive neurosciences, there is an increasing interest in integration of two or more datasets acquired with different imaging techniques with the aim of improving the understanding of the structure and function of the brain. We refer to this type of dataset as multimodal imaging data. In this work, we develop a flexible modeling framework for joint analysis of multimodal imaging data via nested vine copulas, which appropriately characterize the dependence among multiple imaging modalities as well as the complex spatial correlations across different locations. We have three levels of hierarchy. At level 1, we specify the marginal distribution of the modality specific imaging outcome at each location (e.g. voxel or region of interest) in the brain. At level 2, at each location of the brain, we model the joint distribution of the multimodal imaging data via modality-dependent copulas. At level 3, we resort to another location-dependent copula construct the joint distribution of multimodal imaging outcomes over space. The modality-dependent copulas are nested in the location-dependent copula. We study the theoretical properties of the proposed method and develop efficient model inference algorithms from both Bayesian and frequentist perspectives. We illustrate the performance of the proposed method via simulation studies and a joint analysis of resting-state functional magnetic resonance imaging (fMRI) data and diffusion tensor imaging (DTI) data.

High-dimensional pair-copula constructions with financial applications

Thibault Vatter, Thomas Nagler, and Damien Ackerer

Over the last decade, pair-copula constructions (PCCs) have been applied to a wide range of scientific problems, and their success has led to continual advances in related methodology. But the popular open-source software powering most of this applied and methodological research, namely the `VineCopula` package, has numerous shortcomings. In this presentation, we discuss the data structures and algorithms behind `vinecopulib`, a C++ library for vine copula models, as well as `rvinecopulib`, its R API. Notably, redesigned algorithms, efficiently vectorized computations and improved parallelization lead to dramatically shorter runtimes, especially in high dimensions. Furthermore, storing the PCC structure and pair-copulas specifications as truncated triangular arrays, we reduce the memory footprint as well as handle both parametric and nonparametric pairs. As a motivation, we aim at solving problems actually faced by financial institutions, which often monitor around the clock investment portfolios containing several thousands of assets. Using `rvinecopulib` to build high-dimensional risk models that are both scalable and flexible, we obtain accurate risk estimates for thousands of portfolios of thousands of assets within a single framework, while also handling features such as missing data, assets disappearing/appearing over time, etc.

Dynamic regular vine copulas with an application to exchange rates dependence

Alexander Kreuzer, Claudia Czado

Modeling dependence among financial assets is an important research topic as the dependence structure has high influence on the risk associated with a corresponding portfolio. Regular vine copulas have proven as a useful tool in this context. They allow for characteristics like asymmetric tail dependence, which cannot be modeled with a multivariate Gaussian or Student t copula. Usually it is assumed, that the dependence parameters of the regular vine copula remain constant as time evolves. We get rid of this assumption and propose dynamic regular vine copulas. In this dynamic model dependence parameters are described through latent AR(1) processes. Since maximum likelihood estimation is infeasible for these latent AR(1) processes we employ Markov Chain Monte Carlo within a sequential estimation procedure. The approach is illustrated with 25-dimensional exchange rates data, where we find clear evidence for dynamic dependence.

Two-part D-vine copula models for longitudinal insurance claim data

Lu Yang and Claudia Czado

Insurance companies are typically interested in modeling the dependence in policyholders' claims over time in order to better predict future claims. Insurance claim data have their special complexity that they usually follow a two-part mixed distribution: a probability mass at zero corresponding to no claim and an otherwise positive claim from a skewed and long-tailed distribution. In this paper, we propose a two-part D-vine copula model to study the longitudinal mixed claim data. We build two stationary D-vine copulas, one is used to model the dependence in whether or not a claim has occurred, and the other is used to study the dependence in the claim size given occurrence. In addition to the benefit of efficient estimation, the proposed model can predict the probability of making claims and the quantiles of severity given occurrence straightforwardly. We then use our model to investigate the time dependence of insurance claims using a dataset from the Local Government Property Insurance Fund in the state of Wisconsin.

On copula-based collective risk models

Jae Youn Ahn, Rosy Oh, and Woojoo Lee

While the copula method is a popular choice in modelling the dependence, the choice of the proper copula family is much harder in general compared to the choice of the proper marginal distribution families. Especially, in the modelling of dependence between frequency and average severity, we show, by example, that classical copula approach may mislead the dependence between frequency and severity. Alternatively, in this presentation, we provide the copula method which can safely model the dependence among frequency and individual severities. Vine copula representation of the proposed model is also presented for the possible extension of the model.

Pair copula construction for multi-peril ratemaking in nonlife insurance

Peng Shi

In property insurance, a contract often provides the policyholder with protection against damages to the insured properties that arise from a variety of perils. In this paper, we propose a multivariate framework for pricing property insurance contracts with multi-peril coverage in a longitudinal context. The framework easily accommodates different types of data, including continuous, discrete, as well as mixed outcomes. Specifically, the longitudinal observations of each response is separately modeled using pair copula constructions with a D-vine structure. The multiple D-vines are then joined by a multivariate copula. We demonstrate the application of the model using the government property insurance data from the state of Wisconsin in the United States.

Measuring linear correlation between random vectors

Giovanni Puccetti

We introduce a new coefficient to measure linear correlation between random vectors which preserves all the relevant properties of Pearson's correlation in arbitrary dimensions. We build an empirical estimator of the newly defined correlation, give its limiting distribution and illustrate its relevance in some simulation studies. We also give some auxiliary results of independent interest in matrix analysis and mass transportation theory. The paper is available at `https://ssrn.com/abstract=3116066`.

An efficient two-stage estimation in Gaussian copulas

Peter X.K. Song

The bottleneck of applying Gaussian copulas in real-world data analysis is the lack of an efficient procedure to obtain maximum likelihood estimation. Existing two-stage estimation method such as the marginal inference function method, typically, first acquires the estimates of marginal parameters based on independent marginal models, and then obtains the estimation of correlation parameters at the second stage. This kind of method apparently loses some efficiency in estimation. In this talk, I will present a new version of two-stage estimation procedure for the likelihood estimation, where the likelihood equation for the marginal parameters is converted into an optimal estimating equation in connection to the generalized method of moments. As a result, a joint inference on all the model parameters is carried by two steps, each requiring minimal numerical effort. Numerical examples will be illustrated.

[1] Song, P.X.K. (2007). Correlated Data Analysis: Modeling, Analytics, and Applications. *Springer*.

Hierarchical Outer Power Archimedean copulas

Jan Górecki, Marius Hofert, and Ostap Okhrin

Distributions based on hierarchical Archimedean copulas (HACs) became popular as they enable one to model non-elliptical and non-exchangeable dependencies among random variables. Their practical applications reported in the literature are, however, mostly limited to the case in which all generator functions in a HAC are one-parametric, which implies that all properties (e.g., Kendall's tau and tail dependence coefficients) of each bivariate margin of such a HAC is given just by a single parameter. Involving so-called outer power transformations of Archimedean generators in such models, this limitation can be alleviated, which typically allows one to set Kendall's tau and upper-tail dependence coefficient independently of each other. This talk addresses the construction, sampling and estimation of the resulting so-called hierarchical outer power Archimedean copulas.

Factor copula models for mixed data

Sayed H. Kadhem and Aristidis K. Nikoloulopoulos

It is very common in social science to deal with datasets that have diverse types of variables, especially in surveys. Surveys contain combination of variables that are measured on a categorical, discrete or continuous scale. We extend the factor copula models in [1, 2] to the case of mixed continuous and discrete responses. Factor copula models are canonical vine copulas that involve both observed and latent variables, hence they allow flexible (tail) dependence. They can be explained as conditional independence models with latent variables that don't necessarily have an additive latent structure. We focus on important issues that would interest the data analyst, such as model selection and goodness-of-fit. Our methodology is illustrated by analysing the Swiss consumption survey data [3].

[1] Krupskii, P. and Joe, H. (2013). Factor copula models for multivariate data. *Journal of Multivariate Analysis*, 120:85–101.

[2] Nikoloulopoulos, A. K. and Joe, H. (2015). Factor copula models for item response data. *Psychometrika*, 80:126–150.

[3] Irincheeva, I., Cantoni, E., and Genton, M. G. (2012). Generalized linear latent variable models with flexible distribution of latent variables. *Scandinavian Journal of Statistics*, 39(4):663–680.

A multinomial quadrivariate D-vine copula mixed model for diagnostic studies meta-analysis accounting for non-evaluable subjects

Aristidis K. Nikoloulopoulos

Diagnostic test accuracy studies observe the result of a gold standard procedure that defines the presence or absence of a disease and the result of a diagnostic test. They typically report the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). However, diagnostic test outcomes can also be either non-evaluable positives (NEP) or non-evaluable negatives (NEN). We propose a novel model for meta-analysis of diagnostic studies that includes the number of NEP and the number of NEN as separate response categories [1]. We assume independent multinomial distributions for the TP and NEP, and, the TN and NEN, conditional on the latent sensitivity, specificity, probability of NEP and probability of NEN in each study. For the random effects distribution of the latent proportions, we employ a drawable vine copula that can successively model the dependence in the joint tails. Our methodology is demonstrated with an extensive simulation study and applied to data from diagnostic accuracy studies of coronary computed tomography angiography for the detection of coronary artery disease. The comparison of our method with the existing approaches yields findings in the real data application that change the current conclusions.

[1] Nikoloulopoulos, A. K. (2018). A multinomial quadrivariate D-vine copula mixed model for diagnostic studies meta-analysis accounting for non-evaluable subjects. *ArXiv e-prints*. `https://arxiv.org/abs/1812.05915`arXiv:1812.05915.

Copula-based mediation analysis

Wei Hao and Peter X.K. Song

Motivated by pervasive biomedical data, we propose a unified mediation analysis approach to data of mixed types, including continuous, categorical, count variables. We invoke copula models to specify the joint distribution of the outcome variable, the mediator and the exposure variable of interest in the context of generalized linear models. We develop inference procedures to evaluate causal pathways in both aspects of parameter estimation and hypothesis testing for direct and/or indirect effects of the exposure variable on the outcome variable. We compare the performance of the proposed method with other existing method using simulation studies. We illustrate our method via the data from the ”Early Life Exposures in Mexico to ENvironmental Toxicants” study, where we examine the mediation of association between phthalate exposure during pregnancy and offspring's health outcomes at teen years by the onset of offspring's infancy BMI peak being delayed or not.

Regional solar power forecasting with vine copulas for power system applications

Kate Doubleday, José Daniel Lara, Will Kleiber, and Bri-Mathias Hodge

As the deployment of variable and uncertain renewable energy sources like solar photovoltaics (PV) increases, accurate forecasting of renewable power generation is of high priority for economical and reliable power system operations. Transmission (high-voltage) system operators use probabilistic solar power forecasts at either the transmission node or balancing area level for a variety of day-ahead through minute-ahead decisions [2]. Each transmission node usually serves a large geographic region with multiple utility-scale renewable resources; hence, the spatial correlations among the power plants must be accounted for when forecasting the aggregate power generated at a given node. In this presentation, we apply vine copulas to forecast the aggregate production from 11 utility-scale PV power plants in Texas, USA, with hourly resolution up to a day ahead. The individual plant-level forecasts serve as the marginal distributions, modeled with discrete-continuous mixture models obtained through Bayesian model averaging [2]. The discrete component accounts for "clipping" when the power output is limited to its AC power rating [3]. Finally, we analyze the performance of vine copulas in accounting for the spatial correlations in the aggregate PV power forecasts.

[1] R. J. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B.-M. Hodge, U. Cali, and G. Kariniotakis (2017). *Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry.* Energies, 10(9).

[2] R. M. Chmielecki and A. E. Raftery (2011). *Probabilistic Visibility Forecasting Using Bayesian Model Averaging.* Monthly Weather Review, 139(5): 1626–1636.

[3] P. Grana (2017). *Push it to the limit: Rethinking inverter clipping.* Solar Power World,
`www.solarpowerworldonline.com/2017/09/folsom-rethinking-inverter-clipping`.

Machine learning with vine copulas: Application to three supervised classification problems

Diana Carrera

We introduce supervised classification methods based on regular vine (R-vine) copulas. The introduced classifiers can deal with a large number and varied types of features. The algorithms can produce R-vine classifiers that balance their accuracy and complexity. We show the application on two image problems: dune identification and texture segmentation; and a particular instance of the mind reading problem. In dune identification the goal is to detect the presence or absence of sand dunes in the images. In this problem, R-vine copulas are built over gradient histogram features, which describe the directional and periodic characteristics of the dune. The other problem, texture segmentation, pursues the identification of image regions based on their texture. In our approach, R-vine copulas are built on distributions of wavelets features. Finally, we also address the problem of inferring which type of video, among five classes, a subject has watched from the analysis of his/her brain signals. To deal with this problem, we introduce an R-vine-based approach specifically conceived to solve non-binary multi-class classification problems. For each problem class, an R-vine model is estimated from the corresponding instances. Then, each learned model is used to predict the class membership probability of the unlabeled instances, and the class whose corresponding model has assigned the highest probability to the instance is selected. Experimental results show the adequacy of the R-vine copula models for solving classification problems where the interactions between the variables show a great variability.

A copula-based time series model for global horizontal irradiation

Alfred Müller and Matthias Reuber

The importance of renewable energies, especially photovoltaics, in the worldwide electricity generation has increased over the past years. Thus, there is an increasing demand for probabilistic hourly models for local and global PV yields. In this paper we use an indirect modeling approach of local PV yields with irradiation data provided by the Copernicus Atmosphere Monitoring Service. We propose a statistical estimation for lower and upper bounds of global horizontal irradiations with a quantile regression and the peak over threshold method. Moreover, we introduce copula based time series models for the hourly and daily dependence structure. Time dependent parameters of the beta distributed marginals are obtained through a beta regression. We use simple vine copula models in the form of Markov trees to describe the dependence structure. There is empirical evidence for different degrees of upper and lower tail dependence in the data. Therefore we compare different approaches using Gaussian, Gumbel, BB1- and BB7-copulas. Evaluation methods like the continuous ranked probability score (CRPS) and the variogram score (VS) are used to compare the predictive power of the various model approaches.

Vine copulas for uncertainty quantification: why and how

Emiliano Torre, Stefano Marelli, Paul Embrechts, and Bruno Sudret

Systems subject to uncertain inputs produce uncertain responses. Uncertainty quantification (UQ) deals with the estimation of the response statistics of systems for which a runnable computational model is available. Problems of interest are those where the computational model is expensive, making Monte Carlo approaches unfeasible and thus calling for cheaper solutions that require fewer runs.

In these settings, an accurate representation of the input statistics, including their mutual dependencies, is critical to obtain accurate output estimates. For instance, tail dependencies among the inputs may strongly affect failure probabilities in reliability analysis. Failing to capture such forms of correlations may render accurate estimation of the output statistics hopeless, regardless of the UQ method used to carry out the analysis.

The last decade saw a remarkable extension of copula models that can be effectively used to describe multivariate dependence. Among these models, copulas built by tensor product of simple pair copulas (so-called vine copulas) enable a very flexible representation of high-order dependencies. In parallel, novel methods for inference in these models have been proposed.

Here we illustrate how these relatively recent advances in copula modeling can be easily combined with virtually any UQ analysis [1], including those methods that assume the input to have a specific multivariate distribution (such as independent inputs). We showcase the approach on a variety of examples, spanning different simulated problems as well as different UQ techniques used to solve them. The analyses are fully carried out with the UQLab toolbox [2], whose simple syntax is also illustrated.

[1] E. Torre, S. Marelli, P. Embrechts and B. Sudret (2019). *A general framework for data-driven uncertainty quantification under complex input dependencies using vine copulas.* Probabilistic Engineering Mechanics (55): 1-16.

[2] S. Marelli and B. Sudret (2014) *UQLab: A framework for uncertainty quantification in Matlab.* In: Vulnerability, Uncertainty, and Risk (Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), Liverpool, United Kingdom), chapter 257: 2554-2563

Vine copula based post-processing of ensemble forecasts for temperature

Annette Möller, Ludovica Spazzini, Daniel Kraus, Thomas Nagler, and Claudia Czado

To account for forecast uncertainty in numerical weather prediction (NWP) models it has become common practice to employ ensemble prediction systems generating probabilistic forecast ensembles by multiple runs of the NWP model, each time with variations in the details of the numerical model and/or initial and boundary conditions. However, forecast ensembles typically exhibit biases and dispersion errors as they are not able to fully represent uncertainty in NWP models. Therefore, statistical postprocessing models are employed to correct ensembles for biases and dispersion errors in conjunction with recently observed forecast errors. We propose a novel postprocessing approach for temperature forecasts based on D-vine copula quantile regression. It is a multivariate regression approach predicting quantiles of the response (temperature observations) conditioned on a set of predictor variables (the ensemble forecasts), while not making specific assumptions about the shape of the conditional quantiles. It exploits the dependence between observation and predictors, accounting for non-gaussian dependencies in a flexible and data driven way. In a comparative study with temperature forecasts of different forecast horizons from the European Center for Medium Range Weather Forecast (ECMWF) the D-vine postprocessing approach shows to be highly competitive to the state-of-the-art EMOS model, improving over standard EMOS especially for larger forecast horizons. Furthermore, an exploratory data analysis revealed that the dependency between temperature observations and its ensemble forecasts is indeed non-Gaussian, pointing to the need to extend the application of vine copula models for postprocessing, as they allow for more flexibility in the dependence structure than state-of-the-art models.

[1] Möller, A., Spazzini, L., Kraus, D., Nagler, T. and Czado ,C. (2018). *Vine copula based postprocessing of ensemble forecasts for temperature.* arXiv: 1811.02255

Web page

www.vine-copula.org

Books

Kurowicka, D., & Cooke, R. M. (2006). *Uncertainty analysis with high dimensional dependence modelling.* John Wiley & Sons.

Joe, H., & Kurowicka, D. (Eds.). (2011). *Dependence modeling: Vine Copula Handbook.* World Scientific.

Joe, H. (2014). *Dependence modeling with copulas.* Chapman and Hall/CRC. https://doi.org/10.1201/b17116

Czado, C. (2019). *Analyzing dependent data with vine copulas: A practical guide with R.* Lecture Notes in Statistics, Springer International Publishing. https://doi.org/10.1007/978-3-030-13785-4

Software

`VineCopula`: Statistical inference of vine copulas, R package.
http://tnagler.github.io/VineCopula/

`vinecopulib`: A C++ library for vine copula modeling.
https://vinecopulib.github.io/vinecopulib/

`rvinecopulib`: R interface to the `vinecopulib` library.
https://vinecopulib.github.io/rvinecopulib/

`VineCopulaMatlab`: A MATLAB toolbox for vine copulas.
http://maltekurz.github.io/VineCopulaMatlab/