# Modifying copulas for improved dependence modelling

Colette le Roux

# Modifying copulas for improved dependence modelling

by

Colette le Roux

Submitted in partial fulfilment of the requirements for the degree

MSc (Advanced Data Analytics)

In the Department of Statistics

In the Faculty of Natural and Agricultural Science

University of Pretoria

Pretoria

Supervisor: Dr A. de Waal

20 November 2020

# ABSTRACT

Copulas allow a joint probability distribution to be decomposed such that the marginals inform us about how the data were generated, separately from the copula which fully captures the dependency structure between the variables. This is particularly useful when working with random variables which are both non-normal and possibly non-linearly correlated. However, when in addition, the dependence between these variables change in accordance with some underlying covariate, the model becomes significantly more complex.

This research proposes using a Gaussian process conditional copula for this dependence modelling, focusing on time as the underlying covariate. Utilising a Bayesian non-parametric framework allows the simplifying assumptions often applied in conditional dependency computation to be relaxed, giving rise to a more flexible model.

The importance of improving the accuracy of dependency modelling in applications such as finance, econometrics, insurance and meteorology is self-evident, considering the potential risks involved in erroneous estimation and prediction results. Including the underlying (conditional) variable reduces the chances of spurious dependence modelling.

For our application, we include a textbook example on a simulated dataset, an analysis of the modelling performance of the different methods on four currency pairs from foreign exchange time series and lastly we investigate using copulas as a way to quantify the coupling efficiency between the solar wind and magnetosphere for the three known phases of geomagnetic storms.

We find that the Student's $t$ Gaussian process conditional copula outperforms static copulas in terms of log-likelihood, and performs particularly well in capturing lower tail dependence. It further gives additional information about the temporal movement of the coupling between the two main variables, and shows potential for more accurate data imputation.

# DECLARATION

I, Colette le Roux declare that the thesis/dissertation, which I hereby submit for the degree MSc Advanced Data Analytics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE:

DATE: 20 November 2020.

# ACKNOWLEDGEMENTS

I would like to thank the CSIR DSI-Interbursary Support Programme for financial support in funding my studies and making this research possible.

A special thank you to Dr. Stefan Lotz, researcher with SANSA, for your guidance and contribution to the geomagnetic storm application. It has been an honour to work with you.

To Dr. Alta de Waal, thank you for every opportunity you have given me that allowed me to grow. I could not have asked for a better supervisor and I have nothing but respect, admiration and appreciation for you.

Lastly, thank you to my family for your continuous love and support during all the years leading up to this dissertation. I would not have made it this far without you.

*S.D.G.*

# Contents

# List of Figures

# List of Tables

# SYMBOLS

| | |
|---|---|
| $p$ | multivariate dimension/ number of variables |
| $d$ | number of parameters |
| $n$ | sample size |
| $f$ | density function/ probability mass function |
| $F$ | cumulative distribution function |
| $F^{-1}$ | generalized inverse function ($F^{-1}(x) = \inf\{t \in \mathbb{R} : F_p(t) \geq x\}$) |
| $F^{[-1]}$ | pseudo-inverse function |
| $c$ | copula density |
| $C$ | copula |
| $\mathbb{R}$ | set of real values |
| $\mathbb{I}$ | indicator function |
| $\phi$ | standard normal/ Gaussian density function (/elliptical generator function) |
| $\Phi$ | standard normal/ Gaussian cumulative distribution function |
| $\nu$ | degrees of freedom |
| $t$ | Student's $t$ distribution |
| $\rho$ | (Pearson) correlation coefficient |
| $\rho_S$ | Spearman's rank correlation |
| $\rho_\tau/\tau$ | Kendall's tau rank correlation |
| $\lambda$ | coefficient of tail dependence |
| $\varphi$ | Archimedean copula generator function |
| $\mathcal{L}$ | log-likelihood function |

# ABBREVIATIONS

| | |
|---|---|
| i.i.d. | independent and identically distributed |
| RV | random variable |
| CDF | cumulative distribution function |
| PDF | probability distribution function |
| PIT | probability integral transform |
| MLE | maximum likelihood estimation |
| MCMC | Markov chain Monte Carlo |
| AIC | Akaike information criterion |
| BIC | Bayesian information criterion |
| KS | Kolmogorov-Smirnov |
| CvM | Cramèr-von-Mises |
| GP | Gaussian process |
| GPCC | Gaussian process conditional copula |
| EP | expectation propagation |
| FITC | fully independent training conditional approximation |
| CDO | collaterised debt obligation |

# Chapter 1

# Introduction

Copulas are useful for modelling dependence patterns in multivariate data, as well as prediction in regression analysis [2]. For this reason copula-based models are often applied in econometrics, finance, insurance [23, 25, 34, 49, 50] and meteorology [55].

In 2007 and 2008, underestimation of correlations and risks and the misuse of dependence models, such as the Gaussian copula, lead to the financial crisis [40]. This case study highlights the need to improve dependence modelling through both the correlation parameter and choice of model used. Uncertainty from volatilities, heteroskedasticity, extreme values and missing observations all contribute to the difficulty of dependency estimation and prediction. In addition, the assumptions of parametric or Gaussian distributions and linear correlation structures regularly applied in statistical analysis are often violated in practical applications. Much work has been done on relaxing the normal assumptions of the Gaussian copula, to deal with the uncertainties described above [19, 46].

Fully understanding the advantages (and limitations) of copula models helps to avoid model mismatching and broadens the field of possible applications. Before we embark on this research journey, we first give a simple, intuitive overview of copulas as a refresher to the reader.

## 1.1 Copulas in simple terms

Suppose we have two variables with different distributions. It is known that a dependency exists between these two variables, but a mathematical formulation for a joint distribution does not exist. The copula can represent the dependency between these variables, independently of the marginal distributions, allowing the joint distribution to be constructed.

The main tool used in the construction of copulas is the probability integral transform (PIT).

### 1.1.1 Probability integral transform

Starting by generating uniform random variables (RVs) $U_i \sim \mathcal{U}_{[0,1]}$, these variables, also known as **grades**, can be transformed to an arbitrary (univariate) probability distribution

$$X_i = F^{-1}(U_i) \sim f_{X_i}$$

by feeding the grades into the inverse CDF (Figure 1.1) of the desired distribution.

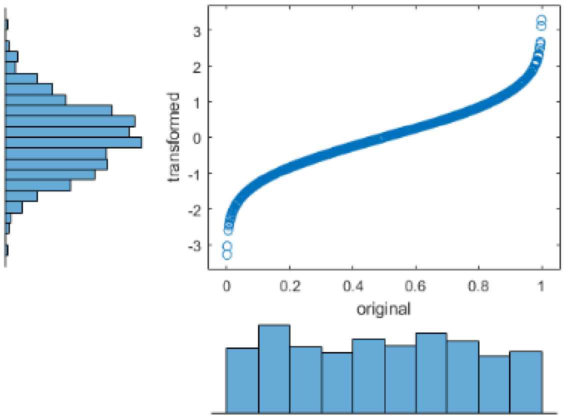


**Figure 1.1:** Inverse CDF of $\mathcal{U}_{[0,1]}$ to $\mathcal{N}(0,1)$ transformation

The opposite of this process can also be applied, by generating random variables from an arbitrary distribution $X_i \sim f_{X_i}$ and feeding them into their own marginal CDFs to transform them to a uniform distribution (Figure 1.2)

$$U_i = F(X_i) \sim \mathcal{U}_{[0,1]}.$$

These uniform random variables can now again be fed into the inverse of a desired CDF to obtain random variables with this new distribution [27].

The method of transforming from an arbitrary distribution to uniform and back is known as the probability integral transform (PIT).

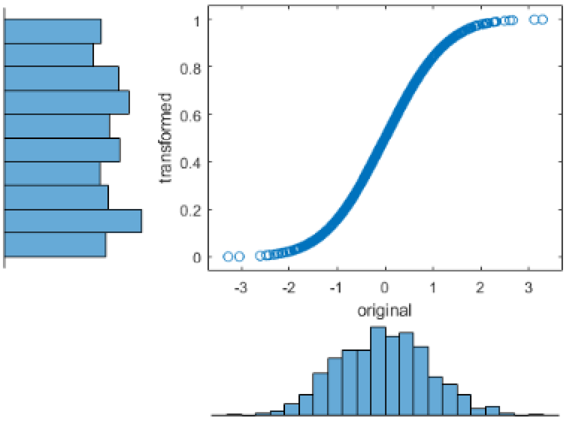
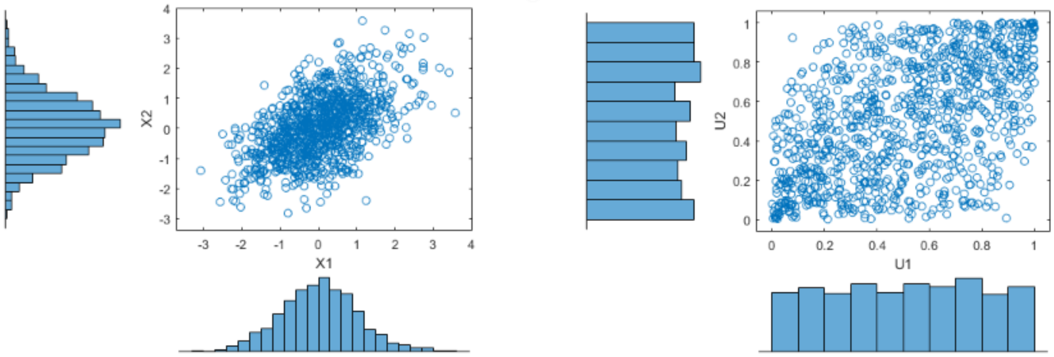


**Figure 1.2:** CDF of $\mathcal{N}(0,1)$ to $\mathcal{U}_{[0,1]}$ transformation

### 1.1.2 Constructing correlation between distributions

The PIT provides us with a tool to switch easily between the uniform and other distributions. Now we can specify a copula which acts as a custom joint probability distribution. We start by simulating RVs from a multivariate distribution and thereby specify the correlation structure between variables.

An example of a correlated multivariate Gaussian distribution with $\mu = [0,\ 0]$ and $\rho = 0.3$ is given in Figure 1.3a. These random variables are then transformed to uniform (Figure 1.3b) and then to the desired (possibly different) probability distributions (Figure 1.4).



**(a)** Correlated multivariate Gaussian

**(b)** Uniform marginals

**Figure 1.3:** Joint plot of 1000 simulated correlated Gaussian random variables and their uniform transformations

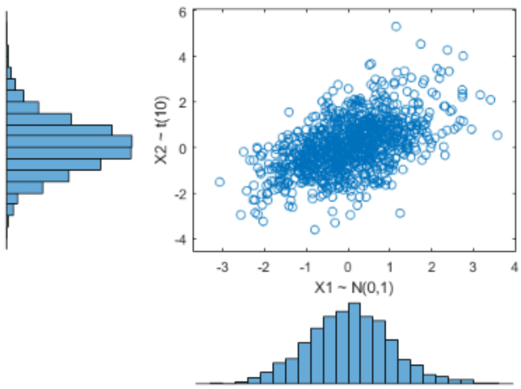


**Figure 1.4:** Joint distribution with correlation

### 1.1.3 Why copulas?

As an example, suppose the variables flood peak and flood volume need to be considered for a flood frequency analysis. (This example is discussed in detail in Chapter 3.) Intuitively, it makes sense that a joint distribution between these two variables exists. Given two completely different marginal distributions, we need to define a custom joint distribution between the two variables. Figure 1.5 illustrates how the joint density under independence differs from that using a Gaussian copula. In the left panel, it is clear how ignoring the underlying dependence between the two variables (red) underestimates the joint density, as opposed to including it using the Gaussian copula (blue).

In this research, we investigate non-parametric copulas, such as the Gaussian process conditional copula (GPCC). Without going into detail (which we do in Chapter 4) the GPCC goes even further to consider the possibility of an additional underlying variable having an effect on the dependence structure between the two main variables. Using time as the conditioning variable, the GPCC in effect becomes a dynamic model. Simply put, the GPCC allows for the contour plot in the right bottom panel of Figure 1.5 to have a different shape at each time point.

Figure 1.6 compares the (conditional) distributions of flood peak obtained from the different methods (independence, Gaussian copula and GPCC) for given values of flood volume.

Keeping the flood volume constant at its median (0.0589) and maximum (0.2125) values, the quantiles from the distribution of possible corresponding flood peak values are summarised in Table 1.1. While, say the 90% confidence interval, will stay constant when the variables are

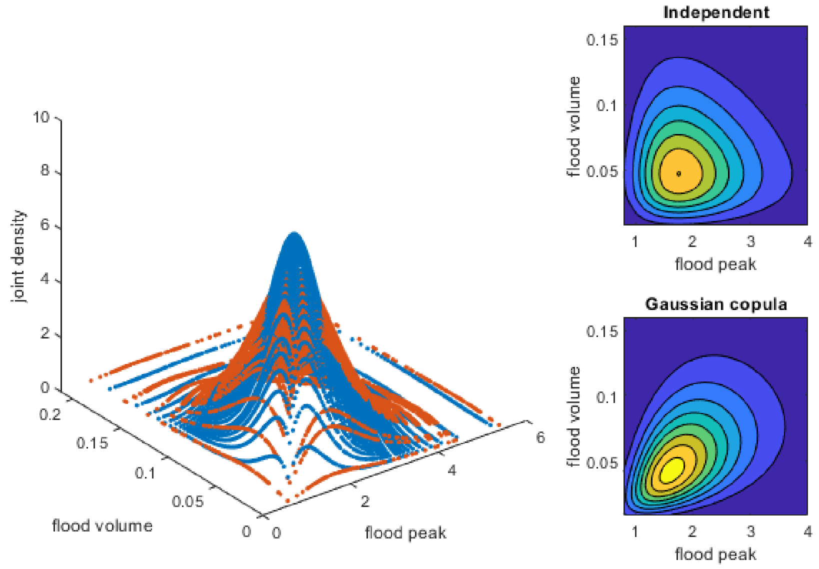


**Figure 1.5:** 3D joint density (left) of flood peak and volume under independence (red) and using the Gaussian copula (blue) and the corresponding contour plots (right) illustrating the difference in dependence structure.

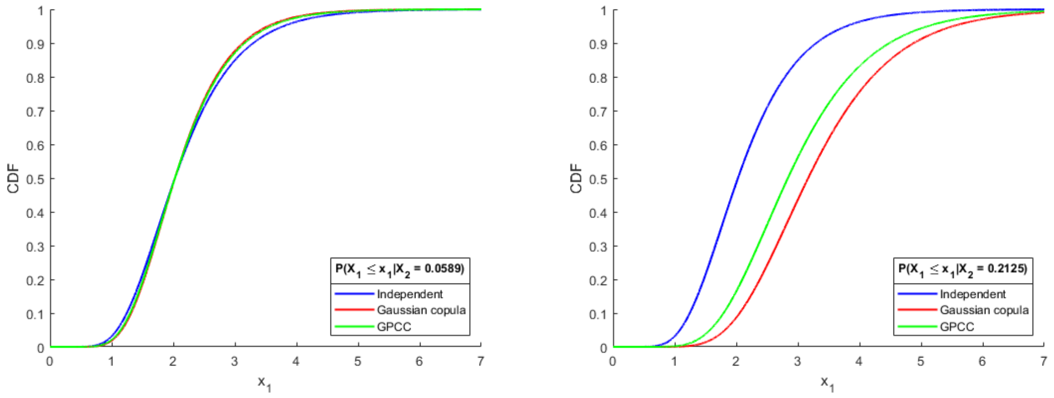


**Figure 1.6:** CDF of flood peak ($y_1$) under independence (blue) and conditional on a given value of flood volume using the Gaussian copula (red) and the GPCC (orange).

assumed to be independent, using a copula and its conditional distribution allows the confidence intervals to shift and change in size. Comparing the three methods, it is seen that, although the results do not differ significantly given the median flood volume (left pane of Figure 1.6), ignoring the underlying dependence in more extreme cases (maximum value of flood volume) may lead to underestimation of the corresponding flood peak value (right pane of Figure 1.6). Using time as the conditioning variable, the GPCC further allows for the copula parameters to change dynamically.

| Independent | Observed $X_1$ | $q_{0.1}$ | $q_{0.5}$ | $q_{0.9}$ | $q_{0.95}$ | $q_{0.99}$ |
|---|---|---|---|---|---|---|
| $X_2 = 0.0589$ | 1.4300 | 1.2462 | 2.0278 | 3.2995 | 3.7877 | 4.9078 |
| $X_2 = 0.2125$ | 2.1600 | 1.2462 | 2.0278 | 3.2995 | 3.7877 | 4.9078 |
| Gaussian copula | Observed $X_1$ | $q_{0.1}$ | $q_{0.5}$ | $q_{0.9}$ | $q_{0.95}$ | $q_{0.99}$ |
| $X_2 = 0.0589$ | 1.4300 | 1.3155 | 2.0280 | 3.1263 | 3.5342 | 4.4495 |
| $X_2 = 0.2125$ | 2.1600 | 2.0547 | 3.1673 | 4.8825 | 5.5203 | 6.9497 |
| GPCC | Observed $X_1$ | $q_{0.1}$ | $q_{0.5}$ | $q_{0.9}$ | $q_{0.95}$ | $q_{0.99}$ |
| $X_2 = 0.0589$ | 1.4300 | 1.3007 | 2.0278 | 3.1615 | 3.5857 | 4.5422 |
| $X_2 = 0.2125$ | 2.1600 | 1.7937 | 2.8380 | 4.4860 | 5.1070 | 6.5133 |

**Table 1.1:** Quantiles of flood peak ($X_1$) for given values of flood volume ($X_2$) obtained from the conditional CDF under independence, a Gaussian copula and the GPCC.

This allows for the possibility of the dependency between the two variables being different, for example during a flood due to a natural disaster compared to a rainy season, even if the flood volume has the same value.

Table 1.1 serves as a motivation for this study: The $99th$ quantile might be underestimated when ignoring underlying dependencies. Although not evident in this dataset, even the Gaussian copula may lead to underestimations with datasets containing non-linear and extreme values such as the data of the 2007 financial crisis case study. The motivation of this study is to investigate copulas which relax parametric and linear assumptions on datasets which pose these challenges.

## 1.2 Aims and objectives

The aim or this study is to improve dependence modelling with copulas by relaxing parametric and linear assumptions. In more detail, this study has the following objectives:

### Tail Dependencies

The two main approaches to handling outliers or missing data are to either remove them, or replace them with some other appropriate value [15], but there are instances, such as in risk-management, where these anomalous observations are of key importance and cannot be eliminated. In these cases, appropriate methods are needed to model tail dependencies.

How well copulas perform in capturing tail (extreme) dependencies as well as dealing with missing observations will be evaluated, as well as whether enhancing the model from a copula to a dynamic copula will increase the accuracy of predictions and estimations.

### Non-parametric, non-Gaussian and non-linear

Most traditional methods for dealing with complex dependency structures assume a Gaussian distribution and linear correlation structure, but these assumptions are often violated in practical applications [55, 19].

Furthermore, non-parametric procedures do not require the form of the distribution to be specified in advance, making it more flexible than parametric models when the underlying distribution is difficult or uncertain, but they are more difficult to interpret and estimate, especially for large datasets [29].

A Bayesian approach will be applied for model improvement when the underlying distribution of the data is non-parametric or non-Gaussian and the dependency structure non-linear.

### Underlying Covariates

While ignoring underlying covariates might yield reasonably accurate models in some instances, time has been found to have an influence on copula parameters when modelling financial data, and bringing this into account therefore leads to improved prediction and estimation [17].

Traditional copulas can be improved with the conditional copula process to take underlying (temporal) covariates that have a potential influence on the correlation structure between variables into account. Introducing a Bayesian framework to the copula also improves the model for the multi-dimensional case.

### Model Complexity

The first problem in high-dimensional dependence structure models is that the computational cost of approximation and parameter estimation increases as the dimension increases [37], making traditional bivariate copula methods, such as MLE [54] and MCMC [45] practically infeasible.

A Bayesian approach can be used to addressed this problem without dimensionality reduction.

## 1.3 Contributions

The contributions of this study can be summarised as follows:

- Theoretical overview of copulas, serving as a concise "cheat sheet" of the most commonly used copulas.
- Description and motivation of dynamic copulas. More specifically the Student's $t$ Gaussian process conditional copula, using time as the conditioning variable.
- Two applications:

  Comparison of the different methods for modelling exchange rate time series dependence;

  Novel application of copulas to capture the coupling in geomagnetic storms.
- A Student's $t$ Gaussian process conditional copula toolbox for MATLAB on GitHub[1].

## 1.4 Outline

**Chapter 1: Introduction**

We start with a simple, intuitive overview of copulas. The motivation, aims and objectives, as well as the contributions of this research are also given.

**Chapter 2: Literature review**

This chapter summarises the most common techniques used and some advantages and disadvantages of the different models discussed in the research.

**Chapter 3: Copulas**

A comprehensive introduction to copulas, including theory on basic concepts and some common copula families, measures of dependence, goodness of fit and estimation. We conclude the chapter with an example application.

**Chapter 4: Gaussian process conditional copulas**

A short discussion on the need of dynamic copulas is given, following theory on conditional copulas, measures of dependence, goodness of fit and estimation. The Gaussian process conditional

[1] GPCC Toolbox:
https://github.com/ColetteLR/GPCC.git

copula is discussed in detail and continuation of the example in Chapter 3 is given.

**Chapter 5: Application**

The first application is a comparison of the static and dynamic copulas on foreign exchange time series data. The second is a novel application of copulas as a way to quantify the coupling efficiency between the solar wind and magnetosphere of geomagnetic storms.

**Chapter 6: Conclusion**

We discuss the main results, contributions and limitations of the research, as well as some possible considerations for future work.

# Chapter 2

# Literature Review

## 2.1 Copulas

In 1940 and 1941, Hoeffding derived many of the basic results for copula functions, and in 1951 Fréchet independently derived many of the same results. It was not however until 1959 that Abe Sklar introduced the name copula [43], derived from the Latin word *copulare*, meaning to link or join, to describe these functions. In rediscoveries by other authors, they have been referred to as uniform representations [20] and dependence functions [27].

### Value Proposition

The main advantage of copulas by design, is to model the dependency structure of a multivariate probability distribution independently of its univariate, uniform marginal distributions. This simplifies the calculation of the multivariate joint distribution when the marginals come from different families of distributions or when the RVs are of mixed types [20, 43], making it a very flexible model [2]. Copulas are therefore of interest to statisticians as a way of studying scale-free measures of dependence and a starting point for constructing families of bivariate distributions [43].

Another advantage is the wide variety of available bivariate copula families, which gives rise to a more accurate and reliable measure of dependence for each data situation. One such parametric family is the elliptical family, comprising the Gaussian and Student's t copula, which are ideal for modelling symmetric dependency structures in high dimensions. These copulas make use of marginal distributions and correlation measures, such as Spearman's rank and Kendall's tau [43]. The Pearson product moment correlation coefficient estimates the linear relationship between marginal distributions, but is inaccurate for non-linear data, in which case the aforementioned two measures are a better alternative [49]. Another family is the Archimedean family, including the Gumbel, Frank, Joe, Clayton and Independent copulas [25], which capture asymmetric tail dependence [49]. These copulas can however not model dependence structures with more than two dimensions [50], and most of them have few parameters, making them less flexible.

The disadvantage that comes with these advantages is that the number of multivariate distributions that have the same underlying copula is not limited, since the dependence structures described by the copula are independent of the marginal distributions. Although the marginals can easily be found from standard univariate methods, the wide variety of dependence patterns that need to be represented by the models makes it less straightforward to know when which copula should be applied [23].

While maximum likelihood is the most efficient estimation method in the fully parametric case [54], a semiparametric copula-based model uses the copula decomposition of the joint distribution to employ a non-parametric model for the marginal distributions and a parametric model for the copula [43].

The most common goodness of fit tests for copula models are the Kolmogorov-Smirnov (KS) and Cramèr-von Mises (CvM) tests [25, 43]. Other alternatives for copula selection include the Akaike information criterion (AIC) [2, 50], Bayesian information criterion (BIC) [54], cross-validation [37] and other Bayesian methods [15].

In the next section, we introduce Gaussian Processes (GPs). GPs combined with copula models can lead to better dependency modelling, specifically in the tail regions of the distributions.

## 2.2 Gaussian Processes

Gaussian processes (GPs) are non-parametric, non-linear regression and classification tools often applied in machine learning, fitting distributions over all possible functions that are consistent with observed data. GPs can be defined by a mean function and covariance matrix consisting of positive definite kernel functions [29] of which one common kernel function used as a measure of similarity is the Radial Basis Function (Gaussian or Squared Exponential kernel function).

The advantage of GPs is that it uses the closure marginalisation and conditioning property of Gaussian RVs for efficient learning and inference [19], and a Bayesian framework for model selection and handling of missing data [17].

Approximation methods include the Dirichlet process prior [45], Laplace's approximation [51], Gibbs sampling [45], FITC approximation [23], Markov chain Monte Carlo (MCMC) approximation [37, 51] and expectation propagation (EP), which can be employed for approximate Bayesian inference [17, 23].

Due to the aforementioned closure property, a GP leads to severe under-estimations in risk-management for non-Gaussian applications [19], such as modelling financial data, wind-speeds and temperatures, and damage from natural disasters. This problem can be addressed with a copula process.

## 2.3 Copula Processes

A copula process (CP) can be used to describe the correlation structure between an arbitrary (finite) set of uniform RVs by fitting a distribution over each data point. Combining a GP with a copula function through a CP, behaviour of the marginals can be separated from the structure of dependence while keeping the non-parametric benefits of the GP, even when working with non-Gaussian distributions. Methods applied in literature to achieve this include the kernel-based copula process (KCP) [19], and the Bayesian Gaussian regression copula framework with discrete, continuous or mixed outcomes [46].

The Gaussian copula process can model the correlations between volatilities at different points in time, include other covariates (such as interest rates for financial data) and handle missing data without difficulty. Another advantage of CPs is that random samples, with arbitrary sizes and marginal distributions, can be generated with desired dependency structures [52]. An example is using Gaussian RVs and transforming them to uniform RVs with an underlying GP dependency structure, specified by a covariance function. These uniform RVs can then again be fed into the inverse CDF of another distribution to obtain RVs from this new distribution, but with the underlying dependency structure still being that of the GP.

Different data collection methods are used with different research situations. Often underlying covariates, such as the spatial or temporal intervals at which the data is collected has an influence on the observed dependence between the main variables.

While ignoring these covariates might yield reasonably accurate results in some instances [23], it has been found that time has an influence on copula parameters when modelling financial data (and therefore potentially in other data as well), and bringing this into account therefore leads to improved prediction and estimation [17].

While the CP allows additional covariates to be included in the model, the conditional copula allows us to condition on these covariates. This enables the modelling of the change in the dependence structure between the main variables in line with changes in the underlying covariate, thus modelling of the effect of underlying covariates.

## 2.4 Conditional Copulas

The conditional copula is a natural extension of a copula, linking joint conditional and marginal conditional distributions [37], and is extremely useful in modelling high-dimensional data.

Since an $n$-dimensional copula is simply a CDF with uniform marginals [41], the copula (which is assumed to be parametric) can be used to estimate a conditional CDF, while the GP is used to specify the parameters of the copula by approximating the non-linear functions of the conditioning variables.

GPs only allow for copulas with one parameter when used to model conditional copulas, but methods have been applied to extend the work to accommodate for copulas with multiple parameters [17], such as the Student's $t$ Gaussian process conditional copula (GPCC-T).

The complexity of estimating the functional relationship between the copula parameter and covariates in many parametric copula family applications lead to the need for more flexible models, such as semiparametric and non-parametric inferential tools [37].

Combining the advantages of a conditional copula approach with the modelling flexibility of Bayesian non-parametrics has been proposed [17], allowing the density of any bivariate conditional copula with continuous or mixed outcomes to be estimated [46].

Many families of parametric copulas are available for different bivariate data situations. However, the choice of parametric modelling families when working with multivariate copulas are much more limited, which is why vine copulas, or pair copula constructions are investigated [23].

Although vine copulas are not discussed in this work, we mention it briefly for reference in future work.

## 2.5 Vine Copulas

Traditional multivariate copulas are limited to cumbersome and inefficient optimisation procedures, and Harry Joe proposed vine copulas, also known as pair copula constructions, to address this issue in 1996 [55].

A vine copula is a hierarchical factorisation of a high-dimensional copula into the product of bivariate copula densities. The pair copulas can have conditional or unconditional distribution functions and be selected from any parametric or non-parametric family, making it a very flexible model [23, 45, 55]. This flexibility makes vine copulas preferable to elliptical copulas for describing complex dependence structures, especially ones with different (asymmetrical) tail dependencies [50]. Vine copulas perform well in describing highly non-linear dependencies [55] and have an advantage over its opponents in dealing with missing and abnormal observations [50].

Examples of vine copulas include the canonical vines (C-vines), drawable vines (D-vines) and regular vines (R-vines), which are a generalisation of all other types [23, 54]. The structure of an R-vine consists of a chain of nested trees with nodes and edges [54].

### Complications with higher dimensions

Maximum likelihood estimation (MLE) is typically used for parameter estimation, but the large number of parameters and complexity of the function for the high-dimensional case makes it difficult to solve with the traditional partial derivative. A common solution to this problem is applying the Nelder-Mead method [54].

The second problem is that stochastic processes form the basis for approximation and sampling methods used in copulas and GPs, among which the most common is MCMC [35]. However, this algorithm becomes practically infeasible, as the data dimensions required for computations are exponential per sampling step [45]. Since the number of variables to be conditioned on increases as one moves deeper into the vine hierarchy, calculating the conditional bivariate copula densities becomes very complex [23].

A maximum spanning tree (MST) algorithm can be employed to select the R-vine which models the variables with higher dependencies (and thus greater influence on the model fit) in the first several trees using a sequential method. As the transformed conditional variables in later trees get closer to being independent, the difference among bivariate copula families become negligible. Although this can still be accepted as a reasonable model, there is no guarantee that the resulting model will be a global optimum, since the trees in the R-vines are modelled separately [50].

The model accuracy and efficiency can be improved and the computational burden lightened by using the truncation method (using independent copulas for higher trees) [45] or simplification method (using Gaussian copulas for higher trees) [50].

Dimensionality reduction methods such as principle component analysis (PCA) and independent component analysis (ICA) assume that variables are linearly correlated and have a Gaussian distribution, but this is often not the case in practical applications. Although many improved methods have been proposed to deal with these constraints, dimensionality reduction may unavoidably lead to loss of information.

In addition, the truncation method enables vine models to be constructed from standard unconditional parametric bivariate copulas, but may cause some of the conditional dependencies in the data to be ignored, leading to an overly simplistic estimation in some data situations [23].

## 2.6 Summary

This chapter provided a literature review of copulas, starting with static copulas and then progressing to dynamic copulas and high dimensional copulas. We highlighted the main contributions, advantages and disadvantages of each approach. In the next chapter, we discuss the basic theory of copulas.

# Chapter 3

# Copulas

In this chapter, we introduce copulas as a method of measuring dependence. Understanding the basics of copulas is crucial before one can move on to more complicated models, including extending the theory to dynamic (time-varying) data. The concepts defined here, are the basic building blocks required for the rest of this dissertation. Figure 3.1 provides an outline of the chapter contents.

## 3.1 Introduction

A natural part of everyday life is trying to minimise uncertainty and understand the risks involved in decisions. The need to control these risks is prevalent in areas such as finance, insurance, econometrics, meteorology, and many more, where underestimation can lead to large scale catastrophes.

While this decision making problem is still simple enough when only one variable is being considered, it becomes more complex when a combination of variables need to be taken into account in the process [5]. It is therefore desirable to understand how these variables co-vary, i.e. how a change in one variable affects the other.

### 3.1.1 Dependency modelling

Dependencies arise when one factor affects more than one variable [9]. Examples of these are

- business cycles, such as the effect of change in season on all wine farms, also leading to seasonal employment.

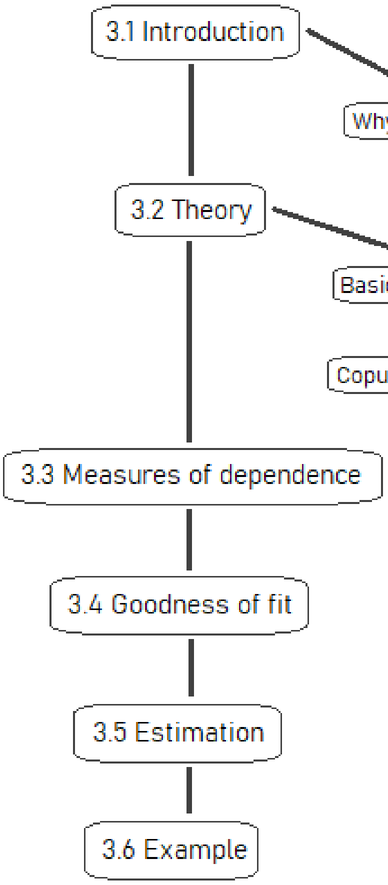


**Figure 3.1:** Schematic illustration of Chapter 3

- concentration of risks in a given sector, say how an investor is affected by the corona virus if he invested in different clothing brands, compared to if he had diversified his risks by investing in different types of business, including food stores and pharmacies.
- extreme events, for example, how a hurricane affects businesses of all kinds.

Note that in all three cases, the individual variables (different farms, clothing brands and companies respectively) may be considered unrelated under normal conditions, but the random factors, such as cycles and extreme events, creates a dependency between them.

### Why do we do dependency modelling?

In a Financial Risk Management setting, risks are measured and managed across a diverse range of activities [3]. From the previous example of concentration of risks, investing in a variety of brands might seem ’safe’ in general (if one brand fails, others might still thrive), but it is the dependencies that arise, or become visible under extreme events where all of these brands are negatively affected at the same time that are of critical importance in order to reduce exposure to the risk of potential losses.

Estimation of these dependencies under (normal) non-volatile conditions may be affected by both the quality and quantity of data available, as well as the complexity of the dependence structure. Understanding dependency structures between different variable pairs are further complicated when the structures are built up from several factors or when dependencies occur at different levels.

An example is how students are clustered within classes under teachers, and these teachers are again clustered within a school. Due to the multilevel (nested) data structure, students within the same class may achieve marks that are more alike than they are to that of students in different classes (i.e. different teachers). Similarly, teachers in the same school may be more similar to each other than the teachers in other schools. Ignoring these relationships within and between groups (classes or teachers and schools) could lead to misleading results.

### Why is correlation important?

Classical methods estimate dependencies in terms of a covariance matrix induced from the data [17]. These covariances are however sensitive to the measuring units of the variables and are not bounded, making it difficult to compare different values. Correlation provides a unitless, bounded measure of the strength and direction of the linear dependence between a variable pair [3]. Similarly to dependencies, correlated data may arise from

- longitudinal studies (multiple measurements on the same variable at different points in time)
- clustering (measurements on variables sharing a common category or characteristic)

Although the variables are independent, the observations are correlated due to the nested data structure. Ignoring correlation and analysing data as if independent may lead to incorrect inferences and inefficient estimation.

### Correlation as dependency modelling technique

The Pearson's correlation coefficient (defined in Section 3.3) is a scalar measure of the strength of the linear association between two variables and is only appropriate when working with normal, or more generally, elliptical distributions [9]. It is bounded on the interval $[-1, 1]$ (whith $\rho = -1$ and $\rho = 1$ corresponding to perfect negative and perfect negative dependence respectively). It is important to note that although independence implies uncorrelated ($\rho = 0$), the converse is not true:

- independent: no relationship
- uncorrelated: no *linear* relationship

Another important fact to bear in mind is that correlation does not imply causation. It is possible to obtain a significant value for correlation while two variables are absolutely uncorrelated. Take for example an increase in ice cream sales accompanied by an increase in drownings at a beach. Eating ice cream is certainly not a cause of drowning, but ice cream sales and swimming are both influenced by sunny weather. Weather results in two uncorrelated events.

### Why/when does correlation not work?

Often in practice the individual variables (marginals) are not well approximated by a normal distribution, but rather a skewed distribution. Furthermore, assuming a joint normal distribution restricts the form of the dependence structure between marginal variables.

Correlation is calculated based on the marginal distributions of the variables which also leads to some pitfalls [3]:

- Not all values on the interval $[-1, 1]$ are attainable. This includes the possibility of having perfect dependence, but a correlation other than $-1$ or $1$.
- Correlation is not invariant under strictly (monotone) increasing transformations of the variables.
- The correlation is restricted to have finite variances, which is not the case for heavy-tailed distributions.

When working with non-elliptical distributions, the correlation coefficient is no longer a satisfactory measure. In this case the marginal distributions and correlations are not enough to determine the joint multivariate distribution, since additional information about the dependence structure is required [3]. This is where the copula comes in handy.

Consider two variables $X_1$ and $X_2$. For these two RVs *independent*, the joint distribution is simply the product of the two marginal distributions:

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2).$$

If they are *dependent* with *normal* marginals, there exists a closed form equation for the correlation known:

$$f_{X_1,X_2}(x_1,x_2,\rho_{12})$$

If they are *dependent* with known marginal distributions, the joint distribution is obtained as

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)C,$$

where the copula $C$ only contains information about the dependence between $X_1$ and $X_2$ (no information about the marginal distributions). This can easily be extended to the multivariate case for $X_1, X_2, ..., X_n$.

### 3.1.2 Copulas as a technique to model dependencies

#### Why it works better than correlation

For independent random variables, the joint distribution can be described by the marginal probability distributions, but when there is a dependency between the variables, some additional information is needed to describe the joint probability distribution in full. Although a multidimensional CDF can be used to represent the joint distribution, this leads to information duplication, since the joint CDF can be used to reconstruct the marginals [21]. This duplication makes the process less time efficient, and a duplication-free method is desirable.

The number of multivariate distributions that have the same underlying copula is unlimited and the marginals can therefore not be obtained based on the copula. A copula is an efficient (duplication-free) method of representing the additional information about the multidimensional distribution with known marginals.

As mentioned before, copulas model the dependency structure of a multivariate probability distribution independently of its univariate, uniform marginal distributions. This simplifies the calculation of the multivariate joint distribution when the marginals come from different families of distributions or when the RVs are of mixed types [20, 43], making it a very flexible model [2]. Copulas are therefore of interest to statisticians as a method of studying scale-free measures of dependence and a starting point for constructing families of bivariate distributions [11, 32, 31].

#### When not to use copulas?

It is important not to apply copula methods blindly in a 'black-box' fashion. The flexibility of the copula approach can be both a virtue and a potential pitfall [56]. If the model is misspecified, it can lead to underestimations in dependencies which can make the difference between landing on Mars and missing the planet entirely.

Another problem is instances where some underlying covariate significantly influences the actual dependencies, for example, dependencies varying over time or being affected by observations of other variables or factors. Since standard copula methods are static and not capable of dealing with conditional dependencies, using a copula to estimate dependencies may be inaccurate under these situations [17].

## 3.2 Theory

In this section, basic copula concepts and theory are introduced. Note that the dependence and goodness of fit measures, as well as estimation methods are defined in terms of copula theory, where applicable.

### 3.2.1 Basic copula concepts

*Sklar's theorem* states that for every multivariate ($p$-dimensional) cumulative distribution function $H$ with marginal distributions $(F_1, ..., F_p)$ there exists a copula $C$ with uniform marginals such that [32, 36, 48]

$$H(x_1, ..., x_p) = \mathbb{P}(X_1 \leq x_1, ..., X_p \leq x_p) = C(F_1(x_1), ..., F_p(x_p)). \tag{3.1}$$

For all marginal distributions continuous, the copula

$$C(u_1, ..., u_p) = H(F_1^{-1}(u_1), ..., F_p^{-1}(u_p)), \tag{3.2}$$

where $u_i = F_i(x_i)$ and $i = 1, ..., p$, will be unique.

For $H$ $p$-times differentiable, the joint density is obtained as

$$h(\mathbf{x}) = \frac{\partial^p}{\partial x_1 \partial x_2 ... \partial x_p} H(\mathbf{x}) = \prod_{i=1}^{p} f_i(x_i) c(F_1(x_1), ..., F_p(x_p)) \tag{3.3}$$

with corresponding copula density [41, 48]

$$\mathbf{c}(u_1, ..., u_p) = \frac{h(F_1^{-1}(u_1), ..., F_p^{-1}(u_p))}{\prod_{i=1}^{p} f_i(F_i^{-1}(u_i))}. \tag{3.4}$$

The *copula analysis* process separates the joint distribution into the copula and marginal distribution:

$$(X_1, \ldots, X_N)^T \sim f_X \mapsto \begin{cases} f_{X_1}, \ldots, f_{X_N} \\ (U_1, \ldots, U_N)^T \sim f_U \end{cases}$$

*Copula synthesis* on the other hand, is the process of combining the copula with the marginals to obtain the joint distribution:

$$\left.\begin{matrix} f_{X_1}, \ldots, f_{X_N} \\ (U_1, \ldots, U_N)^T \sim f_U \end{matrix}\right\} \mapsto (X_1, \ldots, X_N)^T \sim f_X$$

The *survival copula*, $\hat{C}$, is defined as

$$\bar{H}(x_1, ..., x_p) = \mathbb{P}(X_1 > x_1, ..., X_p > x_p) = \hat{C}(\bar{F}_1(x_1), ..., \bar{F}_p(x_p)),$$

where $\bar{H}$ is the joint survival function with marginal survival functions $\bar{F}_i$, $i = 1, ..., p$ [31]. The analysis of survival probabilities are of particular interest in credit risk violations [9].

For $p$ uniform RVs defined on $[0,1]$ with joint distribution function the copula $C$, the joint survival function is $\bar{C}(u_1, ..., u_p) = \hat{C}(1-u_1, ..., 1-u_p)$. From this property, a copula has *radial symmetry* if $C(u_1, ..., u_p) = \hat{C}(1-u_1, ..., 1-u_p)$ [10].

### Conditional probabilities

For some threshold value $q$, the conditional probabilities can be calculated as follows:

$$\begin{aligned}\mathbb{P}(x_1 < q | x_2 < q) &= \frac{c(F_1(q), F_2(q); \theta)}{F_2(q)} \\ \mathbb{P}(x_1 > q | x_2 > q) &= \frac{1 - F_1(q) - F_2(q) + c(F_1(q), F_2(q); \theta)}{1 - F_2(q)}\end{aligned}$$

Values of $q$ in the upper and lower tails are of particular importance [56].

### Conditional distributions

Let $C$ be the joint distribution of $U_1$ and $U_2$ with $U_1$ observed, and $U_2$ to be predicted or estimated. The conditional distribution of the copula [41, 18] is then obtained as

$$\begin{aligned}\mathbb{P}(U_2 \leq u_2 | U_1 = u_1) &= \lim_{\delta \to 0} \frac{\mathbb{P}(U_2 \leq u_2 | U_1 \in (u_1 - \delta, u_1 + \delta])}{\mathbb{P}(U_1 \in (u_1 - \delta, u_1 + \delta])} \\ &= \lim_{\delta \to 0} \frac{C(u_1 + \delta, u_2) - C(u_1 - \delta, u_2)}{2\delta} \\ &= \frac{\partial}{\partial u_1} C(u_1, u_2).\end{aligned}$$

For an $n$-copula $C$, let $C_k(u_1, ..., u_k) = C(u_1, ..., u_k, 1, ..., 1)$ for $k = 2, ..., n-1$. (Note that $C(1, ..., 1, u_i, 1, ..., 1) = u_i$, since the marginal distributions are uniform.) Then for $U_1, ..., U_n$ with joint distribution function $C$, the conditional distribution of $U_k$, given the values of $U_1, ..., U_{k-1}$ is given by

$$\begin{aligned}C_k(u_k | u_1, ..., u_{k-1}) &= \mathbb{P}(U_k \leq u_k | U_1 = u_1, ..., U_{k-1} = u_{k-1}) \\ &= \frac{\partial^{k-1}}{\partial u_1, ..., \partial u_{k-1}} C_k(u_1, ..., u_k) / \frac{\partial^{k-1}}{\partial u_1, ..., \partial u_{k-1}} C_{k-1}(u_1, ..., u_{k-1}).\end{aligned}$$

These conditional distribution concepts can be applied to generate a random variate $(u_1, ..., u_n)^T$ from $C$ [10].

### Empirical copulas

An empirical copula is a non-parametric estimator of a copula:

$$C_n(\mathbf{u}) = H_n(F_{n1}^{-1}(u_1), ..., F_{np}^{-1}(u_p)),$$

with $H_n(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\mathbf{X}_i \leq \mathbf{x})$ and $F_{nj}(\mathbf{x}_j) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(\mathbf{X}_{ij} \leq \mathbf{x}_j)$ for $j = 1, ..., p$ the joint and marginal empirical distribution functions respectively [7].

### Invariance under monotonic transformations

Since a copula is independent of the marginal probability distributions, the dependence structure does not change for any strictly increasing transformations of the random variables [41, 18].

Stated formally, for RVs $X_1, ..., X_p$ with dependence structure defined by the copula $C$, define $T_i : \mathbb{R} \mapsto \mathbb{R}$, $i = 1, ..., p$ as a strictly increasing function. Then the dependence structure of the RVs $T_1(X_1), ..., T_p(X_p)$ is given by the same copula $C$.

This copula therefore captures the non-parametric, scale-invariant nature of the dependence between the $X_i$'s.

### Fréchet-Hoeffding Bounds

All copulas are contained within bounds corresponding to the cases of extreme dependence (illustrated in Figure 3.2). Consider the copula $C(\mathbf{u}) = C(u_1, ..., u_p)$. Then

$$max\{\sum_{i=1}^{p} u_i + 1 - p\} \leq C(\mathbf{u}) \leq min\{u_1, ..., u_p\}$$

defines the Fréchet and Hoeffding bounds [41, 18, 32].

From the upper bound, the *comonotonic copula*, corresponding to perfect positive dependence ($U_2 = U_1$ in the bivariate case), is given by

$$M(\mathbf{u}) := min\{u_1, ..., u_p\}.$$

This copula exists when $X_i = T_i(X_1)$ for $i = 2, ..., p$ and $T_i$ a strictly increasing transformation.

From the lower bound with $p = 2$, the *countermonotonic copula*, corresponding to perfect negative dependence ($U_2 = 1 - U_1$ in the bivariate case), is defined as

$$W(u_1, u_2) = max\{u_1 + u_2 - 1, 0\}.$$

In contrast to these two copulas, the *independence (product) copula* satisfies

$$\Pi(\mathbf{u}) = \prod_{i=1}^{p} u_i.$$

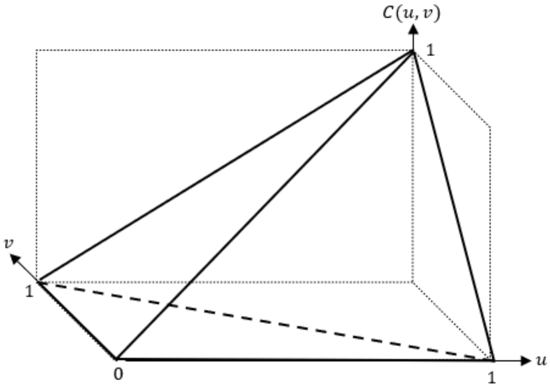


**Figure 3.2:** Fréchet-Hoeffding bounds. The bottom surface and back side is the lower bound $C(u, v) = max\{u + v - 1, 0\}$, while the front side is the upper bound, $C(u, v) = min\{u, v\}$ [41].

### 3.2.2 Some parametric copula families

In this section, some general theory about the different copulas are given, followed by a short summary of important features.

#### Elliptical copulas

Let $\mathbf{X}$ be a $p$-dimensional random vector with $\mu \in \mathbb{R}^p$ and $\mathbf{\Sigma}$ a $p \times p$ symmetric, positive definite matrix. The characteristic function of $\mathbf{X} - \mu$ is a function of the quadratic form $\mathbf{t}^T \mathbf{\Sigma} \mathbf{t}$:

$$\varphi_{\mathbf{X}-\mu}(t) = \phi(\mathbf{t}^T \mathbf{\Sigma} \mathbf{t})$$

For $\mathbf{X}$ elliptically distributed with parameters $\mu$, $\mathbf{\Sigma}$ and $\phi$, denote $\mathbf{X} \sim E_p(\mu, \mathbf{\Sigma}, \phi)$ [10].

Some properties of elliptical distributions [10, 41]:

- The density of $\mathbf{X}$ is of the form $|\Sigma|^{-\frac{1}{2}}g((\mathbf{X}-\mu)^T\mathbf{\Sigma}(\mathbf{X}-\mu))$ for some positive function $g$.
- $p = 1$ produces one-dimensional symmetric distributions with characteristic generator $\phi$.
- If $\Sigma$ is a diagonal matrix, $\mathbf{X}$ has uncorrelated components ($0 < Var(X_i) < \infty$).
- If $\mathbf{X}$ has independent components, then $\mathbf{X} \sim \mathcal{N}_p(\mu, \mathbf{\Sigma})$.
- A linear combination of independent elliptically distributed random vectors with the same $\Sigma$ (up to a constant $c > 0$) remains elliptical.
- The conditional distribution of $\mathbf{X_1}$ given $\mathbf{X_2}$ is also elliptical.
- For random vectors with a joint nonsingular ($\Sigma_{ii} > 0$ for all $i$) elliptical distribution, the linear correlation matrix is denoted by $R$, with $R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$.
- Since an elliptical distribution is uniquely determined by $\mu$, $\mathbf{\Sigma}$ and $\phi$, the copula of a non-singular elliptically distributed random vector is uniquely determined by $R$ and $\phi$.

Although all margins of a multivariate elliptical distribution need to be of the same type, an elliptical copula with different (not necessarily elliptical) types of margins can be chosen to construct a realistic multivariate distribution.

In this case, the copula parameter $R$ can no longer be estimated directly from the data, but its relation to Kendall's tau (3.6) can be used to obtain a more robust non-parametric estimator of the linear correlation ($sin(\frac{\pi\hat{\tau}}{2})$), which can again be estimated directly from the data.

**Gaussian copula**

$$C_\Sigma(\mathbf{u}) = \Phi_\Sigma(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_p))$$

The tractable properties of multivariate normal distributions enable multivariate extremes and other non-normal dependency structures to be modelled. The multivariate normal distribution is also the only elliptical distribution for which uncorrelated implies independence.

The Gaussian copula is asymptotically independent in both tails and has the linear correlation coefficient as dependence parameter.

It is a *comprehensive copula* [41], since it interpolates between the fundamental structures of countermonotonicity, independence and comonotonicity via only one parameter ($\rho = -1, 0$ and $1$ respectively, illustrated in Figure 3.3).

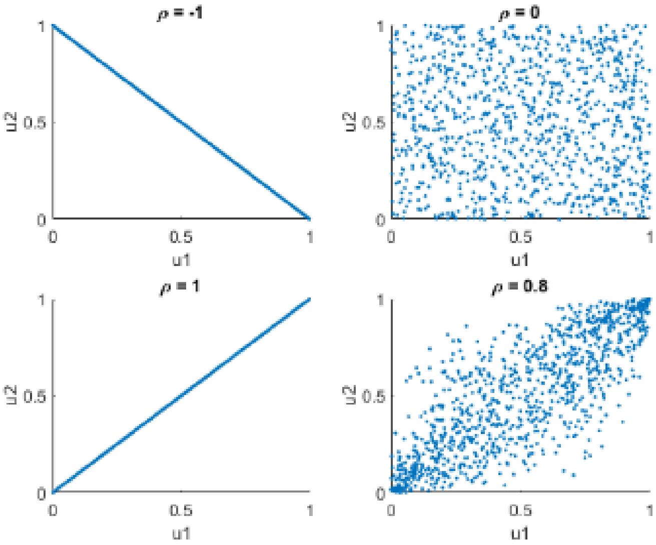


**Figure 3.3:** Simulated Gaussian copula with various levels of $\rho$

**Student's $t$ copula**

$$C_{\nu,\Sigma}(\mathbf{u}) = t_{\nu,\Sigma}(t_{\nu}^{-1}(u_1), ..., t_{\nu}^{-1}(u_p))$$

The $t$-copula exhibits symmetric tail dependence. In contrast to the Gaussian copula, the $t$-copula will exhibit dependence even for zero correlation, since some dependence is introduced by the chi-square variables.

While elliptical copulas are derived from distributions, Archimedean copulas are given explicitly. Although simulation from elliptical copulas are easy, they do not have a closed form expression and are restricted to have radial symmetry, such that asymmetries cannot be modelled. Archimedean copulas are an ideal alternative in the case of asymmetries and a closed form expression for the copula can be obtained.

### Archimedean copulas

Let $\varphi : [0,1] \rightarrow [0,\infty]$ be a continuous, strictly decreasing and convex function, such that $\varphi(1) = 0$. The pseudo-inverse, $\varphi^{[-1]} : [0,\infty] \rightarrow [0,1]$, is defined as

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & 0 \leq t < \varphi(0) \\ 0 & \varphi(0) \leq t < \infty \end{cases}$$

For $C$ a function from $[0,1]^2$ to $[0,1]$, an Archimedean copula is characterised by

$$C(u,v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)), \tag{3.5}$$

where $\varphi$ is referred to as the *generator* of $C$ [7].

If $\varphi(0) = \infty$, then $\varphi^{[-1]} = \varphi^{-1}$ and $\varphi$ is called a *strict generator*. In this case $C(u,v) = \varphi^{-1}(\varphi(u)+\varphi(v))$ will be a *strict Archimedean copula* [10] and $C(u,v) > 0$ for all $(u,v)$ in $(0,1]$ [31].

The main characteristics of the Clayton, Gumbel and Frank copulas are summarised in Table A.1.

In general, with $n \geq 3$, the multivariate Archimedean copula [10] becomes

$$\begin{aligned} C^n(\mathbf{u}) &= \varphi^{[-1]}(\varphi(u_1) + ... + \varphi(u_n)) \\ &= C(C^{n-1}(u_1, ..., u_{n-1}), u_n) \end{aligned}$$

## 3.3 Measures of dependence

Due to different copula functions having specific dependence structures, the values of dependence parameters are not directly comparable across copulas. For this reason, these parameter values need to be converted to some measure of concordance as the basis for testing the agreement between the different methods. In this sub-section we introduce a few important measures of dependence.

| Copula | Generator: $\varphi(t)$ | Copula function: $C_\theta(u,v)$ | Parameter range | Kendall's $\tau$ | Fundamental structures | Tail dependence |
|---|---|---|---|---|---|---|
| Gaussian | | $\Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$ | $-1 \leq \theta \leq 1$ | $\frac{2}{\pi}\arcsin(\theta)$ | $C_{-1} = W$,<br>$C_0 = \Pi$,<br>$C_1 = M$ | asymptotically independent in both tails |
| Student $t$ | | $t_{\nu,\Sigma}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))$ | $-1 \leq \theta \leq 1$ | $\frac{2}{\pi}\arcsin(\theta)$ | | symmetric tail dependence |
| Gumbel | $(-\ln(t))^\theta$ | $\exp[-((-\ln u_1)^\theta + (-\ln u_2)^\theta)^{\frac{1}{\theta}}]$ | $\theta \geq 1$ | $1 - \frac{1}{\theta}$ | $C_{-1} = \Pi$,<br>$\lim_{\theta\to\infty} C_\theta = M$ | upper tail dependence: $2 - 2^{\frac{1}{\theta}}$ |
| Frank | $-\ln(\frac{e^{-\theta t}-1}{e^{-\theta}-1})$ | $-\frac{1}{\theta}\ln(1+\frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1})$ | $\theta \in \mathbb{R}\backslash\{0\}$ | $1 + \frac{4}{\theta}(D_1(\theta) - 1)$ with $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt$ | $\lim_{\theta\to-\infty} C_\theta = W$,<br>$\lim_{\theta\to 0} C_\theta = \Pi$,<br>$\lim_{\theta\to\infty} C_\theta = M$ | satisfies $C(u,v) = \hat{C}(u,v)$ for radial symmetry |
| Clayton | $\frac{t^{-\theta}-1}{\theta}$ | $(max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\})^{\frac{1}{\theta}}$ | $\theta > 0$ | $\frac{\theta}{\theta+2}$ | $C_{-1} = W$,<br>$\lim_{\theta\to 0} C_\theta = \Pi$,<br>$\lim_{\theta\to\infty} C_\theta = M$ | lower tail dependence: $2^{-\frac{1}{\theta}}$ |

**Table 3.1:** Copula families characteristic "look-up" table.

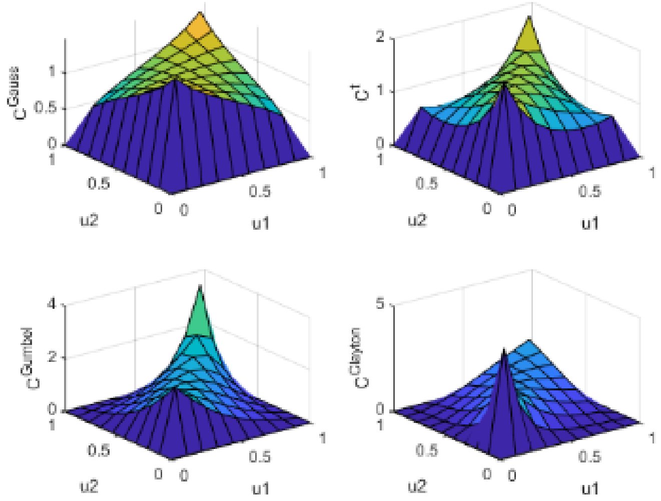


**Figure 3.4:** Different copula densities with the same dependence parameter value

### 3.3.1 Linear (Pearson) correlation

For RVs $X$ and $Y$ with nonzero finite variances, the linear correlation is given by

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}.$$

Under perfect linear dependence, i.e. $Y = aX + b$ with $a \in \mathbb{R}\backslash\{0\}$ and $b \in \mathbb{R}$, it is almost certain that $|\rho(X, Y)| = 1$, otherwise $-1 < \rho(X, Y) < 1$.

From the property

$$\rho(\alpha X + \beta, \gamma Y + \delta) = sign(\alpha\gamma)\rho(XY),$$

$\alpha, \gamma \in \mathbb{R}\backslash\{0\}$ and $\beta, \delta \in \mathbb{R}$, the linear correlation is invariant under strictly increasing linear transformations.

Since the Pearson correlation coefficient only measures the strength of the linear dependence between variables, it is a useful measure of dependence in elliptical distributions [10, 18, 41]. It is not a copula-based measure of dependence, and for any RVs which are not jointly elliptically distributed, uncorrelated in this case does not imply no relationship (independence), since some other dependence structure may still be present.

### 3.3.2 Rank correlation

Rank correlation is a non-parametric correlation estimator which is invariant under strictly increasing transformations. Under continuous marginals, this correlation only depends on the unique copula, and not on the marginals of the joint distribution.

For *Spearman's rank*, the correlation of the ranks can be obtained by taking the linear correlation of the probability-transformed RVs. In the multivariate case, the positive-definite correlation matrix is obtained as

$$\rho_S(\mathbf{X}) := Corr(F_1(X_1), ..., F_d(X_d))$$

with pairwise Spearman's rho correlations $\rho_S(\mathbf{X})_{ij} = Corr(F_i(X_i), F_j(X_j))$.

In the bivariate case of RVs $X_1$ and $X_2$ with continuous marginals, the corresponding copula $C$ can be used to derive the correlation directly [31] as

$$\rho_S(X_1, X_2) = 12 \int_0^1 \int_0^1 (C(u_1, u_2) - u_1 u_2) du_1 du_2 = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3.$$

For a bivariate Gaussian copula, Spearman's rho can be calculated using the Pearson correlation coefficient,

$$\rho_S(X_1, X_2) = \frac{6}{\pi} arcsin \frac{\rho}{2}.$$

For RVs $X_1$ and $X_2$ independent of RVs $\tilde{X}_1$ and $\tilde{X}_2$, but with the same joint distribution, *Kendall's tau* can be calculated as

$$\begin{aligned}\tau(X_1, X_2) :&= E[sign((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2))] \\ &= P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) > 0) - P((X_1 - \tilde{X}_1)(X_2 - \tilde{X}_2) < 0).\end{aligned}$$

This is simply the difference between the probability of concordance and disconcordance of $(X_1, X_2)$ and $(\tilde{X}_1, \tilde{X}_2)$ [31].

For $X_1$ and $X_2$ with continuous marginals and $C$ the copula describing their dependence structure,

$$\tau(X_1, X_2) = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1,$$

and for the bivariate Gaussian copula,

$$\tau(X_1, X_2) = \frac{2}{\pi} arcsin\rho, \tag{3.6}$$

which gives a much more robust estimate of the correlation compared to the Pearson estimator [41, 18].

In both cases, the parameter values are bounded on $[-1, 1]$, and a value of $-1, 0$ and $1$ results in the countermonotonic - , independent - and comonotonic copula respectively.

### 3.3.3 Tail dependence

Whereas correlation measures overall dependence, tail dependence considers the dependence between extreme values, i.e. probability of an extreme event occurring in one variable, given that an extreme event occurs in another variable. Furthermore, for the continuous RVs, the tail dependence is a function of the copula (dependence structure), and is therefore invariant under strictly increasing transformations of the variables [10, 41, 56, 18, 42, 5].

Let some threshold value $q$ represent an extreme event. Then, for RVs $X_1$ and $X_2$ with CDFs $F_1$ and $F_2$, the coefficient of *lower tail* dependence is defined as:

$$\lambda_l := \lim_{q \searrow 0} P(X_2 \leq F_2^{\leftarrow}(q) | X_1 \leq F_1^{\leftarrow}(q)) = \lim_{q \searrow 0} P(X_1 \leq F_1^{\leftarrow}(q) | X_2 \leq F_2^{\leftarrow}(q)),$$

provided that the limit exists. For continuous CDFs, using Bayes' rule,

$$\lambda_l := \lim_{q \searrow 0} \frac{P(X_1 \leq F_1^{\leftarrow}(q), X_2 \leq F_2^{\leftarrow}(q))}{P(X_1 \leq F_1^{\leftarrow}(q))} = \lim_{q \searrow 0} \frac{C(q,q)}{q}.$$

Similarly the coefficient of *upper tail* dependence is

$$\lambda_u := \lim_{q \nearrow 1} P(X_2 > F_2^{\leftarrow}(q) | X_1 > F_1^{\leftarrow}(q)) = \lim_{q \nearrow 1} P(X_1 > F_1^{\leftarrow}(q) | X_2 > F_2^{\leftarrow}(q)),$$

and for continuous CDFs,

$$\lambda_u = \lim_{q \nearrow 1} \frac{1 - 2q + C(q,q)}{1 - q}.$$

For both coefficients, $\lambda_i > 0$ (where $i = l, u$) indicates tail dependence, while $\lambda_i = 0$ indicates that $X_1$ and $X_2$ are asymptotically independent in the relevant tail. For symmetric copulas, $\lambda_l$ and

$\lambda_u$ will be identical.

## 3.4 Goodness of fit measures

The goodness of fit measures are applied to evaluate the performance of different models and accordingly select the model which leads to minimal loss of information from the true unknown distribution [15]. The KS test evaluates the goodness of fit of a distribution function in the non-parametric case, and scoring functions, such as AIC and BIC, are applied to compare models. Discrepancy estimators can be used as scoring functions, where the trade-off in the models will be to find a balance between a good fit and parsimony, or alternatively, between bias and variance.

### 3.4.1 Kolmogorov-Smirnov (KS)

The KS distance is used to test whether a RV $X$ has a particular underlying distribution $F$. This distribution can be estimated using the empirical distribution function, denoted by $S(x) = \frac{1}{n}\sum \mathbb{I}\{x_i \leq x\}$, i.e. the propostion of sample observations less or equal to any point $x$ [15]. Denoting the theoretical distribution under the hypothesis by $F^*(x)$, the hypothesis is defined as:

$$\begin{aligned} H_0 :& F(x) = F^*(x) \\ H_1 :& F(x) \neq F^*(x) \end{aligned}$$

The $KS$ statistic can then be calculated as

$$T_1 = \sup_{-\infty<x<\infty} |S(x) - F^*(x)|.$$

This statistic is simply the supremum of the (uniform) vertical distance between the two functions.

### 3.4.2 AIC & BIC

Although a higher model complexity yields greater model flexibility, and thus a better fit to the observed data, it may also lead to overconditioning of the model [39]. The AIC and BIC scoring functions overcome overconditioning by penalising model complexity, i.e. models with additional parameters [15, 56]. Since these criteria do not have a threshold, their measurement scales are difficult to interpret, but in both cases, the model with the smallest value is preferable.

*Akaike information criterion (AIC)* takes the model complexity into account by adding a penalty term based on the number of parameters in the model, thereby providing a more robust measure of the quality of model prediction:

$$AIC = -2\mathcal{L}(\hat{\xi};\mathbf{x}) + 2d,$$

with $\mathcal{L}(\hat{\xi};\mathbf{x})$ the log likelihood function and $\hat{\xi}$ the corresponding maximum likelihood parameter estimates. AIC assumes that the number of parameters $d$ stays constant as the sample size $n$ increases and does therefore not provide a consistent estimate for the dimension of an unknown model.

*Bayesian information criterion (BIC)* solves the consistency problem of the AIC by taking the sample size into account:

$$BIC = -2\mathcal{L}(\hat{\xi};\mathbf{x}) + dlog(n)$$

As $n$ increases, BIC favours simpler models compared to AIC [15], and is therefore the preferred evaluation method for the performance of different copula models.

## 3.5 Estimation

Once the appropriate marginal distributions and copula have been specified, the corresponding parameters need to be estimated. Methods using the likelihood function and Markov chain Monte Carlo are briefly discussed in this section.

### 3.5.1 Maximum likelihood estimation (MLE)

Maximum likelihood provides the best fit to the observed data by identifying the parameter set which minimises the residuals of the model simulations and observations [39]. Using the joint density in (3.3) with parametric marginal CDFs and an i.i.d. sample $X_{1:n} = (X_1, ..., X_n)$, the copula parameters are estimated by optimising the log-likelihood function

$$\mathcal{L}(\xi;\mathbf{x}) = \sum_{i=1}^{n}\left(\sum_{j=1}^{p} log(f_j(x_{i,j};\phi_j))) + log(\mathbf{c}(F_1(x_{i,1}), ..., c(F_p(x_{i,p})); \theta)\right) \tag{3.7}$$

with respect to the parameters, where $\xi = (\phi, \theta)$ with $\phi$ the marginal parameters and $\theta$ the copula parameters [48].

The problem with MLE is that the number of parameters to estimate increases in line with the number of variables, $p$, increasing the difficulty of the optimisation problem. In addition, misspecification of any of the marginals can introduce biases in the estimation of both the marginals and the copula [18]. These problems can be resolved by applying pseudo-MLE, which uses a two-step approach for estimation.

### 3.5.2 Pseudo-Maximum likelihood estimation

Decomposing (3.7) into the marginal - and copula log-likelihood, the equation can be rewritten as

$$\mathcal{L}(\xi; \mathbf{x}) = m\mathcal{L}(\phi; \mathbf{x}) + c\mathcal{L}(\theta; \mathbf{u}, \phi). \tag{3.8}$$

In the first step, the marginal log-likelihoods ($m\mathcal{L}$) are optimised independently of each other, and in the second step the copula log-likelihood ($c\mathcal{L}$) is optimised conditional on the results from the first step [48].

Pseudo-MLE can follow either a parametric - or a semi-parametric approach [18]. While MLE for both the marginals and copula is the most efficient estimation method in the fully parametric case, a semiparametric copula-based model provides computational tractability, utilising the copula decomposition of the joint distribution to apply a non-parametric model for the marginals and a parametric model for the copula. Estimation is then done via empirical CDFs for the marginals and MLE for the copula [34].

For large values of $d$, the last term in (3.8) may still be difficult to maximise, in which case an additional structure can be imposed on its parameters $\theta$, or ideal starting values for the optimisation can be obtained using a moment-matching approach [18].

### 3.5.3 Markov chain Monte Carlo (MCMC)

Monte Carlo is a method for sampling independent identically distributed (i.i.d.) random numbers from a desired ("target") distribution $p(x)$, using methods such as the inverse transform (or probability integral transform) method and acceptance-rejection method (when the distribution is unknown). If we want to estimate the expected value of some function $f(.)$, we can adjust the accuracy of our estimated statistic by increasing the sample size $n$ in the corresponding formula,

$$\mathbb{E}[f(\mathbf{X})] = \frac{1}{n}\sum_{i=1}^{n} f(X_i),$$

where $\mathbf{X}$ is our matrix of random numbers $X_i$ for $i = 1, ..., n$ and $p(x) = \frac{1}{n}$ since the random numbers are i.i.d..

The Markov chain simulates a sequence of dependent random numbers, each dependent on the previous number in the sequence, with limiting distribution $p(x)$. Naturally, each value in the Markov chain is dependent on the starting value. The first part of the sample in which the sequence stabilises, called the "burn-in" period, can be discarded so that each $X_i$ approaches a stationary distribution, say $p_*(.)$. Let $i = 1, ..., M$ represent the "burn-in" period, the ergodic expected value can be estimated as

$$\frac{1}{n-M}\sum_{i=M+1}^{n} f(X_i). \tag{3.9}$$

By the ergodic theorem of stochastic processes, this will approach the expected value of $f(\mathbf{X})$ calculated with respect to $p_*(.)$ as $n \to \infty$ for any fixed $M$, provided that the correlation between the last observation $f(X_n)$ and the sample mean calculated in (3.9) decreases as $n$ increases.

The Markov chain Monte Carlo method is used to design ergodic Markov chains with $p_*(.)$ equal to $p(x)$. Plotting these random numbers generated by the respective distributions in sequence produces a random walk process. An advantage of the MCMC method is that the integral in the Bayesian applications with posterior distribution as target distribution need not be calculated [44], but can be approximated.

In this case the Bayesian approximation then makes use of the prior distribution and likelihood function to obtain the posterior distribution through the relation $posterior = prior \times likelihood$ or

$$p(\theta \mid x) = p(\theta) \times p(x \mid \theta).$$

The posterior distribution can be viewed as the posterior belief about the data after updating the prior belief with new evidence (observed data). If no information about the data is available prior to the estimation, a uniform distribution can be used, which will be referred to as an *uninformative* prior. In this case the posterior distribution will be equal to the likelihood function and the accuracy of the estimated posterior distribution can be improved by increasing the sample size. The opposite (decreasing the sample size) can be applied when information, and thus an informative prior, is available. Here the posterior distribution will be similar to the prior distribution. When the prior and posterior distribution belong to the same family of distributions, the prior is referred to as a *conjugate* prior.

## 3.6 Example

The following example includes some of the basic copula concepts mentioned above. We use the MATLAB multivariate copula analysis toolbox (MvCAT) with associated dataset 'data2.txt' to repeat the frequency analysis example from [39].

### 3.6.1 Data description

In order to perform a multivariate frequency analysis on flood return periods at the Saguenay River, annual flood peak $[Q(m^3/s)]$ and volume $[V(m^3)]$ data pairs were collected from daily streamflow data. The dataset includes a total of 97 data pairs.

From the joint distribution plot in Figure 3.5 and the summary statistics in Table 3.2, both variables have a leptokurtic, positively skewed distribution. Furthermore, the joint distribution seems more concentrated in the bottom left corner, corresponding to a low flood peak and low flood volume. We therefore expect some positive dependency between the two variables.

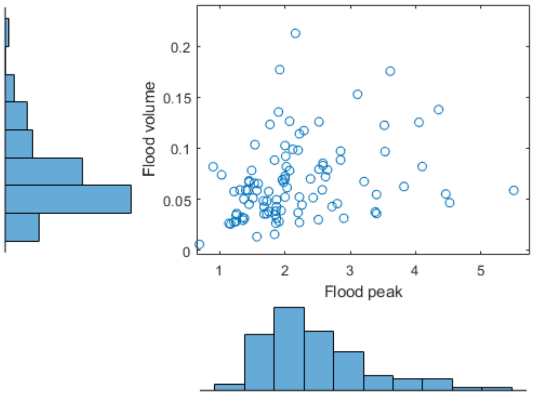


**Figure 3.5:** Joint distribution of flood peak $[Q(m^3/s)]$ and volume $[V(m^3)]$.

| | flood peak $[Q(m^3/s)]$ | flood volume $[V(m^3)]$ |
|---|---|---|
| Mean | 2.1818 | 0.0670 |
| Std. Dev. | 0.8875 | 0.0380 |
| Skewness | 1.2713 | 1.2959 |
| Kurtosis | 4.6539 | 4.9516 |
| Minimum | 0.6900 | 0.0059 |
| $Q_{0.25}$ | 1.5625 | 0.0379 |
| Median | 1.9800 | 0.0589 |
| $Q_{0.75}$ | 2.5725 | 0.0823 |
| Maximum | 5.5000 | 0.2125 |

**Table 3.2:** Summary statistics

### 3.6.2 Dependence measures

The inter-dependency between the two flood variables is measured with different correlation coefficients and is summarised is Table 5.5.

| Correlation type | Correlation Coefficient | p value | Significant at 5%? |
|---|---|---|---|
| Kendall rank | 0.2769 | 0.0001 | Yes |
| Spearman's rank-order | 0.3994 | 0.0001 | Yes |
| Pearson correlation coefficient | 0.2991 | 0.0029 | Yes |

**Table 3.3:** Evaluate dependence between the two input variables.

All of these measures indicate a significant dependence between the the two variables. We will therefore consider both flood peak and volume, as well as their inter-dependency in our frequency analysis model.

### 3.6.3 Estimation

The marginal distribution of each variable is estimated empirically. The best distribution function fit is then selected based on the BIC goodness of fit metric. The parameters are estimated using maximum likelihood, such that the distance between the empirical and modelled probability value is minimised.

For the estimation of the copula parameters, we can either apply a local optimisation algorithm or an MCMC simulation within a Bayesian framework. While local optimisation methods are likely to get trapped in local optima, MCMC provides a more robust estimate of the global optima. The latter also allows prediction uncertainty for the parameters to be measured by approximating the posterior distribution of the copula families [38, 39]. These posterior distributions of the parameters are given in Figure 3.6.

The almost uniform marginal distribution for the second parameter of the $t$ copula indicates that the information in the observed data is not sufficient to constrain the parameters. This is also visible in that the local optimisation and MCMC copula parameters do not coincide, but rather diverge from each other. The parameters of the Clayton and Gumbel copulas converges to the parameter bounds. This means that the optimisation algorithm is trying to improve the fit by going outside the bounds, which is not permitted. These copulas are thus not appropriate for the data [39].

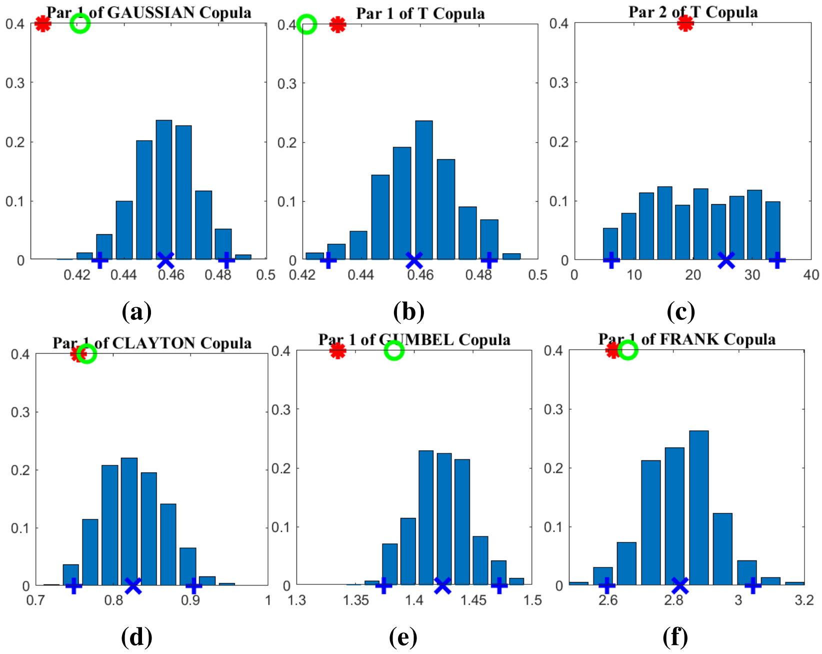


**Figure 3.6:** Posterior distribution of (a) Gaussian, (b-c) Student's $t$, (d) Clayton, (e) Gumbel and (f) Frank copulas derived by a MCMC simulation within a Bayesian framework. The copula value derived by the local optimisation approach (red asterisk), theoretical value (green circle) and the MCMC maximum likelihood parameter (blue cross) are also indicated on the graphs.

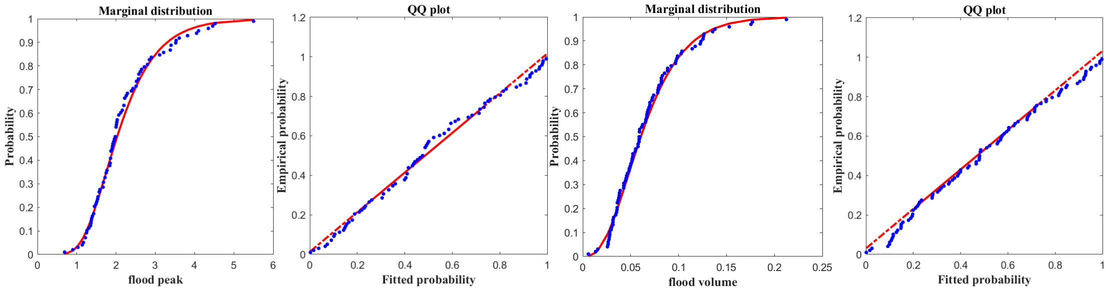
### 3.6.4 Goodness of fit

The chi-square goodness of fit test is applied to test whether the data is in fact sampled from the fitted distribution at a 5% level of significance. The fitted marginal distributions and corresponding estimated parameter values are summarised in Table 3.4. For a visual inspection, the CDF and

QQ-plot of the two variables are given in Figure 3.7, and confirms that the fitted distributions are acceptable.

| | Fitted distribution | Par 1 | Par 2 | Chi-square test |
|---|---|---|---|---|
| Flood peak | Log-Normal | $\mu = 0.7070$ | $\sigma = 0.3798$ | accepted |
| Flood volume | Gamma | $\alpha = 3.3792$ | $\beta = 0.0198$ | accepted |

**Table 3.4:** Evaluate fit of marginal distributions to flood peak and volume data.



**(a)** **(b)** **(c)** **(d)**

**Figure 3.7:** Visual comparison of the fitted distribution (red line) and the empirical distribution (blue dots). The cumulative distribution function and corresponding quantile-quantile plot for the flood peak are given in figures (a) and (b) respectively, and that of the flood volume are given in figures (c) and (d).

In terms of the copula goodness of fit, maximum likelihood, Akaike information criterion (AIC), Bayesian information criterion (BIC), Nash-Sutcliffe efficiency (NSE), and root mean squared error (RMSE) are used as metrics, and are summarised in Table 3.5. Note that an NSE value of 1 and an RMSE value of 0 are considered to be a perfect fit respectively.

| Copula | ML | AIC | BIC | NSE | RMSE |
|---|---|---|---|---|---|
| Gaussian | 421.5835 | **-841.1670** | **-838.5923** | 0.9971 | 0.1276 |
| Student's $t$ | **421.6730** | -839.3460 | -834.1966 | **0.9971** | **0.1275** |
| Clayton | 410.4603 | -818.9206 | -816.3459 | 0.9964 | 0.1431 |
| Gumbel | 416.3809 | -830.7618 | -828.1871 | 0.9968 | 0.1346 |
| Frank | 400.6776 | -799.3553 | -796.7806 | 0.9955 | 0.1583 |
| Independence | 298.1974 | -596.3947 | -596.3947 | 0.9631 | 0.4553 |

**Table 3.5:** Goodness of fit criteria of the fitted copulas: Maximum Likelihood (ML), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Nash-Sutcliffe Efficiency (NSE), and Root Mean Squared Error (RMSE).

While likelihood, NSE and RMSE only focus on minimizing the residuals between observations and model simulations, the AIC and BIC take additional criteria, such as model complexity and number of observations, into consideration. Although the former three metrics all suggest a Student's $t$ copula as the best fit, the estimated degrees of freedom is large enough ($> 25$) that we can switch to the Gaussian copula instead, corresponding to the AIC and BIC suggestion.

The dependency structure between flood peak and flood volume as estimated by the different copula models are plotted in Figure 3.8. It is clear how different copula families can return different dependence structures, even though they are all modelling the same dependence structures. This emphasises the significance of the choice of copula and quantifying the underlying copula model uncertainties [39].

The parameter estimates for the different copulas, along with their 95% uncertainty range as derived by the MCMC simulation, are summarised in Table 3.6. From this it is also seen that the uncertainty is minimised for the Gaussian copula.

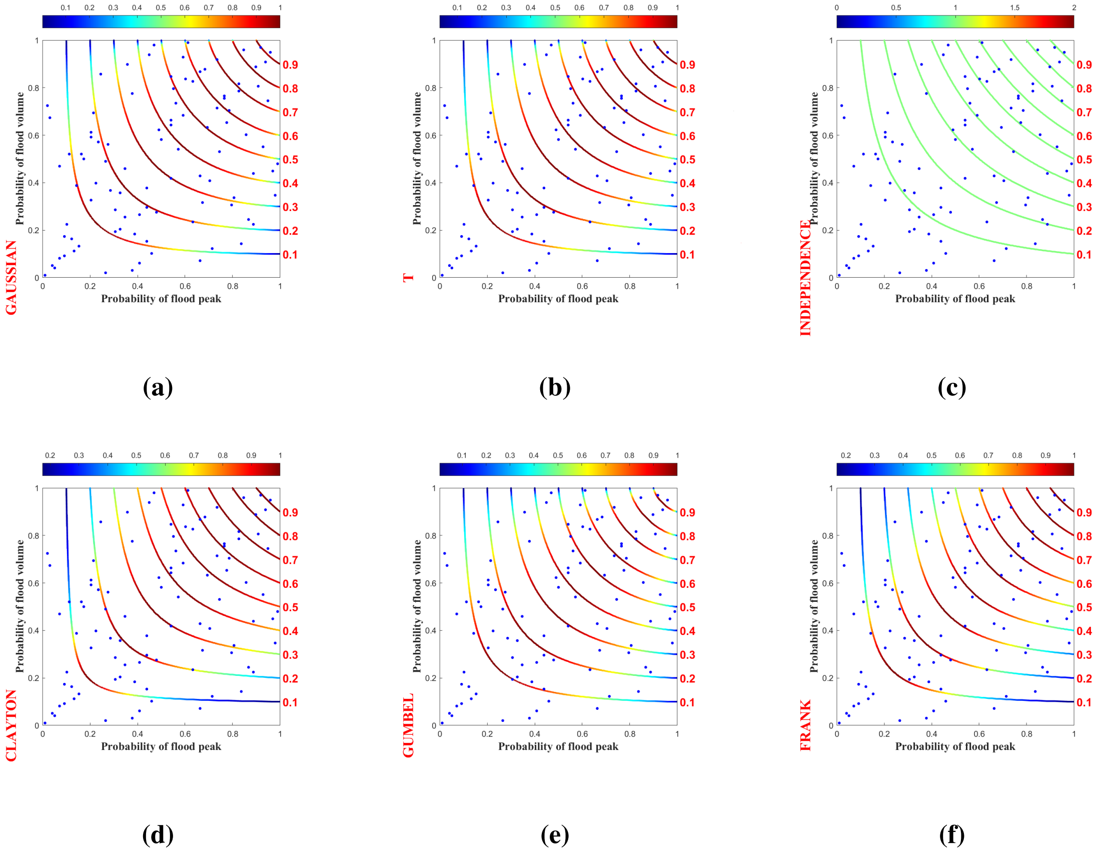
| Copula | Par 1<br>Local | Par 2<br>Local | Par 1<br>MCMC | 95% Par 1<br>Uncertainty<br>Range | Par 2<br>MCMC | 95% Par 2<br>Uncertainty<br>Range |
|---|---|---|---|---|---|---|
| Gaussian | 0.4055 | NaN | 0.4576 | [0.4317, 0.4838] | NaN | NaN |
| Student's $t$ | 0.4320 | 18.7059 | 0.4574 | [0.4260, 0.4841] | 26.2503 | [5.7991, 34.3748] |
| Clayton | 0.7549 | NaN | 0.8257 | [0.7434, 0.9024] | NaN | NaN |
| Gumbel | 1.3350 | NaN | 2.8202 | [2.6299, 3.0367] | NaN | NaN |
| Frank | 2.6168 | NaN | 1.4243 | [1.3802, 1.4693] | NaN | NaN |
| Independence | NaN | NaN | NaN | NaN | NaN | NaN |

**Table 3.6:** Estimated copula parameters.

Choosing the Gaussian copula as the best fit, the joint probability isolines for the copula data space, along with its corresponding return period levels and survival copula data space are given in Figure 3.9. The joint return period (RP) is estimated as

$$RP = \frac{1}{P(Q \geq q_p, V \geq v_p)} = \frac{1}{1 - C(u, v)}$$

such that either flood peak (Q) or flood volume (V), or both exceed a pre-specified threshold value ($q_p$, $v_p$) [39]. Note that this is simply the inverse of the survival copula for the specified threshold values. For a return period of 25 years, the design values of flood peak and volume, based on the most likely design scenario, is $4.382m^3/s$ and $0.1583m^3$ respectively. The design values derived through univariate analysis are $3.945m^3/s$ for flood peak and $0.1424m^3$ for volume. From these results it is clear that ignoring the interactions between these variables can lead to underestimation



(a) (b) (c)

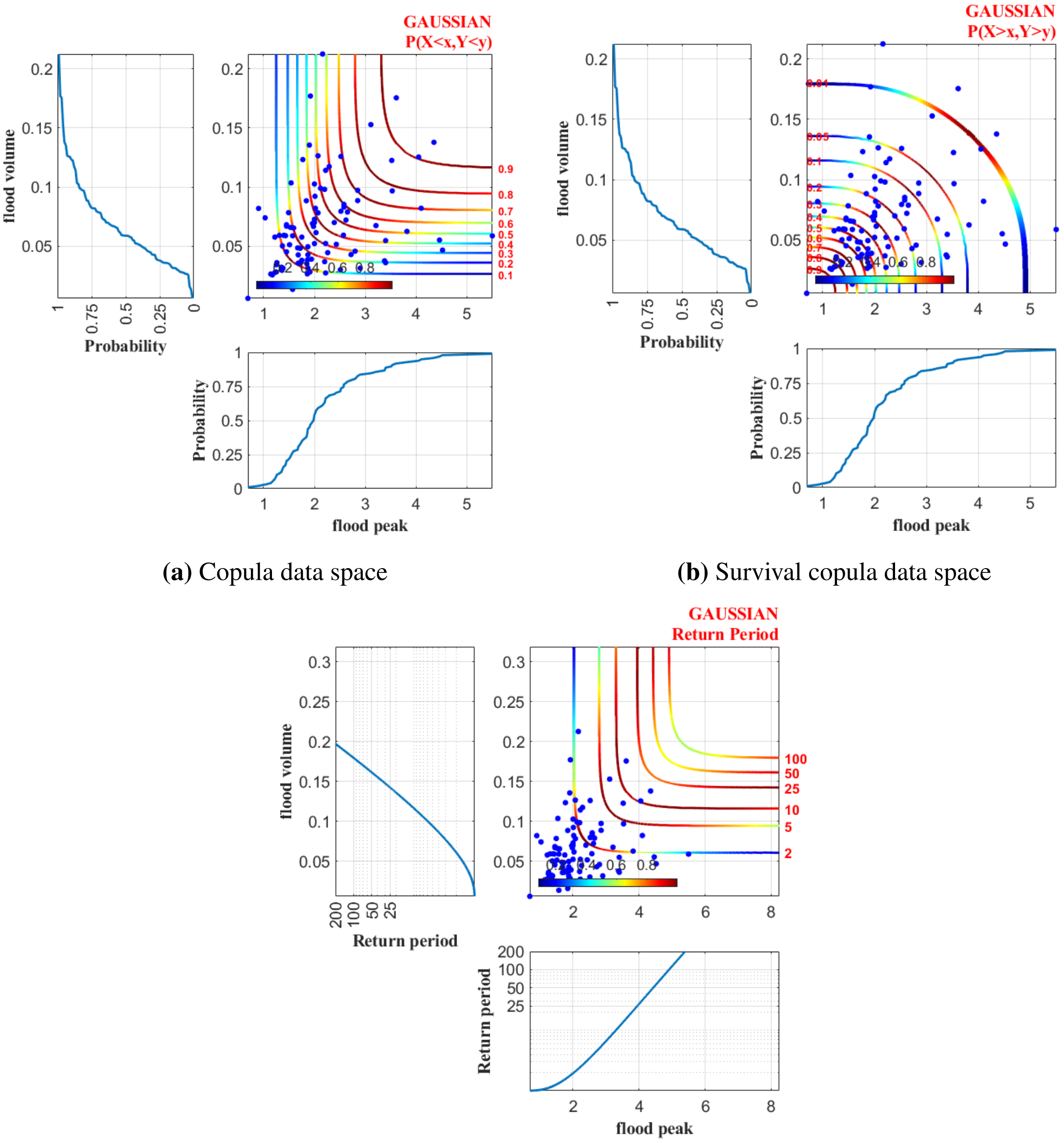
(d) (e) (f)

**Figure 3.8:** Dependence structure of flood peak (x-axis) and flood volume (y-axis) presented in probability space. The (a) Gaussian, (b) Student's $t$, (c) Independence, (d) Clayton, (e) Gumbel and (f) Frank copulas are used to illustrate the bivariate dependence. The observed data is given by blue dots, and the copula joint probability isolines are colour coded according to joint density levels. Blue indicates lower densities, and red higher densities.

of the hazard [38].

## 3.7 Summary

Copulas are the basic building blocks of this report. Although ample other copulas exist, only a few basic ones are mentioned here. These other copulas range from copulas with unique characteristics, such as the Marshall-Olkin copulas incorporating Poisson processes, to mixtures of existing copulas to overcome limitations of the individual families. Due to these individual characteristics, misspecification of models and parameters can lead to under- or overestimation of dependence structures, which could have detrimental effects on interpretation and application of these results.



**(a)** Copula data space

**(b)** Survival copula data space

**(c)** Return period

**Figure 3.9:** Joint probability isolines (a) derived from the Gaussian copula and the corresponding survival copula (b) with the marginal cumulative distributions of flood peak and flood volume. The associated return period isolines and univariate return periods are given in (c). The observed data is depicted by blue dots, and the joint probability and multivariate return period isolines are colour coded according to joint density levels. Blue indicates lower densities, and red higher densities.

The flexibility or generality of copulas compared to other dependency measures can therefore serve both as an advantage and a risky downfall. Understanding the underlying basics of copulas is therefore crucial before starting with applications. In this chapter the focus was on static copulas. Extending these to model dynamic dependence will be investigated in the next chapter.

# Chapter 4

# Gaussian process conditional copula

## 4.1 Introduction

Economic and financial time series are known to exhibit time-varying conditional standard deviations (volatility) and correlations [53]. These time series are typically characterised by a leverage effect and volatility clustering. The former refers to the tendency of negative returns to increase correlation more than positive returns of the same magnitude, thus having an asymmetric dependence [33]. Volatility clustering is simply periods of low and high volatility, which do not usually present any systematic patterns.

A similar problem is observed in geomagnetic storms, where the relationship between the shocked solar wind and the geomagnetic field can be viewed as a highly non-linear, non-stationary transfer function [24]. Fully understanding and quantifying the coupling between the solar wind and the magnetosphere for the known phases of storms is an important task for space physicists striving to provide accurate predictions of geomagnetic storms (see Section 5.2).

### 4.1.1 Why dynamic copulas?

Considering two exchange rates, a bivariate Student's $t$ distribution may be a natural starting point for modelling the joint distribution, since the Student's $t$ distribution provides a reasonable fit to the conditional univariate distribution, but it is restricted by the property that both marginals must have the same degrees of freedom parameter and imposes a symmetric dependence structure [33]. Furthermore, there is no obvious choice of bivariate density when the two variables of interest have very dissimilar marginal distributions. The tendency of co-evolving risks and varying correlation between returns complicates risk management and gives rise to the need for a more flexible model, such as the copula.

Extending the theory of copulas from unconditional to conditional allows its use in the analysis of time-varying conditional dependence. Given the conditional joint distribution, the conditional moments and other dependence measures of interest can be obtained [33].

In this chapter, we consider the relaxation of a parametric copula by introducing the Gaussian process conditional copula, combining the advantages of a conditional copula approach with the modelling flexibility of Bayesian non-parametrics.

The structure of this chapter follows a similar 'flow' to that of Chapter 3: We start with an introduction to the basic theoretical concepts of conditional copulas. Measures of dependence, goodness of fit and estimation methods are discussed in the Section 4.3 to 4.5 before going into more detail about the Gaussian process conditional copula (GPCC) in Section 4.6, which is the main focus of this research study. Again we end the chapter with an example application in of the concepts discussed. This example is a continuation of that at the end of Chapter 3.

## 4.2 Theory

### 4.2.1 Conditional copulas

The conditional copula is introduced when the dependence structure between two variables changes with the value taken by some covariate [14]. A conditional copula links joint conditional and marginal conditional distributions. If $\mathbf{X} \in \mathbb{R}^p$ is a covariate vector, then

$$H_{\mathbf{X}}(y_1, ..., y_k|\mathbf{X}) = C(F_{1|\mathbf{X}}(y_1), ..., F_{k|\mathbf{X}}(y_k)|\mathbf{X}),$$

with all $(y_1, ..., y_k) \in \mathbb{R}^k$ and marginal distribution functions $F_{i|\mathbf{X}}(y_i) = \mathbb{P}(Y_i \leq y_i|\mathbf{X} = \mathbf{x})$ for $i = 1, ..., k$ [37]. If these marginals are continuous in $y$, there exists a copula $C_{\mathbf{X}}$, such that

$$C_{\mathbf{X}}(u_1, ..., u_k) = H_{\mathbf{X}}(F^{-1}_{1|\mathbf{X}}(u_1), ..., F^{-1}_{k|\mathbf{X}}(u_k)),$$

where $F_{i|\mathbf{X}}^{-1}(u) = \inf\{y : F_{i|\mathbf{X}}(y) \geq u\}$ is the conditional quantile function of $Y_i$ given $\mathbf{X} = \mathbf{x}$. The conditional dependence structure of $(Y_1, ..., Y_k)^T$ given $\mathbf{X} = \mathbf{x}$ is fully described by the conditional copula $C_{\mathbf{X}}$ [47].

For $H_{\mathbf{X}}$ $p$-times differentiable, the joint density is obtained as

$$h_{\mathbf{x}}(y_1, ..., y_k|\mathbf{X}) = \frac{\partial^p}{\partial y_1 \partial y_2 ... \partial y_p} H_{\mathbf{X}}(y_1, ..., y_k|\mathbf{X}) = \prod_{i=1}^{p} f_{i|\mathbf{X}}(y_i) c(F_{1|\mathbf{X}}(y_1), ..., F_{p|\mathbf{X}}(x_p)|\mathbf{x}). \tag{4.1}$$

Although the conditional joint distribution $(Y_1, Y_2)|X$ can be computed from the unconditional joint distribution of $(Y_1, Y_2, X)$, assuming $X$ is one-dimensional:

$$F_{12|X}(y_1, y_2|x) = f_x(x)^{-1} \frac{\partial F_{12X}(y_1, y_2, x)}{\partial x} \qquad \text{for } x \in \mathcal{X},$$

for $f_x(x)$ and $\mathcal{X}$ the unconditional density and support of $X$ respectively, the conditional copula of $(Y_1, Y_2)|X$ cannot be computed from the unconditional copula of $(Y_1, Y_2, X)$.

The conditional copula of $(Y_1, Y_2)|X = x$, where $Y_1|X = x \sim F_{1|X}(.|x)$ and $Y_2|X = x \sim F_{2|X}(.|x)$, is defined as the conditional joint distribution function of $U = F_{1|X}(Y_1|x)$ and $V = F_{2|X}(Y_2|x)$ given $X = x$. $U$ and $V$ are the probability integral transforms of $Y_1$ and $Y_2$ given $X$. Here $X$ must be the same for both the marginals and the copula to ensure that the resulting function is a multivariate conditional joint distribution [33, 34].

### Copula-based time series models

For multivariate time series applications, consider some information set $\mathcal{F}_{t-1}$. For $t \in \{1, ..., T\}$, let $\mathbf{Y}_t|\mathcal{F}_{t-1} \sim F(.|\mathcal{F}_{t-1})$, and $Y_{it}|\mathcal{F}_{t-1} \sim F_i(.|\mathcal{F}_{t-1})$. Then, using an extension of Sklar's theorem for conditional joint distributions [33]

$$H(\mathbf{y}|\mathcal{F}_{t-1}) = C(F_1(y_1|\mathcal{F}_{t-1}), ..., F_k(y_k|\mathcal{F}_{t-1})|\mathcal{F}_{t-1}).$$

When the conditional distribution $F_{it}$ is modelled parametrically (ex. Normal, standardised Student's $t$, etc.), it may be modelled as time-varying, but when estimated non-parametrically (using the empirical distribution function), it is assumed constant, such that $F_{it} = F_i$ for all $t$ [34].

### Gaussian process conditional copula

The Gaussian process conditional copula (GPCC) is an extension of the conditional copula model, such that the approximate inference allows for copulas with multiple parameters, instead of just one. Estimation of the conditional copula is done using a general Bayesian non-parametric framework based on Gaussian processes (GPs). For this framework, the copula parameters are unknown non-linear functions of arbitrary conditioning variables. The GPCC is the main focus of this study, and is discussed in more detail in Section 4.6.

### Empirical conditional copula

An empirical estimator for $H_x(y_1, y_2)$ is

$$H_{xh}(y_1, y_2) = \sum_{i=1}^{n} \omega_{ni}(x, h_n)\mathbb{I}\{Y_{1i} \leq y_1, Y_{2i} \leq y_2\},$$

where $\{\omega_{ni}(x, h_n)\}$ is a sequence of weights that smooth over the covariance space and $h_n > 0$ is the bandwidth, which tends to 0 as the sample size increases [14, 47].

From this we obtain the empirical estimator of the conditional copula $C_x(u_1, u_2)$:

$$C_{xh}(u_1, ..., u_p) = H_{xh}(F_{1xh}^{-1}(u_1), ..., F_{pxh}^{-1}(u_p)).$$

## 4.3 Measures of dependence

Suppose we observe $(Y_1, Y_2, X)^T$, but are interested in the relationship of $(Y_1, Y_2)^T$. Ignoring the confounding factor $X$ may distort the true relationship of $(Y_1, Y_2)^T$. Partial correlation coefficient of $(Y_1, Y_2)^T$ given $X$, such as Pearson's or ranked based methods, are used to adjust for the influence of $X$. The remaining question is whether the relationship is the same for "small" (lower quantile) and "large" (upper quantile) values of $X$, which is where we incorporate the conditional copula [14].

The measures of dependence for the conditional copula are calculated similarly to the case with non-conditional copula models, however, the conditional association measures are now functions in the covariate.

Simply substituting the copula with a conditional copula, the population conditional Spearman's rho is

$$\rho_S(x) = 12 \int\int C_x(u_1, u_2) du_1 du_2 - 3.$$

Similarly, the population conditional Kendall's tau of $(Y_1, Y_2)^T$ given $X = x$ is given by

$$\begin{aligned}\tau(x) &= 4 \int\int C_x(u_1, u_2) du_1 du_2 - 1 \\ &= 2\mathbb{P}((Y_1 - \tilde{Y}_1)(Y_2 - \tilde{Y}_2) > 0 | X = \tilde{X} = x) - 1\end{aligned}$$

where $(\tilde{Y}_1, \tilde{Y}_2, \tilde{X})^T$ is an independent copy of random vector $(Y_1, Y_2, X)^T$.

## 4.4 Goodness of fit

### 4.4.1 KS & CvM

Once again, the Kolmogorov-Smirnov (KS) and Cramér-von-Mises (CvM) tests are used, comparing the fitted copula CDF to the empirical copula. These tests rely on the assumption that the true conditional copula is constant. An alternative is using the Rosenblatt transform, which can be used to test both static and time-varying copula models [34]. This simply involves transforming the data to uniform using the PIT method and then applying the appropriate goodness of fit test to the transformed data.

### 4.4.2 Quantile dependence

Comparing the quantile dependence implied by different copulas to the sample quantile dependence gives a visual indication for improvement of the model, compared to the KS and CvM tests which provide no further information if the null of correct model specification is rejected [34]. For quantiles $q \in [0.025, 0.975]$, the lower and upper quantile dependence are given by

$$\begin{aligned}\lambda_L^q &= \tfrac{P(U_1 \leq q, U_2 \leq q)}{q} \quad && 0 < q \leq 0.5 \\ \lambda_U^q &= \tfrac{P(U_1 > q, U_2 > q)}{(1-q)} \quad && 0.5 < q < 1.\end{aligned}$$

These coefficients measure the strength of dependence between two rvs in the joint upper and lower tails of the support of the distribution [13]. We observe lower (upper) tail dependence if $\lambda_L^q$

($\lambda_U^q$) is greater than zero. For either $\lambda_L^q = 0$ or $\lambda_U^q = 0$, the corresponding tail is asymptotically independent, and $\lambda_L^q = \lambda_U^q$ indicates radial symmetry.

Taking the difference between the corresponding tail quantiles gives a clear indication of asymmetry in the dependence structure. Using a bootstrap confidence interval for the difference, the null hypothesis $H_0 : \lambda_q = \lambda_{1-q}$ is not rejected if the confidence interval includes the zero line [13].

## 4.5 Estimation

### 4.5.1 Maximum likelihood estimation

In the fully parametric case, maximum likelihood is the most efficient estimation method. The drawback is the computational burden from the number of parameters to be estimated. This problem can be simplified by estimating the model in stages. This *multi-stage maximum likelihood* (MSML) estimation is less efficient than one-stage (full) MLE.

For a semiparametric approach, *canonical maximum likelihood*, or *pseudo maximum likelihood* makes use of the copula decomposition of a joint distribution, allowing the marginals and copula to be estimated with different methods (using a non-parametric and and parametric model respectively) [34].

### 4.5.2 Generalised FITC approximation

A Gaussian process is a flexible non-parametric modelling approach, with its Bayesian foundation providing good predictive power, as well as an estimate of the variance (i.e. an error bar for prediction). It is, however very time costly, because of the covariance matrix, which is the same size as the number of observations. The fully independent training conditional (FITC) approximation, which is a generalization of the sparse pseudo-input Gaussian process (SPGP) model, serves as a sparse approximation to full GPs to accelerate training and prediction times.

The FITC approximation makes use of all data, yielding a closer approximation to the posterior distribution. By fitting a stable posterior at each iteration, it provides more accurate marginal likelihood estimates and derivatives thereof, allowing for more reliable model selection.

Its ability to locate inducing inputs independently of the training data is a further advantage in finding the sparsest solutions [30].

### 4.5.3 Expectation propagation

Expectation propagation (EP) combines assumed-density filtering (ADF) and loopy belief propagation for a deterministic approximation technique in Bayesian networks. It is a fixed point algorithm, iteratively updating parameters until they stabilise. Since the order of these iterations are not important, the information from later observations can be used to refine the approximations made earlier, so that only the most important information is retained [28].

Given a joint distribution $p(\mathbf{x}, \mathcal{D})$ with observable variables, $\mathcal{D}$, and latent variables, $\mathbf{x}$, from an exponential family $\mathcal{F}$, we want to learn the posterior over $\mathbf{x}$, $p(\mathbf{x}|\mathcal{D})$, as well as the model evidence, $p(\mathcal{D})$.

Here we explain EP step by step:

#### I. Write the posterior as a product

For independent data points with prior $p(\mathbf{x})$ and likelihood $p(\mathcal{D}|\mathbf{x}) = \prod_{i=1}^{n} g_i(\mathbf{x})$, the target distribution

$$\pi(\mathbf{x}) = p(\mathbf{x}|\mathcal{D}) \propto p(\mathbf{x}) \prod_{i=1}^{n} g_i(\mathbf{x})$$

can be written as a product of factors (or *sites*), $\pi(\mathbf{x}) \propto \prod_{i=0}^{n} g_i(\mathbf{x})$, where $g_0(\mathbf{x})$ is the prior.

#### II. Approximate as a product of approximate factors

Usually, the posterior is intractable, and we use an approximate distribution $q(\mathbf{x}) \in \mathcal{F}$. For $\mathcal{F}$ the exponential family, we then have

$$q(\mathbf{x}) = q(\mathbf{x}|\theta) = exp(\theta'\phi(\mathbf{x}) - \Phi(\theta)), \quad \theta \in \Theta,$$

where we refer to $\theta$ as the *natural parameters*, $\Theta$ the *natural parameter space*, $\phi(\mathbf{x})$ the *sufficient statistics* and $\Phi(\theta)$ the *log partition function*. From this, the *moment parameters* can be calculated as $\eta = E_\theta[\phi(\mathbf{x})]$ (the expectation of the sufficient statistics with respect to $p(\mathbf{x}|\theta)$).

EP first approximates $g_i$ with some $\tilde{g}_i$ and then uses an exact posterior with $\tilde{g}_i$. Define this approximate term as the ratio of the new posterior to the old posterior multiplied by a constant:

$$\tilde{g}_i(\mathbf{x}) = \tilde{g}_i(\mathbf{x}|\theta^{(i)}) = Z_i \frac{q(\mathbf{x})}{q^{\setminus i}(\mathbf{x})}$$

where $\tilde{g}_i(\mathbf{x}) \in \mathcal{F}^U$, with $\tilde{g}_i$ the *site approximations* and $\theta^{(i)}$ the *site parameters*. $\mathcal{F}^U$ refers to the *unnormalised exponential family* associated with $\mathcal{F}$ and

$$q(\mathbf{x}) \propto \prod_{i=0}^{n} \tilde{g}_i(\mathbf{x}) = exp(\theta' \phi(\mathbf{x})).$$

**III. Hybridise the true and approximate distribution**

Form a hybrid between the true and approximate distribution by replacing one of the approximate factors with a true factor.

We start by removing the approximate factor to form a *cavity*:

$$q^{\setminus i}(\mathbf{x}) \propto \frac{q(\mathbf{x})}{\tilde{g}_i(\mathbf{x})}.$$

This step can be seen as message deletion, since each site (each bit of likelihood) contributes information to the whole approximation and the cavity removes that contribution. In terms of natural parameters, this is $\theta^{\setminus i} = \theta - \theta^{(i)}$.

Next we project $g_i(\mathbf{x})q^{\setminus i}(\mathbf{x})$, by taking the exact posterior

$$\hat{p}(\mathbf{x}) = Z_i^{-1} g_i(\mathbf{x}) q^{\setminus i}(\mathbf{x})$$

and minimise the KL-divergence $D(\hat{p}(\mathbf{x})\|q(\mathbf{x}))$ for normalising factor $Z_i = \int_{\mathbf{x}} g_i(\mathbf{x})q^{\setminus i}(\mathbf{x})d\mathbf{x}$. The new posterior $q(\mathbf{x}) \in \mathcal{F}$ contains the true factor $g_i(\mathbf{x})$, thus message inclusion, and has the same moments as $\hat{p}(\mathbf{x})$ (moment matching).

**IV. Update the approximate factor**

Update the approximation of the factor $\tilde{g}_i$. That is, find $\tilde{g}_i$ such that $\tilde{g}_i q^{\setminus i}$ has the same moments as $g_i q^{\setminus i}$. This is simply a linear operation in the natural parameters: $\theta^{(i)} = \theta - \theta^{\setminus i}$.

A summary of the above steps is given in Algorithm 1.

**Algorithm 1:** Expectation propagation for approximate Bayesian inference [28]

```
1. Initialise approximations $\tilde{g}_i$
2. Compute the posterior for $\mathbf{x}$ from the product of $\tilde{g}_i$:
       $q(\mathbf{x}) = \frac{\prod_i \tilde{g}_i(\mathbf{x})}{\int \prod_i \tilde{g}_i(\mathbf{x}) d\mathbf{x}}$
3. Until all $\tilde{g}_i$ converge:
    (a) Choose $\tilde{g}_i$ to refine
    (b) Remove $\tilde{g}_i$ from the posterior to get the 'old' posterior $q^{\backslash i}(\mathbf{x})$ by dividing and normalising:
            $q^{\backslash i}(\mathbf{x}) \propto \frac{q(\mathbf{x})}{\tilde{g}_i(\mathbf{x})}$
    (c) Combine $q^{\backslash i}(\mathbf{x})$ and $g_i(\mathbf{x})$ and minimise the KL-divergence to get the new posterior $q(\mathbf{x})$ with normaliser $Z_i$
    (d) Update $\tilde{g}_i = Z_i \frac{q(\mathbf{x})}{q^{\backslash i}(\mathbf{x})}$
4. Use the normalising constant of $q(\mathbf{x})$ as approximation to $p(\mathcal{D})$:
       $p(\mathcal{D}) \approx \int \prod_i \tilde{g}_i(\mathbf{x}) d\mathbf{x}$
```

## 4.6 Gaussian process conditional copula

The notation of this section follows that of the article by Hernández-Lobato et al. [17].

Let $\mathcal{D}_{\mathbf{Z}} = \{\mathbf{z}_i\}_{i=1}^n$ and $\mathcal{D}_{U,V} = \{(u_i, v_i)\}_{i=1}^n$ where $(u_i, v_i)$ is a sample drawn from the assumed parametric copula model $C_{X,Y|\mathbf{z}_i}$. If we let $\theta_i(\mathbf{z}) = \sigma_i[f_i(\mathbf{z})]$ for an arbitrary real function $f_i$ and we let the real line of valid configurations for $\theta_i$ be mapped to the set $\Theta_i$ by a function $\sigma_i$, then the parametric copula model can be described by $k$ parameters $\theta_1, ..., \theta_k$ as $C_{par}[u, v|\theta_1(\mathbf{z}), ..., \theta_k(\mathbf{z})]$, where the $\theta_i$'s may be functions of the conditioning variable $\mathbf{z}$. A parametric model for the conditional copula assumes $C_{X,Y|\mathbf{z}_i} = C_{\theta(\mathbf{z}_i)}$ belongs to a parametric family of copulas and only the parameter $\theta \in \Theta$ varies as a function of $\mathbf{Z}$ [22].

A Bayesian non-parametric analysis can be performed to learn the latent functions $f_1, ..., f_k$ once the parametric form of $C_{par}$ and the mapping functions $\sigma_1, ..., \sigma_k$ have been specified. This is done by setting priors on the functions and determining the posterior distribution given the observed data.

Let $\mathbf{f}_i = (f_i(\mathbf{z}_1), ..., f_i(\mathbf{z}_n))^T$, $\mathbf{m}_i = (m_i(\mathbf{z}_1), ..., m_i(\mathbf{z}_n))^T$ and

$$[\mathbf{K}_i]_{jk} = Cov(f_i(\mathbf{z}_j), f_i(\mathbf{z}_k)) = \beta_i exp\{-(\mathbf{z}_j - \mathbf{z}_k)^T diag(\lambda_i)(\mathbf{z}_j - \mathbf{z}_k)\} + \gamma_i \tag{4.2}$$

for some mean function $m_i(\mathbf{z})$ and $n \times n$ covariance matrix $\mathbf{K}_i$, generated by the squared exponential covariance function in (4.2) with inverse length-scales vector $\lambda_i$, amplitude parameter $\beta_i$ and noise parameter $\gamma_i$. Then the Gaussian process prior distribution for $\mathbf{f}_i$ given $\mathcal{D}_{\mathbf{Z}}$ is

$$p(\mathbf{f}_i|\mathcal{D}_{\mathbf{Z}}) = \mathcal{N}(\mathbf{f}_i|\mathbf{m}_i, \mathbf{K}_i).$$

The posterior distribution for $\mathbf{f}_1, ..., \mathbf{f}_k$ given $\mathcal{D}_{U,V}$ and $\mathcal{D}_{\mathbf{Z}}$, using Bayes' rule, is

$$\begin{aligned} p(\mathbf{f}_1, ..., \mathbf{f}_k|\mathcal{D}_{U,V}, \mathcal{D}_{\mathbf{Z}}) &= \frac{p(\mathcal{D}_{U,V}|\mathbf{f}_1, ..., \mathbf{f}_k) \times p(\mathbf{f}_1, ..., \mathbf{f}_k|\mathcal{D}_{\mathbf{Z}})}{p(\mathcal{D}_{U,V}|\mathcal{D}_{\mathbf{Z}})} \\ &= \frac{[\prod_{i=1}^n c_{par}(u_i, v_i|\sigma_1[f_1(\mathbf{z}_1)], ..., \sigma_k[f_k(\mathbf{z}_k)])] \times [\prod_{i=1}^n \mathcal{N}(\mathbf{f}_i|\mathbf{m}_i, \mathbf{K}_i)]}{p(\mathcal{D}_{U,V}|\mathcal{D}_{\mathbf{Z}})}, \end{aligned} \tag{4.3}$$

where $p(\mathcal{D}_{U,V}|\mathcal{D}_{\mathbf{Z}})$ is the normalisation constant or *model evidence*.

The standard GP prediction formula is used if we want to make predictions about the conditional distribution of $U$ and $V$ given a particular value $\mathbf{z}^*$ of $\mathbf{Z}$:

$$\begin{aligned} p(u^*, v^*|\mathbf{z}^*) = \int & c_{par}(u_i^*, v_i^*|\sigma_1[f_1^*], ..., \sigma_k[f_k^*]) \times p(\mathbf{f}^*|\mathbf{f}_1, ..., \mathbf{f}_k, \mathbf{z}^*, \mathcal{D}_{\mathbf{Z}}) \\ & \times p(\mathbf{f}_1, ..., \mathbf{f}_k|\mathcal{D}_{U,V}, \mathcal{D}_{\mathbf{Z}}) d\mathbf{f}_1, ..., d\mathbf{f}_k d\mathbf{f}^* \end{aligned} \tag{4.4}$$

where $\mathbf{f}^* = (f_1^*, ..., f_k^*)^T$, $f_i^* = f_i(\mathbf{z}^*)$,

$$p(\mathbf{f}^*|\mathbf{f}_1, ..., \mathbf{f}_k, \mathbf{z}^*, \mathcal{D}_{\mathbf{Z}}) = \prod_{i=1}^{k} p(f_i^*|\mathbf{f}_i, \mathbf{z}^*, \mathcal{D}_{\mathbf{z}}),$$

$$p(f_i^*|\mathbf{f}_i, \mathbf{z}^*, \mathcal{D}_{\mathbf{z}}) = \mathcal{N}(f_i^*|m_i(\mathbf{z}^*) + \mathbf{k}_i'\mathbf{K}_i^{-1}(\mathbf{f}_i - \mathbf{m}_i), k_i - \mathbf{k}_i'\mathbf{K}_i^{-1}\mathbf{k}_i),$$

$k_i = Cov[f_i(\mathbf{z}^*), f_i(\mathbf{z}^*)]$ and $\mathbf{k}_i = (Cov[f_i(\mathbf{z}^*), f_i(\mathbf{z}_1)], ..., Cov[f_i(\mathbf{z}^*), f_i(\mathbf{z}_n)])^T$.

Equations (4.3) and (4.4) are approximated using an alternating expectation propagation (EP) algorithm.

### 4.6.1 Alternating EP algorithm for approximate Bayesian inference

We write the joint distribution for $\mathbf{f}_1, ..., \mathbf{f}_k$ and $\mathcal{D}_{U,V}$ given $\mathcal{D}_{\mathbf{Z}}$ as a product of $n+k$ factors:

$$p(\mathbf{f}_1, ..., \mathbf{f}_k, \mathcal{D}_{U,V} | \mathcal{D}_{\mathbf{Z}}) = \left[ \prod_{i=1}^{n} g_i(f_{1i}, ..., f_{ki}) \right] \times \left[ \prod_{i=1}^{k} h_i(\mathbf{f}_i) \right], \tag{4.5}$$

where $f_{ji} = f_j(\mathbf{z}_i)$, $h_i(\mathbf{f}_i) = \mathcal{N}(\mathbf{f}_i | \mathbf{m}_i, \mathbf{K}_i)$ and $g_i(f_{1i}, ..., f_{ki}) = c_{par}(u_i, v_i | \sigma_1[f_{1i}], ..., \sigma_k[f_{ki}])$ Each factor $g_i$ is approximated with an approximate Gaussian factor which may not integrate to one:

$$\tilde{g}_i(f_{1i}, ..., f_{ki}) = s_i \prod_{j=1}^{k} \mathcal{N}(f_{ji} | \tilde{m}_{ji}, \tilde{v}_{ji}) = s_i \prod_{j=1}^{k} e^{-(f_{ji} - \tilde{m}_{ji})^2 / [2\tilde{v}_{ji}]}, \tag{4.6}$$

where $s_i > 0$ and the parameters $\tilde{m}_{ji}$ and $\tilde{v}_{ji}$ are calculated by EP. The factors $h_i$ do not have to be approximated, since they already have a Gaussian form. The product of $\tilde{g}_i$ and $h_i$ is then, up to a normalisation constant, a multivariate Gaussian distribution which approximates the exact posterior $p(\mathbf{f}_1, ..., \mathbf{f}_k | \mathcal{D}_{U,V}, \mathcal{D}_{\mathbf{Z}})$ in (4.3) and factorises across $\mathbf{f}_1, ..., \mathbf{f}_k$. The joint distribution in (4.5) can then be approximated as

$$q(\mathbf{f}_1, ..., \mathbf{f}_k) = \left[ \prod_{i=1}^{n} s_i \prod_{j=1}^{k} \mathcal{N}(f_{ji} | \tilde{m}_{ji}, \tilde{v}_{ji}) \right] \times \left[ \prod_{j=1}^{k} \mathcal{N}(\mathbf{f}_j | \mathbf{m}_j, \mathbf{K}_j) \right], \tag{4.7}$$

We approximate the predictions in (4.4) in two steps:

1. Integrate $p(\mathbf{f}^* | \mathbf{f}_1, ..., \mathbf{f}_k, \mathbf{z}^*, \mathcal{D}_{\mathbf{Z}})$ with respect to $q(\mathbf{f}_1, ..., \mathbf{f}_k)$. This results in a factorised Gaussian distribution $q^*(\mathbf{f}^*)$ which approximates $p(\mathbf{f}^* | \mathcal{D}_{U,V}, \mathcal{D}_{\mathbf{Z}})$.

2. Approximate using Monte-Carlo: sample from $q^*(\mathbf{f}^*)$ and then take the average of $c_{par}(u_i^*, v_i^* | \sigma_1[f_1^*], ..., \sigma_k[f_k^*])$ over the samples.

The resulting conditional copula is semi-parametric, with the dependence between $U$ and $V$ given $\mathbf{Z}$ parametric, and the effect of $\mathbf{Z}$ on the copula non-parametric [23].

EP iteratively updates each $\tilde{g}_i$, as defined in equation (4.6), until convergence. The first step in this process is to compute $q^{\setminus i} \propto q / \tilde{g}_i$ and minimise the Kullback-Leibler (KL) divergence between $g_i q^{\setminus i}$ and $\tilde{g}_i q^{\setminus i}$. That is, we use the KL divergence to minimise the information loss of using $\tilde{g}_i$

instead of $g_i$. This involves updating $\tilde{g}_i$ by using moment matching of the marginal mean and variance of $g_i q^{\setminus i}$ and $\tilde{g}_i q^{\setminus i}$.

Due to the complicated form of $g_i$, the moments cannot be computed analytically. As a solution, an additional approximation is included when computing the moments of $f_{ji}$ with respect to $g_i q^{\setminus i}$ in order to compute the $k$-dimensional integrals. Assume without loss of generality that we want to compute the expectation of $f_{1i}$ with respect to $g_i q^{\setminus i}$ and make the approximation

$$\begin{aligned}\int f_{1i} \times g_i(f_{1i},..., f_{ki}) \times q^{\setminus i}(f_{1i}, ..., f_{ki}) df_{1i}, ..., df_{ki} \approx \\ C \times \int f_{1i} \times g_i(f_{1i}, \bar{f}_{2i}, ..., \bar{f}_{ki}) \times q^{\setminus i}(f_{1i}, \bar{f}_{2i}, ..., \bar{f}_{ki}) df_{1i}\end{aligned} \tag{4.8}$$

where $\bar{f}_{1i}, ..., \bar{f}_{ki}$ are the means of $f_{1i}, ..., f_{ki}$ with respect to the posterior approximation $q$, and the constant $C$ approximates the width of the integral around its maximum in all directions except $f_{1i}$. The one-dimensional integral on the right of (4.8) is computed using numerical quadrature techniques, maximising $q$, instead of $g_i(f_{1i}, ..., f_{ki}) \times q^{\setminus i}(f_{1i}, ..., f_{ki})$ with respect to $f_{2i}, ..., f_{ki}$. Here $q$ is Gaussian and its maximiser is simply its own mean vector. Since $q$ and $g_i(f_{1i}, ..., f_{ki}) \times q^{\setminus i}(f_{1i}, ..., f_{ki})$ both approximate (4.5), they should be similar, and (4.8) is expected to be a good approximation.

Since $q$ factorises across $\mathbf{f}_1, ..., \mathbf{f}_k$, the approximation decouples into $k$ subroutines between which we alternate. Each subroutine is iterated until convergence before re-running the next one [17]. For the $j$th subroutine:

1. Use the means of the approximate distributions generated by the other sub-routines as input to approximate the posterior distribution of $\mathbf{f}_j$.

2. Find a Gaussian approximation to a set of $n$ one-dimensional factors (one factor per data point), such that the $i$-th factor of subroutine $j$ is given by $g_i(f_{1i}, ..., f_{ki})$.

3. Keep $\{f_{1i}, ..., f_{ki}\} \setminus \{f_{ji}\}$ fixed to its current approximate posterior mean estimated by the other subroutines.

Each $j$th EP subroutine optimises the kernel hyper-parameters $\lambda_j$, $\beta_j$ and $\gamma_j$, as well as the pseudo inputs, by maximising the EP of the model evidence, $p(\mathcal{D}_{U,V} | \mathcal{D}_{\mathbf{Z}})$. The generalised FITC approximation is used to speed up the GP related computations.

Algorithm 2 illustrates the process followed in the GPCC estimation. The sequential EP stabilises in 3 to 4 passes, using damped parameter updates to prevent convergence problems [28].

## 4.7 Example

This example is a continuation of the example in Chapter 3. Analysis is therefore done on the same flood data with variables flood peak and flood volume.

### 4.7.1 Estimation

We assume the copula $c(u_1, u_2|\mathbf{z})$ to be described by a parametric copula $c_{par}(u_1, u_2|\theta_1(\mathbf{z}), ..., \theta_k(\mathbf{z}))$. For a conditional Student's $t$ copula, $k = 2$ where $\theta_1$ is the correlation with constraint set $\Theta_1 = [-1, 1]$ and $\theta_2$ is the degrees of freedom with constraint set $\Theta_i = [0, \infty)$. That is,

$$\tau(\mathbf{z}) = \sigma_\tau[f_\tau(\mathbf{z})], \qquad \nu(\mathbf{z}) = \sigma_\nu[f_\nu(\mathbf{z})],$$

with corresponding mapping functions

$$\sigma_\tau(.) = 2\Phi(.) - 1, \qquad \sigma_\nu(.) = exp(.).$$

The parametric copula model is then described as

$$c_{student}(u_1, u_2|\tau(\mathbf{z}), \nu(\mathbf{z})).$$

The objective is to learn the latent real functions $f_1$ and $f_2$ from the data.

Figure 4.1 illustrates the predictions for $\nu(t)$ and $\tau(t)$ of the GPCC-T. The left frame gives the mean parameter value with confidence interval at each time point, and the right frame gives the distribution of the mean parameter values. It can be seen that although the value of $\tau(t)$ changes over time, it stays relatively constant. The values of $\nu(t)$ are more widely spread, but all the predictions are relatively large.

The parameter values at observation 89 might warrant some further investigation to find a possible reason for the change (increase in $\tau(t)$ value and decrease in $\nu(t)$ value).

The one step ahead forecast of the GPCC-T parameters are presented in Figure 4.2, with a

**Algorithm 2:** Gaussian process conditional copula

```
for $\mathbf{X_{train}}$ and $\mathbf{Z_{train}}$ do
    Fit an unconditional $t$ copula to the data:
    Optimise the copula log-likelihood function, $ll = -\sum log(c^t(u_1, u_2))$, and store the
      corresponding $\tau,\ \nu$ and $-ll$.
    Initialise the estimates of $\tau$ and $\nu$.
    Fit a conditional $t$ copula:
    for $i = 1:4$ do
        Estimate $\nu$ given the current estimate of $\tau$:
        for $\tau$ given do
            Initialise the hyper-parameters $(\lambda_i, \beta_i, \gamma_i)$.
            Initialise the gradient optimisation process:
            for EP approximation do
                Initialise the structure with the FITC problem information.
                Initialise the approximate factors and posterior approximation.
                while not converged and $i < 50$ do
                    Refine the first approximate factor using moment matching and check
                      for a positive definite posterior covariance matrix.
                    Refine the second approximate factor.
                    Update the posterior approximation.
                Compute the evidence and its gradients.
            Update the hyper-parameters $(\lambda_i, \beta_i, \gamma_i)$:
            while not converged and iteration $< 50$ do
                Update $(\lambda_i, \beta_i, \gamma_i)$ using gradient evidence.
                Update the corresponding FITC problem information.
                while posterior covariance matrix not positive definite and counter $< 100$
                  do
                    Update $(\lambda_i, \beta_i, \gamma_i)$ using gradient evidence.
                    Update the corresponding FITC problem information.
                Update $(\lambda_i, \beta_i, \gamma_i)$ and optimise gradient using EP approximation.
        Estimate $\tau$ given the current estimate of $\nu$. (Similarly to above.)
        Store the sequence of model evidence, $p(\mathcal{D}_{U,V}|\mathcal{D}_{\mathbf{Z}})$.
for $\mathbf{X_{test}}$ and $\mathbf{Z_{test}}$ do
    Generate predictions for each data point.
    for $\tau$ and $\nu$ do
        Compute the FITC prediction.
        Obtain the new marginals.
        Add the contribution of the prior mean
    Obtain the mean and lower quantiles for $\tau$ and $\nu$.
    Evaluate the test log-likelihood on each data point.
```

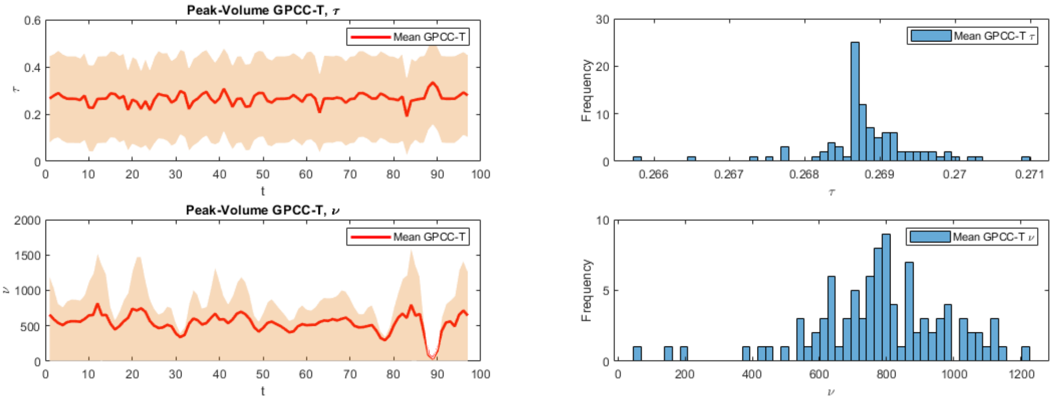


**Figure 4.1:** GPCC-T predictions for $\nu(t)$ and $\tau(t)$ on the flood data. Left is the mean predicted parameter value with confidence interval at each time point and right is the corresponding distribution of the mean predicted parameter values.

rolling-window size of 10 observations respectively. For these forecasts, the value of $\tau(t)$ varies a lot more than for the predicted values in Figure 4.1. In addition, the $\nu(t)$ forecasts are significantly larger compared to those in Figure 4.1.

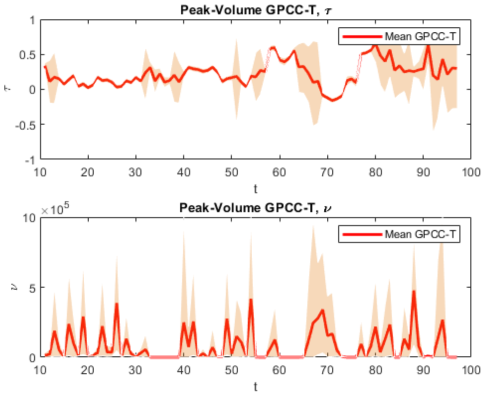


**Figure 4.2:** One-step-ahead GPCC-T forecast for $\nu(t)$ and $\tau(t)$ on the flood data using a rolling-window size of 10 observations.

### 4.7.2 Goodness of fit

Evaluating the test log-likelihood on each data point, the average test log-likelihood of the GPCC-T is $0.1462$, compared to $0.0891$ for the unconditional Gaussian copula.

For an visual interpretation, the estimated quantile dependence is plotted in Figure 4.3, along with its 95% i.i.d. bootstrap confidence interval over $q \in [0.025, 0.975]$.

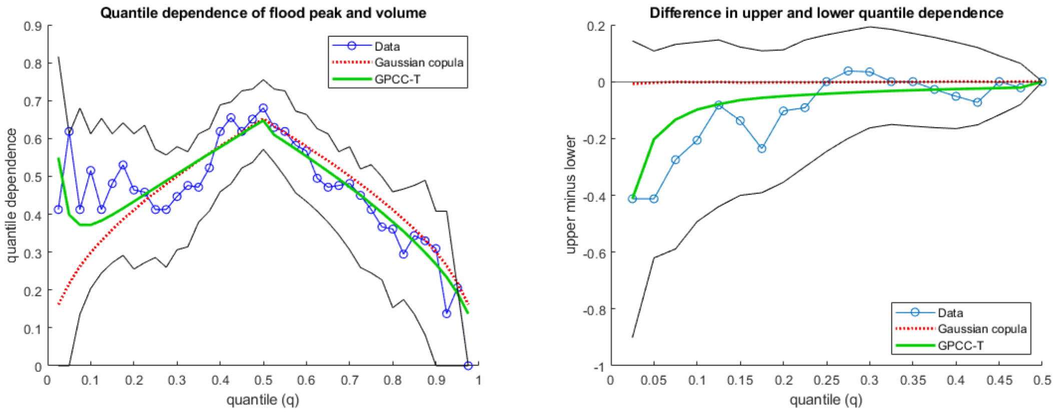


**Figure 4.3:** The left panel shows the estimated quantile dependence between observed flood peak and volume (blue) and 90% bootstrap confidence interval (black), as well as the quantile dependence implied by the (unconditional) Gaussian copula (red) and GPCC-T (green). The right panel gives the difference between corresponding upper and lower quantile tail dependence estimates of all methods and corresponding 90% bootstrap confidence interval of the observed data.

From this plot, divergence from symmetry is visible in the tails (specifically the lower tail). It is clear that while the Gaussian copula is symmetric in both tails and does not allow for tail dependence, the GPCC-T captures the asymmetry in the lower tail. The 'dip' to the right of $q = 0.5$ in the left plot also seems to suggest that the GPCC is able to adjust with the data.

Since the zero line is included in the 90% bootstrap interval on the right of Figure 4.3, the null $H_0 : \lambda_q = \lambda_{1-q}$ cannot be rejected at a 10% level of significance (the difference is significant and the dependence structure is relatively symmetric). The negative difference further show that the lower tail observations are significantly more dependent that those in the upper tail.

The confidence interval for the quantile dependencies become narrower towards $q = 0.5$, compared to the outer tails. This observation corresponds with the results obtained in the introductory example in Chapter 1 (Figure 1.6 and Table 1.1), where the predictions under independence are similar to that of the copulas for the value of flood volume being kept constant at its median, compared to when the maximum value is considered.

# Chapter 5

# Application

Two distinct applications are shown in this chapter. In Section 5.1, a comparison of the static and dynamic (GPCC) copulas discussed before is done on foreign exchange time series data. Section 5.2 is a novel application of copulas as a way to quantify the coupling efficiency between the solar wind and magnetosphere for the three known phases of geomagnetic storms.

All applications are done using MATLAB R2020a software [1], using the MvCAT Version 02.02 [39] toolbox for all static copula applications. We developed algorithms in MATLAB from first principles for the implementation of the GPCC. The code is based on an R software package[1] written by Hernandez-Lobato [17]. All results and graphs in this and previous chapters were produced by this toolbox which can be found on GitHub[2].

## 5.1 Foreign Exchange Time Series

The daily exchange rate of four different currencies, paired with the U.S. dollar, are evaluated from 2006/10/10 to 2010/08/09, yielding a total of 1000 observations [3]. The four currencies are Swiss Franc (CHF), Australian Dollar (AUD), Canadian Dollar (CAD) and South African Rand (ZAR). Since the Swiss franc is a 'safe haven' currency during times of uncertainty, CHF-USD is paired with each of the exchange rate pairs (AUD-USD, CAD-USD and ZAR-USD). The objective of

[1] R code available at:
https://github.com/lopezpaz/gaussian_process_conditional_copulas/tree/master/code

[2] MATLAB GPCC Toolbox:
https://github.com/ColetteLR/GPCC.git

[3] The unprocessed dataset can be found at:
https://www.kaggle.com/thebasss/currency-exchange-rates/data

this application is to model the dependency between the daily logarithmic returns of a currency pair (AUD-USD, CAD-USD and ZAR-USD) and CHF-USD over time as it changes in response to financial conditions [17].

For this case study, a GARCH 'filter' is applied to obtain i.i.d. observations. An AR(1)-GARCH(1,1) model is used to capture the time-dependent scale of the time series observations, such that each marginal model has 7 parameters. This pre-processing of the data improves the flexibility of the fitted copula model. Applying the probability integral transform, the empirical CDF of the standardised residuals from the fitted AR(1)-GARCH(1,1) model is then used to transform the logistic returns into the pseudo-sample of the underlying copula [48]. The performance of the copulas can then be evaluated on the transformed data.

### 5.1.1 Distribution analysis

The time plots of each of the four U.S. dollar pairs are given in the right pane of Figure 5.1, from which we observe a change near the end of 2008 corresponding with the subprime crisis. Although less obvious, the effects of the European sovereign debt crisis can also be seen around June of 2010. The daily logarithmic returns, obtained using

$$r_t = 100(\ln(y_t) - \ln(y_{t-1})),$$

approximates the percentage change in exchange rate at each time point $t$, and are plotted in the left pane of Figure 5.1. From the time plots of the returns, volatility clustering is observed. That is, large (small) fluctuations at time $t$ are follower by large (small) fluctuations at time $t+1$.

Prior to determining the appropriate models (univariate marginal CDFs), we first analyse the distributions using Pearson's moment coefficient of kurtosis and skewness. These values are given in Table 5.1, from which we conclude that the distributions are leptokurtic (since kurtosis is $> 3$), that is, compared to the normal distribution, the tails are heavier (longer and fatter) and the central peak is higher and sharper. This is also visible in the right pane of Figure 5.2. Furthermore, all the currency pairs are negatively skew (since the coefficient of skewness is $< 0$), except for ZAR-USD, which is positively skew.

In the QQ-plot of the returns against the theoretical normal distribution in the left pane of Figure 5.2, the observations do not fall on a straight line, especially in the tails. All of the above

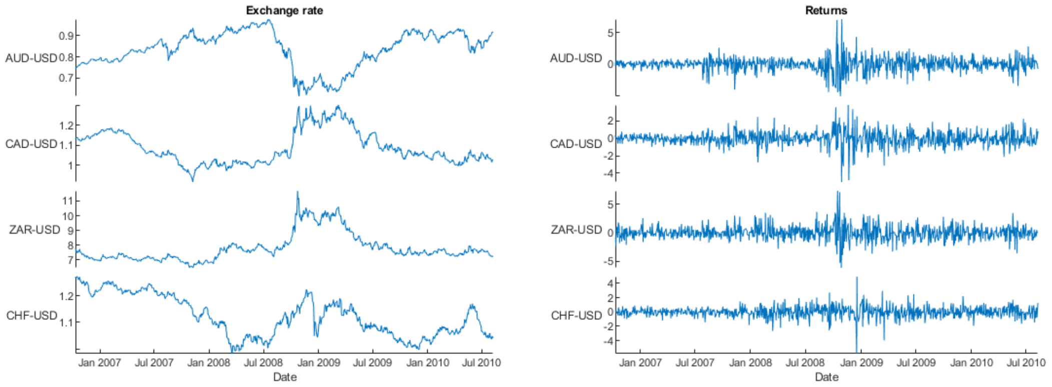


**Figure 5.1:** Daily exchange rate (left) and corresponding returns (right) time plots

| currency pair | skewness | kurtosis |
|---|---|---|
| AUD-USD | -0.0814 | 8.7579 |
| CAD-USD | -0.2463 | 7.8357 |
| ZAR-USD | 0.4401 | 7.0203 |
| CHF-USD | -0.3895 | 10.5756 |

**Table 5.1:** Pearson's moment coefficient of skewness and kurtosis.

indicate that the distributions are not normal and that a skewed distribution with heavier tails is more suitable for these time series.

### 5.1.2 Modelling returns

In order to model the returns, we need a model for the mean level and a GARCH model to capture volatilities.

The univariate marginal CDFs are estimated by assuming an AR(1)-GARCH(1,1) model with skewed Student's $t$-distributed residuals for each series:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \varepsilon_t \tag{5.1}$$

$$\varepsilon_t = \sigma_t z_t, \quad z_t \sim SkT(\nu, \gamma) \tag{5.2}$$

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{5.3}$$

where (5.1) is the mean level and (5.3) is the variance equation with parameter constraints $\omega > 0$,

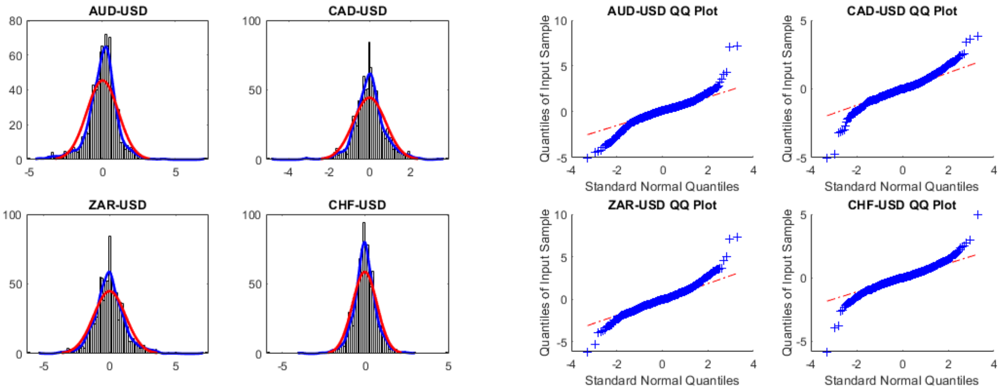


**Figure 5.2:** The left pane shows the histograms of the returns with the kernel density (blue) and normal distribution density curve (red). Right are the QQ-plots of the returns compared to the theoretical normal distribution.

$\alpha \geq 0$ and $\beta \geq 0$. The unconditional variance is

$$\sigma^2 = \frac{\omega}{(1-\alpha-\beta)},$$

such that the error term $\varepsilon_t$ is covariance-stationary when $\alpha + \beta < 1$.

Since the MATLAB Econometric Toolbox does not provide for skewed Student's $t$-distributed residuals, this model is fitted using the Dynamic Copula Toolbox 3.0 [4], and the estimated parameters are summarised in Table 5.2.

The AR(1) process in equation (5.1) yields the autoregressive parameter estimates $\hat{\phi}_0$ at lag 0 and $\hat{\phi}_1$ at lag 1 when fitted to the mean level of the returns.

We see that all variance parameter constraints are met and that the error term is covariance-stationary in all four instances. The conditional variance is therefore mean reverting and stationary. Since the parameter estimate of $\hat{\beta}$ is close to 1 for all return series, the GARCH effect is eminent in modelling the volatility.

The RVs $z_t$ from equation (5.2) follow a skewed Student's $t$ distribution with $\hat{\nu}$ degrees of freedom and skewness parameter $\hat{\tau}$.

[4] Manthos Vogiatzoglou (2020). Dynamic Copula Toolbox 3.0
(https://www.mathworks.com/matlabcentral/fileexchange/29303-dynamic-copula-toolbox-3-0),
MATLAB Central File Exchange. Retrieved July 27, 2020.

| | AUD-USD | CAD-USD | ZAR-USD | CHF-USD |
|---|---|---|---|---|
| $\hat{\phi}_0$ | 0.0524 (0.0229) | -0.0174 (0.0233) | -0.0390 (0.0290) | -0.0286 (0.0193) |
| $\hat{\phi}_1$ | -0.0172 (0.0353) | 0.0301 (0.0342) | -0.0151 (0.0765) | -0.0272 (0.0292) |
| $\hat{\omega}$ | 0.0083 (0.0048) | 0.0036 (0.0023) | 0.0448 (0.0207) | 0.0033 (0.0024) |
| $\hat{\alpha}$ | 0.1201 (0.0246) | 0.0676 (0.0125) | 0.1147 (0.0260) | 0.0397 (0.0130) |
| $\hat{\beta}$ | 0.8781 (0.0255) | 0.9314 (0.0121) | 0.8525 (0.0358) | 0.9556 (0.0150) |
| $\hat{\nu}$ | 9.9655 (2.8617) | 6.0192 (1.1397) | 9.1753 (2.2941) | 5.4471 (0.9840) |
| $\hat{\gamma}$ | -0.1976 (0.0404) | 0.0170 (0.0486) | 0.1018 (0.0454) | -0.0454 (0.0373) |
| $\mathcal{L}$ | -1243.910 | -1036.469 | -1473.761 | -983.474 |
| AIC | 2501.820 | 2086.939 | 2961.523 | 1980.949 |
| BIC | 2536.167 | 2121.286 | 2995.870 | 2015.296 |

**Table 5.2:** Parameter estimates of the AR(1)-GARCH(1,1) model for the daily returns of the respective exchange rates using a skewed Student's $t$-distribution. Standard errors are given in parenthesis. The log-likelihood, AIC and BIC of the fitted models are also included as goodness of fit measures.

### 5.1.3 Estimation

The static and dynamic copula models can now be fitted to the pseudo-sample obtained from the empirical CDF of the standardised residuals from the fitted AR(1)-GARCH(1,1) model.

#### Static copula

A summary of the parameter estimates of the chosen copulas using the MvCAT toolbox [39] is given in Table 5.3. This toolbox provides maximum likelihood, instead of the log-likelihood for goodness of fit, and the optimal copula is selected based on the BIC criteria. AUD-USD paired with CHF-USD is indicated as AUD-CHF, and similarly for the other two exchange rate pairs.

In Table 5.3, $\tau^a$ indicates Kendall's rank calculated from the marginal distributions and $\tau^b$ indicates Kendall's rank derived from the copula correlation parameter ($\theta$) using Table A.1. Comparing these two correlation values, it is clear how ignoring the underlying dependence between each currency pair leads to an underestimation of the dependency ($\tau^a < \tau^b$ in all three cases). Since none of the confidence intervals for $\tau^b$ includes zero, we can conclude that the correlation between each of the CHF currency pairs is significant, with AUD-CHF having a negative correlation and CAD-CHF and ZAR-CHF a positive correlation.

The choice of a Frank copula indicates that the first two currency pairs have more symmetric tail dependencies, compared to the Gumbel copula suggesting an asymmetric structure with greater upper tail dependence for ZAR-CHF.

| | AUD-CHF | CAD-CHF | ZAR-CHF |
|---|---|---|---|
| $\tau^a$ | -0.2084* | 0.1314* | 0.1788* |
| Copula | Frank | Frank | Gumbel |
| $\theta$ | -1.9007 | 1.2101 | 1.2254 |
| $\theta$ c.i. | (-1.9157, -1.8868) | (1.1973, 1.2232) | (1.2229, 1.2277) |
| $\tau^b$ | -0.2041 | 0.1325 | 0.1839 |
| $\tau^b$ c.i. | (-0.2056, -0.2027) | (0.1311, 0.1338) | (0.1823, 0.1855) |
| $\mathcal{ML}$ | 5629.751 | 5705.172 | 5605.421 |
| AIC | -11257.502 | -11408.344 | -11208.843 |
| BIC | -11252.595 | -11403.437 | -11203.936 |

**Table 5.3:** Evaluate dependence between the daily returns of the respective exchange rates. $\tau^a$ indicates Kendall's rank calculated from the marginal distributions and $\tau^b$ is derived from the copula correlation parameter ($\theta$). An asterisk (*) indicates significance of the dependence based on the marginal distributions at a 5% level of significance. Copula selection is done based on the $BIC$ measure for the goodness of fit.

### Gaussian process conditional copula

The GPCC-T predictions for $\nu(t)$ and $\tau(t)$ for each of the time series pairs are plotted in Figure 5.3. From all three currency pairs, it can be seen that the copula parameters do in fact chance over time, and using a static copula copula may not be accurate enough in modelling the dependency structure of these time series.

The sign of the correlation parameter, $\tau(t)$, corresponds to that in Table 5.3, fluctuating around or near the value predicted by the static copula.

The dynamic GPCC-T model provides further information about the movement of the dependency structure. While the parameters of the ZAR-CHF exchange rate pair fluctuates more rapidly (corresponding to the more volatile economic environment), it is clearly visible that the AUD-CHF and CAD-CHF pairs reach a turning point around the beginning of 2008 (at the start of the global recession) and again near the start of 2010 (the European sovereign debt crisis). Comparing this to the dates at which the changes can be observed in the time series in Figures 5.1, it seems that the parameters pick up the effects of the financial events before they are observed in the time series.

Looking at the degrees of freedom parameter, $\nu(t)$, for each series, it appears that the ZAR-CHF return series is more prone to outliers compared to the other two series, with the GPCC-T model least robust to negatively correlated outliers in the CAD-CHF return series [17], with $\nu(t)$

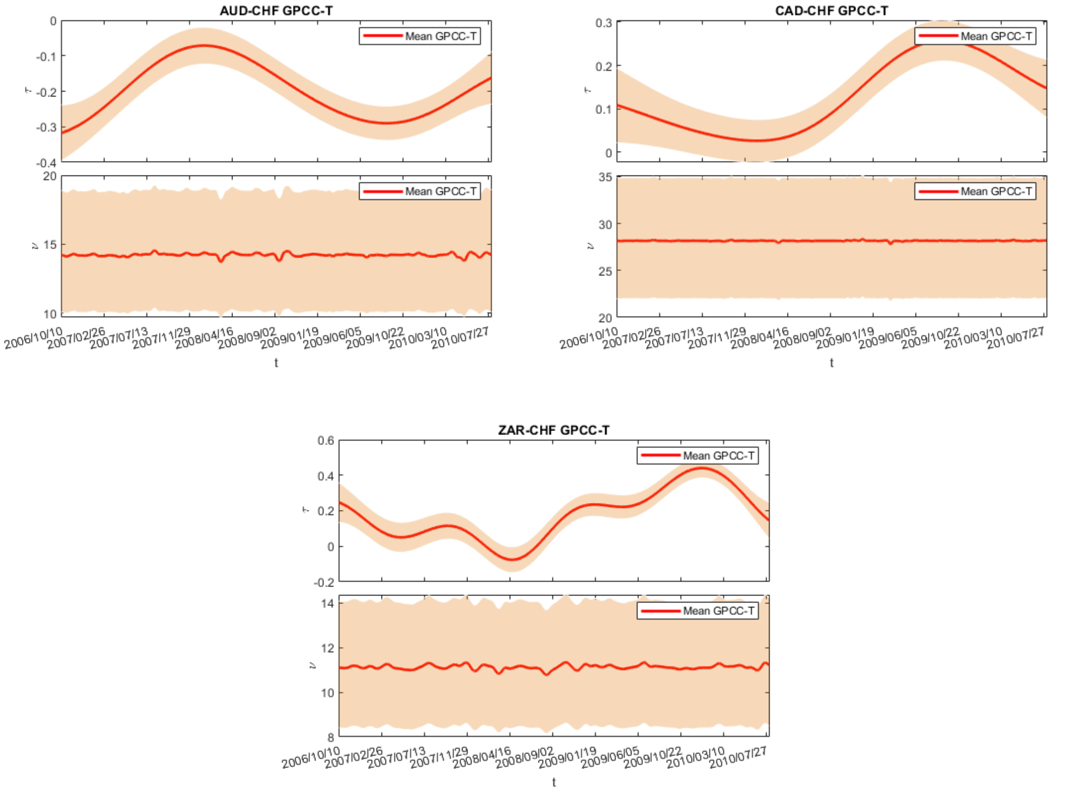


**Figure 5.3:** GPCC-T predictions for $\nu(t)$ and $\tau(t)$ for each of the time series pairs AUD-CHF, CAD-CHF and ZAR-CHF when trained on data from 2006/10/10 to 2010/08/09.

relatively constant at approximately 28 degrees of freedom, compared to approximately 14 and 11 for AUD-CHF and ZAR-CHF respectively.

### 5.1.4 Goodness of fit

The results for the predictive likelihood of each method on the transformed data are shown in Table 5.4. The Gaussian copula is included for interest sake, and the suggested static copula corresponds to the optimal copula suggested by the MvCAT toolbox as summarised in Table 5.3.

Although the Gumbel copula is suggested for the ZAR-CHF pair based on the BIC measure, it seems to be outperformed by the Gaussian copula when considering the average log-likelihood. GPCC-T is seen to be the overall best technique, outperforming the static copulas in all three series.

The quantile dependencies of the three series are plotted in Figure 5.4.

| Method | AUD-CHF | CAD-CHF | ZAR-CHF |
|---|---|---|---|
| Static Gaussian copula | 0.0467 | 0.0189 | 0.0376 |
| Suggested static copula | 0.0497 | 0.0194 | 0.0369 |
| GPCC-T | **0.0616** | **0.0319** | **0.0763** |

**Table 5.4:** Average log-likelihood of the methods (constant and dynamic) on the currency data, with the suggested static copula being that given in Table 5.3.

From the plot of the difference in upper and lower quantile dependence between ZAR and CHF, the volatility of the South African series is once again clear in that, at some quantiles the structure is symmetric (equal to zero), while at other quantiles the difference is significant (zero not included in the 90% bootstrap CI), and the structure therefore asymmetric. We also observe that higher tail quantiles are more dependent that their corresponding lower tail quantiles (above the zero line). This corresponds with the Gumbel copula being suggested by the MvCAT toolbox. Although the Gaussian copula does not capture the dependence structure of this currency pair, it falls withing the 90% CI while the Gumbel copula diverges outside the CI for tail dependencies, which may to some extent explain the log-likelihood results.

### 5.1.5 Conclusion

While the GPCC-T cperforms best for the ZAR-CHF pair in terms of log-likelihood, it does not seem to significantly outperform the static copulas in terms of quantile dependence. This may suggest that the GPCC captures volatility of the dependence, but focuses less on shape/ skewness of the data.

The GPCC-T model outperforms static copulas in all instances of the data considered, showing that the dependence structure does in fact change with time, and this underlying temporal covariate needs to be included in estimation and prediction.

Capturing effects of global events before they are observed in the individual time series, this dynamic model shows potential for predicting financial events and therefore to some extent preventing the large scale effects to follow.

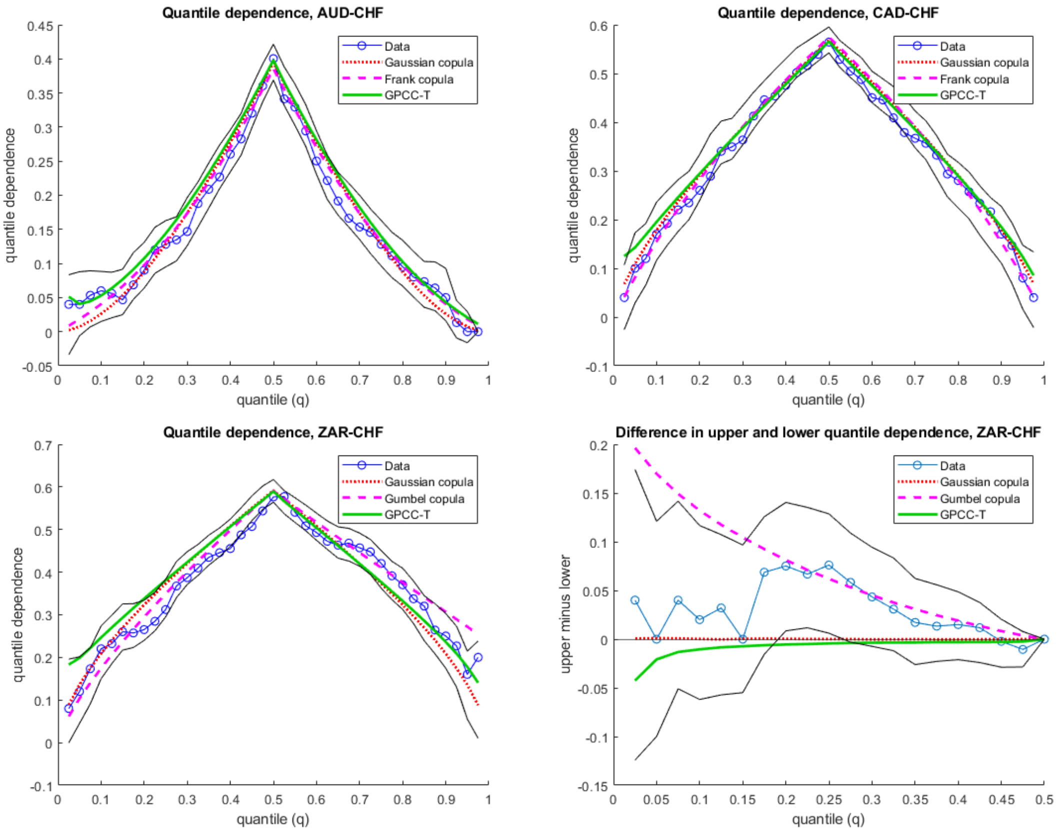


**Figure 5.4:** Observed (blue) currency data quantile dependencies with 90% bootstrap confidence interval (black), as well as the quantile dependencies predicted by each of the models (constant and dynamic copulas).

## 5.2 Geomagnetic storms

Solar activity, through geomagnetic storms, has the ability to cause a number of negative effects on critical technologies such as power grids and various communication systems [12, 4, 6]. Geomagnetic storms are intervals of disturbed geomagnetic field lasting from 10s of hours to multiple days [16]. The most intense storms are caused by energetic plasma from coronal mass ejections impacting the geomagnetic field after propagating the $1.5 \times 10^8$km (= 1AU) via the solar wind to Earth. The relationship between the shocked solar wind and the geomagnetic field can be viewed as a highly non-linear, non-stationary transfer function.

Fully understanding the coupling between the solar wind and the magnetosphere is an important task for space physicists striving to provide accurate predictions of geomagnetic storms. With this in mind we investigate the use of copulas as a way to quantify the coupling efficiency between the solar wind and magnetosphere for the three known phases of storms: onset, main and recovery. Seven intense storms are identified and the dynamic and static copulas between two solar wind parameters ($B_Z$ and $V_{sw}$) and a geomagnetic disturbance index (SYM-H) are calculated.

### 5.2.1 Data sets

Two distinct data sources are utilised in this study.

Measurements of solar wind plasma and the IMF parameters are taken aboard spacecraft orbiting the first Lagrangian point between the sun and the Earth. The time-stamps of these measurements are then shifted in time to account for the propagation time of the plasma from the space craft position to the leading edge of the magnetosphere (the bow shock nose). This time-shifted set then represents the solar wind state right at the time of first interaction with the magnetosphere, enabling direct comparison between space-based and terrestrial data sets, without the need to explicitly account for solar wind propagation speed. This data is averaged to 1-minute averages and published as the High Resolution OMNI data set [8]. In this work we only utilise $V_{sw}$ and $B_Z$; however, the OMNI set provides many more parameters such as density, temperature, full IMF vector ($B_X, B_Y, B_Z$) and various derived parameters such as pressure, plasma $\beta$, etc.

Figure 5.5 depicts a geomagnetic storm from 6 April 2000 to 11 April 2000. The shape of the SYM-H curve in the top panel is typical of a storm driven by a single coronal mass ejection (CME). The second panel shows the solar wind speed $V_{sw}$ and $B_Z$ is plotted in the bottom panel.

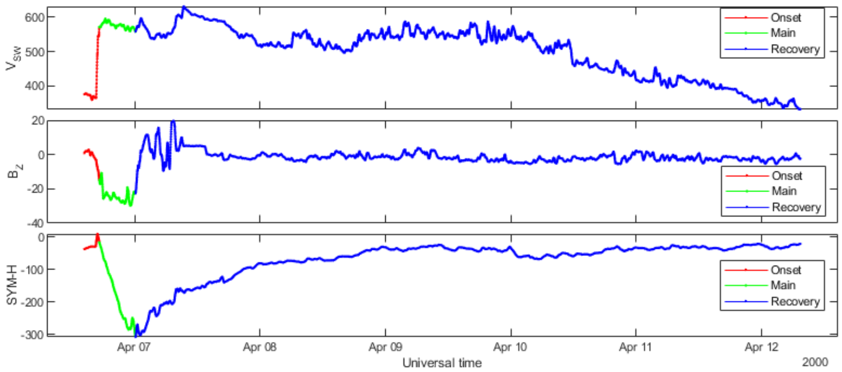


**Figure 5.5:** A storm in April 2000, with clear onset (red), main (green), and recovery (blue) phases.

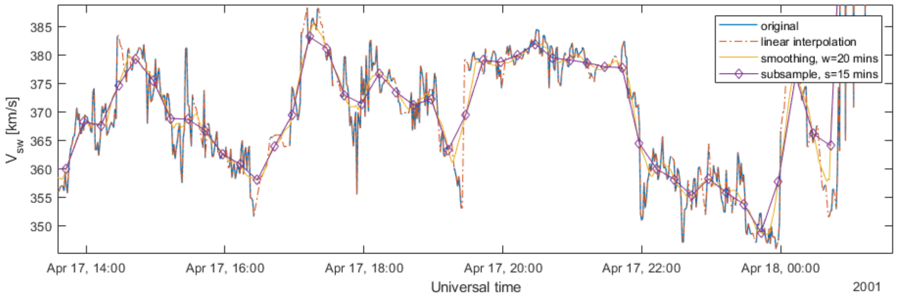


**Figure 5.6:** This figure depicts the data processing applied to the solar wind and SYM-H data sets.

The SYM-H index is supplied by the World Data Centre for Geomagnetism in Kyoto, Japan [26].

In this work 8 intense (SYM-H $\leq -100$nT) storms from 2000 (1) and 2001 (7) are identified according to the process described by [24].

### Data processing

The data selection applied, there is a total of 28 157 minutes per parameter. SYM-H is largely without error, but the solar wind measurements contain many missing values due to instrument error or saturation (28.5% for $V_{sw}$ and 11% for $B_Z$). To combat this we interpolate values linearly, but only for gaps of up to $m = 10$ minutes. Data is also smoothed with a running mean window of $w = 20$ minutes. To reduce data volume we subsample to $s = 15$ minutes. The data processing is illustrated in Figure 5.6.

Before applying the probability integral transform and fitting the copula, the data is normalised ($x - mean/std.dev.$). This offers improvement in the dependency estimation and modelling of tails in the distributions, and removes the need for more complicated copula designs.

## 5.2.2 Analysis

We fitted a copula model SYM-H and $B_Z$ to all the storm phases of all storms in 2001. The aim is to investigate if copulas can quantify the individual dependency structures for each storm phase. For this purpose, we split the storm phases into three separate datasets. Figure 5.7 indicates the distributions of SYM-H, $V_{sw}$ and $B_Z$ for each storm phase.

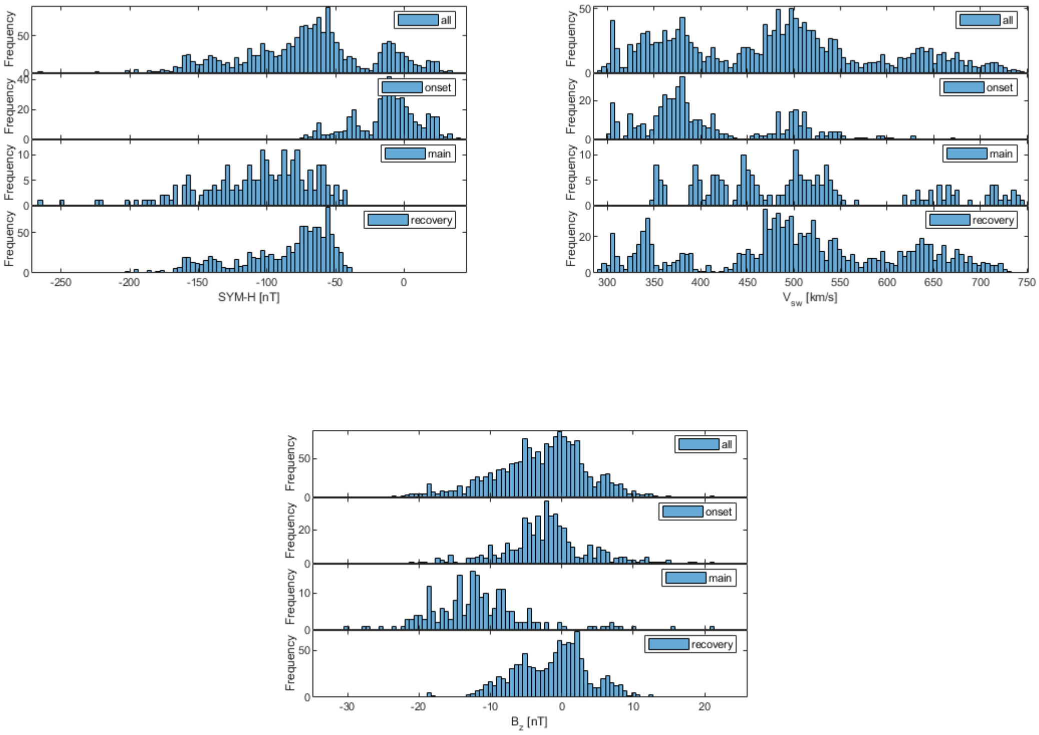


**Figure 5.7:** Distribution of SYM-H, $V_{sw}$ and $B_Z$ for each storm phase.

### 5.2.3 Static copula

In the first part of the analysis, we fit individual copulas to the 7 storms from 2001 and show that the copula functions relating pairs of parameters change significantly when analysed separately by storm phase. We model all nine configurations of variable pairs SYM-H, $B_Z$ and $V_{sw}$ and the three storm phases.

Although many other parametric and non-parametric copulas exist, only the five copula models mentioned in Section 3.2 (Gaussian, Student's $t$, Gumbel, Frank and Clayton) are compared, using the MATLAB package *MvCAT* [39] to perform model selection based on the goodness of fit measures also described in Section 3.2.

Table 5.5 gives the final copula model fitted, along with the Kendall's rank correlation coefficient based in the marginal distributions and that based in the copula. For all variable pairs and corresponding storm phases, apart from the onset and main phase between SYM-H and $V_{sw}$, using the copula yields a larger Kendall's rank value. Bringing the underlying dependence structure into account therefore prevents underestimation of the dependence between variables, particularly in extreme events.

Figure 5.8 is a visual representation of Table 5.5. The lower tail dependence of the Clayton copula is clearly visible in the main and recovery phase between SYM-H and $B_Z$ compared to the upper tail dependence of the Gumbel copula during the onset of the storm.

The characteristic shapes of the different types of copulas highlights the importance of selecting the correct model for each storm phase.

Generally, each of the three pairs exhibit stronger dependence during the main and recovery storm phases than during the onset phase. The increased dependence of SYM-H on $B_Z$ during the main storm phase is expected since it is the prolonged periods of $B_Z < 0$ which enables the strong coupling between the solar wind and magnetospheric plasmas. Solar wind speed $V_{sw}$ has slightly stronger coupling to SYM-H during the main phase, compared to recovery phase, which is also expected. The weak dependence between all three pairs during onset phase may be due to inadequate onset phase identification, as a significant period of quiet time (i.e. no strong coupling) was padded before the onset of most events.

| SYM-H - $B_Z$ | | | |
|---|---|---|---|
| | Onset | Main | Recovery |
| Copula | Gumbel | Clayton | Clayton |
| Kendall's rank$^a$ | 0.287* | 0.286* | 0.257* |
| Kendall's rank$^b$ | 0.288 | 0.348 | 0.260 |
| SYM-H - $V_{sw}$ | | | |
| | Onset | Main | Recovery |
| Copula | Gaussian | Frank | Student's $t$ |
| Kendall's rank$^a$ | -0.115* | -0.105* | -0.314* |
| Kendall's rank$^b$ | -0.112 | -0.051 | -0.481 |
| $B_Z$ - $V_{sw}$ | | | |
| | Onset | Main | Recovery |
| Copula | Gumbel | Gumbel | Frank |
| Kendall's rank$^a$ | 0.026 | 0.049 | 0.058* |
| Kendall's rank$^b$ | 0.053 | 0.093 | 0.089 |

**Table 5.5:** Evaluate dependence between the two input variables. Kendall's rank$^a$ is calculated from the marginal distributions and Kendall's rank$^b$ is derived from the copula correlation parameter. An asterisk (*) indicates significance at a 5% level of significance. Copula selection is done based on the $BIC$ measure for the goodness of fit.

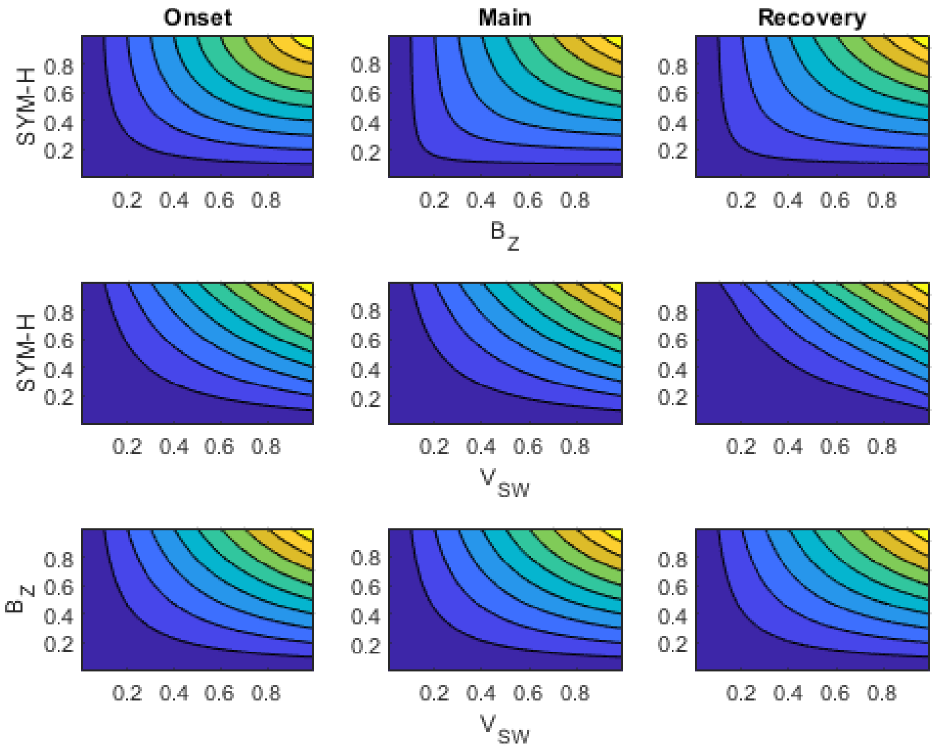


**Figure 5.8:** Contours of the copula joint distribution of $B_Z$ and SYM-H for each storm phase in the probability space.

### 5.2.4 Conditional copula

As mentioned in Chapter 4, we assume the copula $c(u_1, u_2|\mathbf{z})$ to be described by a parametric copula $c_{par}(u_1, u_2|\theta_1(\mathbf{z}), ..., \theta_k(\mathbf{z}))$. For our conditional Student's $t$ copula, $k = 2$ where $\theta_1$ is the correlation with constraint set $\Theta_1 = [-1, 1]$ and $\theta_2$ is the degrees of freedom with constraint set $\Theta_i = [0, \infty)$. That is, $\tau(\mathbf{z}) = \sigma_\tau[f_\tau(\mathbf{z})]$ and $\nu(\mathbf{z}) = \sigma_\nu[f_\nu(\mathbf{z})]$, with corresponding mapping functions $\sigma_\tau(.) = 2\Phi(.) - 1$ and $\sigma_\nu(.) = exp(.)$. The parametric copula model is then described as $c_{student}(u_1, u_2|\tau(\mathbf{z}), \nu(\mathbf{z}))$ and the latent real functions $f_1$ and $f_2$ are learnt from the data.

In Figure 5.9 we show how the copula parameter $\tau$ changes throughout a geomagnetic storm using time as the conditioning variable. We fit a GPCC to each phase of the storm separately for the variable pairs SYM-H - $B_Z$, SYM-H - $V_{sw}$ and $B_Z$ - $V_{sw}$.

We observe that SYM-H and $B_Z$ change from having a mostly negative correlation at the onset of the storm to a mostly positive correlation during the main phase. During recovery this correlation is more erratic and contains more uncertainty (wider confidence interval indicated by the shaded area).

The change in the parameter over the passage of time indicates that there is a temporal influence in the dependence structure between the variable pairs, and taking this into account may improve modelling results.

On the other hand, a constant correlation between variables during a storm phase would indicate that time has a negligible influence on the dependence structure, and that a static copula can be used instead.

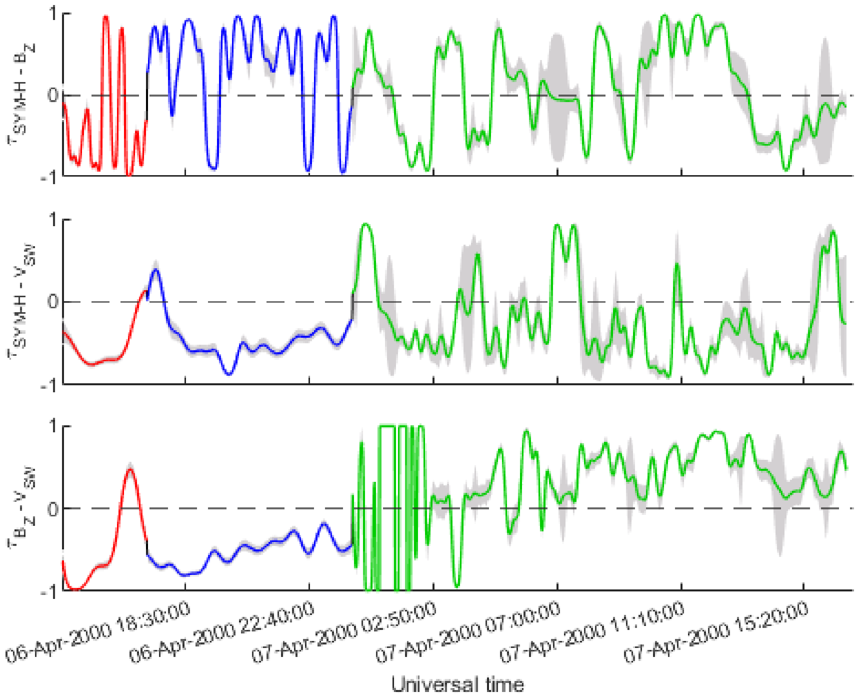


**Figure 5.9:** The copula parameter $\tau$ of the GPCC for SYM-H - $B_Z$, SYM-H - $V_{sw}$ and $B_Z$ - $V_{sw}$ for each storm phase over time.

### 5.2.5 Error correction using copula functions

Spacecraft measuring solar wind and interplanetary magnetic field data often encounter saturation of sensors or other problems, resulting in missing values. It is especially problematic when these problems occur during CME passage or other interesting phenomena.

Here we show that the static copulas above can be used to estimate missing values. Figure 5.10

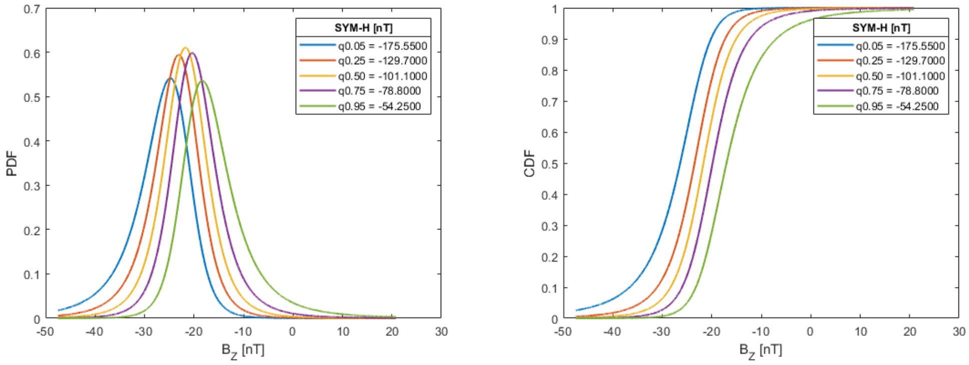


**Figure 5.10:** Density (left) and CDF (right) of $B_Z$ corresponding to respective values of SYM-H for the main phase based on the fitted copula.

gives the PDF and CDF of $B_Z$ for SYM-H fixed at different quantiles. Using the joint distribution obtained from the copula, the PDF over the range of possible values of one (unknown) variable can be obtained for a fixed value of the other (observed) variable. From this, the expected value, as well as a confidence interval, can be computed for these missing values.

In Figure 5.11 the 5th percentile level of SYM-H (=175.55 nT) is used to estimate the corresponding value of $B_Z$ during the event from April 2000. The value of $B_Z$ is estimated by utilising the $q_{0.05}$ distribution depicted as a blue curve in Figure 5.10 and its 90% confidence interval levels. Note that this event (April 2000) was not included in the analysis yielding the distribution functions, i.e. the estimate of $B_Z$ is for an out-of-sample event. The resulting estimate of $B_Z$ (-27.26 nT) does not match the observed value of $B_Z$ (-25 nT) but does fall within the confidence interval, indicated by the red vertical line.

### 5.2.6 Discussion and Conclusions

Geomagnetic storms are capable of causing major damage to various technological systems, and therefore the relevant solar wind parameters and geomagnetic field indicators are closely monitored across the globe to enable forecasts and aid mitigation and planning. Recognising the distinct phases of geomagnetic storms are important to modelling efforts as the coupling between solar wind drivers and the response of the geomagnetic field change significantly from one phase to another.

This work has shown a novel application of statistical copulas to the coupled solar wind – magnetosphere system by analysing copulas between two solar wind parameters ($B_Z$ and $V_{sw}$)

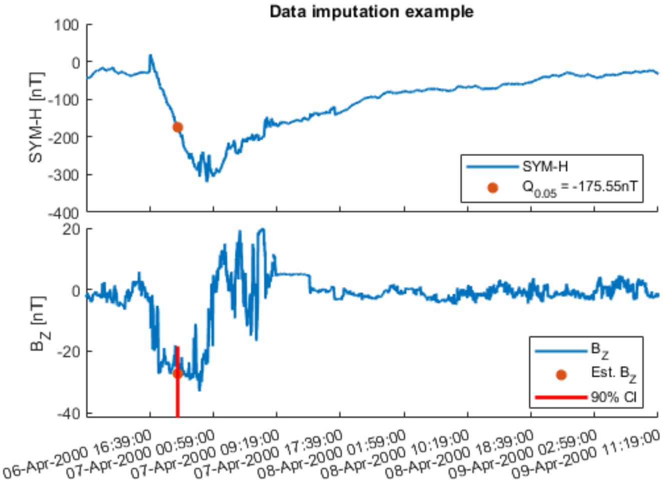


**Figure 5.11:** Utilising the SYM-H $95th$ percentile distribution of $B_Z$ (green curve in Figure 5.10), an estimate for $B_Z$ is calculated and the $90\%$ confidence interval is given.

and a storm time index (SYM-H). Since the solar wind – magnetosphere coupling is highly non-linear, we can view this transfer of information through a Bayesian framework. The parameters of a copula are non-linearly related to a conditioning variable, in our case, time. Another advantage of the Bayesian framework is that it quantifies the uncertainty of the parameter estimations and predictions.

We demonstrated that copulas can be used to confirm what is know from the underlying physical mechanisms: i.e. that the coupling between different solar wind parameters and geomagnetic field differs for different storm phases. For a selection of 7 storm events with fairly simple cause and structure, it was shown that static copula functions behave very differently when conditioned on storm phase. From Figure 5.7 the difference in distributions are clearly seen and the subsequent calculations of copula functions quantify these changes in terms of Kendall's rank.

It was shown that the correlation parameter, $\tau(t)$, of dynamic copula could be used to reliably distinguish between storm phases in real time for input parameters $V_{sw}$ and $B_Z$ versus the output SYM-H.

Copulas are also useful for data imputation (i.e. error correction). Using the copulas calculated from a large set of solar wind and geomagnetic data, missing solar wind observations can be estimated. A simple demonstration of this utility was shown and we believe there is scope for further development and practical application of this method for the space physics community.

It was shown that copulas can be a valuable tool in the analysis of the coupled solar wind – geomagnetic field system, and that there is scope for further exploration of the ideas described in this application.

# Chapter 6

# Conclusion

In this work, the importance and advantages of having the conditional distribution (based on both the static and dynamic copula) in tail dependencies, arising from more extreme events, compared to when more “general” or “average” (ex. mean or median) values are considered, becomes clear.

The GPCC-T model outperforms static copulas in all instances of the financial data considered, showing that the dependence structure does in fact change with time, and this underlying temporal covariate needs to be included in estimation and prediction.

An interesting observation is that the GPCC-T is the best model in terms of log-likelihood, but does not seem to significantly outperform the static copulas in terms of quantile dependence. This may suggest that the GPCC captures volatility of the dependence, but focuses less on shape/ skewness of the data.

In our novel application of copulas to capture the coupling in geomagnetic storms, it was shown that the correlation parameter, $\tau$, of the dynamic (GPCC) copula can be used to reliably distinguish between storm phases in real time.

In both the exchange rate and geomagnetic storm application, the GPCC model shows potential to enable forecasts and aid in mitigation and planning of future events. In the latter application, we also show how copulas can be useful for data imputation (i.e. error correction), thus having great potential for future work for the space physics community.

We wrote a Student’s $t$ Gaussian process conditional copula toolbox for MATLAB based on the R package by Hernández-Lobato et al. [17], which is available on GitHub.

In the broader scientific context, uncertainty from volatilities, heteroskedasticity, extreme values and missing observations all contribute to the difficulty of dependency estimation and prediction. Other models have been considered in a financial setting, but the Student's $t$ GPCC seemed to be the best performing model [17]. In addition, the assumptions of parametric or Gaussian distributions and linear correlation structures regularly applied in statistical analysis are often violated in practical applications. Fully understanding the advantages (and limitations) of copula models helps to avoid model mismatching and broadens the field of possible applications.

The advantage of using the Bayesian framework in the GPCC is that it quantifies the uncertainty of the parameter estimations and predictions, provides further information about the structure/ movement of dependence and relaxes parametric and linear assumptions, providing a more flexible model.

Some limitations of our work is that only a limited number of copulas were considered and that model computations become more time consuming for larger datasets, both for the static and dynamic copula. It is also important to note that pre-processing and parameter initialisation may have a significant influence on estimation results.

A consideration for future work will be to extend the GPCC to higher dimensions, using vine copulas [23], and possibly even creating mixture distributions with other models in order to improve model capabilities.

# Appendix A

# Theory

**Lemma A.0.1.** The random variable $U = F(X)$ is distributed as a $U(0,1)$ random variable. If we let $F^{-1}(u)$, denote the inverse function, that is,

$$F^{-1}(u) = \{x \in \mathbb{R} | F(x) = u\},$$

then the variable $F(U)$ has the same distribution as $X$.

| Copula | Function | Parameter range $(\theta)$ | Kendall's $\tau$ |
|---|---|---|---|
| Gaussian | $\Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$ | $(-1, 1)$ | $\frac{2}{\pi}arcsin(\theta)$ |
| Student t | $t_{\nu,\Sigma}(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2))$ | $(-1, 1)$ | $\frac{2}{\pi}arcsin(\theta)$ |
| Gumbel | $exp[-((-lnu_1)^\theta + (-lnu_2)^\theta)^{\frac{1}{\theta}}]$ | $[1, \infty)$ | $1 - \frac{1}{\theta}$ |
| Frank | $-\frac{1}{\theta}ln(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1})$ | $(-\infty, \infty)$ | $1+\frac{4}{\theta}(D_1(\theta)-1)$ with $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{e^t-1}dt$ |
| Clayton | $(max\{u_1^{-\theta} + u_2^{-\theta} - 1, 0\})^{\frac{1}{\theta}}$ | $(0, \infty)$ | $\frac{\theta}{\theta+2}$ |

**Table A.1:** Copula families and their closed form generating functions

**Algorithm 3:** Simulating from a Gaussian copula

1. Let $X_1$ and $X_2$ be rvs with cdf $F_1(.)$ and $F_2(.)$, and pdf $f_1(.)$ and $f_2(.)$ respectively.
2. Transform to the uniform distribution using the inverse cdf:
   $U_1 = F_1^{-1}(X_1) \sim \mathcal{U}(0,1)$, and $U_2 = F_2^{-1}(X_2) \sim \mathcal{U}(0,1)$.
3. Estimate the distribution of the parameter ($\rho$) using an MCMC simulation within a Bayesian framework.

   Select the maximum likelihood parameter such that $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.
4. The joint distribution is then simply the cdf of the copula,
   $H(x_1, x_2) = C_\Sigma^{Gauss}(u_1, u_2) = \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$,
   with corresponding joint pdf $h(x_1, x_2) = f_1(x_1)\ .f_2(x_2)\ .c_\Sigma^{Gauss}(u_1, u_2)$.
5. Simulate variables from a multivariate normal distribution with correlation structure $\Sigma$. ($(Z_1, Z_2) \sim \mathcal{N}(\mathbf{0}, \Sigma)$).
6. Transform to the uniform distribution using the standard normal inverse:
   $V_1 = \Phi^{-1}(Z_1) \sim \mathcal{U}(0,1)$, and $V_2 = \Phi^{-1}(Z_2) \sim \mathcal{U}(0,1)$.
7. Transform back to the original distributions using the original cdf:
   $Y_1 = F_1(V_1)$ and $Y_2 = F_2(V_2)$.

**Algorithm 4:** Density of Student's $t$ copula i.t.o. Kendall's $\tau$

1. Let $\rho = sin(\frac{\tau\pi}{2})$.
2. From Eq. (3.4), the $t$ copula density is

$$c^T(u_1,u_2) = \frac{\frac{1}{2\pi\sqrt{1-\rho^2}}\left(1+\frac{T_\nu^{-1}(u_1)^2+T_\nu^{-1}(u_2)^2-2\rho T_\nu^{-1}(u_1)T_\nu^{-1}(u_2)}{\nu(1-\rho^2)}\right)^{\frac{-(\nu+2)}{2}}}{t_\nu(T_\nu^{-1}(u_1))t_\nu(T_\nu^{-1}(u_2))}$$

3. The corresponding log-likelihood:

$$\begin{aligned} log(c^T(u_1,u_2)) &= -log(2\pi) - \frac{1}{2}log(1-\rho^2) \\ &- \frac{-(\nu+2)}{2}log\left(1+\frac{T_\nu^{-1}(u_1)^2+T_\nu^{-1}(u_2)^2-2\rho T_\nu^{-1}(u_1)T_\nu^{-1}(u_2)}{\nu(1-\rho^2)}\right) \\ &- log(t_\nu(T_\nu^{-1}(u_1))) - log(t_\nu(T_\nu^{-1}(u_2))) \end{aligned}$$

# Bibliography

[1] MATLAB version 9.8.0.1323502 (R2020a), 2020.

[2] Elif F. Acar, Parisa Azimaee, and Md Erfanul Hoque. Predictive assessment of copula models. *Canadian Journal of Statistics*, 47(1):8–26, 2019.

[3] Tansu Alpcan and Nick Bambos. Modeling dependencies in Financial Risk Management. *Risks and Security of Internet and Systems (CRiSIS), 2009 Fourth International Conference on. IEEE*, pages 113–116, 2009.

[4] Yannick Béniguel and Pierrick Hamel. A global ionosphere scintillation propagation model for equatorial regions. *Journal of Space Weather and Space Climate*, 1(1):A04, 2011.

[5] M Ishaq Bhatti and Hung Quang Do. Recent development in copula and its applications to the energy, forestry and environmental sciences. *International Journal of Hydrogen Energy*, 44(36):19453–19473, 2019.

[6] D. H. Boteler. Assessment of geomagnetic hazard to power systems in Canada. *Natural Hazards*, 23(2-3):101–120, 2001.

[7] Axel Bücher, Holger Dette, and Stanislav Volgushev. A test for Archimedeanity in bivariate copula models. *Journal of Multivariate Analysis*, 110:121–132, 2012.

[8] NASA Goddard Space Flight Center. OMNIwen Plus, 2020.

[9] Martyn Dorey, Phil Joubert, and Coomaren Vencatasawmy. Modelling dependencies : An Overview. *Finance*, (June):19–21, 2005.

[10] Paul Embrechts, Filip Lindskog, and Alexander Mcneil. Modelling Dependence with Copulas and Applications to Risk Management. *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384, 2003.

[11] N I Fisher. No Title. *Encyclopedia of Statistical Sciences*, pages 159–163, 1997.

[12] Nathaniel A. Frissell, Joshua S. Vega, Evan Markowitz, Andrew J. Gerrard, William D. Engelke, Philip J. Erickson, Ethan S. Miller, R. Carl Luetzelschwab, and Jacob Bortnik. High-Frequency Communications Response to Solar Activity in September 2017 as Observed by Amateur Radio Networks. *Space Weather*, 17(1):118–132, 1 2019.

[13] Gabriel Gaiduchevici. A Method for Systemic Risk Estimation Based on CDS Indices. *Review of Economic and Business Studies*, 8(1):103–124, 2016.

[14] Irène Gijbels, Noël Veraverbeke, and Marel Omelka. Conditional copulas, association measures and their applications. *Computational Statistics and Data Analysis*, 55(5):1919–1932, 2011.

[15] Paolo Giudici and Silvia Figini. *Applied Data Mining for Business and Industry*. Wiley, Chichester, U.K., 2009.

[16] W. D. Gonzalez, J. A. Joselyn, Y. Kamide, H. W. Kroehl, G. Rostoker, B. T. Tsurutani, and V. M. Vasyliunas. What is a geomagnetic storm? *Journal of Geophysical Research*, 99(A4):5771, 1994.

[17] José Miguel Hernández-Lobato, James R Lloyd, and Daniel Hernández-Lobato. Gaussian process conditional copulas with applications to financial time series. In *Advances in Neural Information Processing Systems*, pages 1736–1744. MIT Press, 2013.

[18] Martin Hugh. An Introduction to copulas second edition. *lecture notes IEOR E4602*, 1, 2016.

[19] Sebastian Jaimungal and Eddie K H Ng. Kernel-based copula processes. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 628–643, Berlin, Heidelberg, 2009. Springer, Springer.

[20] George Kimeldori and Allan Sampson. Uniform representations of bivariate distributions. *Communications in Statistics–Theory and Methods*, 4(7):617–627, 1975.

[21] Vladik Kreinovich, Hung T Nguyen, Songsak Sriboonchitta, and Olga Kosheleva. Why copulas have been successful in many practical applications: A theoretical explanation based on computational efficiency. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9376:112–125, 2015.

[22] Evgeny Levi and Radu V. Craiu. Bayesian inference for conditional copulas using Gaussian Process single index models. *Computational Statistics and Data Analysis*, 122:115–134, 2018.

[23] David Lopez-Paz, Jose Miguel Hernández-Lobato, and Ghahramani Zoubin. Gaussian process vine copulas for multivariate dependence. In *International Conference on Machine Learning*, volume 28, pages 10–18, 2013.

[24] S. I. Lotz and D. W. Danskin. Extreme Value Analysis of Induced Geoelectric Field in South Africa. *Space Weather*, 15(10):1347–1356, 2017.

[25] Henry Louie. Evaluation of bivariate Archimedean and elliptical copulas to model wind power dependency structures. *Wind Energy*, 17(2):225–240, 2 2014.

[26] Data Analysis Center for Geomagnetism Magnetism and Space. World Data Center for Geomagnetism, Kyoto, 2020.

[27] Attilio Meucci. A Short, Comprehensive, Practical Guide to Copulas. *SSRN Electronic Journal*, (October):22–27, 2012.

[28] Thomas P Minka. Expectation Propagation for Approximate Bayesian Inference. Technical report.

[29] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, Cambridge, Massachusetts, 2012.

[30] Andrew Naish-Guzman and Sean Holden. The generalized FITC approximation. *Advances in Neural Information Processing Systems 20 - Proceedings of the 2007 Conference*, (Ivm):1–8, 2009.

[31] Roger B Nelsen. Properties and applications of copulas: A brief survey. *First Brazilian Conference on Statistical Modelling in Insurance and Finance*, 3:1–18, 2003.

[32] Roger B Nelsen. *An introduction to copulas*. Springer Science & Business Media, New York, 2007.

[33] Andrew J. Patton. Modelling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556, 2006.

[34] Andrew J Patton. A review of copula models for economic time series. *Journal of Multivariate Analysis*, 110:4–18, 2012.

[35] Sheldon M Ross. *Stochastic Processes Second Edition*. John Wiley & Sons, Inc., New York, 1996.

[36] Ludger Rüschendorf. On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, 139(11):3921–3927, 2009.

[37] Avideh Sabeti, Mian Wei, and Radu V Craiu. Additive models for conditional copulas. *Stat*, 3(1):300–312, 2014.

[38] Mojtaba Sadegh, Hamed Moftakhari, Hoshin V Gupta, Elisa Ragno, Omid Mazdiyasni, Brett Sanders, Richard Matthew, and Amir AghaKouchak. Multihazard Scenarios for Analysis of Compound Extreme Events. *Geophysical Research Letters*, 45(11):5470–5480, 2018.

[39] Mojtaba Sadegh, Elisa Ragno, and Amir Aghakouchak. Water Resources Research. *Journal of the American Water Resources Association*, 5(3):2, 1969.

[40] Felix Salmon. The formula that killed Wall Street. *Significance*, 9(1):16–20, 2012.

[41] Thorsten Schmidt. Coping with copulas. *Copulas-From theory to application in finance*, pages 3–34, 2007.

[42] Osvaldo Candido da Silva Filho, Flavio Augusto Ziegelmann, and Michael J Dueker. Modeling dependence dynamics through copulas with regime switching. *Insurance: Mathematics and Economics*, 50(3):346–356, 2012.

[43] M Sklar. Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris*, 8:229–231, 1959.

[44] James C. Spall. Estimation via Markov chain Monte Carlo. *Proceedings of the American Control Conference*, 4(2):2559–2564, 2002.

[45] Lavanya Sita Tekumalla, Vaibhav Rajan, and Chiranjib Bhattacharyya. Vine copulas for mixed data: multi-view clustering for mixed data beyond meta-Gaussian dependencies. *Machine Learning*, 106(9-10):1331–1357, 2017.

[46] Luciana Dalla Valle, Fabrizio Leisen, and Luca Rossini. Bayesian non-parametric conditional copula estimation of twin data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 67(3):523–548, 2018.

[47] Noël Veraverbeke, Marek Omelka, and Irène Gijbels. Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, 38(4):766–780, 2011.

[48] Manthos Vogiatzoglou. Dynamic Copula Toolbox. *SSRN Electronic Journal*, (2009):1–25, 2017.

[49] Tianyang Wang and James S Dyer. A copulas-based approach to modeling dependence in decision trees. *Operations Research*, 60(1):225–242, 2 2012.

[50] Zhao Wang, Weisheng Wang, Chun Liu, Zheng Wang, and Yunhe Hou. Probabilistic forecast for multiple wind farms based on regular vine copulas. *IEEE Transactions on Power Systems*, 33(1):578–589, 2017.

[51] Christopher K I Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT Press, Cambridge, Massachusetts, 2006.

[52] Andrew G Wilson and Zoubin Ghahramani. Copula processes. In *Advances in Neural Information Processing Systems*, pages 2460–2468, Cambridge, 2010. Curran Associates Inc.

[53] Yue Wu, José Miguel Hernández Lobato, and Zoubin Ghahramani. Dynamic covariance models for multivariate financial time series. *30th International Conference on Machine Learning, ICML 2013*, 28(PART 2):1595–1603, 2013.

[54] Dan Xu, Qidong Wei, Elsayed A Elsayed, Yunxia Chen, and Rui Kang. Multivariate degradation modeling of smart electricity meter with multiple performance characteristics via vine copulas. *Quality and Reliability Engineering International*, 33(4):803–821, 6 2017.

[55] Wenjing Zheng, Xiang Ren, Nan Zhou, Da Jiang, and Shaojun Li. Mixture of D-Vine copulas for chemical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 169(July):19–34, 2017.

[56] David M Zimmer. The role of copulas in the housing crisis. *The Review of Economics and Statistics*, 94(2):607–620, 2012.