Titre: Title:	Construction of a Statistical Framework Using Copula Modeling to Assess Sea Level Rise in New-Caledonia, Pacific Ocean
Auteur: Author:	Matheo Kaemo
Date:	2021
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Kaemo, M. (2021). Construction of a Statistical Framework Using Copula Modeling to Assess Sea Level Rise in New-Caledonia, Pacific Ocean [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/9903/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/9903/
Directeurs de recherche: Advisors:	Elmira Hassanzadeh
Programme: Program:	Génie civil

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Construction of a statistical framework using copula modeling to assess sea level rise in New-Caledonia, Pacific Ocean

MATHEO KAEMO

Département de Génie Civil, Géologique et des Mines

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie Civil

Décembre 2021

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitule :

Construction of a statistical framework using copula modeling to assess sea level rise in New-Caledonia, Pacific Ocean

Présenté par **Matheo KAEMO**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

Ahmad SHAKIBAEINIA, président
Elmira HASSANZADEH, membre et directrice de recherche
James GOULET, membre

DEDICATION

Je dédie ce travail à toutes les personnes qui m'ont soutenues durant ma maîtrise en particulier ma famille, ainsi qu'à mes ami(e)s et à mes collègues de Polytechnique pour leurs encouragements.

Également à ma directrice de recherche, Dr. Elmira Hassanzadeh pour sa confiance, sa disponibilité et ses conseils avisés.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Elmira Hassanzadeh for her support and patience during my work. Her advices and guidelines were essential in the achievement of my research and her availability and involvement were highly appreciated and motivated me to go beyond my limits.

I also would like to acknowledge Dr. Ali Nazemi for his ideas and feedback during the presentations of my work and the several meetings we had that really improved its quality.

I would like to thank Polytechnique Montreal along with my colleagues Khalil, Ali, Sarah-Claude, Henrique, and others for their help and support during this special time of COVID-19 pandemic.

I also want to sincerely thank my friend Hamza with whom I shared great times during my research and who helped me keep motivated during the difficult confinement situation.

Finally, I would like to express my sincere gratitude to my family for their encouragement which helped me overcoming all the obstacles that arose during my studies.

RÉSUMÉ

Les mers ont considérablement influencé les activités socio-économiques dans le monde entier. Bien que le niveau de la mer ait varié naturellement au cours de l'histoire, on a observé, au cours des dernières décennies, une élévation significative due aux changements climatiques provoqués par l'homme. Cette élévation du niveau de la mer pourrait causer une série de dommages dans les régions côtières, allant de l'érosion et de la salinisation des sols à la contamination des eaux souterraines et aux inondations. En particulier en Nouvelle-Calédonie, un petit archipel de moins de 300 000 habitants situé à l'est de l'Australie, l'élévation du niveau de la mer peut éroder et faire reculer les rivages, menaçant ainsi les moyens de subsistance des populations locales. Par conséquent, une évaluation correcte de l'élévation du niveau de la mer est nécessaire pour proposer des stratégies d'adaptation efficaces afin de protéger les biens et l'environnement côtiers.

Diverses méthodes ont été développées dans la littérature pour simuler cette variable. Néanmoins, malgré l'influence de la température sur l'élévation du niveau de la mer, leur dépendance n'a pas été entièrement prise en compte dans les modèles disponibles. Dans cette étude, nous avons proposé une méthode utilisant des copules pour simuler l'évolution du niveau de la mer en fonction de la température. Comme étape clé pour la réalisation de ce projet, deux logiciels avec Interface Graphique Utilisateur en MATLAB ont été développés pour les analyses statistiques des données. Le logiciel pour les analyses à une variable comprend une série de codes permettant de quantifier l'autocorrélation, la variabilité intra et interannuelle, les tendances via des méthodes paramétriques et non paramétriques, les points de rupture dans les séries et de réaliser des analyses de fréquence en utilisant une large gamme de distributions. Dans le cadre d'une analyse à deux variables, l'application de copules statistiques permet d'obtenir plusieurs informations liées à la modélisation des dépendances et à la compréhension de l'impact composé des incidents et des risques de défaillance.

Pour mettre en œuvre cette approche en Nouvelle-Calédonie, les données du niveau annuel moyen de la mer (MASEL) et de la température annuelle moyenne (MAT) sur la période 1967-2019 à Nouméa, la ville la plus peuplée, ont été obtenues. Les analyses préliminaires ont montré qu'il existe une autocorrélation et une tendance positive dans ces séries de données. En utilisant une approche simple, la tendance linéaire et autocorrelation des MAT et MASEL ont été retirés pour obtenir les résidus MAT et MASEL indépendants et identiquement distribués. Une dépendance

significative entre les résidus MAT et MASEL a été observée. En utilisant des copules et par échantillonnage conditionnel, une large enveloppe de séries MASEL a été générée. En considérant différentes méthodes pour estimer les paramètres des copules, à savoir la Méthode des Moments et le Maximum de Vraisemblance, ainsi que la sélection des copules et des distributions marginales, un ensemble de six configurations a été défini pour modéliser la structure de dépendance entre les résidus de MAT et MASEL. Nos analyses montrent que tous ces modèles développés sont performants pour représenter les caractéristiques clés de MASEL telles que la moyenne, l'écart-type, les valeurs extrêmes dans la période historique. En utilisant ces modèles, les séries MASEL jusqu'en 2100 ont été simulées.

Les modèles estiment une augmentation considérable du niveau de la mer d'ici la fin du siècle. Nos analyses ont également montré l'importance du traitement des données, c'est-à-dire le maintien ou la suppression de la tendance et de l'autocorrélation des données, car les modèles de copule générés à l'aide des données brutes MAT et MASEL n'ont pas été en mesure de représenter correctement les caractéristiques du niveau de la mer. En outre, l'influence de l'ampleur de la tendance sur les simulations effectuées est déterminée en reproduisant le niveau de la mer avec différentes valeurs de tendance MASEL. Les limites inférieures et supérieures de la tendance ont modifié l'orientation des différentes enveloppes de MASEL, montrant les incertitudes liées au prétraitement des données avant d'appliquer la modélisation copule. Dans l'ensemble, une augmentation maximale de la MASEL d'environ 20 cm est attendue d'ici 2100 dans cet archipel, ce qui pourrait causer des dommages importants aux zones côtières. Les résultats de cette étude peuvent être utilisés par les parties prenantes en Nouvelle-Calédonie pour estimer les dommages possibles associés à l'élévation du niveau de la mer et préparer des stratégies d'adaptation. En outre, le cadre proposé est générique et peut être mis en œuvre pour estimer le niveau de la mer dans d'autres zones côtières en tenant compte des changements de température locale. Les logiciels développés seront disponibles pour le public et pourront être utilisés à des fins éducatives et de recherche.

ABSTRACT

Seas have significantly influenced socio-economic activities across the globe. Although sea levels have been changed naturally throughout the history, over the past few decades a significant sea level rise has been observed due to human-induced changes in climate. Such increases in sea level could cause series of damages in coastal regions ranging from soil erosion and salinization to groundwater contamination and flooding. Particularly in New Caledonia, a small archipelago of less than 300,000 inhabitants located in eastern Australia, sea level rise can retreat and nibble shorelines and threaten livelihood of local people. Therefore, a proper assessment of sea level rise is required to propose effective adaptation strategies to protect coastal assets and environment.

Various methods have been developed in the literature to simulate this variable. Nevertheless, despite the influence of temperature on sea level rise, their dependence have not been fully taken into account in the available models. In this study, we proposed a framework using copulas for simulating the evolution of sea level as a function of temperature. As a key step to fulfill this project, two software packages with Graphical User Interface in MATLAB were developed for statistical analyses of data. The software for univariate analyses includes a series of codes for quantifying autocorrelation, intra and inter-annual variability, trends via parametric and non-parametric methods, break points in series and performing frequency analysis using a wide range of distributions. In the context of a bivariate analysis, the application of statistical copulas allows obtaining several information related to modeling dependencies and understanding compound impact of incidents and risks of failure.

To implement this approach in New Caledonia, data of Mean Annual Sea Level (MASEL) and Mean Annual Temperature (MAT) over the period of 1967-2019 in Nouméa, the most populated city, were obtained. Preliminary analyses showed that there is an autocorrelation and positive trend in these data series. Using a simple approach, the MAT and MASEL were detrended and de-autocorrelated to obtain the Independent Identically Distributed MAT and MASEL residuals. A significant dependence between MAT and MASEL residuals was observed. Using copulas and by conditional sampling a wide envelope of MASEL series were generated. Considering different methods to estimate copula's parameters, namely Method of Moments and Maximum Likelihood, as well as selection of copulas and marginal distributions, a set of six configurations has been

defined to model the dependence structure between MAT and MASEL residuals. Our analyses show that all these developed models perform well in representing key MASEL characteristics such as mean, standard deviation, extreme values in the historical period. Using these models, MASEL series until 2100 were simulated.

The models estimate a considerable rise of sea level by the end of century. Our analyses also showed the importance of data processing, i.e., keeping or removing trend and autocorrelation from data since generated copula models using raw MAT and MASEL data were not able to represent the characteristics of sea level correctly. In addition, the influence of trend magnitude on the performed simulations is determined by reproducing sea level with different MASEL trend values. The lower and upper bounds of trend modified the orientation of the different envelopes of MASEL, showing the uncertainties related to preprocessing data before applying copula modelling. Overall, a maximum increase of MASEL of about 20 cm is expected by 2100 in this archipelago, potentially causing significant damage to coastal areas. The results of this study can be used by stakeholders in New Caledonia to estimate possible damages associated with sea level rise and prepare adaptation strategies. Moreover, the proposed framework is generic and can be implemented to estimate sea level in other coastal areas considering the changes in local temperature. The developed software packages will be available for public to be used for educational and research purposes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AIC Akaike information criterion

BIC Bayesian information criterion

CDF Cumulative distribution function

GCM General circulation model

GEV Generalized extreme value

GUI Graphic user interface

IID Independence identically distributed

IQR Inter quartile range

MAD Median absolute deviation

MASEL Mean annual sea level

MAT Mean annual temperature

ML Maximum likelihood

MoM Method of moments

PDF Probability distribution function

RCM Regional circulation model

SNHT Standard normal homogeneity test

$C, \bar{C}$ Copula and survival copula

LL Loglikelihood

RP Return period

X, Y Random variables representing the time series of observations

p Risk of failure

$u, v \,/\, \bar{u}, \bar{v}$ Marginal distributions and survival distributions of X and Y

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

1.1 Background

Seas give rhythm to the life of billions of people around the world. Considered as a transport axis, they also play key roles in global trade. Moreover, coastal ecosystems are rich and diverse and contribute to food security across various scales. Indeed, industrial fishing and aquaculture feed 55 million people (FAO, 2020) every year and provide a non-negligible income for countries with an opening to ocean waters. Although natural changes in sea levels have been observed in the past thousand years, since the mid 19th century, a significant rise in global sea level has been recorded, which can be due to impact of changing climate conditions (Nordell, 2003). Such increments in sea level can threaten human and environmental health. For instance, intrusion of seawater in land can contaminate soil and groundwater along with freshwater resources, which can directly impact ecosystems and water supply for various activities such as agriculture, aquaculture, or industries. Erosion of shorelines induced by sea level rise can be also an issue for coastal landscape and area. Depending on the magnitude of sea level rise, low-lying areas could be submerged and retreated causing deterioration of local infrastructure. Addressing these issues requires proper monitoring and assessment of sea level rise at the local scale.

1.2 Case study and problem definition

In the Pacific Ocean, where the countries are constituted in the form of archipelagos totalling more than 20 000 islands, sea level is particularly important as major part of the population lives close to the shorelines. In addition to impacts on economy and environment, sea level has defined the culture and organization of traditional societies as well as spatial repartition of population in territories. Therefore, increases in sea level can cause unwanted relocation of tribes, disturb local dynamics (Nunn et al., 2016), and increase tension between communities on occupation of the territory (Ash and Campbell, 2016). The relocation of shoreline infrastructures and villages becomes one of the main consequences of sea level rise in the region (Nunn and Mimura, 1997). Regarding the environment, mangroves located in coastal areas are the most endangered species under pressure due to sea level rise and their spatial density is gradually disappearing in the territories (Lovelock et al., 2015).

Among these sensitive areas, New-Caledonia (Figure 1.1) a small archipelago located in Southwest Pacific, which is one of the top 25 global biodiversity hotspots (Myers et al., 2000). With an area of 18 575 km², New-Caledonia is composed of a main island named “Grande Terre” representing 89.7% of the total area and several smaller islands distributed around. The total population is 271 960 inhabitants in 2020 of which a third live in the capital, Nouméa, located in the Southwest. The Economic Exclusive Zone is 1 740 000 km² and contains the second most important coral reefs in the world with 1 500 km of continuous barrier reef surrounding the Grande Terre and source of an extraordinary biodiversity. In coastal areas, mangroves act as nurseries for marine species and cover a total area of 351 km². Mainly located in the west coast, they are very sensitive to human activities and endangered by coastal urbanization and are now under environmental protective measures. Regarding the topology of the island, west coast is characterized by flat lands and steep shores and is separated from east coast via a range of mountains. Furthermore, the nickel contained in these mountains allows the archipelago to occupy an important place among ore exporters with 25% of global resources of nickel only located in the Grande Terre. For more general information on New-Caledonia, see Payri et al., (2019).

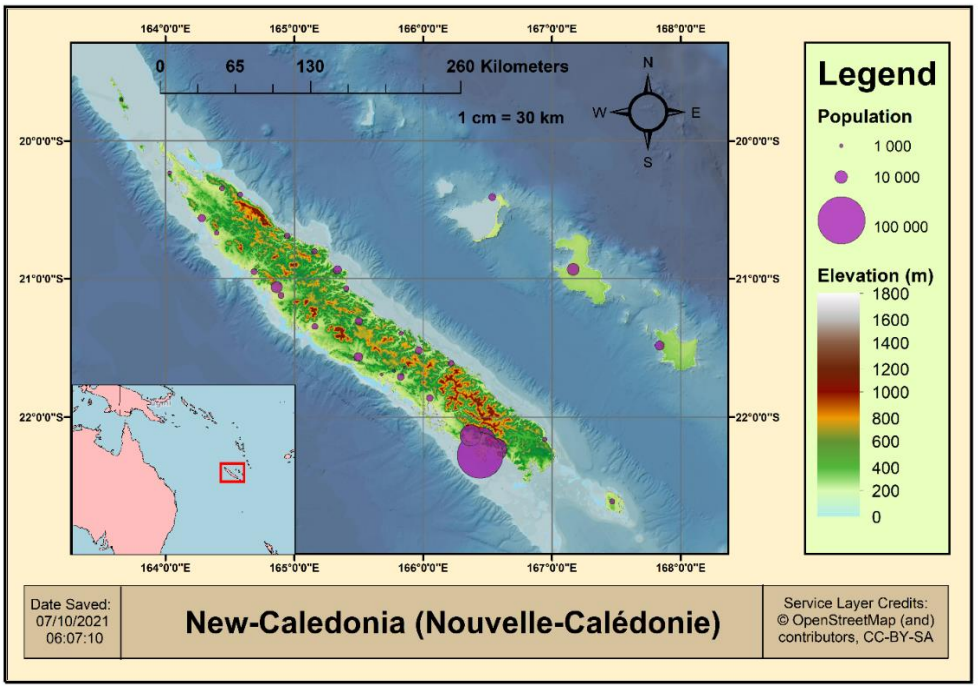

Figure 1.1 Location of New-Caledonia in the Southwest Pacific area with population repartition (ISEE, 2021)

The impact of climate change has already been observed across the archipelago. Thus, over the period 1967-2019, temperature varied between 20.5 °C and 26.4 °C with an overall recorded increase of 2.1°C on the annual average temperature. This magnitude is 48% higher than the global mean temperature augmentation. In parallel, number of hot days (number of days above the 99^{th} percentile of overall temperature distribution) is increasing at the rate of more than 6 days per decade. In the same period sea level has been increased considerably too, continuation of which can damage society, culture, and environment in this area. Therefore, in this study, we aim to simulate sea level rise in New Caledonia to provide accurate estimation of hazards and inform local stakeholders about the upcoming changes in their territory. The literature on statistical analysis of hydroclimate variables and sea level modeling, as well as risks of sea level rise in the Pacific region are detailed in Chapter 2. The objectives of this thesis and the organization of the remaining chapters are explained in Chapter 3.

CHAPTER 2 LITTERATURE REVIEW

This chapter is dedicated to the literature review and aim to provide relevant information to the research project. Also, this review will allow to situate the work achieved within the range of existing studies on the subject, justifying the problematic of the study. It is structured around three themes. The first theme will focus on the recovery of observations and the analyses used to characterize the hydroclimate variables, from univariate to multivariate considerations with an emphasize on copula modeling to represent joint distribution. The second theme will detail the different models currently used to simulate sea level at different scales from process-based approaches extracted from climate models to semi-empirical modeling. These themes show the different methods used to statistically represent variables in hydroclimate context, while highlighting the shortcomings of existing models in the case of sea level assessment. Thus, it allows to better locate the position of this work in the current research field. The third theme will finally review different approaches considered for quantification of risk and estimation of sea level rise in the Pacific region.

2.1 Regression and characterization methods for hydroclimate variables

2.1.1 Recovery of hydroclimate data

Hydroclimate data are often retrieved through using ground-based sensors or from satellite measurements. Thus, common sensors aim to provide estimation of the measurand with a tolerance interval directly provided by the manufacturer and depending on the device's characteristics (Despax, 2016; Goffart et al., 2016). Located in hydrometeorological stations, the data are recovered at regular time intervals. Moreover, high-frequency sensors have been developed and provide observations at small timestep from hours to less than a second. The acquisition of these data is done in national or regional climate centers and mostly available online on dedicated websites. The most common hydrological and meteorological variables considered for climate studies are precipitation (Barbero and Moron, 2011), temperature (Whan et al., 2014), sea level (Turki et al., 2015) and streamflow (Leta et al., 2018). However, there are several issues related to validity domain of these observations considering the implementation of the sensors. For instance,

regarding the sea level, measurement in coastal areas does not properly account for offshore sea level and need to be placed in stable zones in order not to be impacted by tectonic activity that can falsify the measures (Cazenave and Nerem, 2004). Thus, all these data can easily be measured locally through sensors network and are accessible by local stakeholders. However, the development of satellite measurements since few decades is now a new source of observations and provide data at global scale (Cazenave and Nerem, 2004) allowing to overcome the issues faced by conventional measurement systems. Precipitation, temperature, wind intensity and direction or sea level are variables that can be measured through satellite network (Schmetz et al., 2002; Wick et al., 2002; Magnes et al., 2020). Especially for sea level, Jason 1, 2 and 3 satellites along with the TOPEX/Poseidon satellite allow measurements at the global scale. Even though the data collected are reliable and accurate, satellite measurement possesses sources of error. Thus, the stability of the orbits of each satellite is an important factor in the quality of the observed data and a validation step is required to ensure the reliability of measurements (Popp et al., 2020). Moreover, considering the large area covered by satellites, the data collected are retrieved at a grid scale and need to be downscaled for local assessment, which is a source of error for the observations (J. Chen et al., 2011; F. Chen et al., 2019; Zorzetto and Marani, 2019).

2.1.2 Univariate modeling and distribution

The modeling of univariate time series is often considered in hydroclimate analyses (Helsel and Hirsch, 1992). From common models to sophisticated approaches, there is a plethora of methods to characterize and analyze hydroclimate observations. Considering the evolution of the variables over time, linear trend (Hess et al., 2001) and rank-based method as Sen's slope estimator (Sen, 1968) are usually applied with their associated significance test (Mann, 1945; Santer et al., 2000). More advanced methods can also be applied to model time series using spectral analysis. Thus, Fourier transform (El Maghraby et al., 2016) and wavelet analysis (Drago and Boxall, 2002) are used in hydroclimate studies to characterize time series considering their harmonics from frequency spectrum (Meyers et al., 1993; Zaporozhets et al., 2018). The advantage of these methods lies in their ability to model any non-periodic signal into a composition of periodic signals whose characteristics are determined from the analysis of the harmonics of the initial time series. All these above-noted methods aim to represent a time series, but they do not consider intrinsic features of

the variable and the possibility that an observation at a given timestep is function of other observations at previous timesteps. Thus, specific modeling is considered to overcome this consideration and characterize time series with the lagged values. A common approach known as Auto-Regressive Integrated Moving Average provide models that simulate a variable at a given timestep using a linear regression applied with the precedent values (Box et al., 2015). However, they can only be used on time series where variance is constant over the timespan (McCabe and Tremayne, 1995; Milionis, 2004). The development of non-parametric method based on probabilistic method to simulate the evolution of hydroclimate variables over time is also widely used. For instance, Gaussian Process Regression (Grbić et al., 2013; Sun et al., 2014) is considered to simulate the output variable and simultaneously estimating the associated uncertainty for each prediction. If both approaches show interesting advantages regarding the modeling of time series, their limitations prompted the use of different methods in the modeling of hydroclimate variables. Thus, statistical approach is usually considered to study these variables through their distributions to estimate the non-exceedance probability directly from observations (Maity, 2018). Both parametric and non-parametric methods (Helsel and Hirsch, 1992) can be used considering different models that provide an estimate of the Probability Distribution Function (PDF) and the Cumulative Distribution Function (CDF). These characterizations of hydroclimate time series are considered for further risk assessment and multivariate modeling (Moftakhari et al., 2017).

2.1.3 Multivariate analysis

The multivariate analysis is utilized when the influence of several drivers needs to be considered to simulate the evolution of a variable. The classic multiple linear regression aims to model a dependent variable based on linear combination of independent variables (Holder, 1985). Nevertheless, this method includes several restrictive assumptions related to the linear properties of the model, such as independence between observations, absence of correlation between and within variables, homoscedasticity and normal distribution of residuals (Maity, 2018). These hypotheses need to be verified prior of applying a multiple linear regression. To overcome the issue of correlation between variables, Principal Component Analysis can be considered to modify a set of correlated variables into a set of uncorrelated components, linear combination of these variables, using a transformation vector (NCARF, 2013.). Thus, the model is built by adding principal

components according to their ability of explaining the variance (Maity, 2018). However, this modeling is limited as it becomes difficult to interpret the principal components when lots of variables are included, especially if some of them are qualitative. To consider cyclic variation over time, multivariate frequency analysis can be applied in order to estimate the harmonic distribution of several drivers. The Fourier transform is the most common method, aiming to model a multivariate time series in a layer of overlaid trigonometric functions (Z. Zhang et al., 2018). Moreover, to consider time dependence in data, the wavelet analysis provides more robust model in harmonic analysis. However, even though these above-described methods are performing robust modeling of hydroclimate variables, they do not consider the dependence between the different drivers which may lead to misinterpretation of the variable of interest or underestimation of risk for hazard assessment studies (Moftakhari et al., 2017). Thus, copula modeling is an interesting alternative option to simulate hydroclimate variables considering their dependence (Favre et al., 2004; Dupuis, 2007). Applicable with 2 variables or more, copulas provide an estimation of the joint probability calculated from marginal distributions and whose parameter is a function of dependence between variables. In the bivariate consideration which is the most common approach in hydrology (Salvadori and De Michele, 2007), the dependence is assessed through the Kendall's τ (Kendall, 1938) or Spearman's ρ_s (Hauke and Kossowski, 2011). These are non-parametric estimators that aim to quantify the correlation between two sets of data based on the rank of observations. Results obtained from copula modeling can be used in further risk assessment studies, similar to univariate non-exceedance probability.

2.2 Statistical modeling through copula

2.2.1 Scope of copula modeling

The application of copula has been particularly important in risk analysis for financial market and insurance and were especially developed in the economic field (Cherubini et al., 2004; D'Amico and Petroni, 2018). Their development in this sector for insurances, bank and other financial institutions is justified by the need to properly evaluate the hazard level associated to a multivariate event. Thus, these entities are capable to predict potential losses related to a situation and be able to conduct an optimal management policy (Frees and Wang, 2006; Patton, 2009). However, their

mathematical development and the consideration of dependence in the establishment of the model allowed them to be applied in various fields. Especially in civil engineering and hydrology, copulas have been considered to analyze situations related to extreme events (Favre et al., 2004; Ward et al., 2018), flood hazards (Salleh et al., 2016; Bevacqua et al., 2019) and other risk assessment studies related to compound events (Moftakhari et al., 2017; Sadegh et al., 2018). In structure monitoring, copulas have also been recently used to evaluate the health of bridges (Y. Zhang et al., 2020). More generally, copulas appear to be widely considered for risk assessment studies and simulation of hydroclimate variables (Nazemi and Elshorbagy, 2012; Yavuzdoğan and Tanır Kayıkçı, 2020).

2.2.2 Variety of models and possible analyses

Different families of copulas can be considered for modeling purpose. The two most common families are: Archimedian (including in particular Frank, Gumbel and Clayton models) and Metaelliptical (Gaussian and t-student copula). The Archimedian copulas have symmetric and explicit functions to represent the dependence structure. They are obtained from a generator function (Nelsen, 2006) with a specific expression for each copula. Thus, a multitude of copula models can be created using different generator functions (Naz et al., 2020). Metaelliptical copulas are also used for multivariate analyses and has been developed from the models of elliptical distribution families, including the multivariate normal distribution (Genest et al., 2007). Other copulas are also considered to perform multivariate analysis (Durrleman et al., 2000). Then, different possible analyses are possible with copula modeling. Risk study can be performed using the joint probability extracted from copula in order to quantify the bivariate return period associated to a multivariate event (Salvadori et al., 2011; Moftakhari et al., 2017). Thus, quantification of hazard is directly provided by the copula model. Simulation of variable is also another interesting approach related to the use of copula. Indeed, it allows to estimate the evolution of a given variable over time based on other drivers and their associated dependence (Bernardi et al., 2018; Yavuzdoğan and Tanır Kayıkçı, 2020). This method turns out to be particularly effective for simulation purpose and projection.

2.3 Sea level modeling

2.3.1 Physical (or process-based) models

Physical modeling of sea level aims to provide estimations that rely on the consideration of multiple drivers and their interactions through a dynamic model. Thus, General Circulation Models (GCMs) have been developed to simulate the energy and mass exchange between atmosphere and ocean using both prognostic and diagnostic equations to properly model different phenomena involved in climate variations (Adem, 1962; Manabe et al., 1965). The Atmosphere-Ocean GCMs are coupled models from atmospheric and oceanic GCMs and aim to simulate climate response from greenhouse gas emissions at the global scale through simulation of several drivers such as thermal expansion of oceans, ice and land sheets melting, etc. Sea level but also temperature and precipitation are some of the data that can be retrieved from these models' simulations and can be used for further assessment (Gregory et al., 2001; Rahmstorf, 2012). Typical resolution of GCMs is few hundred kilometers, e.g., 300 km × 300 km. Thus, GCMs can not properly represent the local specifications and processes (Giorgi, 2019) that affect sea level fluctuations. To estimate the values at a smaller scale, statistical or dynamic downscaling approaches can be used (Schmidli et al., 2006). Regional Climate Models (RCMs) are products of dynamic downscaling and have a resolution of about 15 km × 15 km (Dickinson et al., 1989; Campbell et al., 2011). Thus, sea level assessments can be realized in specific areas. However, downscaling approaches have their own limitations and uncertainties in the modeling of sea level drivers (Rahmstorf, 2012) can impact the overall quality of physical models (J. Chen et al., 2011). Moreover, data generated from are not often accessible and understandable by local stakeholders. Specific process-based models have also been developed at global scale to simulate global mean sea level but still remain complex (Church et al., 2013b). At local scale, process-based models can be individually developed to represent the evolution of hydroclimate variables in a specific area, considering local climate conditions (Fatichi et al., 2016). Based on equations specific to the studied phenomenon, these models remain nevertheless complex (Clark et al., 2015). In addition, these models require high quality data for various variables, which are not often available in various regions.

2.3.2 Semi-empirical models

Semi-empirical methods aim to provide a framework for sea level estimation from temperature datasets. Gornitz et al., (1982) provide an estimation of sea level trend through a 5-years moving window, thus roughly approximating the rate of sea level using a semi-empirical method. Then, a more developed consideration from Rahmstorf, (2012) has been used based on the statement that augmentation of temperature directly implies sea level is increasing at a faster rate. Hence, the resulting deterministic equation then allows to estimate sea level considering an equilibrium point, from which sea level is varying in response to temperature variation (Moore et al., 2013; Orlić and Pasarić, 2013). The modeling of sea level is realized through a first-order differential equation where the rate of sea level increase is provided by the first derivative (Rahmstorf, 2007). Moreover, more advanced models have been developed to also consider short-term fluctuations of sea level (Vermeer and Rahmstorf, 2009). Specific terms for temperature in this equation allow to model the input signal to which sea level is reacting. To determine the coefficients in the equation, model is usually calibrated and validated using respectively historical data from ground-based stations and satellite altimetry in order to build a robust model for projecting sea level (Rahmstorf et al., 2012; Moore et al., 2013). The consideration of temperature from observed data using climate stations or from simulated data through climate models is possible in this approach (Horton et al., 2008; Vermeer and Rahmstorf, 2009; Jevrejeva et al., 2010). Although they are simple model to develop, the principal limitation lies on the consistency in the estimation of the parameters in future projection (Rahmstorf, 2012). Indeed, the parameters of the defined deterministic equation have been estimated and calibrated with historical observations but the relationship between sea level and temperature may not hold in the future. Thus, it may not properly simulate future sea level.

2.4 Sea level rise analysis in the Pacific area

The Pacific area encompasses more than 20 000 islands. Due to the poor density of population, most of these islands are not monitored and only 10% of them are equipped with climate stations. In this area, sea level fluctuations are associated with thermal expansion of oceans (Becker et al., 2012). Over the period 1993-2009 in the Western Pacific, an increasing sea level ranging between the rate of 1.5 mm/year and 4.9 mm/year depending on the tide gauges' location has been observed using linear regression with univariate and multivariate considerations (Church et al., 2006; Becker

et al., 2012; Aucan et al., 2017). Using moving-window analysis a trend in sea level in the area (Hamlington et al., 2021) with a magnitude close to 2 mm/year has been observed using both tide gauges and satellite altimetry observations. More advanced methods such as Bayesian network have been implemented to estimate coastal hazard related to sea level (Sahin et al., 2019) but it remains anecdotal in the region. Linear trend remains the most used method to account for the phenomenon of sea level rise. However, considering specific characteristics related to sea level such as its decadal or interannual variability allows to approach the situation from a different angle. Thus, with specific indices related to Pacific region (X. Zhang and Church, 2012), linear trend application and physical interpretations (X. Zhang and Church, 2012) provide more robust results. Nevertheless, considering the large area covered by the Pacific Ocean, tide gauge observations cannot be used to properly assessed these climate indices in offshore areas. Satellite altimetry appears to be an interesting alternative to estimate sea level in these areas far from the coast. In New-Caledonia, most of the studies carried out only focus on the meteorological variables directly impacting the land activities and the coral reef lagoon condition such as precipitation and temperature (Fischer et al., 2004; Ouillon et al., 2005; Barbero and Moron, 2011). Also, the assessment of tropical cyclones and their impact on hydrological response of watershed mainly focused on streamflow evolution in different rivers across the Grande-Terre (Terry et al., 2008; Terry and Wotling, 2011) but no in-depth study regarding sea level considerations has been carried out at the local scale.

A topologic study has been led in New-Caledonia to estimate the loss of territory under several sea level rises arbitrary scenarios and to classify the shoreline into different threat levels regarding sea level rise (Bellard et al., 2013). Thus, a loss of land between 5% and 10% is expected under sea level rise scenarios ranging between 1 m and 3 m. In the surroundings areas, classification of coastal hazard has been implemented in order to assign threat levels related to sea level rise in different locations along the shoreline (Sahin et al., 2019). Associated to a probability of occurrence of an event, the return period can be estimated in univariate and bivariate analysis using non-exceedance probability (Shiau, 2003; Salvadori et al., 2011).

2.5 Conclusions

This literature review aims to present the different methods employed to model sea level, emphasizing the lack of statistical methods to properly assess sea level rise considering its possible dependence on other drivers, notably temperature. In univariate simulations, from both empirical observations and satellite measurements, the characterization of hydroclimate variables remains unreliable in terms of projection and accuracy regarding the complexity of the underlying process that generate sea level rise. If multivariate analysis tried to overcome these issues, models are still complex and not especially accessible by local stakeholders. A similar conclusion is drawn from process-based and semi-empirical models that have been widely used to simulate sea level at different scales. A compromise between complexity of the models and their accuracy is found in using copula modeling. Thus, estimations of sea level rise considering its possible dependence on other drivers through copulas provide an interesting framework for risk assessment and simulations. In the Pacific region where coastal areas are particularly vulnerable and dependent to marine condition, hazard studies and sea level simulations can provide important information for local decision-makers regarding land management.

CHAPTER 3 ORGANIZATION OF THE WORK

The Pacific area and especially New-Caledonia will have to deal with sea level rise over the next few decades. In order to protect coastal ecosystem, infrastructure, and more generally support existing cultural developments and socio-economic activities in the shoreline, there is a need to assess sea level rise and its associated risk in the area. As noted, due to complexity of sea level dynamics, there is a need for multivariate modeling considering the dependence between sea level and its key influencing factor. Here we considered the impact of changes in temperature as it is the main driver of sea level rise in the 21st century. A statistical framework based on copula is applied to estimate sea level rise in New-Caledonia.

Thus, the specific objectives of this study are:

(1) Developing generic software packages with a Graphic User Interfaces (GUI) for univariate and bivariate statistical analyses of hydroclimatic data with a greater goal of modeling joint dependence between variables.

(2) Using the codes developed in Objective (1), to assess evolution of sea level rise based on changes in temperature in New-Caledonia considering the ground-based datasets.

Reaching these objectives will provide a general statistical framework that could be used in other locations or for analyses in different domains and contribute to increase the knowledge regarding sea level situation in New-Caledonia. The remaining Chapter 4 and Chapter 5 are respectively dedicated to each of these objectives as below.

Chapter 4 presents a comprehensive set of analyses that can be implemented for univariate and bivariate considerations during this research project. In brief, a series of MATLAB codes are developed to analyze (1) the data time series ranging from characterizing variability, autocorrelation and parametric distribution to detecting trends as well as detection of change points; (2) the data in the frequency domain e.g., to quantify the dependence between variables and use this information to develop a copula model. Also, a description of the bivariate analysis realized with copula modeling is provided. Different copulas are presented, along with a description of the bootstrap resampling method used to select the best copulas and their application in risk assessment and simulations. Examples of applying these codes are provided for our case study. It should be

noted that a user manual to employ these codes are provided in a separate document. A GUI is also developed for utilising these available codes as well, for which a manual is provided. Examples for using this GUI are also presented in this chapter.

Chapter 5 presents the framework based on copula analysis to estimate sea level rise in New-Caledonia. In brief, observed sea level and temperature data is obtained over the timespan 1967-2019. The data series are preprocessed. Autocorrelation and trend are quantified and removed from dataset to obtain Independence identically distributed residuals. Thus, residuals are considered for the copula-based bivariate modeling to estimate sea level as a function of temperature. The model is then used to simulate sea level rise until 2100 at an annual scale. A set of analysis are provided to better understand the impact of data preprocessing on sea level estimations, as well as understating risks of passing critical sea level threshold in the future.

This chapter is a submitted paper to the target journal “Science of Total Environment”.

Chapter 6 presents a general discussion, describing the limitations of this work and points of improvements. Finally, **Chapter 7** summarizes the thesis and provides recommendation for further developments.

CHAPTER 4 CONSIDERED METHODS FOR UNIVARIATE AND BIVARIATE STATISTICAL ANALYSIS

In this chapter, different methods used to develop MATLAB codes for univariate and bivariate analysis are provided in Sections 4.1 and 4.2. Moreover, the GUI of the developed software dedicated to these analyses is presented in Section 4.3.

4.1 Univariate analysis of hydroclimate variable

4.1.1 Characterization and autocorrelation in the collected data

4.1.1.1 Quality and characteristics assessment

As an initial step to understand the quality and range of data, the ratios of missing data, as well as minimum, zero, and maximum values are found, – see Figure B.1. Consideration of zeros along with maximum and minimum are important to study extreme events especially in hydrology, (e.g., finding the number of dry days, maximum precipitation, etc.). This simple calculation can assist in making decisions for handling missing data e.g., to remove the problematic days or years, or to fill missing data by for instance interpolation. The ratio of zeros in a dataset is estimated in the same way.

4.1.1.2 Autocorrelation

Assessment of autocorrelation is an important step in analysis of time series as it provides insight on the correlation between different subsets of the same dataset and estimate if an observation at time t depends on the previous observations. More specifically, estimation of lag-k autocorrelation in a time series provides information regarding the correlation between two equal-length subsets of this time series spaced by k observations. Three indices are available to assess the correlation: Pearson's ρ, Spearman's ρ, and Kendall's τ (Kendall, 1938; Chok, 2010; Hauke and Kossowski, 2011) using the formula described in Equation 4.1 for two random variables X and Y with their respective ranks rg and standard deviations σ.

$$\rho_{Pearson} = \frac{COV(X,Y)}{\sigma_X . \sigma_Y},$$
$$\rho_{Spearman} = \frac{COV\big(\mathrm{rg}(X), \mathrm{rg}(Y)\big)}{\sigma_{\mathrm{rg}(X)} . \sigma_{\mathrm{rg}(Y)}}, \quad (4.1)$$
$$\tau_{kendall} = \frac{2}{n(n-1)} \sum_{i<j} signe\big(x_i - x_j\big) . signe\big(y_i - y_j\big),$$

where n corresponds to the number of observations in the dataset. i, j and k are moving indices of the sample in each time series X (or Y). A critical threshold of 5% for the p-value is considered for the significancy of the results. Thus, for each lag, a value of correlation is calculated and compared with the boundaries of confidence interval (see Figure B.2) defined in Equation 4.2. In this formula, z represents the CDF of the standard normal distribution, α is the significance level (here 5%) and n is the length of the dataset:

$$CI = \pm \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \quad (4.2)$$

Thus, a correlation value outside the boundaries of the confidence interval is considered as significant and should be removed to get independent residuals.

4.1.2 Monotonous trend tests

Monotonous trend tests aim to estimate the magnitude and significance of the overall trend of a time series (Hess et al., 2001). The purpose is to understand the general behavior of a variable, regarding its evolution over a study period. Here, two methods are considered to estimate the trend: a classic linear trend and a rank-based method using Sen's slope. Each method has its own test to estimate the significance of the trend. For the linear trend, t-student test is applied to determine whether the calculated trend over the timespan is significant or not whereas Sen's slope uses Mann-Kendall test for the significance. The p-value of each trend is determined using a bootstrap resampling method (Yue and Pilon, 2004).

4.1.2.1 Linear trend estimation and significance

Linear trend estimation provides general information regarding the increase or decrease of a variable over the timespan – see Figure B.3. For the time series X_t, this trend is estimated using least-square method, which allows to estimate the intercept β_{t0} and slope β_{t1} of the regression model (along with their respective confidence intervals through the covariance matrix) for the trended values $\hat{X}_t$ and the residuals ϵ_t (Şen, 2012; Mudelsee, 2019). The significance of this trend is directly provided by the t-student test where the null hypothesis is a slope equal to 0 (Santer et al., 2000). In this test, a significant threshold α of 5% is commonly considered for the p-value. Estimation of trended values is provided by Equation 4.3 below:

$$X_t = \underbrace{\beta_{t0} + \beta_{t1}.t}_{\hat{X}_t} + \epsilon_t, \tag{4.3}$$

4.1.2.2 Sen's slope estimator and significance

Sen's slope (Sen, 1968) is used to estimate the trend in the time series – see Figure B.4. This robust rank-based method is considered as it is not sensitive to outliers. The significance of the slope is calculated with the Mann-Kendall test (Mann, 1945; Kendall, 1976). This non-parametric test is based on the difference of signs between consecutive observations. Sen's slope and statistic value for the Mann-Kendall test are detailed in Equation 4.4 below:

$$\begin{aligned} S &= \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \text{signe}\left(x_j - x_k\right), \\ a &= \text{median}\left(\frac{x_k - x_i}{k - i}\right), i < k, \end{aligned} \tag{4.4}$$

where x is the considered time series of length n, i, j and k are moving indices. Same as linear trend, a significant threshold α of 5% is selected for the p-value of the Mann-Kendall test.

4.1.2.3 P-value calculation from bootstrap resampling

The calculation of p-value for each approach is realized using the bootstrapping procedure proposed by Yue and Pilon, (2004). First, an ensemble of bootstrap samples is generated. A trend

is calculated for each sample using either parametric or rank-based method. The empirical CDF is determined for the set of bootstrapped trends estimated and the p-value is provided by the non-exceedance probability calculated at the trend magnitude for the complete dataset – see Figure 4.1. A p-value above 0.5 (resp. below 0.5) indicates that the trend is positive (resp. negative). In one-tailed test, if the value is below the critical threshold of 0.05 (resp. above 0.95), the trend is negative (resp. positive) and significant. A p-value close to 0.5 reveals that the data does not show any trend.

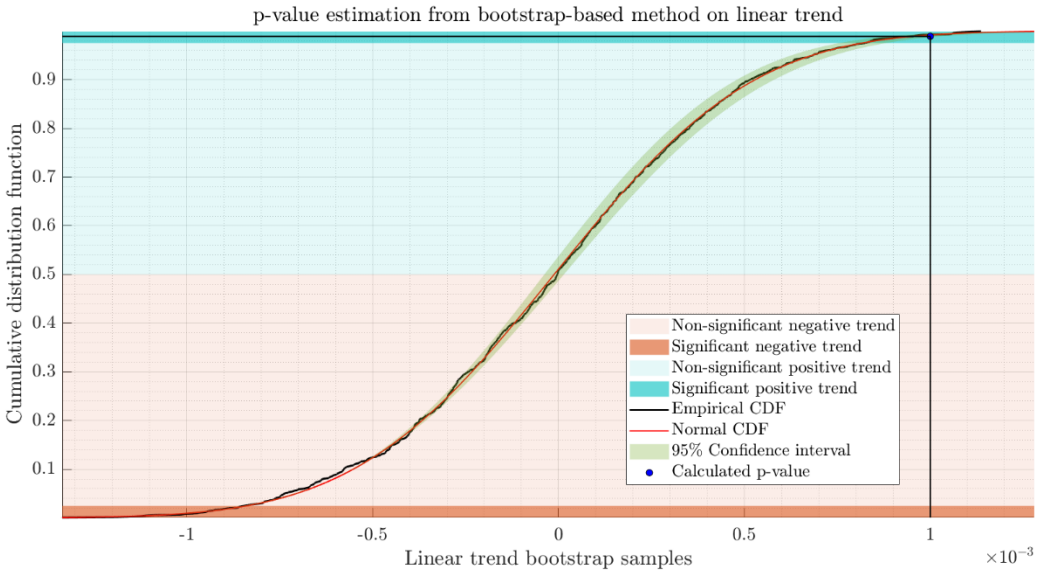

Figure 4.1 p-value estimation from bootstrap resampling method (Yue and Pilon, 2004)

4.1.3 Variability assessment

The variability of the considered time series is analyzed through 3 different approaches. First, general statements using statistical indices are possible in order to characterize the behaviour of the time series and their evolution over the study period. This analysis aims to provide general information on the dataset using trend and moving window analyses of standard deviation, Inter Quartile Range (IQR) and Median Absolute Deviation (MAD) defined in Equation 4.5 below. Then, especially in hydroclimate analysis, intra annual and inter annual variability can be assessed to provide information regarding the variation of the considered time series over the timespan or within a year.

In rank-based analyses, IQR and MAD are used to quantify the variability of a variable X around its median value:

$$\begin{aligned} IQR &= Q_{75} - Q_{25}, \\ MAD &= \text{median}\big(X - \text{median}(X)\big), \end{aligned} \tag{4.5}$$

where Q_{25} and Q_{75} are respectively the first and third quartile of the variable of interest. If IQR only aims to determine the range of the middle 50% of the data, MAD quantify the median distance of the data around their median. Assessing the evolution of each index over a given timespan allows to estimate the evolution of the variability of the underlying process influencing the variables (see Figure B.5). A moving window analysis can also be used to determine the evolution of the variability over moving period within the overall timespan. Intra and inter annual (resp. monthly) variability are performed on the dataset of observations in order to quantify the intrinsic variability and providing information regarding the annual (resp. monthly) behaviour of the variable – see Figure B.6. Thus, for each year (resp. month) Sen's slope can be calculated to estimate the magnitude of the annual (resp. monthly) trend variable – see Figure B.7. Thus, the magnitude of the variation of the timeseries across the year (resp. month) can be quantified.

4.1.4 Parametric distributions

4.1.4.1 Set of considered distributions

In order to characterize the variables in a frequency domain, parametric distributions can be fitted to observations to for instance estimate an exceedance probability for a given observation or to perform risk-based assessment (Johnson et al., 1995; Cowan, 1998). In this project, a set of 11 parametric distribution available with MATLAB is considered. Details regarding the PDF and parameter definitions are available in Table 5.1. This set is composed of the following distributions:

(1) **Generalized Extreme Value** (**GEV**): GEV distribution has 3 parameters and is considered to model extreme values (smallest or largest) in a dataset of IID variables. Physically it is built on the combination of three simpler distribution models that aims to describes tails (Bali, 2003).

(2) **Extreme Value** (or **Gumbel**): this model with 2 parameters allows properly simulating the smallest value from a distribution where the tail is decreasing exponentially (e.g., normal distribution). By inversing the sign of observations, this distribution can be used to model the largest value (Nadarajah and Kotz, 2004).

(3) **Generalized Pareto**: Generalized Pareto is a method that can be used to model the tails, similarly as the GEV distribution. It has 3 parameters and is based on the combination of three basic distributions (Peng and Welsh, 2001; Del Castillo and Daoudi, 2009).

(4) **Loglogistic**: Loglogistic distribution is appropriate to model events with a distribution that has first an initial rate increase followed by a rate decrease often used in survival analysis where data must be positive.

(5) **Lognormal**: Lognormal distribution, also called Glaton distribution, is considered when $\log(x)$ follows a normal distribution (Siano, 1972). As Loglogistic distribution, x must have strictly positive values.

(6) **Normal**: This distribution is used when the Central Limit Theorem is justified (Kwak and Kim, 2017). It basically implies that the sum of samples from independent identically distributed variables converges to a normal distribution when the sample size increase.

(7) **Weibull**: Weibull distribution is adapted to model break strength in materials analysis and often used in reliability and lifetime assessment (Hallinan Jr, 1993).

(8) **Rician**: Rician distribution is often used in communication to represent signal data using a set of two parameters. It has been developed to model Rician fading in signal theory.

(9) **Nakagami**: Nakagami distribution has also been developed for communication application using two parameters. This distribution is related to the gamma law and is often used in wave propagation.

(10) **t Location Scale**: this distribution can be used to model data with more important tails than normal distribution (more outliers). If the shape parameter ν tend to infinite value, then the t Location Scale tend to normal distribution whereas small values of ν implies important tails.

(11) **Gamma**: Gamma distribution is a 2-parameter distribution, appropriate to represent the sum of exponentially distributed random variables (Lukacs, 1955).

For more information related to distribution, see Cowan (1998).

4.1.4.2 Criterions for goodness-of-fit estimation

The estimation of the parameters for the different models described in Table 5.1 allows to fit a parametric distribution to a set of observations. In this project Maximum Likelihood (ML) approach is used to estimate distribution parameters and their confidence intervals (Nguyen et al., 2020). The loglikelihood value resulting from this approach is then used to estimate the goodness-of-fit. Two criterions are considered: the Bayesian Information Criterion (BIC, Equation 4.6) and the Akaike Information Criterion (AIC, Equation 4.7). The BIC highlights simple models with a low number of parameters that still provide accurate results (Burnham and Anderson, 2002; Neath and Cavanaugh, 2012) whereas AIC emphasizes on the models with high accuracy and less consideration for the complexity (Sakamoto et al., 1986; Cavanaugh and Neath, 2019).

$$\mathrm{BIC} = -2.LL + k.\ln(n), \tag{4.6}$$

$$\mathrm{AIC} = 2.\mathrm{k} - 2\ln(LL), \tag{4.7}$$

with k showing number of parameter(s) in the model and n as the number of data points. LL represents the loglikelihood associated with the considered variable and parametric model. Thus, the goodness-of-fit of each distribution can be quantified and compared – see Figure B.8. For more details on AIC and BIC, see Burnham and Anderson (2002).

4.1.5 Change point detection

Homogeneity break tests are utilized to determine the location of a change point in time series, characterized by a break in the mean (Reeves et al., 2007; Aminikhanghahi and Cook, 2017). Thus, a mean value can be estimated before and after the calculated change point and its significance is assessed using p-value with a critical threshold of 5%. A total of 4 homogeneity break tests, namely Pettitt test, Buishand test, Standard Normal Homogeneity Test (SNHT) and CUSUM test are implement in this module (Pohlert, 2016). These tests detect a single change point in time series except for the CUSUM test that allows the detection of multiple change points (Tartakovsky et al., 2014). The only trend break test used in this module is the sequential Mann-Kendall test (Bari et al., 2016) whose purpose is to detect change points in the trend in a time series. These tests are explained below.

4.1.5.1 Pettitt test

Pettitt test (Pettitt, 1979) is a rank-based test that was developed to find a single change point in a time series (see Figure B.9), especially efficient for break point located in the middle of the time series (Kang and Yusof, 2012). The test is structured around a score called "test statistic" noted S_n calculated from the rank of the variables using Equation 4.8. Finding of the maximum for the absolute value of $S_n$indicates the location of the change point in the time series. The associated p-value is calculated with an approximate formula. According to Pettitt (1979), this approximation of p-value is efficient until two decimals for a p-value below 0.5:

$$\begin{aligned} S_n &= \sum_{i=1}^{t} \sum_{j=t+1}^{n} \text{signe}\left(X_i - X_j\right), \\ K &= \max(|S_n|), \\ p &\approx 2.\exp\left(\frac{-6.K^2}{n^3 + n^2}\right), \end{aligned} \tag{4.8}$$

where S_n is the test statistic, n is the number of observations in X; K is the location of the change point; p is the approximate p-value associated with the Pettitt test.

4.1.5.2 Buishand test

Buishand test (Buishand, 1982) allows find a single change-point in a time series (see Figure B.10). Opposite to Pettitt test, Buishand test is based on the assumption that the variable is normally distributed (Kang and Yusof, 2012) as it is not a rank-based test. However, it is built with the same structure as Pettitt test in the way that a statistic value S_n is calculated (Equation 4.9), and the location of the change point is based on the maximum (or minimum) of the statistic value S_n. From a practical point of view, Buishand test is efficient to detect change point in the middle of the time series.

$$S_n = \sum_{i=1}^{n} (x_i - \hat{x}) \,, 1 \leq i \leq n, \tag{4.9}$$

$$R = \frac{\max(S_n) - \min(S_n)}{\sigma},$$

where S_n is the adjusted partial sum, corresponding to the statistic value for the test and R is the test statistic; $\hat{x}$ represent the mean value of observations and σ the standard deviation. Significance of the break point is estimated by comparing the value of $R/\sqrt{n}$ to the critical p-values provided by Buishand (1982).

4.1.5.3 SNHT test

Standard Normal Homogenized Test (Hawkins, 1977; Alexandersson, 1986) allows locate a single change point in a time series (see Figure B.11). Similar to Buishand test, SNHT is based on the assumption that the variables are normally distributed (Kang and Yusof, 2012). Moreover, a test statistic value T_y is calculated, and its maximum value correspond to the change point location noted K in Equation 4.10 (Pohlert, 2016). SNHT is efficient to locate change point close to the beginning or the end of a time series. Physically, the test statistic T_y allows to compare the mean from first y years extracted from z_1 with the mean of the last $(n - y)$ years extracted from z_2:

$$\begin{gathered} z_1 = \frac{1}{y}\sum_{i=1}^{y}\frac{x_i - \hat{x}}{\sigma}\ ;\ \mathrm{z}_2 = \frac{1}{n-y}\sum_{i=y+1}^{n}\frac{x_i - \hat{x}}{\sigma}; 1 \le y \le n, \\ T_y = y.z_1^2 + (n-y).z_2^2, \\ K = \max(T_y), \end{gathered} \tag{4.10}$$

where $\hat{x}$ represent the mean value of observations and σ the standard deviation. Significance estimation is based on comparison between K and critical test statistic value depending on the number of observations (see Figure B.12) and provided by Khaliq and Ouarda, (2007).

4.1.5.4 CUSUM test

CUSUM test has been developed to locate multiple change points in a time series. A change point is considered as such when a small change in mean is detected (see Figure B.13). This detection is based on a simple procedure using the two formulas described in Equation 4.11 below (MATLAB-Cusum-test, 2021). An area of 10 standard deviations σ centered around 0 is defined as a bandwidth

based on MATLAB recommendations. Two scores are calculated: the upper and lower cumulative sum, respectively U_i and L_i. A small change in mean value μ_x is detected each time U_i (resp.L_i) crosses the boundary of +5 σ (resp. -5 σ).

$$U_i = \begin{cases} 0, i = 1 \\ \max\left(0, U_{i-1} + x_i - \mu - \frac{1}{2}n.\sigma\right), i > 1, \end{cases}$$
$$Li = \begin{cases} 0, i = 1 \\ \max\left(0, L_{i-1} + x_i - \mu + \frac{1}{2}n.\sigma\right), i > 1, \end{cases} \quad (4.11)$$

where n is the number of standard deviations from the targeted mean calculated from the sample of observations.

4.1.5.5 Sequential Mann-Kendall test

Unlike the previous tests, sequential Mann-Kendall is a trend break test (see Figure B.14) that allows locating change point in the trend of a time series (trend shift). Two scores u and u' are computed using the initial time series and the reverse one following the calculation steps of the procedure described below in Figure 4.2 (Sayemuzzaman and Jha, 2014; Bari et al., 2016). After applying this procedure, a sequential progressive u and sequential backward u' are obtained. Comparing these values with the z-score (obtained from the inverse CDF of normal distribution) calculated from the time series of observations x_i allows to detect change points (Bisai et al., 2014). Indeed, each time u and u' are crossing each other, a change point is detected. If these intersections are located inside the boundaries defined by the z-score, the change point is significant, otherwise the change point is considered as not significant.

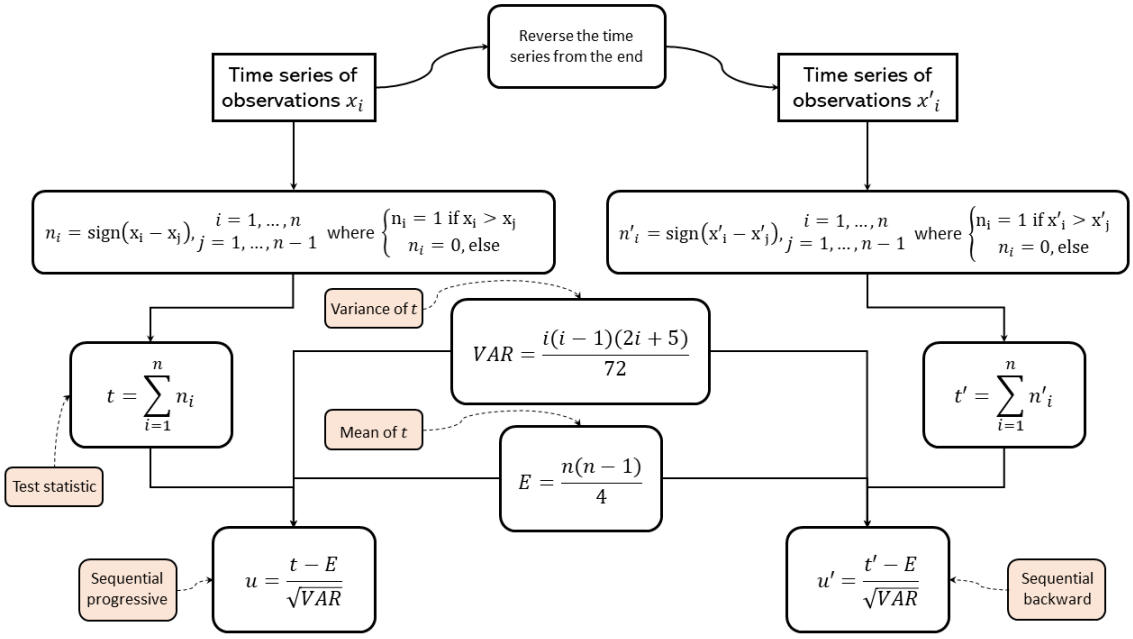

Figure 4.2 Procedure to calculate the sequential progressive and sequential backwards scores to estimate change point location (Bari et al., 2016)

4.2 Bivariate analysis of hydroclimate variables

4.2.1 Joint dependence assessment through copula-based modeling

Considering two random variables X and Y that are continuous with their associated marginal CDFs F_X and F_Y, the joint CDF $F_{XY}(x, y)$ can be assessed with copula by Equation 4.12 (Sklar, 1959):

$$F_{XY}(x, y) = C\big(F_X(x), F_Y(y)\big) = C(u, v), \tag{4.12}$$

In this project, a set of five parametric copulas, namely Gaussian, t, Clayton, Frank, and Gumbel are considered for bivariate modeling. Part of Metaelliptical copulas, Gaussian and t copulas are considered when the variables present an elliptical contour whereas Clayton, Frank and Gumbel copulas are part of Archimedian family characterized by explicit and symmetric function to describe the dependence structure (Genest et al., 2007). Archimedian copulas are determined from a generator function (Nelsen, 2006) specific to each copula. Thus, the joint probability is estimated from the copula modeling (Salvadori et al., 2016):

$$C(u,v) = P(u < u^*, v < v^*), \tag{4.13}$$

where u^* and v^* are the corresponding percentile thresholds for the marginal CDFs of the random variables X and Y and P represent the joint CDF of the considered copula.

If both variables are totally independent (Kendall's τ close to 0) or totally dependent (Kendall's τ close to 1), the formulation of the joint probability is provided by Equation 4.14:

$$\begin{aligned} C(u,v) &= u * v \ \rightarrow \text{independence}, \\ C(u,v) &= \min(u,v) \rightarrow \text{dependence}, \end{aligned} \tag{4.14}$$

Defined through parametric equations, copulas' parameter can be estimated with two different methods: Maximum Likelihood (ML) method through the Inference Functions for Margins (IFM), and Method and Moments (MoM). In the bivariate case, loglikelihood and copula's parameter θ are calculated using Equation 4.15:

$$\begin{aligned} LL &= \sum_{i=1}^{n} \log\big(c(u_i, v_i|\theta)\big), \\ \hat{\theta} &= \text{argmax}(LL), \end{aligned} \tag{4.15}$$

where n is the number of observations, u_i and v_i are the non-exceedance probabilities of X and Y and c is the copula PDF. In the case of fully parametrized ML, the parameters of marginal distributions of X and Y are simultaneously fitted with θ through a maximization of loglikelihood. In MoM, copula's parameter θ is directly calculated from the value of Kendall's τ, thus representing the dependence between the considered variables. Performance assessment of the model is realized using goodness-of-fit AIC and BIC criterions defined in Section 4.1.4.2. Table 4.1 defines the joint PDF and estimation of copula's parameter from Kendall's τ magnitude between the variables:

Table 4.1 Copula functions and their parameters used to estimate the dependency between two random variables

Copula name	Joint density function	Parameters	Parameter $\boldsymbol{\theta}$ range	Expression of Kendall's $\boldsymbol{\tau}$
Gumbel	$\exp\left(-((-\ln(u))^{\theta}+(-\ln(v))^{\theta})^{\frac{1}{\theta}}\right)$	1 parameter: $\theta \geq 1$	$[1,+\infty[$	$=1-\frac{1}{\theta}$
Clayton	$(uv)^{-\theta-1}(\theta+1)(u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}-2}$	1 parameter: $\theta \geq 1$ and $\theta \neq 0$	$[-1,+\infty[\setminus\{0\}$	$=\frac{\theta}{\theta+2}$
Frank	$\frac{\theta e^{-\theta(u+v)}(e^{-\theta}-1)}{e^{-\theta(u+v)}-e^{-\theta u}-e^{-\theta v}e^{-\theta}}$	1 parameter: $\theta \neq 0$	$\mathbb{R}^*$	$=1-\frac{4}{\theta}\left[\frac{1}{\theta}\int_0^{\theta}\frac{t}{e^t-1}dt-1\right]$
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left(-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right)dxdy$	1 parameter: $-1<\theta<1$	$[-1,1]$	$=\frac{2}{\pi}\arcsin(\theta)$
t (student)	$\int_{-\infty}^{t_v^{-1}(u)}\int_{-\infty}^{t_v^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\left(1+\frac{x^2-2\theta xy+y^2}{v(1-\theta^2)}\right)^{-\frac{v+2}{2}}dydx$	2 parameters: $-1<\theta<1$ v	$\theta \in [-1,1]$ $v \in \mathbb{R}^+$	$=\frac{2}{\pi}\arcsin(\theta)$

In terms of frequency of events, each copula can be used in a specific context depending on the location of the maximums of their joint PDF. Thus, Clayton copula can be used for events with high-frequency considering that its PDF shows a maximum peak for low-percentiles of each margin. On the opposite, Gumbel copula is efficient to represent joint probability of infrequent events whereas Frank copula is more suitable for events equitably distributed in terms of occurrence (Figure 4.3). Elliptical copulas are especially efficient in representing joint events with high and low frequencies.

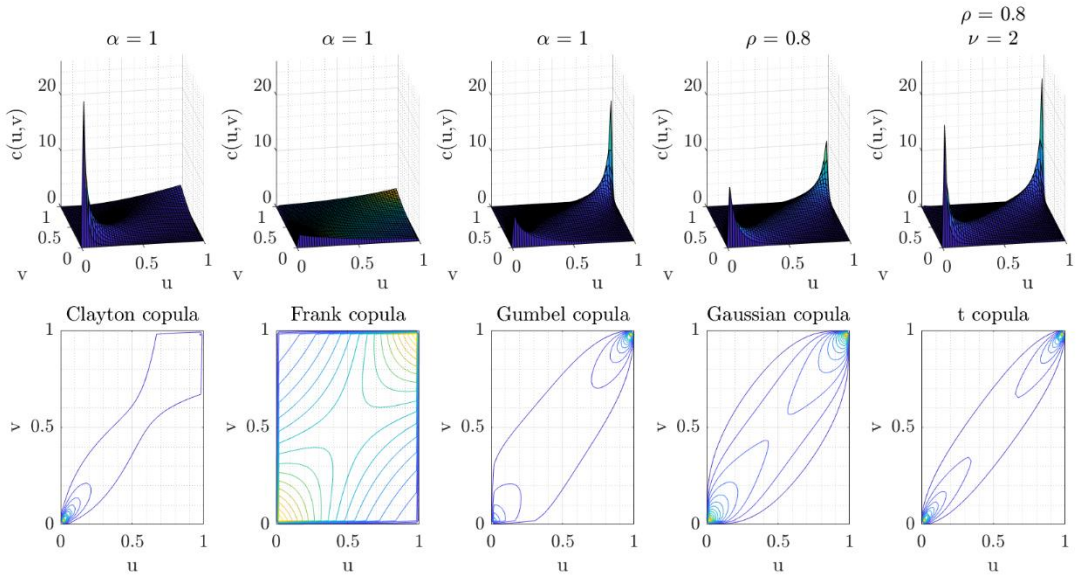

Figure 4.3 As an example, copula probability density function 3-dimensionnal representation (top row) along with their associated respective contours (bottom row)

4.2.2 Selection of copula and confirmation process using bootstrap resampling method

4.2.2.1 Selection of the copula

The copulas are selected based on their ability to represent the dependence between the variables and their simplicity and accuracy. The testing and validation of copula models rely on these considerations. Here, the dependence based on the copula is quantified by the simulated value of Kendall's τ, noted τ_{sim}, which is calculated from the copula's parameter θ_{cop} (Table 4.1). The simplicity and accuracy are measured with AIC and BIC. As an example, in Figure B.15 (resp. Figure B.16), the assessments for each copula models are realized considering ML (resp. MoM). Depending on the method used to estimate θ_{cop}, the testing and validation are either based on representing the dependence with the calculation of τ_{sim} or rely on AIC and BIC to estimate their accuracy. In ML consideration, the best copula is selected based on its ability to preserve the dependence between variables. Therefore, the copula that has τ_{sim} closed to the empirical value of Kendall's τ, noted τ_{emp} calculated from the initial data, is selected. The different models are tested on their AIC and BIC value. In the case of using MoM, the testing of copula is based on the value

of τ_{sim}. Here, the τ_{emp} is used to estimate θ_{cop} (Table 4.1) and the resulting value of τ_{sim} is compared to τ_{emp} to test each copula models. Therefore, the most suitable copula is selected based on the minimum BIC value (validation). Thus, the bivariate model built can be used to calculate the joint probability through 3-dimensional surface of the copula CDF.

4.2.2.2 Bootstrap method for validation

In order to further validate the performance of copulas, a bootstrap resampling method is employed in this project. This approach consists of resampling from the observed datasets a large number of times and it allows generating variability in the results (Bland and Altman, 2015). Considering two variables X and Y selected, the bootstrap method applied here create a 1000 bootstrap samples X_{boot} and Y_{boot}, obtained by randomly picking values in the dataset of X and Y. Once the set of bootstrap samples is created, for each bootstrap sample, copulas performance is assessed through calculating BIC or estimating Kendall's τ (depending on the method used to estimate copula's parameter θ_{cop} – see Section 4.2.1). In total depending on the considered number of samples, an envelop of Kendall's tau values and BIC values are found, which can be shown by boxplots. In case of using MoM to estimate copula's parameter, the boxplot of BIC values for each copula are compared. The copula that shows the lowest median value of BIC is considered as the most suitable copula. In the case of using ML method, the boxplot with similar median value to the real Kendall's τ is the most suitable model to assess the dependence. The wideness of each boxplot can also be used to compare with the range of simulated Kendall's τ values to see if each copula can correctly catch the variability. This procedure is explained in Figure 4.4 and an example is provided in Figure B.17.

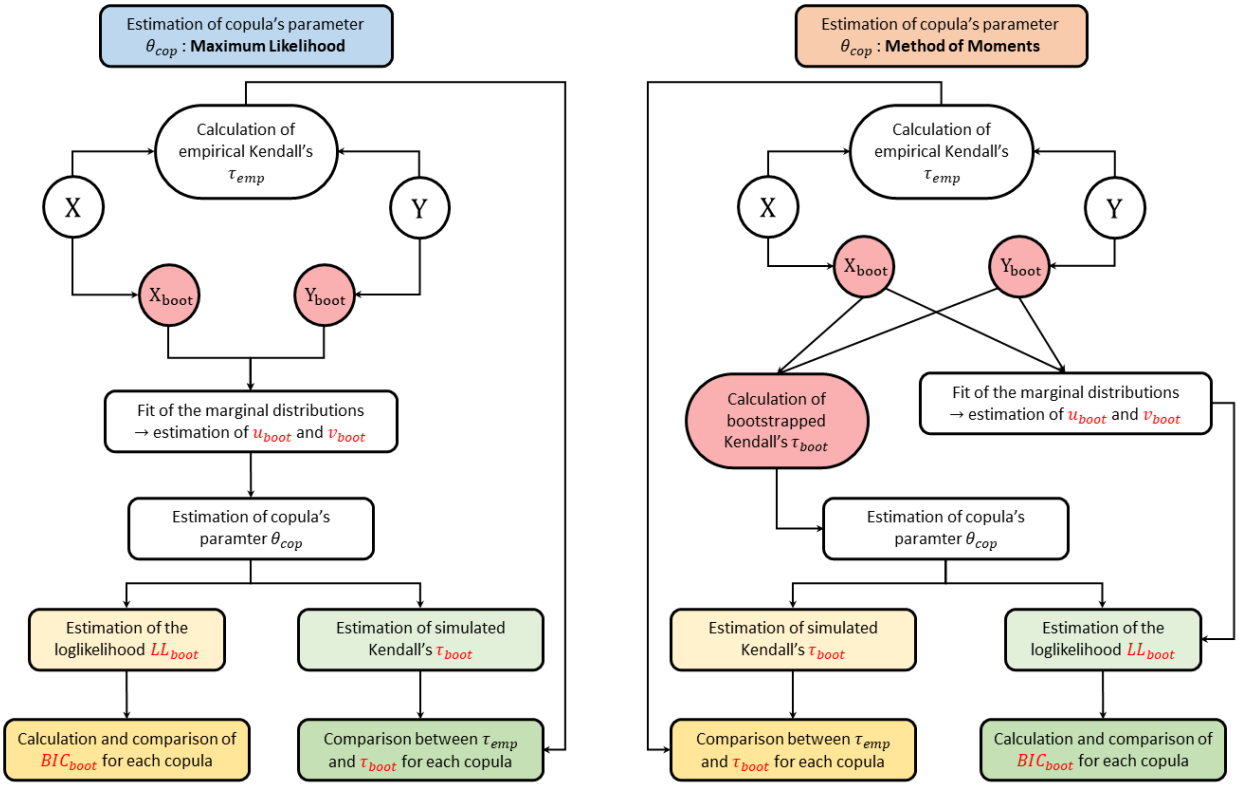

Figure 4.4 Flowchart of the bootstrap resample procedure considering Maximum Likelihood and Method of Moments for the estimation of θ_{cop}

4.2.3 Risk assessment under copula modeling

4.2.3.1 Consideration of bivariate hazard scenarios

In the context of copula, the risk assessment is based on the calculation of bivariate return period using arbitrary thresholds for the marginal distributions. Indeed, for these given percentiles of X and Y, a joint probability can be computed and considered to estimate the return period of two events of fixed frequency. Formulation of the return period depends on the scenario assumed for the copula modeling. Two different scenarios, namely OR and AND are considered as hazard scenarios in this study (Salvadori et al., 2016). The physical meaning behind these scenarios lies in the manner to consider events in a hazardous situation. Thus, in OR situation, an event is considered risky if only one of the two variables exceed a given threshold. For instance, at the

junction between two rivers, a flood hazard can be caused by the water level of one river or both (Bender et al., 2016). In AND scenario, an event is considered risky if both variables exceed their respective given thresholds. In the previous example, it would mean that both rivers must be considered for a situation to be hazardous. However, whatever the scenario considered, the framework remains the same and a joint probability is calculated with its associated critical threshold. Then, this information allows to quantify the joint return period and risk failure. The implication of the considered scenario only changes the way of calculating these different quantities. In the case of bivariate analysis with OR scenario, the joint probability is directly estimated from the copula model (Salvadori et al., 2013) using Equation 4.12 (Section 4.2.1) and the associated critical bivariate threshold α_{OR} (Salvadori et al., 2016) is provided by the Equation 4.16:

$$\alpha_{OR} = 1 - C(u^*, v^*), \tag{4.16}$$

Thus, the joint probability is expressed by the copula $C(u^*, v^*)$ and provide isolines of probability (see Figure 4.5 below) that can be considered for risk assessment through the estimation of isolines of bivariate return period (see Section 4.2.3.2). These isolines of probability (green line) correspond to a critical bivariate threshold. Applying inverse CDF to each marginal distributions on the isoline if joint probability allows considering it in the initial space of variables (see Figure B.18) in order to identify hazardous events.

In the case of bivariate analysis with AND scenario, the joint probability and its associated threshold α_{AND} (Salvadori et al., 2016) are estimated from the survival copula function $\bar{C}$ applied to survival marginal distribution $\overline{F_X}$ and $\overline{F_Y}$ (Salvadori et al., 2013) and is estimated with Equation 4.17:

$$\begin{aligned} \bar{C}\left(\overline{F_X}(x), \overline{F_Y}(y)\right) = P(X > x^*, Y > y^*) = \bar{C}(\bar{u}, \bar{v}), \\ \alpha_{AND} = \bar{C}(\bar{u}^*, \bar{v}^*), \end{aligned} \tag{4.17}$$

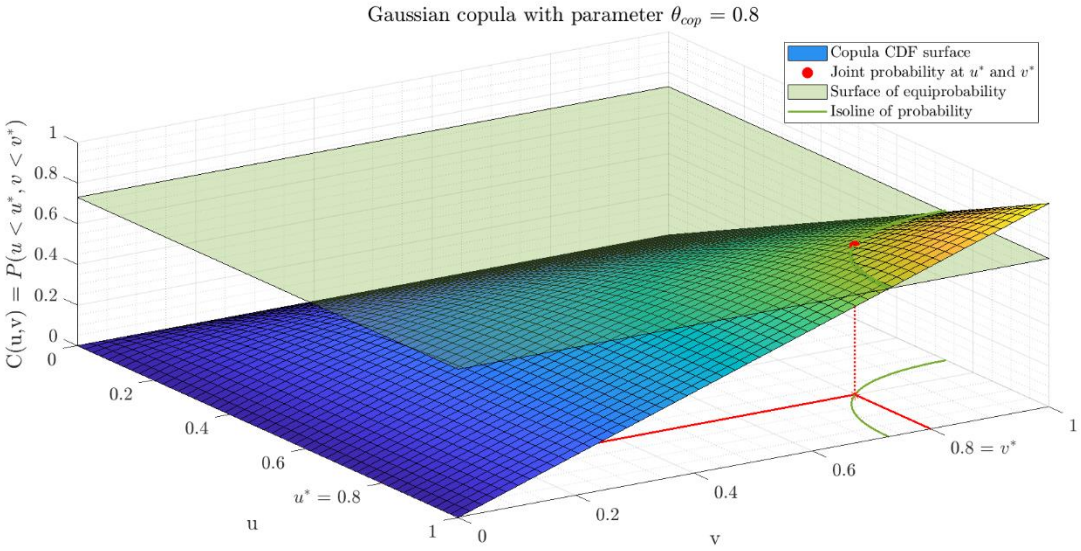

Figure 4.5 As an example, isoline of joint probability for percentile thresholds of 0.8 on both marginal CDFs u and v

The relationship between the function and survival function for marginal distributions and the relationship between copula and survival copula (Salvadori and De Michele, 2004) are provided by Equation 4.18:

$$\bar{u} = 1 - u, \qquad \bar{C}(\bar{u}, \bar{v}) = C(u, v) + 1 - u - v, \tag{4.18}$$

For more information on hazard scenarios and theoretical background regarding copula, see Salvadori and De Michele (2004).

4.2.3.2 Risk of failure probability and bivariate return period

Once the joint probability and its associated bivariate threshold have been calculated, risk failure and return period can be estimated. More specifically, a risk of failure represents the evolution of risk over time corresponding to a specific value of the joint probability. It is directly calculated after estimating the joint probability for a given scenario and copula (Salvadori et al., 2016; Moftakhari et al., 2017). The bivariate risk failure, under each scenario of OR and AND is defined in Equation 4.19:

$$p_{OR} = 1 - \left(C(u^*, v^*)\right)^T = 1 - (1 - \alpha_{OR})^T,$$
$$p_{AND} = 1 - \left(1 - \bar{C}(\bar{u}^*, \bar{v}^*)\right)^T = 1 - (1 - \alpha_{AND})^T, \quad (4.19)$$

where u^* and v^* are the corresponding percentile thresholds of the random variables X and Y. T is the timespan considered to calculate the risk of failure probability p_{OR} or p_{AND}.

The bivariate return period RP (Salvadori et al., 2011; Salvadori et al., 2013; Bender et al., 2016; Moftakhari et al., 2017) can be estimated using the relationship existing between copula (resp. survival copula) and return period in the case of OR (resp. AND) scenario through Equation 4.20:

$$RP_{OR} = \frac{1}{1 - C(u^*, v^*)} = \frac{1}{\alpha_{OR}},$$
$$RP_{AND} = \frac{1}{\bar{C}(\bar{u}^*, \bar{v}^*)} = \frac{1}{\alpha_{AND}}, \quad (4.20)$$

Thus, evolution of the bivariate return period can be assessed over the study timespan and provide information related to the risk of apparition of an event (Moftakhari et al., 2017).

The obtained results can be compared to extreme opposite cases where return periods for total dependence and total independence are calculated. In the case where both variables are independent (resp. totally dependent), the bivariate return period RP_{IND} (resp. RP_{DEP}) can be estimated from u and v directly using Equation 4.21 (Salvadori et al., 2011; Bender et al., 2016; Moftakhari et al., 2017):

$$RP_{IND} = \frac{1}{1 - u^*.v^*},$$
$$RP_{DEP} = \frac{1}{1 - \min(u^*, v^*)}, \quad (4.21)$$

Thus, risk failure (resp. return period) calculated with copula are logically comprised between risks failures (resp. return periods) from independence and total dependence – see Figure B.19. Calculation of risk of failure and return period allows estimating the importance of an event

regarding a specific area. Consideration of its magnitude can provide further information about its impact on a given location.

4.2.3.3 Classification of risky event

The classification process relies on a simple consideration of a joint probability value associated to each event and compared to isolines of joint probability. These isolines provide a critical hazard level based on percentile thresholds of each distribution used to compute the value of copula CDF. Indeed, hazard areas are delimited with the space between these isolines for different percentile thresholds of X and Y (see Figure 4.6). Thus, for k percentile thresholds considered, $k + 1$ hazard areas are defined by the spaces between isolines themselves and spaces between isolines and the range of each variable as described in Figure 4.6. Initially determined in the copula space (left panel), these contours of joint equiprobability are transferred in the original space of the variables X and Y using their respective inverse CDF.

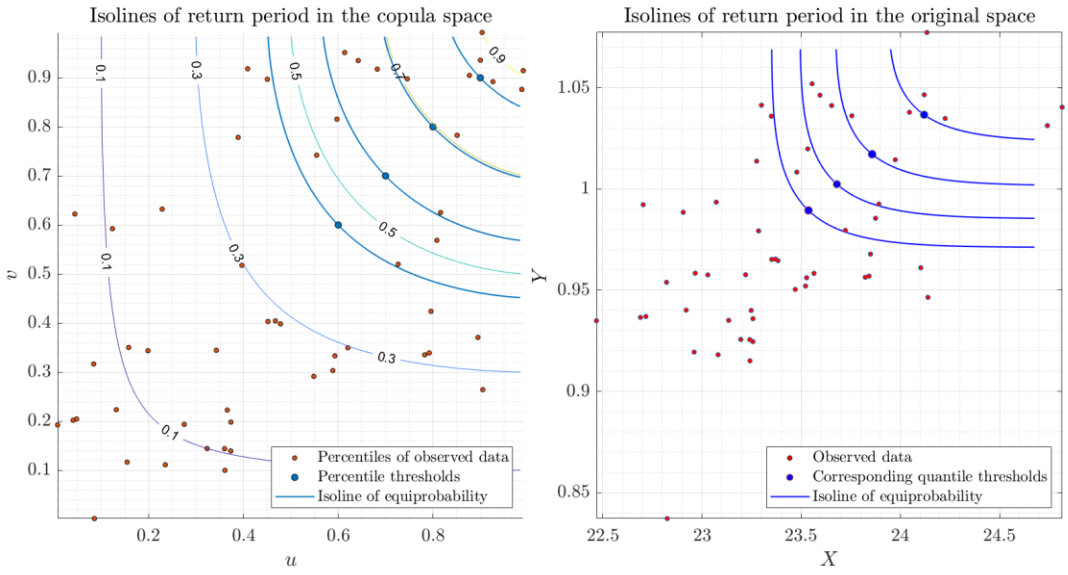

Figure 4.6 As an example, isolines of joint probability for percentile thresholds of 0.6, 0.7, 0.8 and 0.9 (blue points) in the copula space (left panel) and transfer into the original space of the observed data (right panel)

The delimitation of these hazard areas provides a class for each event (e.g., classes 1 to 5 in the figure above). Thus, at each timestep, a hazard level is associated to an event and the evolution of

this hazard, considering copula – see Figure B.20. This evolution of the risk class over time is analyzed to estimate the conditional probability of each hazard level at time $t+1$ considering a given level at time t in each case (Figure B.20 – bottom graph). Thus, a matrix M that estimate the number of times that an event pass from class C at time t to class C' at time $t+1$ divided by the number of observations is built using Equation 4.22. This matrix M directly provides the class conditional probability associated to risk changes.

$$M = \left[\begin{array}{ccccc} \overbrace{\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \end{array}}^{\text{Class at time t+1}} & & & & \\ M_{1\to1} & \cdots & \cdots & \cdots & M_{1\to5} \to 1 \\ \vdots & \ddots & \cdots & \cdots & \vdots \to 2 \\ \vdots & \vdots & M_{3\to5} & \cdots & \vdots \to 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \to 4 \\ M_{5\to1} & \cdots & \cdots & \cdots & M_{5\to5} \to 5 \end{array} \right] \quad \text{with } M_{i\to j} = \frac{\#\,\overset{\text{events}}{\widehat{e}}\left(\overset{\text{class i at time t}}{\widehat{C_i^t}} \to \overset{\text{class j at time t+1}}{\widehat{C_j^{t+1}}} \right)}{\text{total \#e}}, \quad (4.22)$$

where $(i,j) \in [\![1,5]\!]$

4.2.4 Stochastic conditional sampling procedure

4.2.4.1 Preprocessing methods

For simulation purpose, preprocessing steps are required in the case you need Independent Identically Distributed (IID) variables prior of applying the conditional sampling procedure. Two methods are considered to get Independent variables and Identically Distributed variables. Removing the trend in the variable time series provides Identically Distributed residuals whereas removing the autocorrelation generates Independent residuals. Combining these two methods with first removing the autocorrelation and the trend provides overall IID residuals.

Autocorrelation analysis provides information regarding the lag to consider and its significance through a lag-plot – see Section 4.1.1.2. Statistical indices related to autocorrelation are provided in Equation 4.1 and confidence interval formula for the lag-plot is defined in Equation 4.2. To remove an identified lag-k autocorrelation from a time series X containing n observations, a multiple linear regression model is built from the $k+1$ subsets created after splitting this time series – see Equation 4.23. Resolution of this equation provides an estimation of the parameters β_a for the autocorrelation model along with the independent residuals ϵ_a

$$\underbrace{\begin{bmatrix} x_{k+1} \\ x_{k+2} \\ \vdots \\ x_n \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & x_1 & x_2 & x_3 & \cdots & x_k \\ 1 & x_2 & x_3 & x_4 & \cdots & x_{k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-k} & x_{n-k+1} & x_{n-k+2} & \cdots & x_{n-1} \end{bmatrix}}_{X} \cdot \underbrace{\begin{bmatrix} \beta_{a0} \\ \beta_{a1} \\ \vdots \\ \beta_{ak} \end{bmatrix}}_{\beta_a} + \underbrace{\begin{bmatrix} \epsilon_{a1} \\ \epsilon_{a2} \\ \vdots \\ \epsilon_{an-k} \end{bmatrix}}_{\epsilon_a}, \tag{4.23}$$

The independent residuals can be reautocorrelate using the autocorrelation model parameters β_a and the initialization vector $[x_1, x_2, x_3, \dots, x_k]$ in the linear model for Y.

To obtain Identically Distributed residuals, it is necessary to remove the trend from the time series. For this purpose, we consider the residuals extracted from the linear trend (Equation 4.24) fitted to the time series using the least square method in Equation 4.3:

$$\epsilon_{tr} = X_{tr} - \underbrace{\beta_{\mathrm{tr0}} + \beta_{tr1}\mathrm{t}}_{\hat{X}_t}, \tag{4.24}$$

where X_{tr} is the considered time series presenting a trend, β_{tr} are the parameters of the linear trend fitted to X_{tr} and ϵ_{tr} are the detrended residuals. The trend in the time series can be retrieved from the identically distributed residuals ϵ_{tr} using the trend model parameters β_{tr} and ϵ_{tr}.

4.2.4.2 Conditional sampling procedure explanations

The conditional sampling procedure aims to simulate the evolution of the variable Y based on observation of X using the dependence between variables through the copula model. It is a cycle of calculation realized at each timestep e.g., repeated 1000 times, to provide a range of simulations for the predictand variable Y. The cycle is described below and refers to the Figure 4.7:

Consider the quantile of X at time step t.

(1) Calculate the associated percentile u^* (e.g, 0.7 in Figure 4.7 below) using the marginal distribution of X and definition of the intersection plan (red surface) passing through u^*.
(2) Extract the CDF of Y provided by the copula CDF surface estimated at u^* (red line).
(3) Pick a random percentile from this CDF (star dot).
(4) Estimate the corresponding quantile of Y using the inverse CDF of v^* (dashed red lines).

Repeating this cycle of calculation for each time step provides a simulation of Y over study period and repeating this procedure 1000 times generates a range of simulations of Y shown in Figure

B.21. Thus, for each simulation of Y, an interval of two standards deviations centered around the mean can be described to simulate the behaviour of the time series within its range of variations, defined by the maximum and minimum of amplitudes of simulations (Figure B.21).

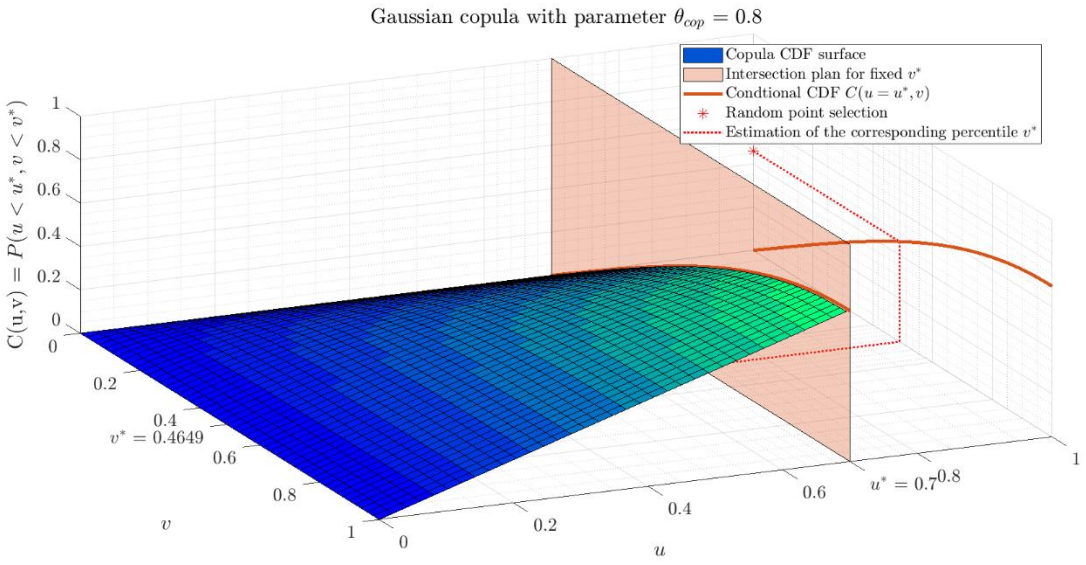

Figure 4.7 Graphical illustration of steps 1 to 4 in the cycle of the conditional sampling procedure with a percentile value of $u^* = 0.7$

4.3 Developed MATLAB toolboxes for analysis

The analyses previously described for univariate and bivariate time series are implemented in two separate software packages using MATLAB (2020a). Each package has a GUI that help the user obtain the results without diving through specific functions or scripts. These GUIs are built with specific buttons dedicated to a given interaction, such as pushbuttons to run a function, toggle buttons to select different settings, etc. By interacting with these interfaces, the user can run all the needed analysis and directly get the results through graphical visualization (with figures) or in text files. For more information related to constructing GUI, see Smith (2006). Apart from using the GUI, the function scripts are well documented and explained. Therefore, the functions in each software can be accessed and customized depending on users' projects.

4.3.1 Univariate analysis: Statistical Analysis Software

The “Statistical Analysis Software” is built around a control center which allows the user to load a file in MATLAB workspace, modify or save a downloaded dataset. A single panel that contains all the analysis possibly performed with this software is accessible via the pushbutton “Statistical Analysis” and allows perform all the analysis described in Section 4.1 – see Figure B.22. Modification of the dataset is visually assessed and the setup for operations allow selecting a timestep and a method for each variable (Figure 4.8). The main panel contains a set of seven analysis that allows the user estimating the autocorrelation, analyzing variability and the trend along with fitting parametric distributions and detecting change points in a considered time series. As an option, there is the possibility to create a customize configuration file where specific analysis can be applied to the selected variable. This software aims to provide a straightforward and user-friendly interface to analyze a univariate time series and obtaining graphical results that summarize the relevant information. More information about the software and a description of the codes used are available in the downloading platform GitHUB.

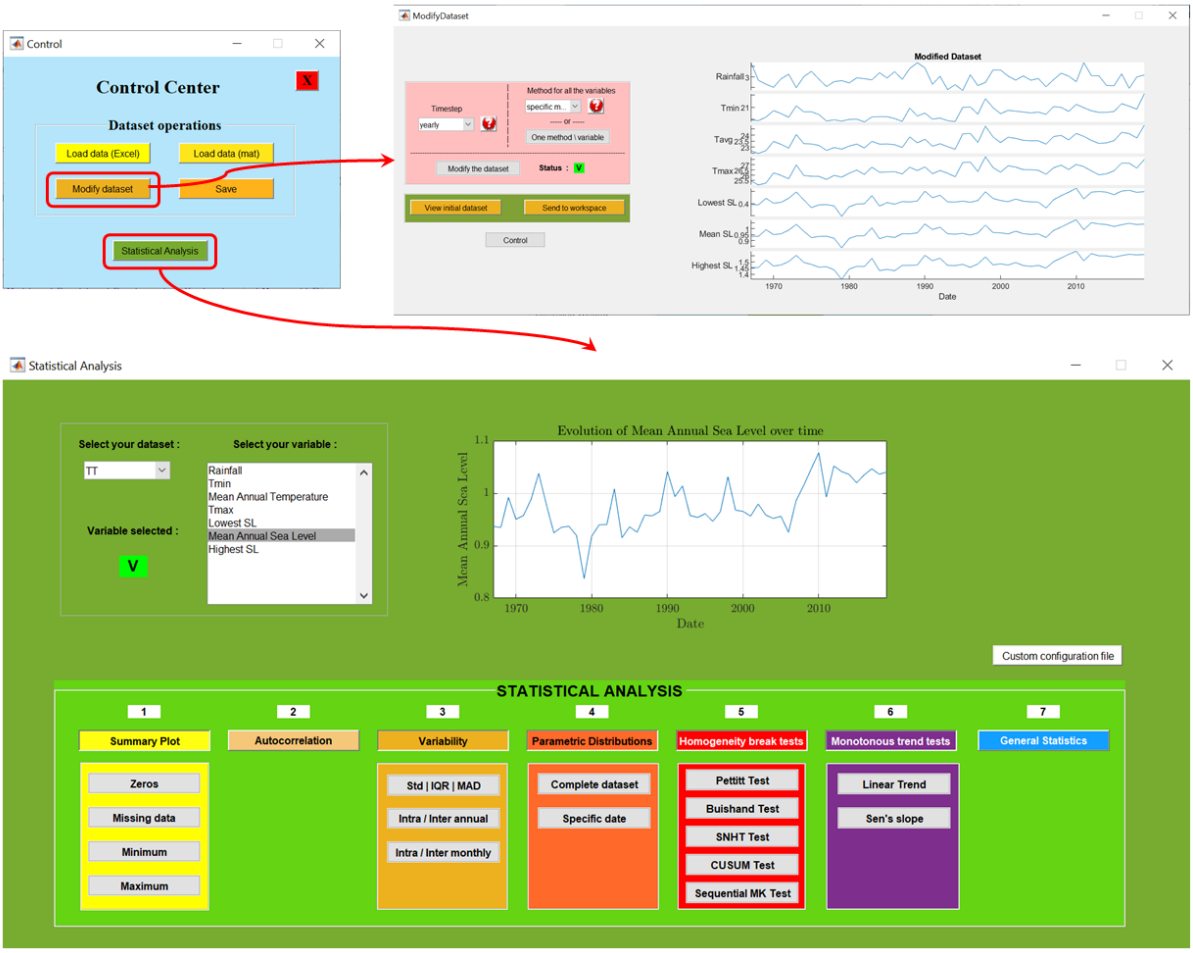

Figure 4.8 Developed graphical user interface for Statistical Analysis Software with Control Center (top left), panel for modification of the input variable (top right) and main panel with the set of available analyses (bottom)

4.3.2 Bivariate analysis: Copula Analysis Software

The Copula Analysis Software is also built around a control center, corresponding to the main interface, in which all the 4 analyses (calibration / validation of the copula model, confirmation using a bootstrap approach, simulation through conditional sampling and risk assessment) are available (Figure B.23). The top left panel named “Copula” focuses on selecting marginal CDFs and copulas using different parameter estimation methods described in the Section 4.2.1 along with complementary analysis for estimation of Kendall’s τ. The bottom left panel named “Bootstrap” can be used to verify and confirm the choice of the marginal CDFs and copula – see Section 4.2.2.

The bottom right and top right panels respectively “Risk” and “Sampling” are analysis for risk assessment and stochastic sampling of the predictand variable Y (Section 4.2.3 and 4.2.4) that consider the copula model selected with the two previous analysis panel. Besides the 4 pushbuttons related to bivariate analysis, 2 pushbuttons named “Load data” and “Save” can be used to respectively load a dataset and save the current dataset of observations used in the workspace as a new dataset that will be stored in the folder “Data”. Moreover, in each panel, a pushbutton “Control” allows the user to directly return to the “Control Center” and closing the current window after finishing the analysis (Figure 4.9). This software also aims to present a simple and straightforward interface that can be easily used and understood by the user in the context of bivariate analysis. More information about the software and a description of the codes used are also available in the platform GitHUB.

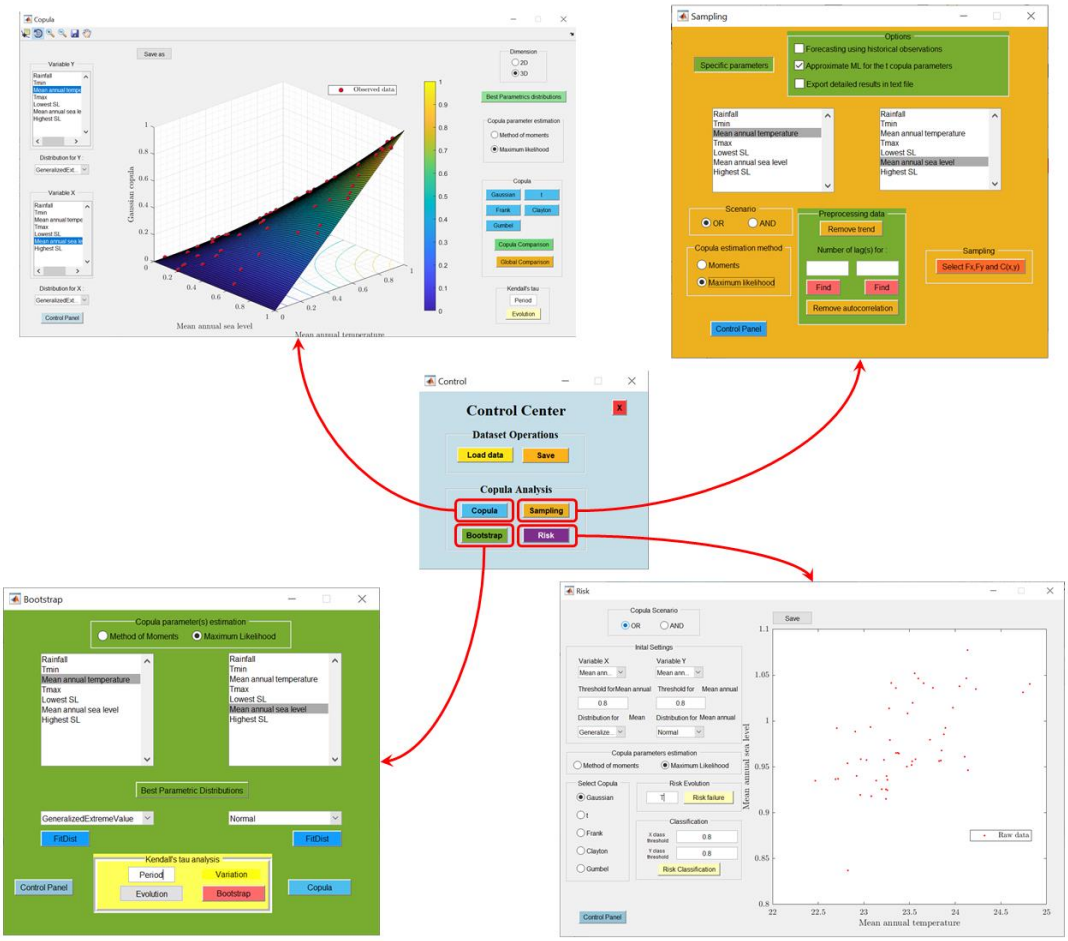

Figure 4.9 Developed graphic user interface for Copula Analysis Software to develop the copula model (top left), confirm the choice of copula (bottom left), simulate a variable (top right) and perform a risk assessment (bottom right)

CHAPTER 5 ARTICLE 1: A LOCALLY RELEVANT FRAMEWORK FOR ASSESSING THE RISK OF SEA LEVEL RISE UNDER CHANGING CONDITIONS: APPLICATION IN NEW CALEDONIA, PACIFIC OCEAN

Matheo Kaemo[1], Elmira Hassanzadeh[1*], Ali Nazemi[2]

[1]*Department of Civil, Geological, and Mining Engineering, Polytechnique Montréal, Montréal, QC, Canada*

[2]*Department of Building, Civil, and Environmental Engineering, Concordia University, Montréal, Canada*

[*]Corresponding author: elmira.hassanzadeh@polymtl.ca

This chapter was submitted to the journal *Science of the Total Environment.*

CONTRIBUTION OF THE MASTER STUDENT

The master student obtained the climate data from the Meteo France institute, performed the analysis and graphical visualization of the results along with writing the draft of the article. Dr Elmira Hassanzadeh and Dr Alireza Nazemi proposed the research methodology. All co-authors participate to the validation of the results and editing and reviewing the manuscript.

ABSTRACT

Sea level rise is a key consequence of warming climate and can be seen across coastal regions globally. At this juncture, assessing climate change-induced sea level rise is urgently needed, particularly in island nations, where the threats of sea level rise are immediate and there is a high level of unpreparedness in dealing with its adverse impacts. Here, we propose a stochastic simulator to link changes in Mean Annual Temperature (MAT) to Mean Annual Sea Level (MASEL) at the local scale to facilitate accessibility of the models to stakeholders and to provide a basis for bottom-up impact assessment through what-if scenarios, developed based on the linkage between local temperature and sea level. For this purpose, we decompose the MAT and MASEL signals into their linear trend and autocorrelation components as well as independent and identically distributed residual terms. We further explore interdependencies between trend and residual terms of MAT

and MASEL. If such dependencies exist, scenarios of sea level can be synthesized based on the trend and residual terms of temperature. We use linear regression to link trends of MAT and MASEL, and bivariate copulas to formulate joint dependencies between their residuals. This allows stochastic sampling of MASEL conditioned to trend and random variability in MAT. We showcase the application of this framework by retrospective and prospective simulations of MASEL in Nouméa, the capital city of New Caledonia in Pacific Ocean. We set up six different model configurations for developing the stochastic sampler, each including various parametric options. By selecting the best setup from each configuration, we provide a multi-model stochastic projection of MASEL, assuming the persistence in current long-term trend in MAT and MASEL. We demonstrate how such simulations can be used for a bottom-up and risk-based impact assessments and discuss sources of uncertainty in future projections of MASEL.

Keywords: Sea level rise; Bottom-up approach; Copulas; Conditional sampling; Risk-based impact assessment; New Caledonia

5.1 Introduction

Coastal regions have been historically hubs for human development and socio-economic activities. Currently around 30% of the world's population resides in coastal areas, and 80% of world trade is transported from and to coasts (Sirimanne et al., 2019). Coastal regions are also key to food security. On the one hand, shorelines' neighboring lowland deltas delineate some of the world's most fertile lands, supplying calorie requirements of highly populated areas, such as southeastern Asia. On the other hand, coasts are home to rich aquatic ecosystems and are the main contributors to fisheries globally (Friess et al., 2019). Concentration of people, food and human activities create cultural and historical landmarks that are cherished by the community, and bring assets, infrastructures, businesses, and indeed geopolitical interests in the shorelines.

The other side of the coastlines' glory is vulnerability, particularly when one of the fundamental characteristics of coastal regions, i.e., the sea level, is subject to irreversible changes. Water level fluctuations are in fact among the main features in seas and oceans, caused by tides, wind and other atmospheric drivers, along with tsunamis, volcanic activities, earth astronomical cycle, and intense solar radiation (Joussaume, 1993; Nesme-Ribes et al., 2000; Jacques and Le Treut, 2004; Blumberg and Bruno, 2018). Melting ice sheets and warming oceans due to the

anthropogenic climate change, however, have caused irreversible increases in sea levels globally (Hansen et al., 2010, Churchill et al., 2019; Nordell, 2003; NOAA, 2021; J. Chen et al., 2013). Gradual increase in average sea level also intensifies the magnitude and frequency of maximum annual water level as it is well-documented in some data rich regions such as New York (Talke et al., 2014; Sweet et al., 2017). This can consequently result in coastal flooding and nibbling as well as deterioration of infrastructure (Mondal, 2013; Anderson et al., 2017). In addition, inundation during high tides and storm surges degrades agriculture along shorelines, contaminates drinking water supplies and other land resources. Continuation of sea level rise will inevitably result into mass migration: Even a modest Sea Level Rise (SLR) of half meter could potentially threaten the 10% of the world's population living within 10 meter above the sea level (Pilkey et al., 2016).

Vulnerabilities associated with the SLR are manifested differently in developed and developing countries. In developed countries, coasts are filled with expensive properties and assets. Organization for Economic Cooperation and Development reports that around 60% of the world's assets that are exposed to SLR are concentrated in three developed and wealthy countries, the US, Japan, and the Netherlands (Pilkey et al., 2016). This leads into a situation in which only a slight change from "normal" conditions can lead in to devastating financial consequences. However, due to concentration of socio-economic activities, abandoning shorelines are extremely unpopular for taxpayers. The nature of vulnerability is largely different in developing countries, where SLR is also a trigger for some other sources of vulnerability such as poverty, pollution, and overpopulation (Wilby, 2017). While in these regions, SLR can endanger health and livelihood of people, yet the population is the least prepared to address or even assess the risks of SLR. Such circumstances can lead into major devastations, including large mortality in vulnerable population, and raise some profound ethical concerns due to inequality in a highly unjust social context.

Assessing vulnerabilities of coastal regions to SLR requires reliable mathematical models, with which the dynamics of sea level under changing climatic can be quantified in time and space. The most widely used approach to project SLR in the context of climate change is through the use of Earth System models that represent the physical principles, e.g., conservation of mass, energy and momentum, between elements of land, ocean and atmosphere (Wada et al., 2010; Church et al., 2013a; Slangen et al., 2017). However, due to limitations in representing complex processes such as penetration of heat into deep ocean, glacial dynamics as well as energy and water exchange

between ocean, land and atmosphere, there are large uncertainties in the sea level projections (Rahmstorf, 2007). In addition, Earth System models are relevant at large scales; and therefore, they are not able to accurately describe the dynamics of sea level at smaller scales, where decision-making takes place. Thus, the consideration of downscaling the observations at a designated location for local analysis is necessary. However, lack of resources in developing countries, where such projections are urgently needed, prevents the application of these models in real-world impact assessments in some of the most vulnerable parts of the globe.

Changes in the sea level can be also estimated using data-driven models that are forced by future climate projections (Rahmstorf, 2010; Bittermann et al., 2013; Leta et al., 2018) as well as astronomical tides (Walsh et al., 2012; G. Salvadori et al., 2016), and/or storm surges (Khanal et al., 2019). These models are particularly suitable for impact assessment at the local scale and can range from simple trend models (Chatfield, 2000; Hess et al., 2001; Kozłowski et al., 2018) to complicated statistical and machine learning models (Vermeer and Rahmstorf, 2009; Niedzielski et al., 2009; Imani et al., 2014, 2019; Ardabili et al., 2019) and can be further linked to downscaled global projection of sea level and climate for future projections. Regardless of the straightforwardness and reasonable accuracy of these models during historical periods, like any other top-down scenario-led impact assessment (see IPCC, 2014), their future estimates are subject to large uncertainty due to the inherent uncertainty in climate projections (Nazemi and Wheater, 2014; Nazemi et al., 2020).

From a broader perspective, vulnerability to changing climate should be assessed through models that are not only accurate, but also accessible. Having an impact model, accessible to a wide range of stakeholder, can lead to collaborations on regional adaptation plans (Hassanzadeh et al., 2019; Zammali et al., 2021). This has resulted into the advent of bottom-up assessment schemes that are scenario-neutral and evaluate the impact of changing climate through a set of what-if narratives that can be derived from many sources, including future climate projections, perturbations of historical climate conditions and/or the local knowledge. Such bottom-up impact assessments have been used frequently in the context of water resource management (Hassanzadeh et al., 2016; Zandmoghaddam et al., 2019) and have been recently used in the context of energy production (Amir Jabbari and Nazemi, 2019) as well as assessing the impact of climatic changes on freeze-thaw (Hatami and Nazemi, 2021). Within this context, having stochastic impact models

are highly favorable as they provide a basis for risk-based assessments (Hassanzadeh et al., 2016, 2019).

Stochastic approaches to study SLR have started to appear in the literature, however, they rather considered independence between at least some of the involving variables (see e.g. Bevacqua et al., 2019) and/or did not consider the nonlinear dependence structure between sea level and hydroclimate variables (Moftakhari et al., 2017). Here we propose the use of copulas, a generic framework to represent interdependence (see Genest and Favre, 2007), to condition SLR to changing temperature. Copulas have been successfully used in various field from finance (Cherubini et al., 2004; Patton, 2009, D'Amico and Petroni, 2018, Frees and Wang, 2006) to engineering (Y. Zhang et al., 2020; Liu and Fan, 2020) to hydrology and climatology (Yang et al., 2019; Nazemi and Elshorbagy, 2012; Aghakouchak, 2014). The main advantage of copulas is in the facts that modeling joint and marginal characteristics can be separate, and no presumption should be made in the nature of dependencies (Embrechts et al., 2001; Nelsen, 2006; Sadegh et al., 2018; Anderson et al., 2019; Naz et al., 2020). Application of copulas to sea level predictions has been reported in the recent work of Yavuzdoğan and Tanır Kayıkçı (2020), who used copulas as a autoregressive timeseries model to predict sea level anomalies in the Black Sea; however to the best of our knowledge, copulas have not yet been applied to describe the linkage between the sea level and climate variables.

The objective of this study is to simulate the evolution of sea level at the local scale using copulas, in which the anomalies in sea level are described based on the anomalies in local temperature. We do recognize that the causes of local SLR can go well beyond the local scale; however, building models that describe a global phenomenon using local information that are meaningful to stakeholders can enhance accessibility, trust, and uptake of adaptation strategies (Stedman et al., 2004). This is particularly relevant to communities that are under immediate threats with critical consequences that require immediate actions. One of such regions are small islands of Pacific that have experienced simultaneous increases in temperature and sea level (Manton et al., 2001; Church et al., 2006; Becker et al., 2012; X. Zhang and Church, 2012; Nunn, 2013). The 2014 World Health Organization's report to IPCC mentions that due to SLR the livelihood of this regions will be disrupted by death, injury and illness (Pilkey et al., 2016). In New Caledonia, our case study, the positive trend in temperature (0.25°C/decade during 1961-2011; Whan et al., 2014) has

coincided with a positive trend of SLR (3.2 mm/year in the period 1993-2009; Church and White, 2011). This not only threatens infrastructure and socio-economic activities, but also endangers particular way of life in local Indigenous tribes and can cause disputes between traditional societies (Ash et al., 2016).

While building our stochastic impact model, we pay a particular attention to satisfy the underlying requirements of copula modeling. In particular, copula modeling is built upon the use of random variables that are independent and identically distributed (iid; see Bedoui and Dbabis, 2008; Bücher and Volgushev, 2013), which is not the case in a warming world, in which the sea level increases temperature. As a result, in line with some recent recommendations (Q. Zhang et al., 2013; Wani et al., 2019; Tootoonchi et al., 2020), we attempt to remove the trend and autocorrelation in the timeseries of temperature and sea level, in case they are found significant. The rest of the paper is organized as the following: Section 5.2 presents the methodological elements of our proposed model. Section 5.3 introduces our case study and available data. We configure our experimental setup in Section 5.4. Results are presented in Section 5.5 and are further discussed and implemented for impact assessment in Section 5.6. Finally, Section 5.7 concludes this study and provides some further remarks.

5.2 Methodology

Our specific objective is to build a stochastic model to represent the link between the Mean Annual SEa Level (MASEL) and the Mean Annual Temperature (MAT) and to assess the risk of SLR under changing temperature conditions. We choose copulas as the core sampling algorithm behind our stochastic sampler; yet we put in perspective that copulas are suitable for describing the interdependencies between *iid* random variables. To ensure the existence of iid condition, we analyze the trend and autocorrelation and remove them, if needed (Şen, 2012; Mudelsee, 2019; Lomnicki and Zaremba, 1957; Abdulhafedh, 2017), and use copulas to describe the dependence between detrended and de-autocorrelated residuals of MASEL and MAT (Hofert, 2012). By understanding the link between trends of MASEL and MAT, as well as extending or perturbing the joint dependencies between historical trends and/or dependence structure between MASEL and MAT, MASEL can be projected by conditional sampling obtained through copulas (see Wang et al., 2019). Below, we describe the methodological elements of this framework in more details.

5.2.1 Copulas

Copulas aim at providing a flexible framework for representing the joint dependencies between random variables. In brief, for two continuous iid random variables X and Y with marginal Cumulative Density Functions (CDFs) of u and v, the joint probability $F_{XY}(x, y)$ can be presented by a copula as $F_{XY}(x, y) = C(u, v)$ – see the proof given by Sklar (1959). As a result, the representation of marginal and joint distributions becomes mutually independent. This adds tremendously to the flexibility of copulas (Jaworski et al., 2010; Bedoui and Dbabis, 2008; Genest et al., 2009). Commonly used copulas can be classified into two different families of Archimedean (e.g., Frank, Gumbel, Clayton) and Metaelliptical (e.g., Gaussian, t copula) – see Nelsen (2006) for a detailed review of copula structures and their formulations.

To measure the dependence between MAT and MASEL, we use a nonparametric rank correlation coefficient measure, namely Kendall's τ (Kendall, 1938; Genest and Favre, 2007; Bonett and Wright, 2000). In contrast to Pearson's correlation coefficient that aims to describe linear linkages, Kendall's τ aims at quantifying whether observations have similar ranks. Kendall's τ can be formulated as the following:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} sign(x_i - x_j).sign(y_i - y_j), \tag{5.1}$$

where n is the length of dataset, i, j and k are moving indices of the sample in X and Y, respectively. τ ranges between -1 and +1: Kendall's τ values of +1 and -1 indicates that observations have fully the same or fully opposite ranks, and τ=0 reveals no association between the observations (Brossart, 2018). The significance of Kendall's τ can be estimated by a p-value.

5.2.2 Analysis and removal of trend

Non-parametric Mann-Kendall test (Mann, 1945; M. Kendall, 1976) is commonly used to detect trends and their significance in various hydroclimatic variables (Manton et al., 2001; Whan et al., 2014; Nazemi et al., 2017; Jehanzaib and Kim, 2020) and employed here to analyze the MAT and MASEL time series. In contrast to multiple regression, Man-Kendall test is a rank-based method; therefore, it is not sensitive to outliers.

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign\left(x_j - x_k\right), \tag{5.2}$$

where x is the considered time series of length n. i, j and k are moving indices of the sample in x. S is the statistical value of Mann-Kendall test. We use the robust rank-based Sen slope estimate (Sen, 1968) to quantify the magnitude of trend:

$$a = \text{median}\left(\frac{x_k - x_i}{k - i}\right), i < k, \tag{5.3}$$

In the case of significant trends in MAT and/or MASEL, the detrending should be made. Figure 5.1 shows a schematic illustration of detrending. In brief, least-square method is applied to estimate trended values $\hat{X}_t$ by finding the intercept β_{t0} and slope β_{t1} parameters of the regression model for the observed sample X_t. Accordingly, the detrended residuals ϵ_t can be also retrieved.

$$X_t = \underbrace{\beta_{t0} + \beta_{t1} t}_{\hat{X}_t} + \epsilon_t \,, \tag{5.4}$$

$$\epsilon_t = X_t - \widehat{X_t} \,, \tag{5.5}$$

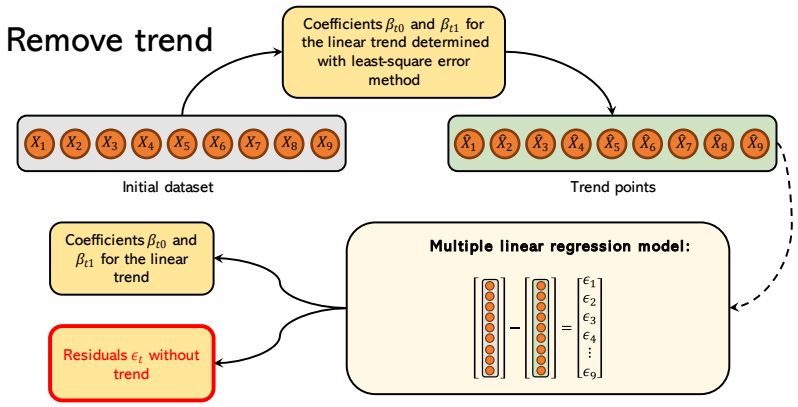

Figure 5.1 The procedure of trend removal from a given dataset using the least-square method

5.2.3 Analysis and removal of autocorrelation

Autocorrelation is common in hydroclimate data (Yue et al., 2002; X. L. Wang, 2008). Autocorrelation within random variables can be formally investigated by Pearson's ρ correlation coefficient at any lag $k < n - 2$ and n is the number of samples. We implement the commonly used confidence interval of normal serial correlation at a particular significance level to identify the number of lags with significant autocorrelation.

$$CI = \pm \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \tag{5.6}$$

where z is the cumulative distribution function of the standard normal distribution, and α is the significance level. If the autocorrelation is significant, then we use a multiple linear regression model to remove the autocorrelation. Accordingly, for a given series with a lag-k autocorrelation, the dataset is split into $k + 1$ subsets to create the predictor matrix X and the target vector Y as shown in Equation 7. Figure 5.2 depicts a schematic illustration of de-autocorrelating process for removals of lag-1 and -2 autocorrelations. By using the least square estimator, parameters β_a for the autocorrelation model and associated residuals ϵ_a can be found.

$$\underbrace{\begin{bmatrix} x_{k+1} \\ x_{k+2} \\ \vdots \\ x_n \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & x_1 & x_2 & x_3 & \cdots & x_k \\ 1 & x_2 & x_3 & x_4 & \cdots & x_{k+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n-k} & x_{n-k+1} & x_{n-k+2} & \cdots & x_{n-1} \end{bmatrix}}_{X} \cdot \underbrace{\begin{bmatrix} \beta_{a0} \\ \beta_{a1} \\ \vdots \\ \beta_{ak} \end{bmatrix}}_{\beta_a} + \underbrace{\begin{bmatrix} \epsilon_{a1} \\ \epsilon_{a2} \\ \vdots \\ \epsilon_{an-k} \end{bmatrix}}_{\epsilon_a}, \tag{5.7}$$

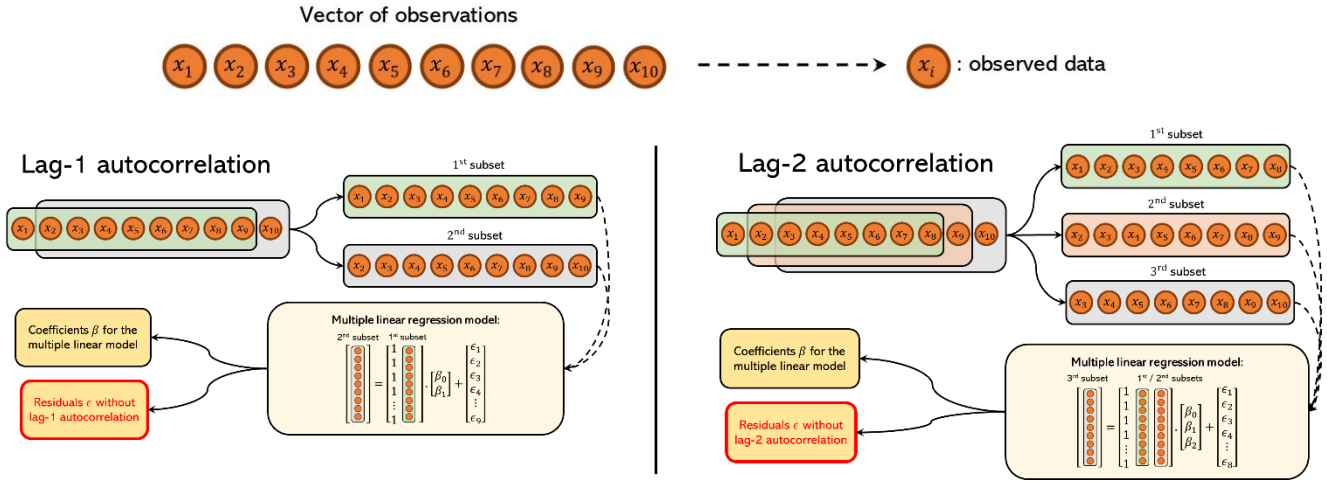

Figure 5.2 A schematic illustration for removing lag-1 and -2 autocorrelation from a dataset using a multiple linear regression.

5.2.4 Stochastic sampling approach

If the joint dependence between two random variables along with their marginal distributions are known, one variable can be sampled based on the other one (see Hatami and Nazemi, 2021). For the case of MAT and MASEL, and if the trend and autocorrelation are significant, the data should be first filtered; and, the remaining *iid* residuals of MAT and MASEL can be used for fitting a parametric copula. The fitted parametric copula can be then used for conditional sampling of MASEL residuals, using the known residuals of MAT at every timestep. By adding the historical trend and autocorrelation in the randomly reconstructed residuals of MASEL, a stochastic realization for the actual timeseries of MASEL under retrospective conditions can be generated. This procedure can be repeated for a number of times (1000 times in this study) to provide an estimation for the CDF of MASEL, given a known value of MAT – see Figure 5.3 below.

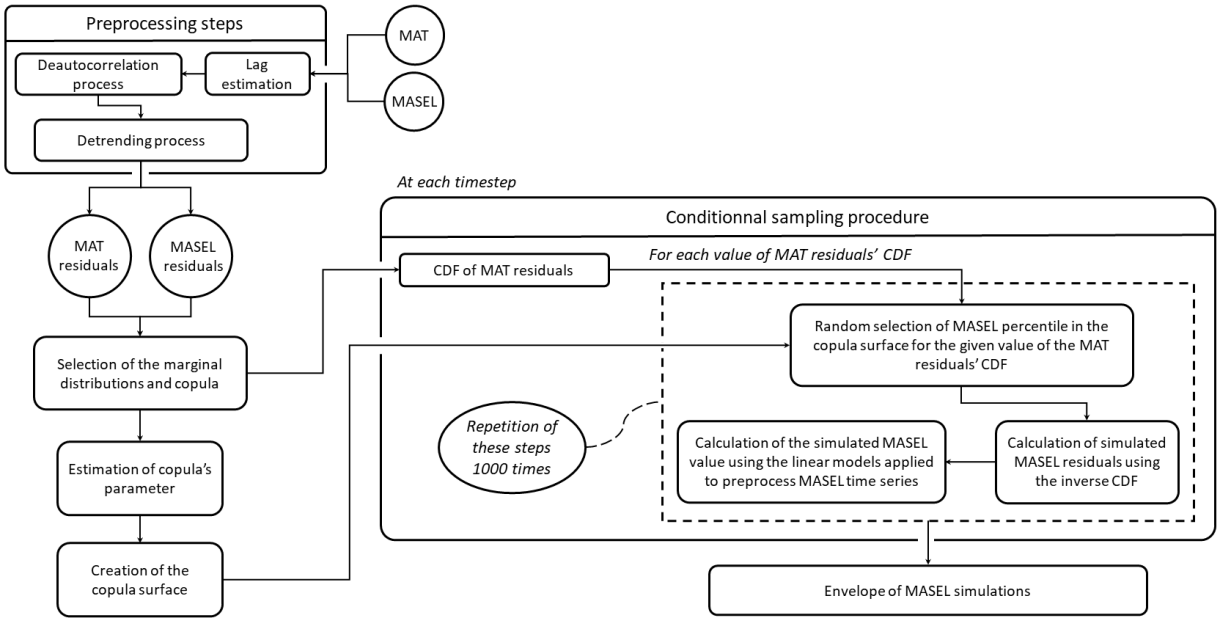

Figure 5.3 A schematic view at the proposed stochastic sampling framework for estimating MASEL based on MAT in the retrospective mode

The same sampling approach can be used in the projection mode to estimate future MASEL based on scenarios of trend in MAT as well as joint dependency between MASEL and MAT residuals. Scenarios of trend in MAT can be made based on available downscaled projections of local MAT obtained from regional and/or global models, past observations of historical trends, and/or stakeholders' assumptions about future changes in temperature. Accordingly, the trend in MASEL should be estimated based on the trend in MAT. For this, a moving window methodology similar to what proposed in Nazemi et al. (2017) and further used in Hatami et al. (2019) and Zaerpour et al. (2021) is utilized. In brief, a moving window with a specific size can be considered and sled over the MAT and MASEL residuals year-by-year. Accordingly, the trend in MAT and MASEL in each window can be estimated and the dependence between corresponding trends in MASEL and MAT, obtained from multiple windows of data, can be analyzed. If the dependence is found significant, then a regression model can link the trend in MASEL to the trend in MAT. By assuming that marginal and joint CDFs of MASEL and MAT remains unchanged, residuals of MASEL can be conditionally reconstructed from random samples for residual of MAT. From these residuals, then, the projected future realizations for MASEL can be inferred based on the considered scenarios

of trend in MAT (from which the trend in MASEL is estimated), and the autocorrelation in MASEL. The inferred probability distribution can be used for a risk-based impact assessment. We show an example of such a bottom-up approach in our considered case study.

5.3 Case study and available data

New Caledonia (In French: Nouvelle-Calédonie) is a small French-speaking archipelago in the southeast of Pacific Ocean. With a total area of around 18,000 km² that includes 3,400 km of shorelines, New Caledonia is home to nearly 300,000 inhabitants that are unequally distributed over the land. The majority of the population resides in the main island of "Grande Terre", which includes the capital city of Nouméa that is home to almost 30% of the total population. A mountain range passes through the Grande Terre and splits the island in to the west and east sides. West coast of the Grande Terre is characterized by large flat lands, while east coast includes rocky and steep shorelines. The Grande Terre's lagoon is a UNESCO World Heritage site (UNESCO, 2021) and is surrounded by a 1600 km coral reef, which is considered as one of the 25 global hotspots of biodiversity. New Caledonia contains approximately one fourth of global nickel resources, making mining the main driver of the economy. After mining, merchant and transportation form ~46% of the New Caledonian GDP, most of which is passing through Nouméa's port (Bunel, 2017). New Caledonia has a tropical climate, characterized by a wet season from May to September and a dry season from November to February. The long-term average annual temperature is around 24°C and total precipitation ranges between 1000 mm and 1800 mm per year (MétéoFrance, 2021). The dry season is the time for major cyclonic events that can cause severe damages in the east coast due to floods and landslides.

Inhabitants of New Caledonia include indigenous "Kanak" people that are spread mainly in the East coast of Grande Terre and Loyalty islands. The link to the land is critical for Kanak people and their special way of life. European settlers include (1) the descendants of former exiled prisoners that are mainly spread in the west coast, raise livestock, and have a similar way of life as ranchers in the western US and Canada; and (2) free settlers arrived in the island during the French colonial government of the 3rd Republic to develop the mining activity and live mainly in Nouméa. They have the same way of life as modern French society. New Caledonia also includes an important community of Polynesian immigrants who came mainly for work. The Polynesian

community also has a strong connection to the land, but their traditions and ways of life are different from Kanaks.

It is not a surprise that New Caledonia is socially and politically divided. The island suffers from the bitter memories of past colonization that reduced the population of Kanak people quite significantly, as well as heated political debates between separatists and loyalists, who have very different views regarding the economy, environment, and socio-economic activities. This socio-political landscape majorly complicates decision making particularly around climate change adaptation. Although SLR can have multiple negative impacts on New Caledonia (Le Cozannet et al., 2014; Duvat, 2019; Le Duff et al., 2020), so far SLR has been seen as local issue relevant to few Indigenous tribes in the shoreline, particularly in the islands surrounding the "Grande Terre" and not a threat that can endanger the faith of the island and its people. Bellard et al., (2013) estimate that an increase of sea level from 1 m to 3 m can lead to 5% to 10% loss of the "Grande Terre" due to erosion. Land erosion and coastal flooding can be a major threat to the livelihood and assets of people who just live a few meters away from the sea (Lincke et al., 2021). Continuation of SLR and loss of shorelines can force communities to relocate, which for the case of Indigenous community may lead to territorial disputes (Nunn, 2013). SLR can also deteriorate functioning of the port and cause disruptions in trading activities. Frequent flooding and increasing soil salinity can be also a danger to the fragile coastal ecosystem (Bellard et al., 2013; Ellison, 1993; Lovelock et al., 2015).

Different methodologies are used to estimate the SLR in the Pacific Islands. Among these, analyzing the linear trend is common to characterize the changes in the sea level (X. Zhang and Church, 2012; Hamlington et al., 2021), temperature (Mote and Salathé, 2010) and sea-surface pressure (Walsh et al., 2012). For instance, a 20 cm sea level rise by 2100 is estimated by extrapolating MASEL records since 1900 (Parris et al., 2012). Other approaches such as Empirical Orthogonal Functions are also used to quantify the signal of change in sea level data, obtained from TOPEX/Poseidon satellite (Church et al., 2004; Nerem et al., 2018). Advanced semi-empirical methods are also used to estimate MASEL in the Pacific Islands as a function of GCM-based temperature, showing an overall increase of 71 cm in some regions by 2100 (Horton et al., 2008). It should be however noted that there is a wide discrepancy in currently available estimations and large uncertainties exist in future projections (Zhao et al., 2019). The propose stochastic approach

that inherently results in a range of estimations is therefore favourable. In addition, as local data are used as a basis of impact assessment, it can develop a sense of belongingness and trust between stakeholders, which can increase the chance of uptaking adptation strategies.

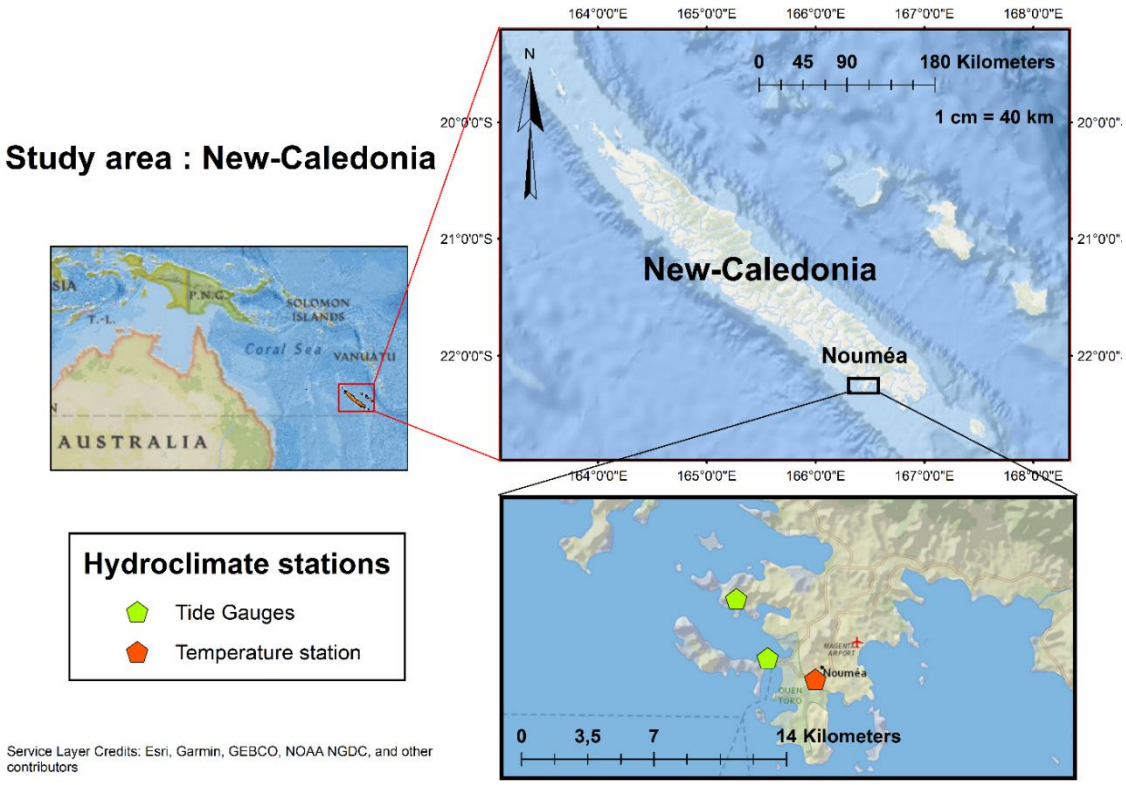

Figure 5.4 The map of New Caledonia, including the considered in-situ gauging stations used for developing the copula-based impact assessment model.

Daily temperature and sea level data at the capital city of Nouméa from 1967 to 2019 are obtained from MétéoFrance (http://www.meteo.nc/nouvelle-caledonie/climat/releves) and the reference network for tide gauge observations (http://refmar.shom.fr/fr/partenaires/producteurs-de-donnees/reseau-maregraphique-de-nouvelle-caledonie) respectively, from which MAT and MASEL data are derived, that show significant positive dependence with a Kendall's τ of 0.42. Figure 5.5 shows the data, in which the top row displays the MAT and MASEL series, their trends and associated *p*-values. Our analyses based on the Mann-Kendall test and Sen's slope show that there is a significant positive trend in the MAT (~ 2°C/century) and the MASEL (~ 1.75 mm/year)

in Nouméa during the study period. In addition, both MAT and MASEL have respectively lag-3 and lag-4 autocorrelations, respectively – see Figure 5.5 bottom row.

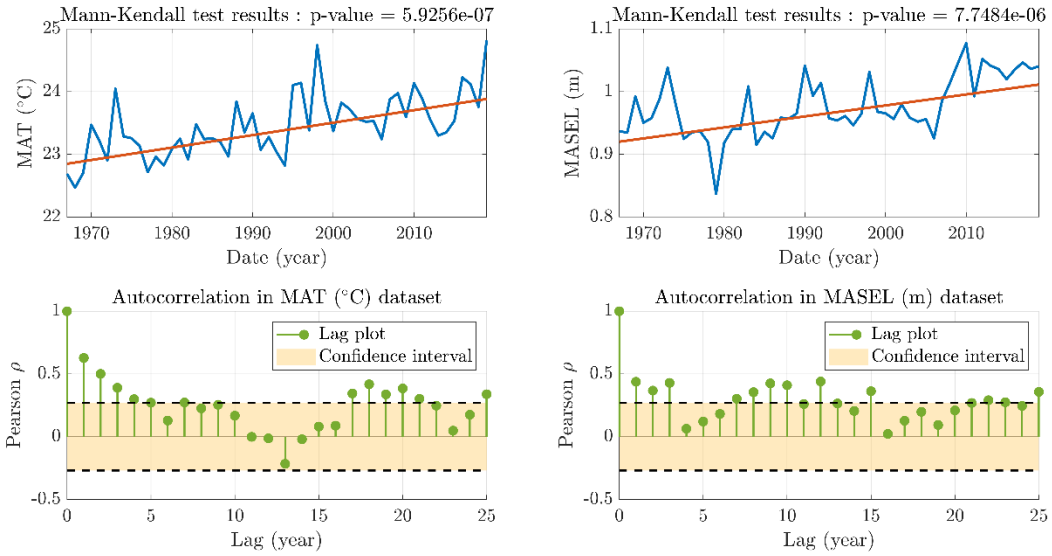

Figure 5.5 Results of trend (top) and autocorrelation (bottom) analyses in MAT (left) and MASEL (right) data at New Caledonia.

The results obtained by the analyses of trend and autocorrelation in Figure 5.5 necessitate de-trending and de-autocorrelating MAT and MASEL data and implementing copulas to represent the dependence between MAT and MASEL residuals. Before explaining how copula models are configured and set up, here we briefly look at the linkage between trends in MAT and MASEL using the moving window methodology. Figure 5.6 summarizes this analysis, where in each panel the dependence between trend in MAT (x axis) and trend in MASEL (y-axis) is shown at a particular window size (from 10 years in top left to 35 years in bottom right). In all panels, empty dots show the magnitudes of trends between MAT and MASEL in a particular time episode and the blue dot shows the trends observed from 1967-2019 and shown in Figure 5.5. Our analyses show that in all window sizes the dependence between MAT and MASEL is revealed by a pair of linear equations that describes the dependence between MAT and MASEL below (blue fitted line) and above (orange fitted line) the median of MAT, identified in each panel by a vertical dashed line. In majority of multi-decadal moving windows, the linear dependence between MAT and MASEL is strongly significant (p-value ≤ 0.1) and shows negative and positive relationships between local and MAT and MASEL below and above the median of MAT, respectively. We

emphasize however that this dependence should not be mistaken by causality as the cause of trend in SLR in Nouméa is well beyond the trend in local temperature; however, we argue that the trend in local temperature itself contains the information about the global cause of SLR.

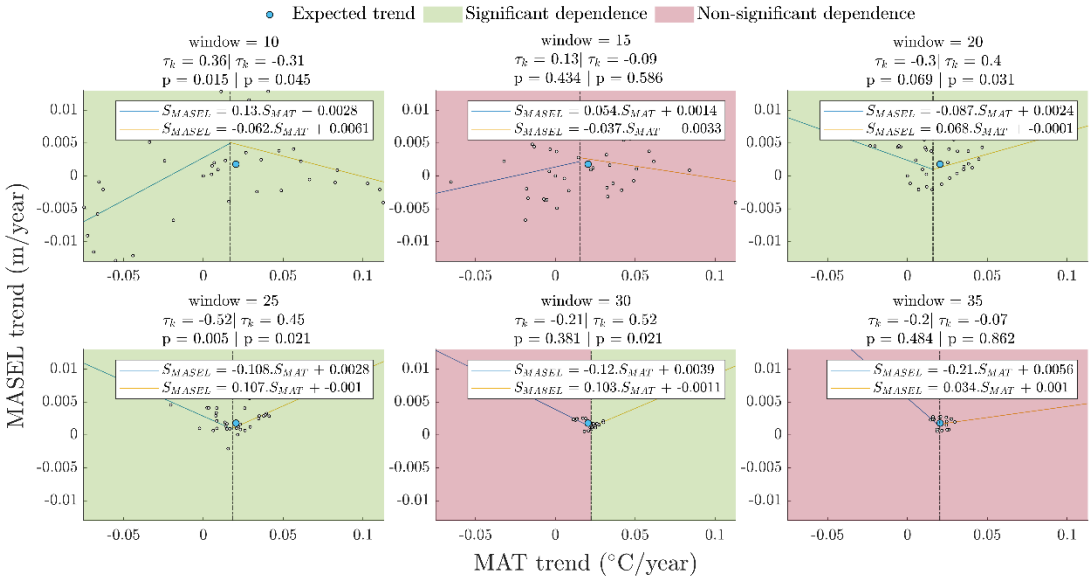

Figure 5.6 The dependence between trend in MAT and MASEL described as a pair of linear equations below and above the median of MAT across a range of window sizes. Strongly significant dependencies between trends in MAT and MASEL are shown with red color.

5.4 Model configuration and experimental setup

The skill of copula models is highly sensitive to the choice and parameterization of marginal and joint probability distributions (Wei et al., 2014; Ko et al., 2019). Similar to Zandmoghaddam et al., (2019), we take a reductionist approach by developing a wide range of modeling hypotheses using different configurations of marginal and copula distributions. After parameterization, a formal resampling study is used to falsify developed modeling hypotheses to a level that no other model can be falsified based on the available information. This results into a set of non-falsified models that can be then used as a basis for the projections of MASEL and impact assessment. For marginal distributions, we consider the Gamma, Extreme Value, Generalized Extreme Value (GEV), Normal, Rician, Nakagami, Generalized Pareto, Loglogistic, Lognormal, Weibull and t location scale distributions. The formulations and parametrizations of these distributions are presented in Table 1. We only use Maximum Likelihood Method (MLM) to find the parameters of marginal

distributions and their associated confidence intervals (see Nguyen et al., 2020). Here, we use Bayesian Information Criterion (BIC; Neath and Cavanaugh, 2012) and Akaike Information Criterion (AIC, Burnham and Anderson, 2002) as the Goodness-of-Fit (GoF) for likelihood, with which we compare and falsify marginal distributions and copula models. BIC emphasizes on finding distributions (or copula models) with the highest accuracy and the lowest number of parameters. AIC only considers the accuracy and do not penalize the number of parameters. BIC and AIC can be estimated as:

$$\mathrm{BIC} = -2.LL + k.\ln(n), \tag{5.8}$$

$$\mathrm{AIC} = 2.\mathrm{k} - 2.LL, \tag{5.9}$$

where k and n are the number of model parameter(s) and data points, and LL represents the loglikelihood value of the parametric distribution and/or copula.

Table 5.1 Parametric probability distributions considered for marginal distributions and their associate parameterizations

Number	Distribution name	Probability density function	Parameter(s)
1	Gamma	$\frac{1}{b^a\Gamma(a)}x^{a-1}e^{-\frac{x}{b}}$	a : shape parameter b : scale parameter
2	Extreme Value	$\frac{1}{\sigma}\exp\left(\frac{x-\mu}{\sigma}\right)\exp\left(-\exp\left(\frac{x-\mu}{\sigma}\right)\right)$	μ : location parameter σ : scale parameter
3	Generalized Extreme Value	$\frac{1}{\sigma}\exp\left(-\left(1+k\frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}}\right)\left(1+k\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{k}}$	k : shape parameter σ : scale parameter μ : location parameter
4	Normal	$\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	μ : mean σ : standard deviation
5	Rician	$I_0\left(\frac{x.s}{\sigma^2}\right).\frac{x}{\sigma^2}e^{-\frac{x^2+s^2}{2.\sigma^2}}$	s : noncentrality parameter σ : scale parameter
6	Nakagami	$2.\left(\frac{\mu}{\omega}\right)^{\mu}.\frac{1}{\Gamma(\mu)}.x^{2\mu-1}.e^{-\frac{\mu}{\omega}x^2}$	μ : shape parameter ω : scale parameter
7	Generalized Pareto	$\frac{1}{\sigma}\left(1+k\frac{x-\theta}{\sigma}\right)^{-1-\frac{1}{k}}$	k : tail index (shape) σ : scale parameter θ : threshold parameter

Table 5.1 Parametric probability distributions considered for marginal distributions and their associate parameterizations (suite)

8	Loglogistic	$\frac{1}{\sigma.x}\frac{e^{\frac{\log(x)-\mu}{\sigma}}}{\left(1+e^{\frac{\log(x)-\mu}{\sigma}}\right)^2}$	μ : mean of logarithmic values (LV) σ : scale parameter of LV
9	Lognormal	$\frac{1}{\sigma\sqrt{2\pi}x}\exp\left(\frac{-(\log(x)-\mu)^2}{2\sigma^2}\right)$	μ : mean of LV σ : standard deviation of LV
10	Weibull	$\frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}\right)^b}$	a : scale parameter b : shape parameter
11	t Location Scale	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left[\frac{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right]^{-\frac{\nu+1}{2}}$	μ : location parameter σ : scale parameter ν : shape parameter

We also consider five parametric copulas, namely, Gaussian, t, Frank, Clayton, and Gumbel for modeling joint relationships between MASEL and MAT residuals. These structures together, they can represent a wide range of dependence structures among random variables (see Joe, 1997). Table 2 shows the formulations of the joint density functions for these copulas along with their parameterizations, validity ranges, and their links to Kendall's τ, which provides a formal approach for parametrizing copulas known as the Method of Moments (MoM; Genest and Favre, 2007; Abdi et al., 2017). Apart from MoM, we implement MLM to estimate copula parameters based on maximizing the likelihood function of a given joint distribution function (Bedoui and Dbabis , 2008; Weiß, 2011; Jane et al., 2016). The definition of the overall parametric model is based on considering both joint density function and marginal. Following the approach indicated for the selection of marginal distributions, we consider minimizing AIC and BIC functions to identify the most suitable copula model.

Table 5.2 Copula functions and their parameters used to estimate the dependency between MASEL and MAT residuals

Copula name	**Joint density function**	**Parameters**	**Parameter θ range**	**Expression of Kendall's τ**
Gumbel	$\exp\left(-\left((-\ln(u))^\theta+(-\ln(v))^\theta\right)^{\frac{1}{\theta}}\right)$	1 parameter: $\theta \geq 1$	$[1,+\infty[$	$=1-\frac{1}{\theta}$

Table 5.2 Copula functions and their parameters used to estimate the dependency between MASEL and MAT residuals (suite)

Clayton	$(uv)^{-\theta-1}(\theta+1)\left(u^{-\theta}+v^{-\theta}-1\right)^{-\frac{1}{\theta}-2}$	1 parameter: $\theta \geq 1$ and $\theta \neq 0$	$[-1,+\infty[\setminus\{0\}$	$=\frac{\theta}{\theta+2}$
Frank	$\frac{\theta e^{-\theta(u+v)}\left(e^{-\theta}-1\right)}{e^{-\theta(u+v)}-e^{-\theta u}-e^{-\theta v}e^{-\theta}}$	1 parameter: $\theta \neq 0$	$\mathbb{R}^*$	$=1-\frac{4}{\theta}\left[\frac{1}{\theta}\int_0^{\theta}\frac{t}{e^t-1}dt-1\right]$
Gaussian	$\int_{-\infty}^{\Phi^{-1}(u)}\int_{-\infty}^{\Phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left(-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right)dxdy$	1 parameter: $-1<\theta<1$	$[-1,1]$	$=\frac{2}{\pi}\arcsin(\theta)$
t (student)	$\int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{t_\nu^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\left(1+\frac{x^2-2\theta xy+y^2}{\nu(1-\theta^2)}\right)^{-\frac{\nu+2}{2}}dydx$	2 parameters: $-1<\theta<1$ ν	$\theta \in [-1,1]$ $\nu \in \mathbb{R}^+$	$=\frac{2}{\pi}\arcsin(\theta)$

Marginal and joint distributions are setup using three different schemes. In the first setup, marginal distributions for MASEL and MAT residuals are chosen based on the user's *priori* knowledge of the considered marginal random variable (here GEV distribution after recommendation of Moftakhari et al., 2017), and the best copula to represent the joint distributions is found using MLM (Configuration 1) or MoM (Configuration 4). In the second setup, the best marginal distributions for variables are primarily chosen based on the considered GoF measures; and accordingly, the best copula is identified independently from the margins using MLM (Configuration 2) or MoM (Configuration 5). In the third setup, the best marginal and copula distributions are simultaneously chosen from the available pool of marginal and copula distributions based on the overall modeling capability. Again, we parameterize copulas once with MLM (Full likelihood function; Configuration 3) and once with MoM (Configuration 6). Note that each of these six configurations includes multiple modeling hypotheses. See Nazemi and Elshorbagy (2012) and references therein for methodological details of copula parametrization. Within each modeling configuration, we rigorously compare the accuracy of parallel modeling hypotheses based on the BIC measure and/or ability of the model to preserve the empirical Kendall's τ. If MLM (Configurations 1 to 3) is implemented as the basis of parameterization, then reproduction of Kendall's τ is taken as the basis for model accuracy and the reproduction of GoFs are only taken as extra evidence for the accuracy

of the model. In contrast, if MoM is implemented, then the performance measures are taken as the notion of efficiency for copula models.

We also use a formal bootstrapping method, similar to what proposed by Genest et al. (2009) and Bland and Altman (2015), to address the robustness of copula models. In brief, we randomly draw samples of MAT and MASEL pairs with same as the size as observed data (53 pairs in this study) using the fitted copula functions. We then re-parameterize the copula function using randomly sampled pairs as well as MLM (Configurations 1 to 3) or MoM (Configurations 4 to 6) and estimate the Kendall's τ and the BIC measures accordingly. We repeat this sampling procedure 1000 times to provide an empirical range for the simulated BIC and Kendall's τ. A modeling hypothesis can be falsified through comparison with other copula models, if it has more deviation from empirical statistics, and/or it demonstrates more variation in simulated statistics through resampling.

5.5 Results

The data is first detrended and de-autocorrelated. Figure 5.7 (top row) clearly shows that there is no significant trend (orange line) in the detrended MAT and MASEL residuals (blue line) based on the Mann-Kendall test. In addition, Figure 5.7 (bottom row) shows that that the autocorrelations in MAT and MASEL are also removed efficiently, as no Pearson correlation beyond the confidence intervals can be observed in the residuals. As a result, the residuals of MAT and MASEL are truly *iid* variables, and their dependency can be represented through copulas.

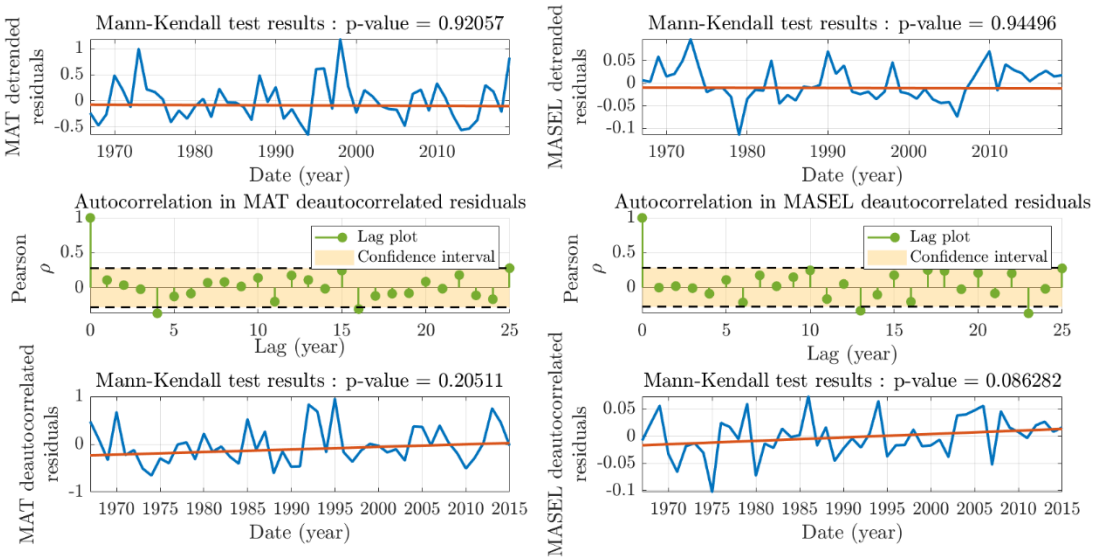

Figure 5.7 The effect of removing trend (top), autocorrelation (middle) and trend from de-autocorrelated residuals (bottom) of MAT (left) and MASEL (right).

We develop copula models to represent the dependence between MAT and MASEL residuals under the six configurations, introduced in Section 5.4. Figure 5.8 summarizes the results with relation to marginal distributions of MAT and MASEL residuals. First and third columns from left show the estimated AIC and BIC measures in all considered probability distributions in Table 1 for MAT and MASEL residuals, parameterized under the six considered configurations. In each column/row, the optimal distributions based on the GoFs are shown by green bars. In the second and forth column from left, the empirical CDFs for the MAT and MASEL residuals are compared with optimal parametric distributions and their 95% confidence intervals.

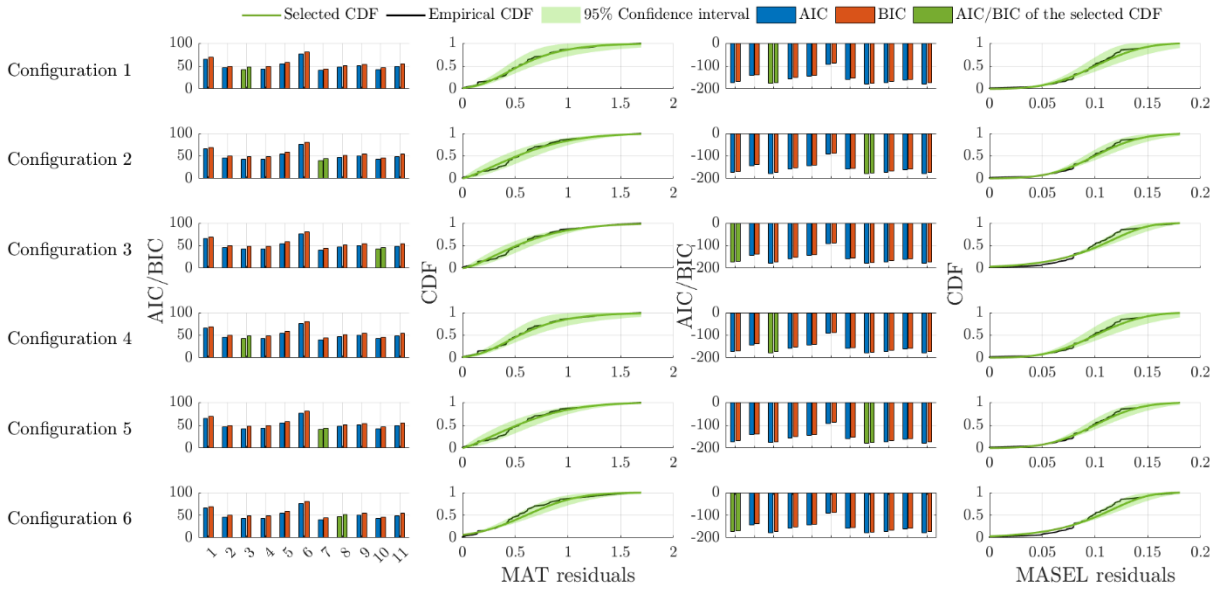

Figure 5.8 Performance of utilized 11 distributions for fitting residuals of MAT and MASEL (1st and 3rd columns from left) under six considered model configurations, along with the efficiency of best fitted distributions in covering the empirical margins of MAT and MASEL residuals (2nd and 4th columns from left).

As noted in Section 5.4, in Configurations 1 and 4, the GEV distribution is chosen based on a *priori* recommendation. In Configurations 2 and 5, the Nakagami and Normal distributions are selected for MAT and MASEL residuals, respectively. For Configurations 3 and 6, where the identification of marginal distributions is integrated with copula parameterization, Weibull and Extreme Value distributions are selected for MAT and MASEL residuals in the context of Configuration 3 respectively. In contrast in Configuration 6, Normal and Extreme Value distributions are the best marginal fit to MAT and MASEL residuals, respectively – see the associated parameterizations of these distributions in Table 3. By visual inspection, it can be revealed that overall, these identified distributions mostly cover well the empirical distribution, except in Configurations 3 and 6 where the Extreme Value distribution does not fully cover the MASEL residuals within its 95% confidence interval.

Table 5.3 Selected marginal distributions and their associated parameterizations for MAT and MASEL residuals under the considered six configurations.

Configuration number	Marginal distribution for MAT residuals	Distribution's parameter(s) for MAT	Marginal distribution for MASEL residuals	Distribution's parameter(s) for MASEL
1 &4	GEV	$\begin{cases} k = 0.038 \\ \mu = 0.426 \\ \sigma = 0.29972 \end{cases}$	GEV	$\begin{cases} k = -0.282 \\ \mu = 0.079 \\ \sigma = 0.035 \end{cases}$
2 & 5	Nakagami	$\begin{cases} \mu = 0.598 \\ \omega = 0.529 \end{cases}$	Normal	$\begin{cases} \mu = 0.092 \\ \sigma = 0.036 \end{cases}$
3	Weibull	$\begin{cases} a = 0.661 \\ b = 1.814 \end{cases}$	Extreme Value	$\begin{cases} \mu = 0.108 \\ \sigma = 0.023 \end{cases}$
6	Normal	$\begin{cases} \begin{cases} \mu = 0.551 \\ \sigma = 0.335 \end{cases} \end{cases}$	Extreme Value	$\begin{cases} \mu = 0.108 \\ \sigma = 0.025 \end{cases}$

After identifying marginal distributions, we turn to assessing the performance of copula models. Figure 5.9 shows the best fitted parametric copulas among a pool of candidates based on the specific criteria with which the optimal copula model is tested and selected (left column) and the unseen criteria with which the models are validated (right column). For Configurations 1 to 3, MLM is applied to parameterize and to select copulas' parameters; and therefore, models are validated using the ability to preserve the empirical Kendall's tau dependence (τ_{emp}) between MAT and MASEL residuals. In contrast for Configurations 4 to 6, in which MoM is used for copula parameterization and model selection, AIC/BIC criteria are used for model validation.

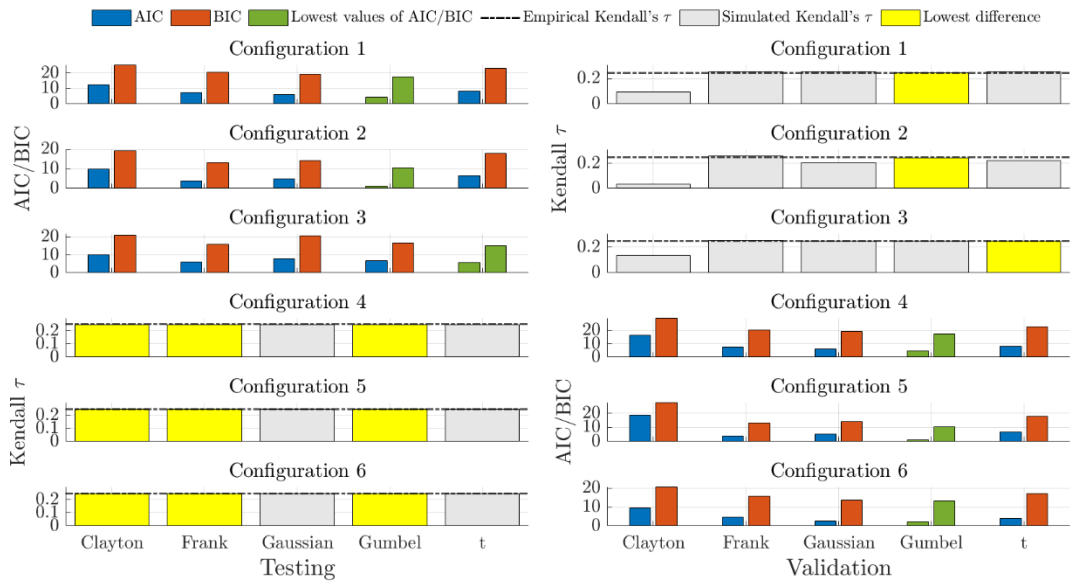

Figure 5.9 The performance of copula models during parameterization and model selection (testing; left column) as well as validation with unseen criteria (validation; right column).

For Configurations 1 and 2, Gumbel copula is the most efficient parametric joint dependence model based on both AIC/BIC and representing the empirical dependence. Respectively, the Gumbel copula is able to capture the τ_{emp} between MAT and MASEL residuals with relative errors of 0.6% and 1.8% in Configurations 1 and 2. In Configuration 3, t copula provides the best results in terms of AIC/BIC measures, as well as preserving the τ_{emp} (relative error 0.08%); however it should be noted that the differences between various copulas in preserving τ_{emp} are almost unnoticeable. In Configurations 4 to 6, in which MoM and accordingly τ_{emp} is used for parameterizing copulas, different copulas show comparable performance during testing and selection period; however, when validated with AIC/BIC measures, Gumbel copula stands out.

The robustness of copula models is further inspected using the bootstrap approach. For Configurations 1 to 3, in which the MLM is used for copula parameter estimation, the reconstruction of empirical Kendall's tau based on 1000 randomly sampled pairs of MAT and MASEL residuals is taken as a notion of robustness. For Configurations 4 to 6, in which MoM is employed for copula parametrization, the resampled BIC values are used to address the robustness of copula models. The results of this experiment are summarized in Figure 5.10. Left and right panels show the results during the testing and selection as well as validation. In each panel,

boxplots show the resampling results obtained by each copula model, in which boxes show the interquartile range; whiskers are the 95% confidence interval; red stars are the outliers; and the black dashed lines within each box show the median values obtained from 1000 resampling. Solid black lines depict the values of BIC as well as τ_{emp} during model parameterization and validation, reported in Figure 5.9. The results of this resampling clearly reveal that the previously identified copulas based on GoFs and/or τ_{emp} provide the most robust modeling options that are able to replicate the initial modeling efficiency through resampling. We also repeat the resampling process and realize that the findings are not sensitive to reinitializing and/or random seeds of the resampling.

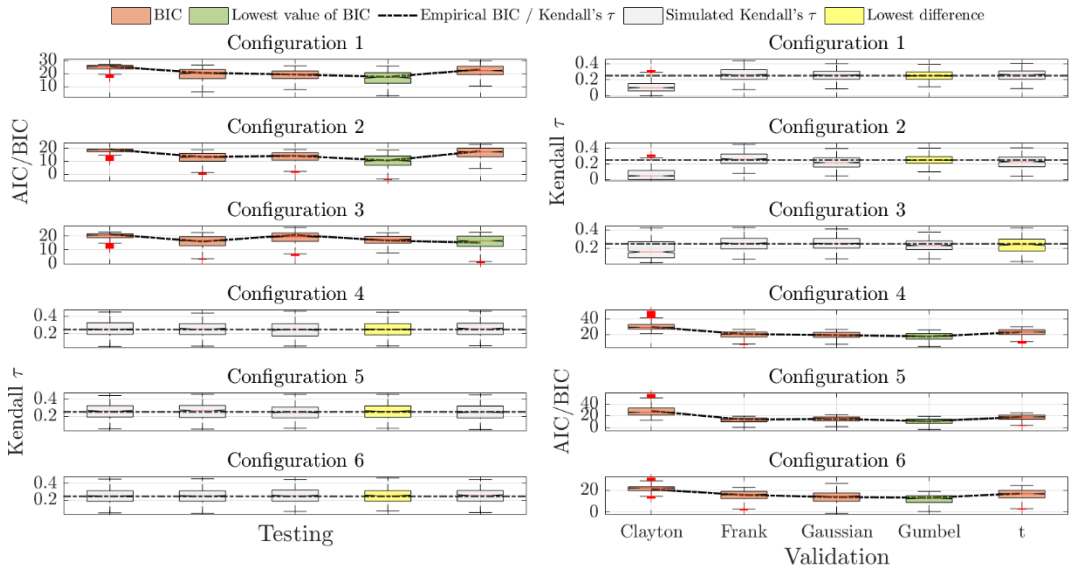

Figure 5.10 The result of bootstrapping approach for evaluating of the robustness of copula models under six considered configurations based on replications of Kendall's τ and BIC.

Having the copulas models parametrized, validated, and evaluated, the MASEL residuals can be generated based on residuals of MAT through conditional resampling. During the historical period, when MAT residuals over the simulation period are known, sampling can be done through a uniform random generation over the joint probability distribution. During projection, the sampling is slightly different, as first MAT residuals should be sampled using the parametrized marginal distributions. The autocorrelation and trend can be then rebuilt in the sampled MASEL residuals to generate MASEL timeseries. Figure 5.11 shows the results of this attempt using the best copula

setups, obtained using Configurations 1 to 6, during historical period (1967-2019) as well as the rest of 21st century (projection period: 2020-2100). The projected timeseries are generated by assuming the same trend and autocorrelation structure in MAT and MASEL during the historical period. The simulated range (grey envelope) is obtained by 1000 resampling. The solid red and blue lines show the observed and expected MASEL timeseries, respectively; and the orange region identifies one standard deviation around the expected MASEL time series. As it can be clearly seen, the envelopes of simulated MASEL obtained by all configuration is able to capture the observed MASEL timeseries during the historical period.

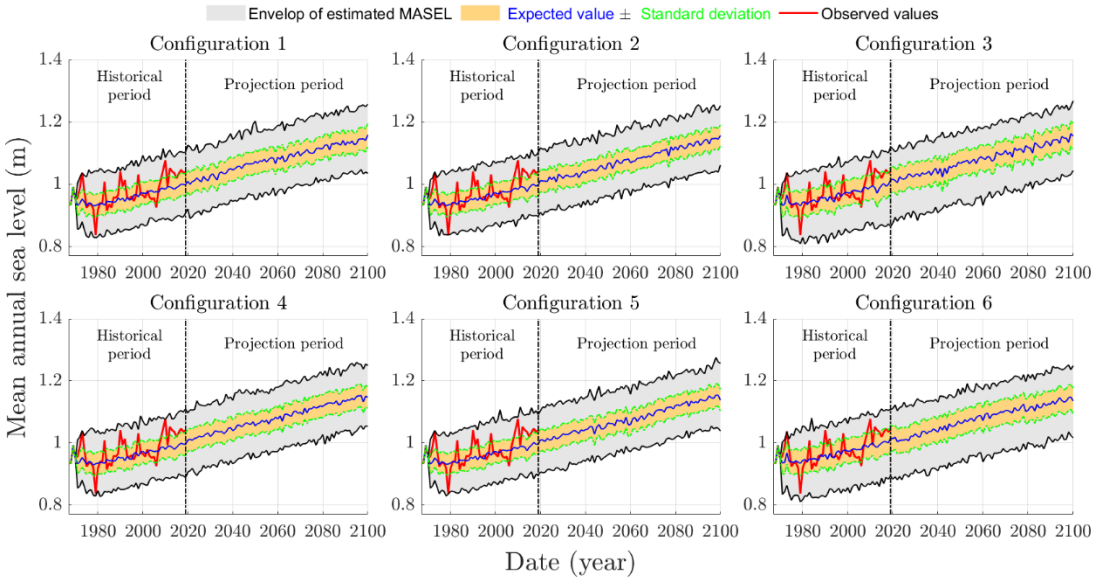

Figure 5.11 Observed MASEL (red line) along with the envelope of simulated MASEL (grey envelope), and associated expected values (blue) and ranges for standard deviation (orange envelope) during the historical (1967-2019) and future periods (2020 to 2100) obtained by the best setups in Configurations 1 to 6.

5.6 Discussion and impact assessment

An inclusive approach has been taken to develop alternative modeling hypotheses by mixing-and-matching multiple representations of marginal and joint dependencies that are parameterized through six configurations. This attempt has provided some solid experimental results, with which we can draw conclusions on how the marginal and copula models can be optimally setup and used. First and foremost, our findings reveal that the sequential selection of marginal and copula models

results in improved GoFs measures for both marginal and joint representations. On the one hand, priori selections of the margins (Configurations 1 and 4) clearly result in sub-optimal choice of margins (see Figure 5.8) and consequently sub-optimal choice of the copula model (see Figure 5.9). On the other hand, simultaneous selection of marginal and joint distributions (configurations 3 and 6) results in sub-optimal choice of both marginal (Figure 5.8) and joint models (Figure 5.9). Second, copula parameterization using MoM presents similar or even better GoF in comparison to configurations in which MLM is used for model parametrization. In addition, copulas parameterized using MoM exhibit more robust options and are able to replicate the empirical dependency and original fi well.

One key aspect of our methodology is removing the trend and autocorrelation from MAT and MASEL and working with residuals instead of the original data. We argue that this is an essential step in the conditional sampling in the presence of significant trends in considered variables. To support our argument, we repeat the model setups and configurations using the original data and the procedure introduced in Section 5.4. Similarly, we generate historical and future projections of MASEL using the 1000 conditional samples and estimate the range of potential MASEL. Figure 5.12 shows the envelope of simulated MASEL (grey) in each configuration along with the expected timeseries of MASEL (blue line) and the range for standard deviation (orange) during the historical and future periods. Observed historical values are shown by red lines. As it can be clearly witnessed, the envelopes of simulated MASEL cannot fully capture the extreme low and/or high values during the historical period. In addition, none of the configurations are able to reconstruct the historical trend in MASEL, which is shown to be significant (see Figure 5.5). Due to this, a relatively stationary progression in MASEL is projected until 2100, which can mislead the impact assessment.

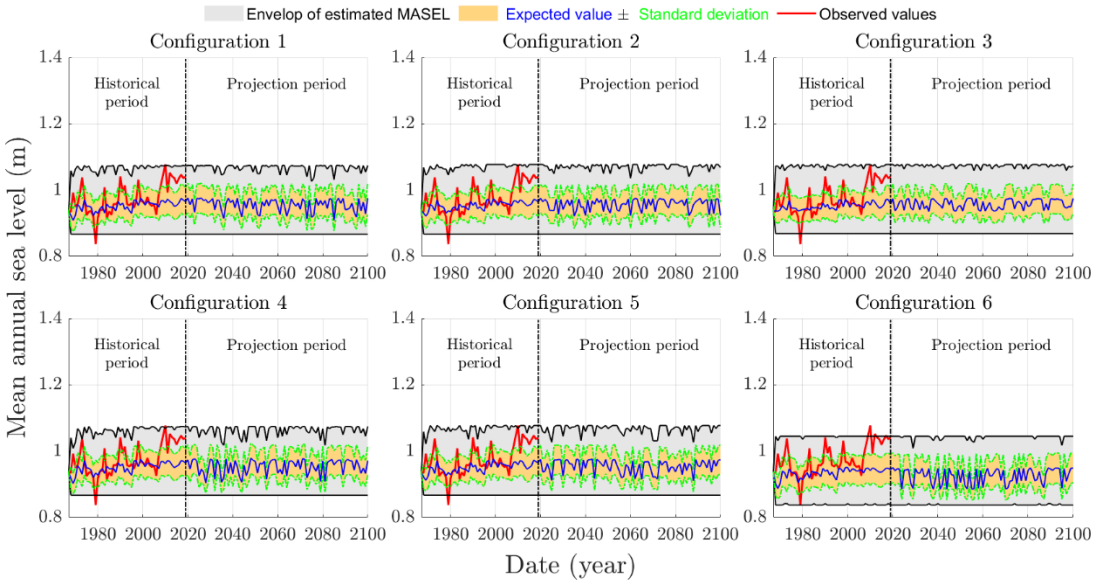

Figure 5.12 The envelope of simulated MASEL (grey envelope), and associated expected values (blue) and ranges for standard deviation (orange envelope) during the historical (1967-2019) and future periods (2020 to 2100) using Configurations 1 to 6 when autocorrelations and trends are not removed from data. Historical observations of MASEL are shown with red lines.

Another key question here is the choice of configuration that should be used for impact assessment. As previously noted, the accuracy and robustness of best models from each configuration are almost the same (see Figures 5.9 and 5.10); Alternatively, here we look into the differences between MASEL simulations obtained by the six model configurations (see Figure 5.11) and assess how different configurations result in different projections of SLR. For this purpose, we look at the characteristics of projections at the end of 21st century, when the simulations are terminated. The results are summarized in Table 4, showing that the differences between projections are also marginal. In addition, the inter-model variability between simulations obtained by the six configurations is still within the 95% confidence interval of one of the models (i.e. Configuration 3 – see Table 4). Accordingly, instead of discarding any of the six configurations, we embrace the uncertainty in MASEL projections coming from different configurations and consider all of them as an ensemble for impact assessment. In that sense and based on the multi-model mean ensemble, an increase of 15.7 % in MASEL is projected in 2100 compared to the baseline year 2019. This percentage of SLR is in accordance with the projections of the US Geological Survey through extrapolation of tide records (Parris et al., 2012). Given the topology of the island with long profile

and flat coasts (Bellard et al., 2013), such an increase in MASEL can lead to coastal erosion and affect community assets and other socio-economic activities as discussed in Section 5.3.

Table 5.4 Statistical characteristics of simulated MASEL in 2100 under each configuration

Configuration N°	Expected value	95% confidence interval of simulated MASEL (% of SLR compared to benchmark year 2019)		
		Lower bound	Median	Upper bound
Configuration 1	1.146 m	1.033 m (13.58 %)	1.146 m (14.50 %)	1.255 m (15.37%)
Configuration 2	1.148 m	1.043 m (13.21 %)	1.147 m (14.69 %)	1.239 m (16.05%)
Configuration 3	1.134 m	1.028 m (11.94 %)	1.136 m (14.17 %)	1.238 m (18.96 %)
Configuration 4	1.133 m	1.044 m (12.06 %)	1.131 m (13.13 %)	1.238 m (16.59 %)
Configuration 5	1.155 m	1.049 m (16.49 %)	1.156 m (15.48 %)	1.276 m (18.75 %)
Configuration 6	1.146 m	1.036 m (13.02 %)	1.149 m (14.01 %)	1.252 m (16.42 %)

Using the ensemble of projections obtained from the six considered configurations, a probabilistic look can be provided on exceeding a given MASEL during the simulation period and/or exceeding a MASEL level in a given year. Left panel in Figure 5.13 shows the non-exceedance probabilities of MASEL in four reference years of 2040, 2060, 2080, and 2100, each with a unique color. For each year, the envelop of CDF contains 6000 realizations obtained by the six configurations, each having 1000 simulations. The solid black lines reflect the mean ensemble. The right panel in Figure 5.13 shows the non-exceedance probability of future years reaching to four reference levels of 1.10 m, 1.15 m, 1.20 m, and 1.25 m, each identified with a unique color. The envelopes of CDFs are again identified by 6000 realizations and the solid black lines is the mean ensemble. Such probabilistic risk profiles (see Hassanzadeh et al., 2016) can provide a tool for decision makers, with which a target decision value – in this case sea level – can be chosen based on the risk that can be accepted for exceeding a certain MASEL in a particular year (see also Hassanzadeh et al., 2019). For instance, based on the right panel of Figure 5.13 and in the year 2100, there is more than 90% chance for MASEL to exceed 1.10 m, around 50% chance to exceed 1.15 m, around 8% chance to reach 1.20 m, and almost negligible chance to reach 1.25 m. Using this information and the level of risk that can be accepted for the problem in hand, one can choose a MASEL value for design of flood defense system, dicks, and levees and or re-calculating the borderline for new land developments and/or socio-economic activities (Le Cozannet et al., 2014).

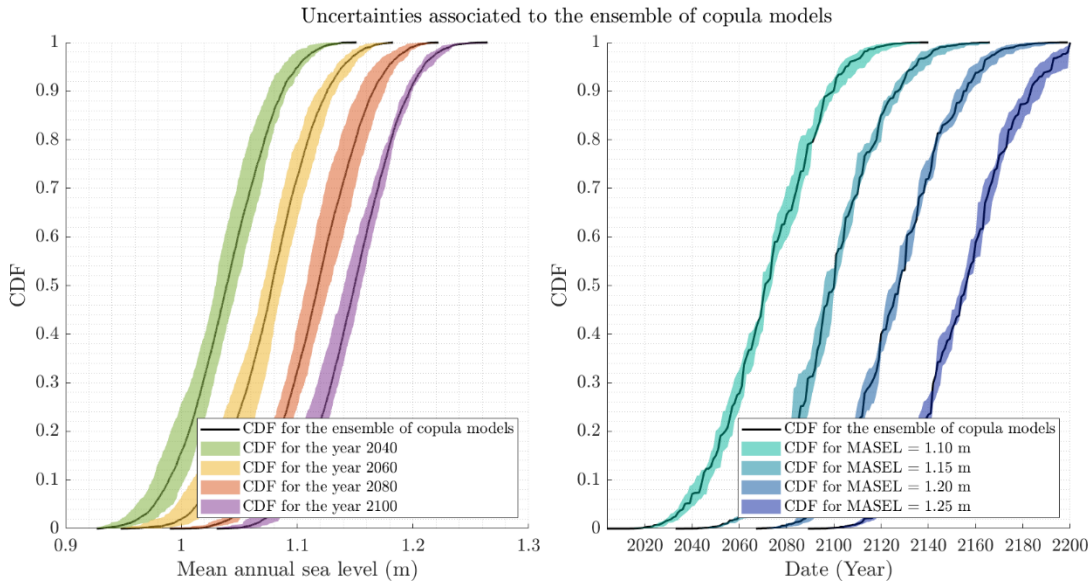

Figure 5.13 Non-exceedance probabilities of MASEL in years 2040, 2060, 2080 and 2100 (left panel), as well as the non-exceedance probabilities of given year passes MASEL thresholds of 1.1 m, 1.15 m, 1.2 m and 1.25 m (right panel).

We have further analyzed the impact of uncertainties in estimation of MASEL trends in the historical period on the developed models and projections of sea level in the future. Figure 5.14 shows the simulated MASEL using the lower and upper bounds of estimated trend slope in MASEL under the considered 6 configurations. The panels show that the simulated MASEL series are highly sensitive to estimated trend values. Indeed, the range of projections significantly diverge given the lower and upper bounds, particularly by the end of century. Same as previous observations, the models are less sensitive to the choice of configuration.

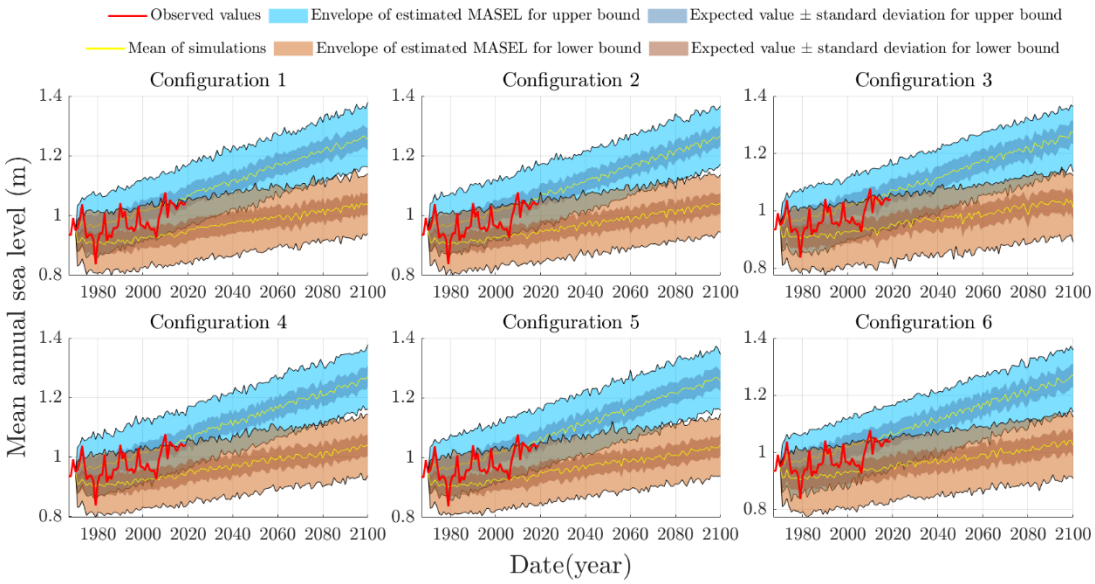

Figure 5.14 The envelopes of simulated MASEL during the historical (1967-2019) and future (2020-2100) periods using the six considered configurations and considering lower and upper bounds of historical trend in MASEL. Mean ensembles are shown with solid yellow lines and the ranges of one standard deviation around the mean are shown with darker shades. The observed MASEL data are shown with solid red lines.

To showcase how the total projection uncertainty in SLR is contributed by the uncertainty in identifying the trend in MASL as well as the random variability due to stochastic sampling, we look at the CDF and PDF of MASEL in reference years of 2040, 2060, 2080, and 2100, simulated using the expected value as well as the 95% confidence interval of the trend in MASEL and compare them with the CDFs and PDFs of simulated MASEL in the year 2019 as a benchmark year. Figure 5.15 summarizes the results. Top row shows the total uncertainty in CDFs of the sea level in the four reference years obtained by the uncertainty in trend, represented by the 95% confidence interval. In each panel the yellow range represent the uncertainty in the CDFs due to the uncertainty in the trend. The green and red CDFs show the non-exceedance probability of the sea level under lower and upper bounds of the 95% confidence interval of trend, respectively; and solid and dashed CDFs show the non-exceedance probability of the sea level in year 2019 as well as under the expected historical trend. To decompose the total uncertainty into the portions initiated from uncertainty in the trend and the random variability, in bottom row we compare the corresponding PDFs of the sea level in the benchmark year 2019 (dark solid line), with the future

PDFs of the sea level in the considered reference years, obtained by the lower and upper bounds as well as the most likely trends of MASEL. This analysis not only provides a vivid picture of the total uncertainty and how it is composed of natural variability and the uncertainty in the trend, but also shows how the total uncertainty can slightly increase during the projection period. Analyzing of the total uncertainty shows the uncertainty in future sea level is almost equally contributed by the uncertainty in the trend and the random variability in MASEL residuals, contributed from the random variability in the MAT residuals and the joint dependency between residuals of MASEL and MAT. This can be understood by the range of each PDF compared to the distance between the modes of PDFs for lower and upper bounds

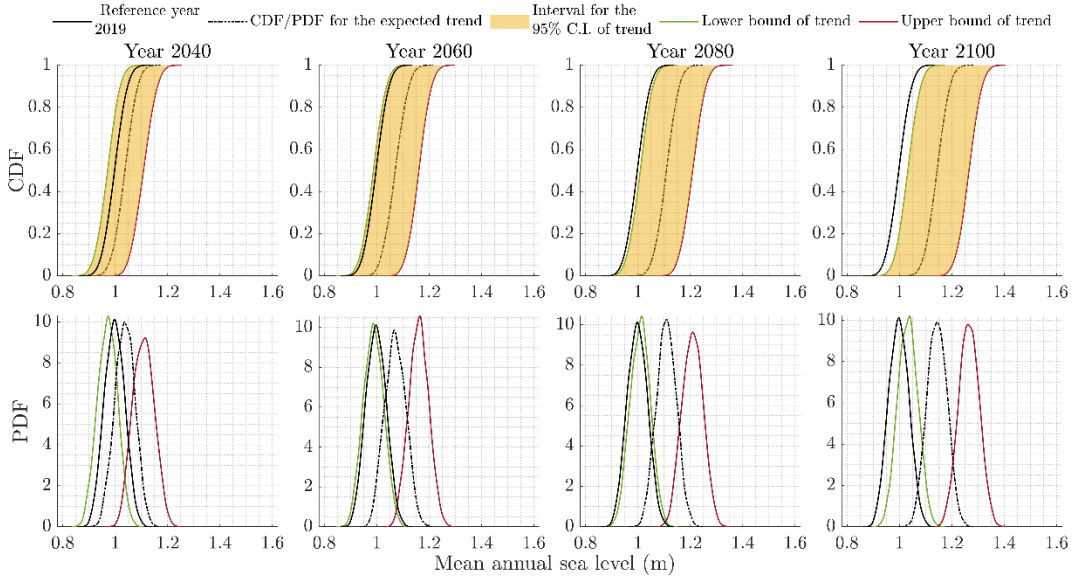

Figure 5.15 Risk profiles of MASEL in reference years of 2040, 2060, 2080 and 2100 provided as CDFs (top row) and PDFs (bottom row) and compared with corresponding risk profile in the benchmark year 2019.

5.7 Summary, conclusions and further remarks

SLR is one of the immediate consequences of warming climate, caused by both increasing freshwater inflows and thermal expansions of seas and oceans. Continuation of SLR can result in global disruptions in ecosystems and socio-economic activities. Current methodologies for predicting the SLR are often based on the IPCC's top-down scenario-led framework that use Earth

System models in conjunctions with future scenarios for emission or concentration of greenhouse gases. However, the results of such assessments are often associated with a large uncertainty, which limits the accessibility and applicability of top-down projections in practice. At this juncture, methodologies based on so-called bottom-up impact assessments that can provide a basis for looking at the vulnerability to SLR through information and are more sensible to decision makers become favorable. Such approaches to impact assessment allow exploring the modes of vulnerability through what-if scenarios and facilitate identifying critical thresholds beyond which the system under consideration becomes vulnerable. In this context, probabilistic impact models are particularly suitable as they provide an opportunity for a risk-based assessment in light of information available about the system under consideration.

Here we attempt to build up such a probabilistic impact model, which can provide a basis for bottom-up impact assessments in the context of SLR. The idea is to represent the evolution in the local SLR subject to local temperature that is sensible to stakeholders. While we do recognize that the causes of SLR are indeed global and well beyond the evolution in the local temperature; however, the local temperature itself is affected by the global warming and therefore can be used as a proxy to describe the local SLR with. To do so, we decompose MAT and MASEL into trends and random variability and attempt to link the corresponding components in the two signals. We use linear regression to explore the linkage between trends in MAT and MASEL and implement copulas to represent the joint dependencies between residuals of MAT and MASEL. This provide an opportunity for stochastic sampling of MASEL conditioned to MAT and can be extended into the future projections by preserving or perturbing trends and the dependence between MAT and MASEL.

We showcase the application of this approach in New Caledonia, a French-speaking archipelago in the southeast of Pacific Ocean, where SLR not only is associated with deterioration of infrastructure and destruction of socio-economic activities, but also threatens a unique culture and way of life of Kanaks, the Indigenous people of New Caledonia. Our results based on in-situ observations clearly demonstrate that there is a strong linear linkage between trends of MAT and MASEL across the majority of timescale considered. We also show that copulas are also able to represent the nonlinear relationships between MAT and MASEL residuals and can provide a basis for conditional sampling of MASEL residuals based on the residuals of MAT. Our analyses based

on a number of configurations for marginal and joint distributions show that while the sequential identification of marginal and joint characteristics can marginally improve the overall GoF measures, the difference between best setups in each configuration is rather marginal when it comes to the simulation of SLR under retrospective and prospective conditions. As a result, we propose using a multi-model ensemble simulation for impact assessment and show the practical utility of such simulations in risk-based analyses of SLR. We additionally show that decomposing the MAT and MASEL signals into components of trend and random variability is deemed necessary. By exploring the total uncertainty in the simulations, we demonstrate that the uncertainty is almost equally contributed from the uncertainty in the trend and random variability.

Despite our current findings much more can be done. Obviously, the applicability of this approach should be verified in other locations and the possibility of projecting SLR in light of global temperature and/or carbon concentrations should be explored. Moreover, whether this approach is also relevant to simulation of maximum annual sea level is not yet clear. Last but not the least, we should clearly state that we do not intend to bring a new methodology against the current top-down approach. In fact, we believe that top-down and bottom-up approaches to vulnerability assessment are complementary and together they can provide an improved confidence about the future projections of SLR that are inherently uncertain. We hope our study can trigger more attempts toward improving preparedness in fight against adverse impacts of SLR.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Alexandre Peltier for his support in providing the data for Nouméa station. This research is supported by NSERC-DG (RGPIN-2020-05563), held by Elmira Hassanzadeh. This paper is dedicated to the islands known as New Caledonia and the islanders known as New Caledonian

CHAPTER 6 GENERAL DISCUSSION

This chapter presents a discussion on copula modeling and its application for assessment of sea level rise in New Caledonia. Advantages and limitations of copulas are presented and potential of this approach in statistical analysis is explained. The potential impact of sea level rise in New Caledonia and the complementary analyses that could be performed locally are explained. Finally, possible improvements of the software are discussed.

6.1 Consideration of copula for bivariate modeling

Copulas are efficient in estimating the joint probability of different drivers, considering their dependence. However, the application of copula requires IID variables (Section 4.2.4.1). As it was shown in this thesis, preprocessing is an important step to quantify and remove trend and autocorrelation before simulating a variable. The copula is then applied considering the residuals and the simulated values are retrieved from the residuals thanks to the models used to preprocess each dataset. Employing a set of copulas enables finding the best bivariate model but requires methods to determine the most suitable copula to use too.

6.1.1 Extensions and complementary analyses

This study has been the first attempt to simulate sea level rise using copulas considering its influencing factors. Nevertheless, this work can be extended and improved in future studies. For instance, in this study parametric marginal distributions and copulas are used for modeling sea level. Therefore, a set of non-parametric considerations can be used to understand their impact on the quality of the developed model can be studied. For example, Empirical copula, described as a joint empirical probability, allows estimating the bivariate probability considering the rank of the observed data. Moreover, a maximum pseudo-likelihood method (Joo et al., 2020) can be applied in complement to the IFM and MoM to construct the copula models. In this method, empirical marginal CDFs are considered instead of parametric distributions and ML method is employed to estimate copula's parameter. Moreover, in this study only temperature is considered as the key influencing factor for sea level. In future studies, a range of hydroclimatic variables such as river streamflow at the outlet of a watershed, precipitation, or temperature at larger scale (regional or global) can be considered and using dependence analysis two dominant factors can be selected.

Accordingly, trivariate copulas (Salvadori and De Michele, 2007) can be implemented to consider these influencing drivers on sea level. Nevertheless, it should be mentioned that application of these copulas is based on specific families and is still anecdotical compared to bivariate copulas given their complexity (De Michele et al., 2007; Balistrocchi and Bacchi, 2011).

6.2 Analysis of sea level rise in New-Caledonia

6.2.1 Coastal impact

The application of copula modeling in New-Caledonia allows simulating sea level rise until 2100. The magnitude of sea level in 2100 can be used for topologic studies in order to assess the percentage of land impacted by such increase, similar to the work of Bellard et al., (2013). Thus, consequences of sea level rise can be presented in specific locations, which is useful for local decision-makers. Moreover, the method of event classification can be used to provide a threat level for different locations along the shoreline. This allows identifying vulnerability of local infrastructure to sea level rise. However, it should be noted that this classification approach depends on the availability of sea level and temperature data across the entire shoreline. This means that efforts should be in place to install and collect hydroclimatic data in different parts of the archipelago. Since ground data with high quality are not currently available spatially, satellite data for sea level and temperature can be used in copula analysis. Then, future sea level projection can be simulated based on future temperature projection from GCM under several scenarios.

6.2.2 Potential improvements

The study realized to simulate sea level rise in New-Caledonia can be completed by adding a flood hazard study from compound events (Moftakhari et al., 2017) considering several drivers– method described in Section 4.2.3. For instance, assessing this risk of flood in the area from sea level and streamflow datasets could provide indication regarding the apparition of at least one hazardous scenario during the timespan. Moreover, the consideration of specific return periods can be used to determine bivariate risk of failure and its evolution over time. However, it is necessary that tide gauges and streamflow sensors are located in the same area (e.g., at the outlet of a watershed) so the dependence between streamflow and sea level is assessed locally. Another improvement consists in using trivariate copula, adding a supplementary variable compared to bivariate copula,

and estimating the risk of flood in specific locations. For instance, the definition of flooh hazard from sea level rise, rainfall and streamflow datasets can be interesting, especially in watershed with low-lying outlet where the unit hydrograph shows an important peak for streamflow. The application of copula modeling to different locations in the archipelago can be considered to determine the risk associated to sea level in specific area in order to inform local populations. However, given the low quality of each dataset regarding sea level (Table A.1) except for Nouméa, assessing its impact and simulating its values will not be based on long-term time series in New-Caledonia. These improvements are realistic to implement locally regarding the availability of the data and the methods to apply and can provide supplementary information to the analysis realized previously.

6.3 Improvements in the developed software

6.3.1 Upgrading the analyses

The provided functions in the developed software packages allow performing a wide range of analysis of hydroclimatic data; nevertheless, the methods can still be upgrade, and new approaches can be added. For instance, a bootstrap resampling can be used to estimate the 95% confidence interval for the parameters of marginal distributions and copula. Moreover, the consideration of this confidence interval for θ_{cop} provides a range for isolines of joint probability (Figure 6.1) which can be used to determine uncertainties in the risk assessment process. Thus, for a given bivariate critical threshold $C(u^*, v^*)$, two isolines from the surfaces of 95% confidence interval of θ_{cop} can be extracted to model its uncertainties. For a given value of X, a range of Y can be quantified from the bivariate joint probability. These complementary analyses provide an interesting material for copula modeling and can extend the ensemble of considerations already used in this project.

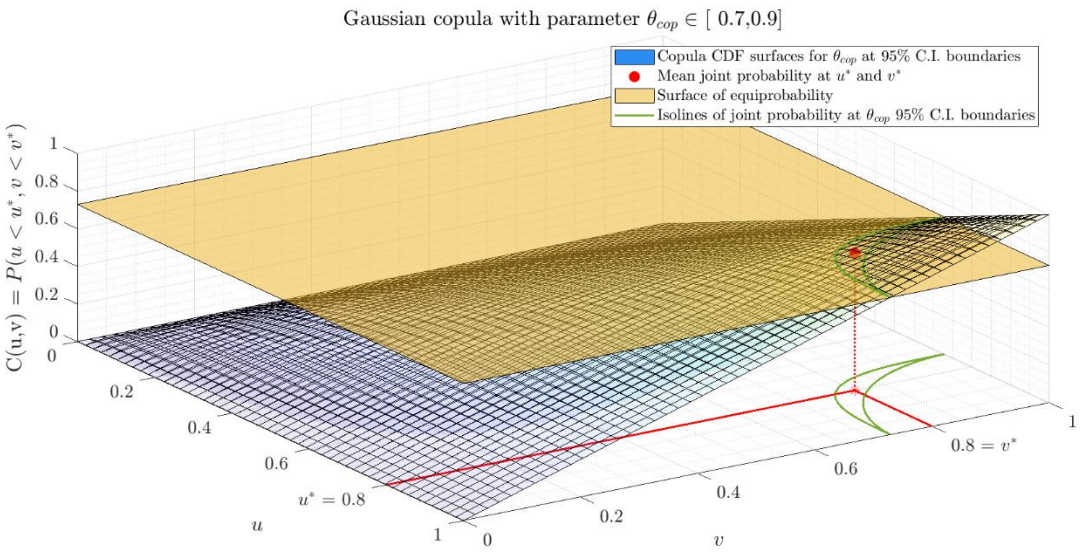

Figure 6.1 As an example, isolines of joint probability considering the 95% confidence interval of [0.7,0.9] of copula's parameter θ_{cop}

Moreover, different copula models can be added to the existing set to expand the possible choices. Regarding the confirmation for the value of marginal CDFs and copula parameters, a training / validation process can be implemented in order to select the most suitable value for each parameter in complement with the bootstrap approach needed to estimate the confidence interval. Thus, with these considerations, a more robust and complete model can be developed. Then, in terms of risk of failure for copula, different scenarios in complement to OR and AND, namely Kendall and Survival Kendall scenarios (Salvadori et al., 2016) can be used based on stochastic consideration. They allow to calculate different probabilities and provide supplementary information regarding the occurrence of at least one hazard scenario over a given timespan. Finally, if data series include missing points, adding an option to fill the missing data after developing and using copula model can be a relevant option to consider larger ensemble of datasets without acknowledging their quality. The GUI can be also improved in the future based on the comments received by general users to make it more effective or moving to the MATLAB App Designer instead of the basic GUI.

6.3.2 Extension to other programming languages

In order to make these software packages freely available to public, they can be converted into other open access languages such as Python. This also allows global contributors to update and extend these codes if it is needed. Indeed, Python language enables performing these analyses

online without downloading the software. Thus, the GUI can be replaced by the interfaces developed in the internet browser which could lead to a more enhanced disposition of buttons. A conversion of these codes in R language can also be considered given that this programming is dedicated to statistical analysis and could be more relevant to implement these codes. In both cases, conversion methods need to be developed to transfer the MATLAB script into the adapted language.

6.4 Conclusions

Copula appears to be an interesting tool in bivariate analysis to simulate the evolution of a variable considering its dependence with another driver. The method shows a robust framework with a complete set of analysis that can be upgraded with supplementary analysis for risk assessment. The case study of New-Caledonia provides another example of copula application in the field of hydrology but new regarding the studies performed in this area of the Pacific, helping to improve the knowledge related to sea level rise in this place. Thus, through the modeling, several improvements can be planned to extend the current studies to different drivers or locations. Finally, the two software packages developed during this project can be improved adding new analyses and considerations for copula modeling especially. Their extension to new language of programming appears to be a wise choice regarding their open-source nature.

CHAPTER 7 CONCLUSION AND RECOMMENDATIONS

7.1 Summary of the work

This project essentially aims to apply methods to analyze univariate and bivariate time series in the context of hydrology and more specifically to estimate sea level rise. In the Pacific region, sea level rise appears to be a critical issue considering the topology of the islands and the repartition of the population in these territories. The application of appropriate methods to characterize time series has been developed to provide local stakeholders information related to this issue. Using general approaches such as estimation of autocorrelation, variability and trend or fit of parametric distributions along with detection of change points, the purpose was to provide a set of relevant analyses to characterize hydroclimate variables. Thus, quantification of autocorrelation through Pearson's ρ, Spearman's ρ or Kendall's τ allow estimate the independent behaviour of a dataset. Inter and intra annual variability can also be purchased in order to represent the evolution of the variable over time, within a year or a period, and show any pattern. The estimation of linear and rank-based trends is considered to evaluate the overall behaviour of the variable (increase or decrease over the timespan). The detection of change points in the time series appears to be an interesting complementary analysis to locate the time where the mean changed and estimate the significancy of the change. Finally, the fit of parametric distribution allows find the most suitable model to represent the variable and can be used to estimate the return period associated to each observation.

In bivariate consideration, the application of copula modeling allowed interesting analyses of risk and simulation of variable considering the dependence between both drivers. The calibration and validation of copula based on simulated Kendall's τ_{sim} and AIC and BIC values provide the most suitable model to perform analysis considering the dependence between the variables. The application of a bootstrap resampling method to confirm and emphasize the choice of copula show the robustness of the employed approach to build the model. Then, the risk assessment estimated from isolines of joint probability allow quantify the evolution of the risk of failure and perform a classification of each event from observations. The simulation of a variables through a conditional sampling procedure provides an information related to the evolution of the variable over time and its possible future projection considering the dependence with another key driver. Considering the

multiple approaches and the different possible nesting between analyses, the organization of the set of MATLAB codes in software turns out to be the most effective way to perform quick calculation with a simple and user-friendly interface. This decision became the main structure of the project. Thus, the development of these software is part of a research field essentially based on climate data regarding the scientific institutes located in the archipelago. They can contribute to the improvement of methods already implemented locally and aim to be used as a complementary tool by local stakeholders. The case study of sea level rise in New-Caledonia shows the potential application of copula modeling and the free-access to these software aims to encourage scientist to use and develop the existing codes. It also emphasizes the need to monitor and properly estimate sea level rise in this region to provide accurate information to local population.

7.2 Conclusion of the study

The analysis of sea level in New-Caledonia shows a dependence between this variable and temperature at the local scale. Even though sea level rise is induced by the global variation of temperature, its estimation from local temperature is easier to catch from local stakeholders as they can directly access the information through the climate stations. Thus, with a significant dependence of 41.65% estimated from Kendall's τ, the application of copula modeling appeared to be relevant to simulate sea level in this area. Moreover, the preprocessing of each dataset to obtain IID residuals was required considering that both variables showed significant trend and autocorrelation. By considering several different configurations in the modeling process, an ensemble of copula models has been constructed in the study to estimate sea level rise in the area. The models project an overall maximum increase of almost 20 cm by the end of the century, considering. Moreover, the removal of autocorrelation and trend appears to be necessary given that copula modeling based on raw datasets show envelopes of scenarios that does not fully catch the range of variation of MASEL. Thus, the preprocessing is more influential on the performance of modeling than the different manners to build the copula model through the selection of marginal distributions and copula, which emphasizes the robustness and simplicity of copula modeling. Considering the ensemble of copula models, the analysis of CDF provides information related to the uncertainties of the ensemble and its influence on the results. The consideration of specific MASEL thresholds and their associated CDFs allow find the expected year when these values will be reached. It shows the uncertainty from the ensemble of model to properly assess the cumulative

probability of sea level. In complement, the estimation of MASEL's CDFs for specific year provides insights regarding the non-exceedance probability and can show its shift to higher values over time. Finally, replicating this analysis with lower and upper bounds of MASEL trend demonstrates the uncertainties related to the preprocessing step in the consideration of trend. The resulting range of MASEL simulations is extended, which illustrates the influence of MASEL trend uncertainties on the results. This consideration is interesting to represent the uncertainties associated to this key component and its impact on the modeling of MASEL. The calculation of CDFs and PDFs for specific years under these boundaries shows a shift to higher MASEL value over time, which correspond to the existing trend and the worst-case consideration simulate a MASEL increase of 28 cm in the area by 2100. More generally, apart from highlighting the importance of preprocessing variables, this study shows that New-Caledonia will face sea level rise over the next few decades with a magnitude up to 20 cm. These results can be used by local stakeholders to prepare adapted policies considering the different scenarios of trend and could help them anticipate the changes and impact on the shorelines.

7.3 Recommendations

This project can be upgraded in terms of considered analyses, software development and also the scope of the copula modeling applied. First, the representation of the envelope of scenarios defined by an empirical confidence interval of 95% could be more appropriate than using extremes of simulations. Then, more generally, an interesting improvement would be to consider uncertainties in the copula modeling through the 95% confidence interval in estimating the copula's parameter. Thus, the analysis of risk would consider a range of isolines of joint probability instead of a single value. As developed in the Chapter 5, the consideration of the 95% confidence interval for the trend in the preprocessing step allows finding lower and upper bounds for the envelope of simulations. The generalization of this process to each parameter in the copula modeling can lead to an interesting set of analyses where the uncertainties could be quantified through this interval. Moreover, the existing pool of copulas and marginal distributions can be expanded to have a wider choice in the modeling process. Apart from these improvements, the developed software can be adapted to other programming languages such as R and Python to reach wider group of users. Moreover, upgrading the quality of the interface by switching to the MATLAB App Designer could be a positive first step as it would make the GUI more user-friendly and up to date with the latest

interface component of MATLAB. Finally, the application of the copula modeling considered in this project to estimate sea level in different islands in the Pacific region could be relevant. The use of this method for future projection could provide important information related to sea level rise in these areas. With a framework already built and data accessible through national institutes, such analysis could provide a regional view of the issue.

REFERENCES

Abdi, A., Hassanzadeh, Y., Talatahari, S., Fakheri-Fard, A., & Mirabbasi, R. (2017). Parameter estimation of copula functions using an optimization-based method. Theoretical and Applied Climatology, 129(1), 21-32.

Abdulhafedh, A. (2017). How to detect and remove temporal autocorrelation in vehicular crash data. Journal of transportation technologies, 7(2), 133-147.

Adem, J. (1962). On the theory of the general circulation of the atmosphere. *Tellus, 14*(1), 102-115.

Aghakouchak, A. (2014). Entropy–copula in hydrology and climatology. Journal of Hydrometeorology, 15(6), 2176-2189.

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of climatology, 6*(6), 661-675.

Aminikhanghahi, S., & Cook, D. J. (2017). A Survey of Methods for Time Series Change Point Detection. *Knowl Inf Syst, 51*(2), 339-367. doi:10.1007/s10115-016-0987-z

Amir Jabbari, A., & Nazemi, A. (2019). Alterations in Canadian hydropower production potential due to continuation of historical trends in climate variables. *Resources, 8*(4), 163.

Anderson, D. G., Bissett, T. G., Yerka, S. J., Wells, J. J., Kansa, E. C., Kansa, S. W., . . . White, D. A. (2017). Sea-level rise and archaeological site destruction: An example from the southeastern United States using DINAA (Digital Index of North American Archaeology). *PloS one, 12*(11), e0188142.

Anderson, M. J., de Valpine, P., Punnett, A., & Miller, A. E. (2019). A pathway for multivariate analysis of ecological communities using copulas. Ecology and evolution, 9(6), 3276-3294.

Ardabili, S., Mosavi, A., Dehghani, M., & Várkonyi-Kóczy, A. R. (2019). *Deep learning and machine learning in hydrological processes climate change and earth systems a systematic review.* Paper presented at the International Conference on Global Research and Education.

Ash, J., & Campbell, J. (2016). Climate change and migration: the case of the Pacific Islands and Australia. *Journal of Pacific Studies, 36*(1), 53-72.

Aucan, J., Merrifield, M. A., & Pouvreau, N. (2017). Historical Sea Level in the South Pacific from Rescued Archives, Geodetic Measurements, and Satellite Altimetry. In *Pure and Applied Geophysics* (Vol. 174, pp. 3813-3823).

Bali, T. G. (2003). The generalized extreme value distribution. *Economics letters, 79*(3), 423-427.

Balistrocchi, M., & Bacchi, B. (2011). Modelling the statistical dependence of rainfall event variables by a trivariate copula function. *Hydrology & Earth System Sciences Discussions, 8*(1).

Barbero, R., & Moron, V. (2011). Seasonal to decadal modulation of the impact of El Niño-Southern Oscillation on New Caledonia (SW Pacific) rainfall (1950-2010). In *Journal of Geophysical Research Atmospheres* (Vol. 116, pp. 1-17).

Bari, S. H., Rahman, M. T. U., Hoque, M. A., & Hussain, M. M. (2016). Analysis of seasonal and annual rainfall trends in the northern region of Bangladesh. *Atmospheric Research, 176*, 148-158.

Becker, M., Meyssignac, B., Letetrel, C., Llovel, W., Cazenave, A., & Delcroix, T. (2012). Sea level variations at tropical Pacific islands since 1950. *Global and Planetary Change, 80*, 85-98.

Bedoui, R., & Dbabis, M. B. (2008). Copules et mesures du risque bidimensionnelles : application pratique aux hedge funds. In (pp. 1-29).

Bellard, C., Leclerc, C., & Courchamp, F. (2013). Potential impact of sea level rise on French islands worldwide. *Nature Conservation, 5*, 75.

Bender, J., Wahl, T., Müller, A., & Jensen, J. (2016). A multivariate design framework for river confluences. *Hydrological Sciences Journal, 61*(3), 471-482.

Bernardi, M., Durante, F., Jaworski, P., Petrella, L., & Salvadori, G. (2018). Conditional risk based on multivariate hazard scenarios. In *Stochastic Environmental Research and Risk Assessment* (Vol. 32, pp. 203-211): Springer Berlin Heidelberg.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., & Widmann, M. (2019). Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. *Science advances, 5*(9), eaaw5531.

Bisai, D., Chatterjee, S., Khan, A., & Barman, N. (2014). Application of Sequential Mann-Kendall Test for Detection of Approximate Significant Change Point in Surface Air Temperature for Kolkata Weather Observatory, West Bengal, India. *International Journal of Current Research, 6*, 5319-5324.

Bittermann, K., Rahmstorf, S., Perrette, M., & Vermeer, M. (2013). Predictability of twentieth century sea-level rise from past data. *Environmental Research Letters, 8*(1), 014013.

Bland, J. M., & Altman, D. G. (2015). Statistics notes: bootstrap resampling methods. *bmj, 350*.

Blumberg, A. F., & Bruno, M. S. (2018). *The urban ocean: the interaction of cities with water*: Cambridge University Press.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika, 65(1), 23-28.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*: John Wiley & Sons.

Bücher, A., & Volgushev, S. (2013). Empirical and sequential empirical copula processes under serial dependence. Journal of Multivariate Analysis, 119, 61-70.

Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology, 58*(1-2), 11-27.

Bunel, M. (2017). Décomposition du PIB par habitant, formation supérieure et capital humain en Nouvelle Calédonie, quelle économie pour la Nouvelle-Calédonie après la période référendaire?

Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. *Model selection and multimodel inference, 2*.

Campbell, J. D., Taylor, M. A., Stephenson, T. S., Watson, R. A., & Whyte, F. S. (2011). Future climate of the Caribbean from a regional climate model. *International Journal of Climatology, 31*(12), 1866-1878.

Cavanaugh, J. E., & Neath, A. A. (2019). The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. *Wiley Interdisciplinary Reviews: Computational Statistics, 11*(3), e1460.

Cazenave, A., & Nerem, R. S. (2004). Present-day sea level change: Observations and causes. *Reviews of Geophysics, 42*(3).

Chatfield, C. (2000). *Time-series forecasting*: CRC press.

Chen, F., Gao, Y., Wang, Y., Qin, F., & Li, X. (2019). Downscaling satellite-derived daily precipitation products with an integrated framework. *International Journal of Climatology, 39*(3), 1287-1304.

Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of hydrology, 401*(3-4), 190-202.

Chen, J., Wilson, C., & Tapley, B. (2013). Contribution of ice sheet and mountain glacier melt to recent sea level rise. *Nature Geoscience, 6*(7), 549-552.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula methods in finance*: John Wiley & Sons.

Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data.* University of Pittsburgh,

Church, J. A., Clark, P., Cazenave, A., Gregory, J. M., Jevrejeva, S., Levermann, A., . . . Nunn, P. (2013a). Sea level change. Climate change 2013: the physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1137-1216.

Church, J. A., Monselesan, D., Gregory, J. M., & Marzeion, B. (2013b). Evaluating the ability of process based models to project sea-level change. *Environmental Research Letters, 8*(1), 014051.

Church, J. A., & White, N. J. (2011). Sea-level rise from the late 19th to the early 21st century. Surveys in geophysics, 32(4), 585-602.

Church, J. A., White, N. J., Coleman, R., Lambeck, K., & Mitrovica, J. X. (2004). Estimates of the regional distribution of sea level rise over the 1950–2000 period. Journal of climate, 17(13), 2609-2625.

Church, J. A., White, N. J., & Hunter, J. R. (2006). Sea-level rise at tropical Pacific and Indian Ocean islands. Global and Planetary Change, 53(3), 155-168.

Churchill, S. A., Inekwe, J., Smyth, R., & Zhang, X. (2019). R&D intensity and carbon emissions in the G7: 1870–2014. *Energy Economics, 80*, 30-37.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., . . . Brekke, L. D. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research, 51*(4), 2498-2514.

Cowan, G. (1998). *Statistical data analysis*: Oxford university press.

D'Amico, G., & Petroni, F. (2018). Copula based multivariate semi-Markov models with applications in high-frequency finance. In *European Journal of Operational Research* (Vol. 267, pp. 765-777): Elsevier B.V.

De Michele, C., Salvadori, G., Passoni, G., & Vezzoli, R. (2007). A multivariate model of sea storms using copulas. *Coastal Engineering, 54*(10), 734-751.

Del Castillo, J., & Daoudi, J. (2009). Estimation of the generalized Pareto distribution. *Statistics & Probability Letters, 79*(5), 684-688.

Despax, A. (2016). Incertitude des mesures de débit des cours d'eau au courantomètre. Amélioration des méthodes analytiques et apports des essais interlaboratoires. Université Grenoble Alpes,

Dickinson, R. E., Errico, R. M., Giorgi, F., & Bates, G. T. (1989). A regional climate model for the western United States. *Climatic change, 15*(3), 383-422.

Drago, A., & Boxall, S. (2002). Use of the wavelet transform on hydro-meteorological data. *Physics and Chemistry of the Earth, Parts A/B/C, 27*(32-34), 1387-1399.

Dupuis, D. (2007). Using copulas in hydrology: Benefits, cautions, and issues. *Journal of hydrologic engineering, 12*(4), 381-393.

Durrleman, V., Nikeghbali, A., & Roncalli, T. (2000). Which copula is the right one? *Available at SSRN 1032545*.

Duvat, V. K. (2019). A global assessment of atoll island planform changes over the past decades. Wiley Interdisciplinary Reviews: Climate Change, 10(1), e557.

El Maghraby, E., Abu Khadra, S., & Eissa, H. (2016). Using the Fast Fourier Transform Technique for Climate Time Series Decomposition. *Arab Journal of Nuclear Sciences and Applications, 49*(3), 78-85.

Ellison, J. C. (1993). Mangrove retreat with rising sea-level, Bermuda. Estuarine, Coastal and Shelf Science, 37(1), 75-87.

Embrechts, P., Lindskog, F., & McNeil, A. (2001). Modelling dependence with copulas. Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich, 14.

FAO. (2020). *La situation mondiale des pêches et de l'aquaculture 2020. La durabilité en action.* Retrieved from Rome, Italy:

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., . . . Ebel, B. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of hydrology, 537*, 45-60.

Favre, A. C., El Adlouni, S., Perreault, L., Thiémonge, N., & Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research, 40*(1).

Fischer, M., Dewitte, B., & Maîtrepierre, L. (2004). A non-linear statistical downscaling model: El Niño/Southern Oscillation impact on precipitation over New Caledonia. In *Geophysical Research Letters* (Vol. 31).

Frees, E. W., & Wang, P. (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics, 38*(2), 360-373.

Friess, D. A., Rogers, K., Lovelock, C. E., Krauss, K. W., Hamilton, S. E., Lee, S. Y., . . . Shi, S. (2019). The state of the world's mangrove forests: past, present, and future. *Annual Review of Environment and Resources, 44*, 89-115.

Genest, C., & Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering, 12*(4), 347-368.

Genest, C., Favre, A. C., Béliveau, J., & Jacques, C. (2007). Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resources Research, 43*(9).

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics, 44(2), 199-213.

Giorgi, F. (2019). Thirty years of regional climate modeling: where are we and where are we going next? *Journal of Geophysical Research: Atmospheres, 124*(11), 5696-5723.

Goffart, J., Rabouille, M., & Mendes, N. (2016). *Impact des incertitudes de mesure des variables météorologiques sur le processus de comparaison mesure/simulation en simulation thermique dynamique.* Paper presented at the Conférence IBPSA France. Marne-la-Vallée.

Gornitz, V., Lebedeff, S., & HANSEN, J. (1982). Global sea level trend in the past century. *Science, 215*(4540), 1611-1614.

Grbić, R., Kurtagić, D., & Slišković, D. (2013). Stream water temperature prediction based on Gaussian process regression. *Expert systems with applications, 40*(18), 7407-7414.

Gregory, J., Church, J., Boer, G., & Dixon, K. (2001). Comparison of results from several AOGCMs for global and regional sea-level change 1900–2100 Cited by me. In *Climate Dynamics*.

Hallinan Jr, A. J. (1993). A review of the Weibull distribution. *Journal of Quality Technology, 25*(2), 85-93.

Hamlington, B., Frederikse, T., Thompson, P., Willis, J., Nerem, R., & Fasullo, J. (2021). Past, Present, and Future Pacific Sea-Level Change. *Earth's Future, 9*(4), e2020EF001839.

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. In *Reviews of Geophysics* (Vol. 48, pp. 1-29).

Hassanzadeh, E., Elshorbagy, A., Wheater, H., & Gober, P. (2016). A risk-based framework for water resource management under changing water availability, policy options, and irrigation expansion. *Advances in Water Resources, 94*, 291-306.

Hassanzadeh, E., Nazemi, A., Adamowski, J., Nguyen, T. H., & Van-Nguyen, V. T. (2019). Quantile-based downscaling of rainfall extremes: Notes on methodological functionality, associated uncertainty and application in practice. Advances in Water Resources, 131, 103371.

Hatami, S., & Nazemi, A. (2021). A statistical framework for assessing temperature controls on landscape Freeze-Thaw: Application and implications in Québec, Canada (1979–2016). *Journal of hydrology*, *603*, 126891.

Hatami, S., Nazemi, A. & Amirjabbari, A. Evolving Trends of Rain over Precipitation in Canadian Cold Season During the late 20th Century. in Canadian Society for Civil Engineering (pp. 1–5), Laval. https://www.csce.ca/elf/apps/CONFERENCEVIEWER/conferences/2019/pdfs/PaperPDFversion_8_0513041903.pdf (retrieved on 2021-11-22).

Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data.

Hawkins, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American statistical association*, *72*(357), 180-186.

Helsel, D. R., & Hirsch, R. M. (1992). Statistical methods in water resources. In *Statistical methods in water resources*.

Hess, A., Iyer, H., & Malm, W. (2001). Linear trend analysis: a comparison of methods. *Atmospheric Environment*, *35*(30), 5211-5222.

Hofert, M. (2012). A stochastic representation and sampling algorithm for nested Archimedean copulas. Journal of Statistical Computation and Simulation, 82(9), 1239-1255.

Holder, R. (1985). Multiple regression in hydrology: Institute of hydrology.

Horton, R., Herweijer, C., Rosenzweig, C., Liu, J., Gornitz, V., & Ruane, A. C. (2008). Sea level rise projections for current generation CGCMs based on the semi-empirical method. *Geophysical Research Letters*, *35*(2).

Imani, M., You, R.-J., & Kuo, C.-Y. (2014). Forecasting Caspian Sea level changes using satellite altimetry data (June 1992–December 2013) based on evolutionary support vector regression algorithms and gene expression programming. *Global and Planetary Change*, *121*, 53-63.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

ISEE. (2021). Retrieved from https://www.isee.nc/population/recensement/structure-de-la-population-et-evolutions

Jacques, G., & Le Treut, H. (2004). *Le changement climatique*: Unesco.

Jane, R., Dalla Valle, L., Simmonds, D., & Raby, A. (2016). A copula-based approach for the estimation of wave height records through spatial correlation. Coastal Engineering, 117, 1-18.

Jaworski, P., Durante, F., Hardle, W. K., & Rychlik, T. (2010). Copula theory and its applications (Vol. 198): Springer.

Jehanzaib, M., & Kim, T.-W. (2020). Exploring the influence of climate change-induced drought propagation on wetlands. Ecological Engineering, 149, 105799.

Jevrejeva, S., Moore, J. C., & Grinsted, A. (2010). How will sea level respond to changes in natural and anthropogenic forcings by 2100? *Geophysical Research Letters, 37*(7).

Joe, H. (1997). Multivariate models and multivariate dependence concepts: CRC Press.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2* (Vol. 289): John wiley & sons.

Joo, K., Shin, J.-Y., & Heo, J.-H. (2020). Modified maximum pseudo likelihood method of copula parameter estimation for skewed hydrometeorological data. *Water, 12*(4), 1182.

Joussaume, S. (1993). Climate from yesterday until tomorrow. Climat d'hier a demain.

Kang, H. M., & Yusof, F. (2012). Homogeneity tests on daily rainfall series. *Int. J. Contemp. Math. Sciences, 7*(1), 9-22.

Kendall, M. (1976). Rank Auto Correlation Methods, Griffin. In: Oxford.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81-93.

Khaliq, M. N., & Ouarda, T. B. (2007). On the critical values of the standard normal homogeneity test (SNHT). *International Journal of Climatology: A Journal of the Royal Meteorological Society, 27*(5), 681-687.

Khanal, S., Ridder, N., de Vries, H., Terink, W., & van den Hurk, B. (2019). Storm surge and extreme river discharge: a compound event analysis using ensemble impact modeling. *Frontiers in Earth Science, 7*, 224.

Ko, V., Hjort, N. L., & Hobæk Haff, I. (2019). Focused information criteria for copulas. Scandinavian Journal of Statistics, 46(4), 1117-1140.

Kozłowski, E., Kowalska, B., Kowalski, D., & Mazurkiewicz, D. (2018). Water demand forecasting by trend and harmonic analysis. *Archives of Civil and Mechanical Engineering, 18*(1), 140-148.

Kwak, S. G., & Kim, J. H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology, 70*(2), 144.

Le Cozannet, G., Garcin, M., Yates, M., Idier, D., & Meyssignac, B. (2014). Approaches to evaluate the recent impacts of sea-level rise on shoreline changes. Earth-science reviews, 138, 47-60.

Le Duff, M., Dumas, P., Allenbach, M., & Cohen, O. (2020). An orientation for coastal disaster risks management and prevention policy in a global warming context: Case study in Ouvea (New Caledonia). Marine Policy, 117.

Leta, O. T., El-Kadi, A. I., & Dulai, H. (2018). Impact of climate change on daily streamflow and its extreme values in pacific island watersheds. *Sustainability, 10*(6), 2057.

Lincke, D., & Hinkel, J. (2021). Coastal Migration due to 21st Century Sea-Level Rise. Earth's Future, 9(5), e2020EF001965.

Liu, Y. F., & Fan, X. P. (2020). Dynamic reliability prediction for the steel box girder based on multivariate Bayesian dynamic Gaussian copula model and SHM extreme stress data. Structural Control and Health Monitoring, 27(6), e2531.

Lomnicki, Z., & Zaremba, S. (1957). On the estimation of autocorrelation in time series. The Annals of Mathematical Statistics, 28(1), 140-158.

Lovelock, C. E., Cahoon, D. R., Friess, D. A., Guntenspergen, G. R., Krauss, K. W., Reef, R., . . . Swales, A. (2015). The vulnerability of Indo-Pacific mangrove forests to sea-level rise. *Nature, 526*(7574), 559-563.

Lukacs, E. (1955). A characterization of the gamma distribution. *The Annals of Mathematical Statistics, 26*(2), 319-324.

Magnes, W., Hillenmaier, O., Auster, H.-U., Brown, P., Kraft, S., Seon, J., . . . Fischer, D. (2020). Space Weather Magnetometer Aboard GEO-KOMPSAT-2A. *Space Science Reviews, 216*(8), 1-36.

Maity, R. (2018). *Statistical methods in hydrology and hydroclimatology*: Springer.

Manabe, S., Smagorinsky, J., & Strickler, R. F. (1965). Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly weather review, 93*(12), 769-798.

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, 245-259.

Manton, M. J., Della-Marta, P. M., Haylock, M. R., Hennessy, K. J., Nicholls, N., Chambers, L. E., . . . Yee, D. (2001). Trends in extreme daily rainfall and temperature in southeast Asia and the south Pacific: 1961-1998. In *International Journal of Climatology* (Vol. 21, pp. 269-284).

MATLAB-Cusum-test. (2021). Retrieved from https://www.mathworks.com/help/signal/ref/cusum.html

McCabe, B. P., & Tremayne, A. R. (1995). Testing a time series for difference stationarity. *The Annals of Statistics*, 1015-1028.

MétéoFrance. (2021). Retrieved from http://www.meteo.nc/nouvelle-caledonie/climat/releves

Meyers, S. D., Kelly, B. G., & O'Brien, J. J. (1993). An introduction to wavelet analysis in oceanography and meteorology: With application to the dispersion of Yanai waves. *Monthly weather review, 121*(10), 2858-2866.

Milionis, A. E. (2004). The importance of variance stationarity in economic time series modelling. A practical approach. *Applied Financial Economics, 14*(4), 265-278.

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Compounding effects of sea level rise and fluvial flooding. *Proceedings of the National Academy of Sciences, 114*(37), 9785-9790.

Mondal, D. (2013). Assessing Social Vulnerability to Coastal Hazards: An Examination on Sagar Island of Sundarban Delta. *Research Journal of Humanities and Social Sciences, 4*(2), 210-215.

Moore, J. C., Grinsted, A., Zwinger, T., & Jevrejeva, S. (2013). Semiempirical and process-based global sea level projections. *Reviews of Geophysics, 51*(3), 484-522.

Mote, P., & Salathé, E. (2010). Future Climate in the Pacific Northwest. Climatic change, 102(1), 29-50. doi:10.1007/s10584-010-9848-z

Mudelsee, M. (2019). Trend analysis of climate time series: A review of methods. *Earth-science reviews, 190*, 310-322.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature, 403*(6772), 853-858.

Nadarajah, S., & Kotz, S. (2004). The beta Gumbel distribution. *Mathematical Problems in engineering, 2004*(4), 323-332.

Naz, S., Jamil, S. S., & Iqbal, M. J. (2020). Archimedean copula-based bivariate flood-frequency analysis on Sukkur, Pakistan. *Arabian Journal of Geosciences, 13*(7), 1-13.

Nazemi, A., & Elshorbagy, A. (2012). Application of copula modelling to the performance assessment of reconstructed watersheds. *Stochastic Environmental Research and Risk Assessment, 26*(2), 189-205.

Nazemi, A., & Wheater, H. S. (2014). Assessing the vulnerability of water supply to changing streamflow conditions. Eos, Transactions American Geophysical Union, 95(32), 288-288.

Nazemi, A., Wheater, H. S., Chun, K. P., Bonsal, B., & Mekonnen, M. (2017). Forms and drivers of annual streamflow variability in the headwaters of Canadian Prairies during the 20th century. Hydrological Processes, 31(1), 221-239.

Nazemi, A., Zaerpour, M., & Hassanzadeh, E. (2020). Uncertainty in Bottom-Up Vulnerability Assessments of Water Supply Systems due to Regional Streamflow Generation under Changing Conditions. *Journal of Water Resources Planning and Management, 146*(2), 04019071.

NCARF. (2013., Last modified 22 Jul 2103). The Climate Data Guide: Empirical Orthogonal Function (EOF) Analysis and Rotated EOF Analysis. Retrieved from https://climatedataguide.ucar.edu/climate-data-tools-and-analysis/empirical-orthogonal-function-eof-analysis-and-rotated-eof-analysis.

Neath, A. A., & Cavanaugh, J. E. (2012). The Bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics, 4*(2), 199-203.

Nelsen, R. (2006). Springer series in statistics, An introduction to copulas. In: Springer Science+ Business Media, New York, NY.

Nerem, R. S., Beckley, B. D., Fasullo, J. T., Hamlington, B. D., Masters, D., & Mitchum, G. T. (2018). Climate-change–driven accelerated sea-level rise detected in the altimeter era. Proceedings of the National Academy of Sciences, 115(9), 2022-2025.

Nesme-Ribes, E., Thuillier, G., & Pecker, J.-C. (2000). *Histoire solaire et climatique*: Belin Paris.

Nguyen, V. A., Zhang, X., Blanchet, J., & Georghiou, A. (2020). Distributionally robust parametric maximum likelihood estimation. *arXiv preprint arXiv:2010.05321*.

Niedzielski, T., & Kosek, W. (2009). Forecasting sea level anomalies from TOPEX/Poseidon and Jason-1 satellite altimetry. In *Journal of Geodesy* (Vol. 83, pp. 469-476).

NOAA. (2021). Climate Change: Global Sea Level. Retrieved from https://www.climate.gov/news-features/understanding-climate/climate-change-global-sea-level

Nordell, B. (2003). Thermal pollution causes global warming. *Global and Planetary Change, 38*(3-4), 305-312.

Nunn, P. D. (2013). The end of the Pacific? Effects of sea level rise on Pacific Island livelihoods. *Singapore Journal of Tropical Geography, 34*(2), 143-171.

Nunn, P. D., Kumar, L., Eliot, I., & McLean, R. F. (2016). Classifying pacific islands. *Geoscience Letters, 3*(1), 1-19.

Nunn, P. D., & Mimura, N. (1997). Vulnerability of South Pacific island nations to sea-level rise. *Journal of coastal research*, 133-151.

Orlić, M., & Pasarić, Z. (2013). Semi-empirical versus process-based sea-level projections for the twenty-first century. *Nature Climate Change, 3*(8), 735-738.

Ouillon, S., Douillet, P., Fichez, R., & Panché, J.-Y. (2005). Enhancement of regional variations in salinity and temperature in a coral reef lagoon, New Caledonia. *Comptes Rendus Geoscience, 337*(16), 1509-1517.

Parris, A. S., Bromirski, P., Burkett, V., Cayan, D. R., Culver, M. E., Hall, J., ... Obeysekera, J. (2012). Global sea level rise scenarios for the United States National Climate Assessment.

Patton, A. J. (2009). Copula–based models for financial time series. In *Handbook of financial time series* (pp. 767-785): Springer.

Payri, C. E., Allain, V., Aucan, J., David, C., David, V., Dutheil, C., . . . Pestana, G. (2019). New Caledonia. In *World Seas: an Environmental Evaluation* (pp. 593-618): Elsevier.

Peng, L., & Welsh, A. (2001). Robust estimation of the generalized pareto distribution. *Extremes, 4*(1), 53-65.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 28*(2), 126-135.

Pilkey, O. H., Pilkey-Jarvis, L., & Pilkey, K. C. (2016). *Retreat from a rising sea*: Columbia University Press.

Pohlert, T. (2016). Non-parametric trend tests and change-point detection. *CC BY-ND, 4*.

Popp, T., Hegglin, M. I., Hollmann, R., Ardhuin, F., Bartsch, A., Bastos, A., . . . Buchwitz, M. (2020). Consistency of satellite climate data records for Earth system monitoring. *Bulletin of the American Meteorological Society, 101*(11), E1948-E1971.

Rahmstorf, S. (2007). A semi-empirical approach to projecting future sea-level rise. *Science, 315*(5810), 368-370.

Rahmstorf, S. (2010). A new view on sea level rise. *Nature Climate Change, 1*(1004), 44-45.

Rahmstorf, S. (2012). Modeling sea level rise. *Nature Education Knowledge, 3*(10), 4.

Rahmstorf, S., Perrette, M., & Vermeer, M. (2012). Testing the robustness of semi-empirical sea level projections. *Climate Dynamics, 39*(3-4), 861-875.

Reeves, J., Chen, J., Wang, X. L., Lund, R., & Lu, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of applied meteorology and climatology, 46*(6), 900-915.

Sadegh, M., Moftakhari, H., Gupta, H. V., Ragno, E., Mazdiyasni, O., Sanders, B., . . . AghaKouchak, A. (2018). Multihazard scenarios for analysis of compound extreme events. *Geophysical Research Letters, 45*(11), 5470-5480.

Sahin, O., Stewart, R. A., Faivre, G., Ware, D., Tomlinson, R., & Mackey, B. (2019). Spatial Bayesian Network for predicting sea level rise induced coastal erosion in a small Pacific Island. *Journal of environmental management, 238*, 341-351.

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *Dordrecht, The Netherlands: D. Reidel, 81*(10.5555), 26853.

Salleh, N., Yusof, F., & Yusop, Z. (2016). Bivariate copulas functions for flood frequency analysis. In *AIP Conference Proceedings* (Vol. 1750).

Salvadori, G., & De Michele, C. (2004). Frequency analysis via copulas: Theoretical aspects and applications to hydrological events. *Water Resources Research, 40*(12).

Salvadori, G., & De Michele, C. (2007). On the use of copulas in hydrology: theory and practice. *Journal of hydrologic engineering, 12*(4), 369-380.

Salvadori, G., De Michele, C., & Durante, F. (2011). On the return period and design in a multivariate framework. In *Hydrology and Earth System Sciences* (Vol. 15, pp. 3293-3305).

Salvadori, G., Durante, F., & De Michele, C. (2013). Multivariate return period calculation via survival functions. In *Water Resources Research* (Vol. 49, pp. 2308-2311).

Salvadori, G., F. Durante, C. De Michele, M. Bernardi, & Petrella, L. (2016). A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. In *Water Resources Research* (Vol. 5, pp. 2-2).

Santer, B. D., Wigley, T., Boyle, J., Gaffen, D. J., Hnilo, J., Nychka, D., . . . Taylor, K. (2000). Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *Journal of Geophysical Research: Atmospheres, 105*(D6), 7337-7356.

Sayemuzzaman, M., & Jha, M. K. (2014). Seasonal and annual precipitation time series trend analysis in North Carolina, United States. *Atmospheric Research, 137*, 183-194.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., & Ratier, A. (2002). An introduction to Meteosat second generation (MSG). *Bulletin of the American Meteorological Society, 83*(7), 977-992.

Schmidli, J., Frei, C., & Vidale, P. L. (2006). Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology: A Journal of the Royal Meteorological Society, 26*(5), 679-689.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association, 63*(324), 1379-1389.

Şen, Z. (2012). Innovative trend analysis methodology. *Journal of hydrologic engineering, 17*(9), 1042-1046.

Shiau, J. (2003). Return period of bivariate distributed extreme hydrological events. *Stochastic Environmental Research and Risk Assessment, 17*(1), 42-57.

Siano, D. B. (1972). The log-normal distribution function. *Journal of Chemical Education, 49*(11), 755.

Sirimanne, S. N., Hoffman, J., Juan, W., Asariotis, R., Assaf, M., Ayala, G., . . . Premti, A. (2019). *Review of maritime transport 2019*. Retrieved from

Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. *Publ. inst. statist. univ. Paris, 8*, 229-231.

Slangen, A. B., Meyssignac, B., Agosta, C., Champollion, N., Church, J. A., Fettweis, X., . . . Palmer, M. D. (2017). Evaluating model simulations of twentieth-century sea level rise. Part I: Global mean sea level change. *Journal of climate, 30*(21), 8539-8563.

Stedman, R. C., D. J. Davidson, and A. Wellstead (2004), Risk and climate change: Perceptions of key policy actors in Canada, Risk Anal., 24(5), 1395–406, doi:10.1111/j.0272-4332.2004.00534.x.

Sun, A. Y., Wang, D., & Xu, X. (2014). Monthly streamflow forecasting using Gaussian process regression. *Journal of hydrology, 511*, 72-81.

Sweet, W., Horton, R., Kopp, R., & Romanou, A. (2017). Sea level rise.

Talke, S. A., Orton, P., & Jay, D. A. (2014). Increasing storm tides in New York harbor, 1844–2013. *Geophysical Research Letters, 41*(9), 3149-3155.

Tartakovsky, A., Nikiforov, I., & Basseville, M. (2014). *Sequential analysis: Hypothesis testing and changepoint detection*: CRC Press.

Terry, J., Kostaschuk, R. A., & Wotling, G. (2008). Features of tropical cyclone-induced flood peaks on Grande Terre, New Caledonia. *Water and Environment Journal, 22*(3), 177-183.

Terry, J., & Wotling, G. (2011). Rain-shadow hydrology: Influences on river flows and flood magnitudes across the central massif divide of La Grande Terre Island, New Caledonia. In *Journal of hydrology* (Vol. 404, pp. 77-86): Elsevier B.V.

Tootoonchi, F., Haerter, J. O., Räty, O., Grabs, T., Sadegh, M., & Teutschbein, C. (2020). Copulas for hydroclimatic applications–A practical note on common misconceptions and pitfalls. Hydrology and Earth System Sciences Discussions, 1-31.

Turki, I., Laignel, B., Kakeh, N., Chevalier, L., & Costa, S. (2015). A new hybrid model for filling gaps and forecast in sea level: application to the eastern English Channel and the North Atlantic Sea (western France). In *Ocean Dynamics* (Vol. 65, pp. 509-521).

UNESCO. (2021). Retrieved from https://whc.unesco.org/en/list/1115/

Vermeer, M., & Rahmstorf, S. (2009). Global sea level linked to global temperature. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 106, pp. 21527-21532).

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S., & Bierkens, M. F. P. (2010). Global depletion of groundwater resources. *Geophysical Research Letters, 37*(20), n/a-n/a. doi:10.1029/2010gl044571

Walsh, K. J., McInnes, K. L., & McBride, J. L. (2012). Climate change impacts on tropical cyclones and extreme sea levels in the South Pacific—A regional assessment. *Global and Planetary Change, 80*, 149-164.

Wang, X. L. (2008). Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. Journal of applied meteorology and climatology, 47(9), 2423-2444.

Wang, Z., Wang, Z., Yu, S., & Zhang, K. (2019). Time-dependent mechanism reliability analysis based on envelope function and vine-copula function. Mechanism and Machine Theory, 134, 667-684.

Wani, O., Scheidegger, A., Cecinati, F., Espadas, G., & Rieckermann, J. (2019). Exploring a copula-based alternative to additive error models—for non-negative and autocorrelated time series in hydrology. Journal of hydrology, 575, 1031-1040.

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I. D., Hendry, A., Muis, S., . . . Wahl, T. (2018). Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries. In *Environmental Research Letters* (Vol. 13).

Wei, P., Lu, Z., & Song, J. (2014). Moment-independent sensitivity analysis using copula. Risk Analysis, 34(2), 210-222.

Weiß, G. (2011). Copula parameter estimation by maximum-likelihood and minimum-distance estimators: a simulation study. Computational Statistics, 26(1), 31-54.

Whan, K., Alexander, L., Imielska, A., McGree, S., Jones, D., Ene, E., . . . Kumar, R. (2014). Trends and variability of temperature extremes in the tropical Western Pacific. *International Journal of Climatology, 34*(8), 2585-2603.

Wick, G. A., Bates, J. J., & Scott, D. J. (2002). Satellite and skin-layer effects on the accuracy of sea surface temperature measurements from the GOES satellites. *Journal of Atmospheric and Oceanic Technology, 19*(11), 1834-1848.

Wilby, R. L. (2017). *Climate change in practice: topics for discussion with group exercises*: Cambridge University Press.

Yang, A., Si, Y., Fan, Y., Kong, X., & Lin, X. (2019). Correlation study of rainfall and runoff in Xiangxi River based on Archimedean copula function. In IOP conference series: earth and environmental science (Vol. 223, No. 1, p. 012055). IOP Publishing.

Yavuzdoğan, A., & Tanır Kayıkçı, E. (2020). A copula approach for sea level anomaly prediction: a case study for the Black Sea. *Survey Review*, 1-11.

Yue, S., & Pilon, P. (2004). A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection/Une comparaison de la puissance des tests t de Student, de Mann-Kendall et du bootstrap pour la détection de tendance. *Hydrological Sciences Journal, 49*(1), 21-37.

Zaerpour, M., Hatami, S., Sadri, J., & Nazemi, A. (2021). A global algorithm for identifying changing streamflow regimes: application to Canadian natural streams (1966–2010). Hydrology and Earth System Sciences, 25(9), 5193-5217.

Zammali, M. K., Hassanzadeh, E., Shupena-Soulodre, E., & Lindenschmidt, K. E. (2021). A generic approach to evaluate costs and effectiveness of agricultural Beneficial Management Practices to improve water quality management. Journal of Environmental Management, 287, 112336.

Zandmoghaddam, S., Nazemi, A., Hassanzadeh, E., & Hatami, S. (2019). Representing local dynamics of water resource systems through a data-driven emulation approach. *Water resources management, 33*(10), 3579-3594.

Zaporozhets, A., Redko, O., Babak, V., Eremenko, V., & Mokiychuk, V. (2018). Method of indirect measurement of oxygen concentration in the air. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, 5*, 105-114.

Zhang, X., & Church, J. A. (2012). Sea level trends, interannual and decadal variability in the Pacific Ocean. *Geophysical Research Letters, 39*(21).

Zhang, Y., Kim, C.-W., Zhang, L., Bai, Y., Yang, H., Xu, X., & Zhang, Z. (2020). Long term structural health monitoring for old deteriorated bridges: a copula-ARMA approach. *Smart Structures and Systems, 25*(3), 285-299.

Zhang, Q., Xiao, M., Singh, V. P., & Chen, X. (2013). Copula-based risk evaluation of droughts across the Pearl River basin, China. Theoretical and Applied Climatology, 111(1), 119-131.

Zhang, Z., Zhang, K., & Khelifi, A. (2018). *Multivariate time series analysis in climate and environmental research*: Springer.

Zhao, J., Fan, Y., & Mu, Y. (2019). Sea level prediction in the Yellow Sea from satellite altimetry with a combined least squares-neural network approach. Marine geodesy, 42(4), 344-366.

Zorzetto, E., & Marani, M. (2019). Downscaling of rainfall extremes from satellite observations.

Water Resources Research, 55(1), 156-174.

APPENDIX A SUPPLEMENTARY MATERIAL RELATED TO CHAPTER 3

In this project, observations from meteorological stations and tide gauges have been retrieved from partnership signed with a national institute and from data available online. Thus, temperature and precipitation datasets have been provided by the stations network of Meteo France in New-Caledonia represented in red triangles in Figure A.1 below. Sea level observations (blue squares) have been retrieved from the REFMAR network (Réseau de REFérence des observations MARégraphiques in blue) under the supervision of the SHOM (Service Hydrographique et Océanographique de la Marine). Datasets from a total of 8 and 9 stations have been considered for sea level and temperature/precipitation respectively – see Figure A.1. The characteristics of each hydroclimate stations are detailed in the Table A.1. Considering the quality of each dataset through the missing data, it appears that stations located in Nouméa are the most reliable observations for all the hydroclimate variables. Regarding sea level especially, the other tide gauges present a ratio of missing data too important to be considered for analysis. In meteorological stations, every dataset shows few missing data (under 2.5%). Thus, with the highest population density in the territory, Nouméa is the place considered with interest for this study.

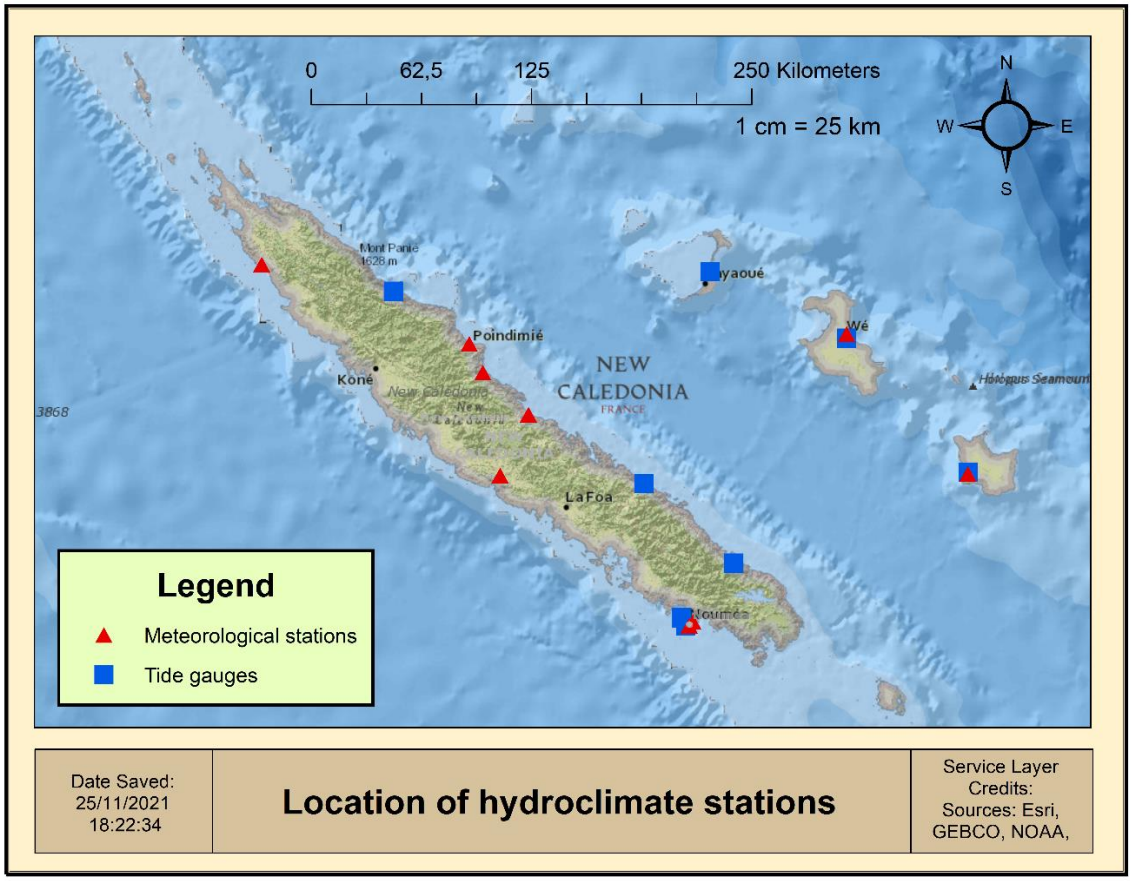

Figure A.1 Hydroclimate stations in New-Caledonia retrieved from Meteo France (red) and REFMAR (blue)

Table A.1 Characteristics of the hydroclimate stations provided in New-Caledonia (stations located in Nouméa are highlighted)

TEMPERATURE DATASETS (▲)						
Station Name	**Record length**	**Timestep**	**Indices**		**Missing data**	
NOUMÉA	1950-2020	Daily	Min	Max	0%	0%
MAGENTA	1964-2020	Daily	Min	Max	0.034%	0.034%
NESSADIOU	1967-2020	Daily	Min	Max	0.42%	0.42%
LA ROCHE	1970-2020	Daily	X	Max	X	0.79%
HOUAILOU	1952-2020	Daily	X	Max	X	1.038%
PONERIHOUEN	1969-2020	Daily	Min	X	4.21%	X
POINIDIMIE	1964-2020	Daily	Min	Max	0.34%	0.29%
LIFOU	1960-2020	Daily	Min	Max	0.036%	0.05%
KOUMAC	1951-2020	Daily	Min	Max	0.024%	0.02%

RAINFALL DATASETS (▲)			
Station Name	**Record length**	**Timestep**	**Missing data**
HOUAILOU	1952-2020	Daily	1.26%
KOUMAC	1951-2020	Daily	0.004%
NOUMEA	1950-2020	Daily	0%
LIFOU	1960-2020	Daily	1.66%
POINDIMIE	1964-2020	Daily	2.45%
PONERIHOUEN	1952-2020	Daily	0.052%

Table A.1 Characteristics of the hydroclimate stations provided in New-Caledonia (stations located in Nouméa are highlighted) (suite)

SEA LEVEL DATASETS (■)			
Station Name	**Record length**	**Timestep**	**Missing data**
POINTE-CHALEIX	1967-2005	Hourly	4.65%
NUMBO	2001-2020	Hourly	21.50%
OUINNE	1981-2020	Hourly	76.20%
THIO	1967-2020	Hourly	83.85%
MARE	1982-2020	Hourly	86.38%
LIFOU	1969-2020	Hourly	85.44%
HIENGHÈNE	1983-2020	Hourly	85.34%
OUVÉA	1981-2020	Hourly	93.75%

APPENDIX B SUPPLEMENTARY MATERIAL RELATED TO CHAPTER 4

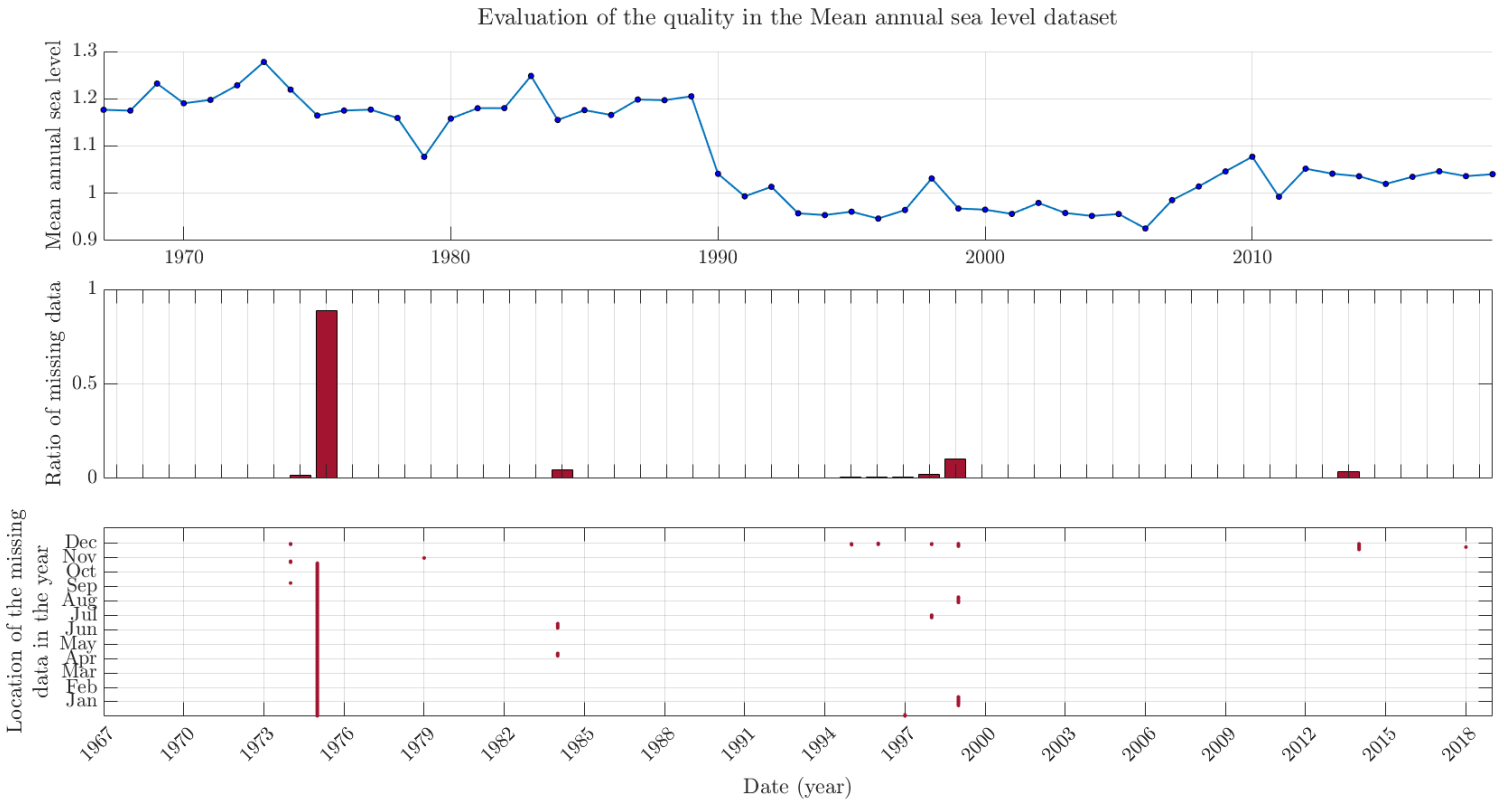

Figure B.1 Estimation of the quality for the dataset of mean sea level considered at an annual timestep

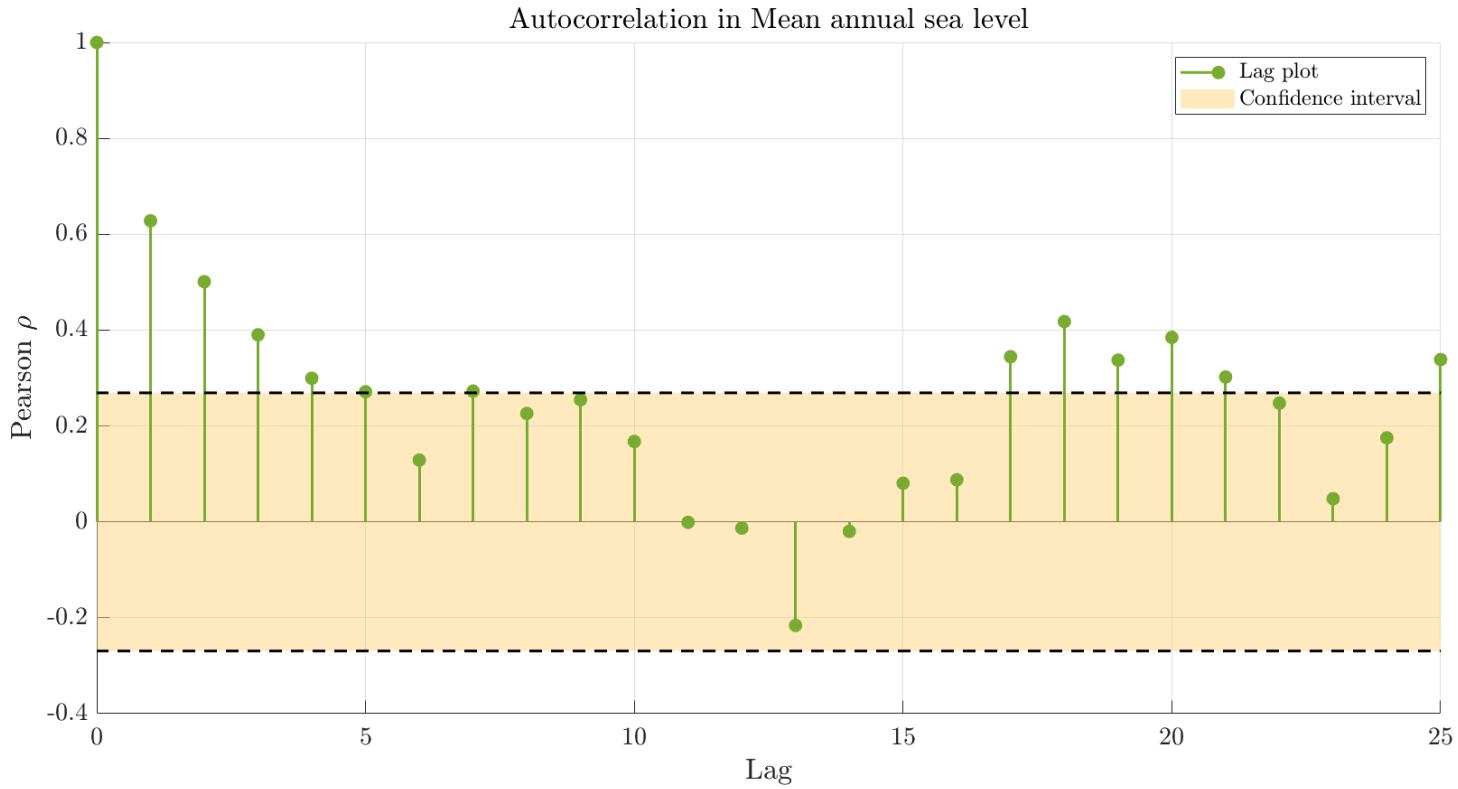

Figure B.2 Lag-plot used to estimate the lag-k autocorrelation in the mean annual sea level time series using Pearson's ρ

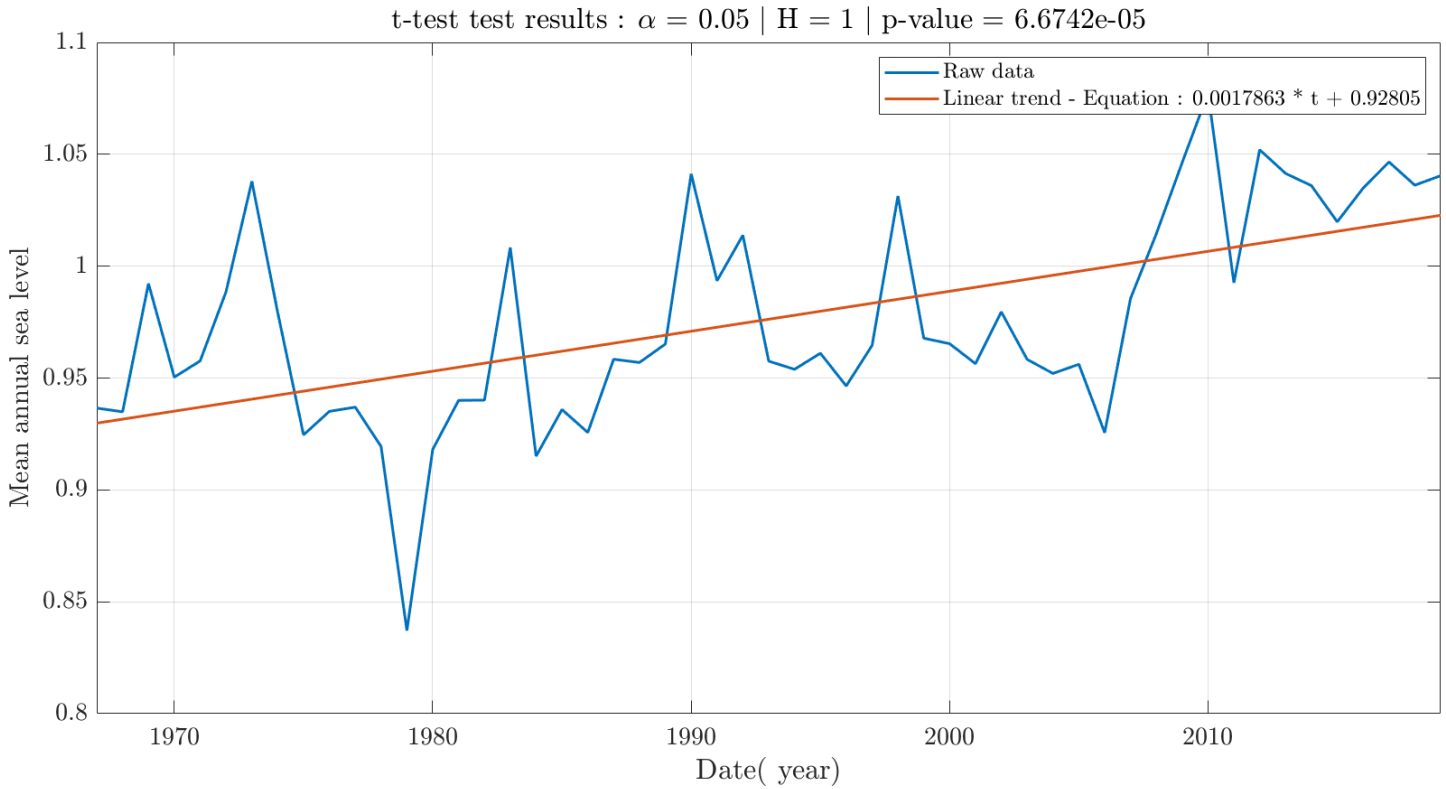

Figure B.3 Linear trend estimation (orange line) with bootstrapped p-value estimation

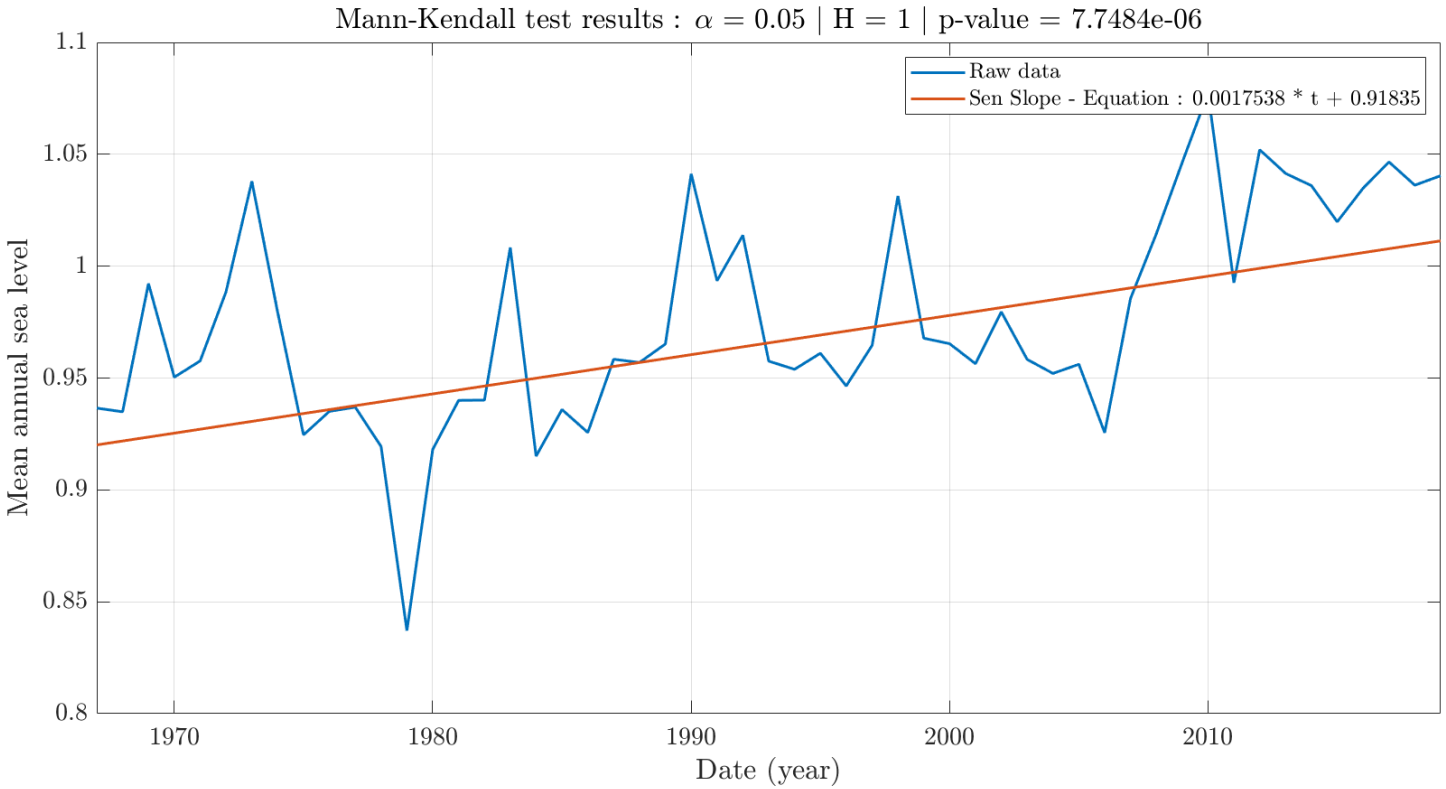

Figure B.4 Sen's slope estimation (orange line) with Mann-Kendall statistical test p-value

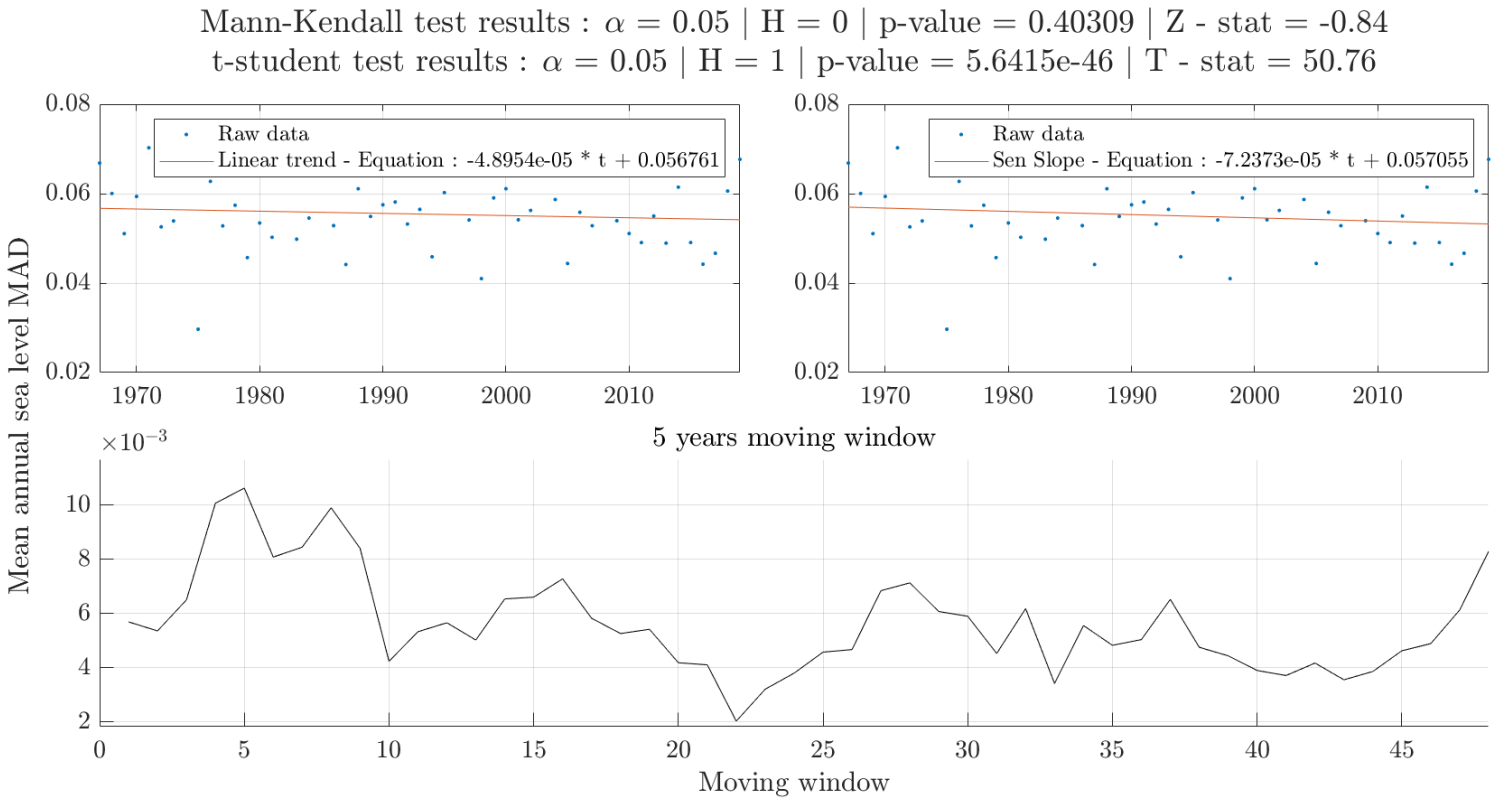

Figure B.5 Analysis of mean annual sea level variability through MAD with trend estimation (on top row, linear trend in the left graph and Sen's slope in the right graph) and moving window evolution (bottom row)

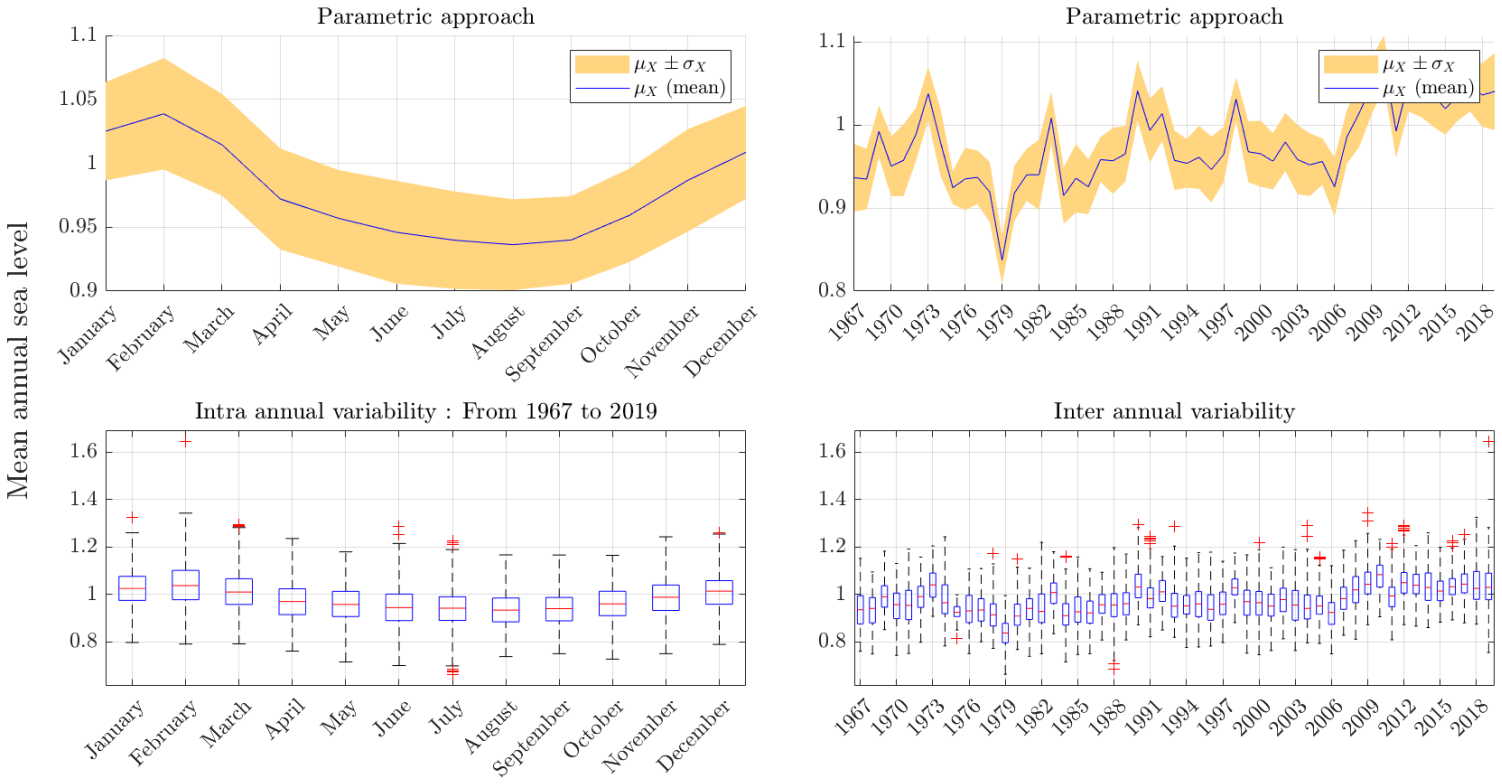

Figure B.6 Intra annual (left columns) and inter annual (right columns) variability assessment of mean annual sea level using general statistic methods (top rows) and non-parametric methods (bottom rows)

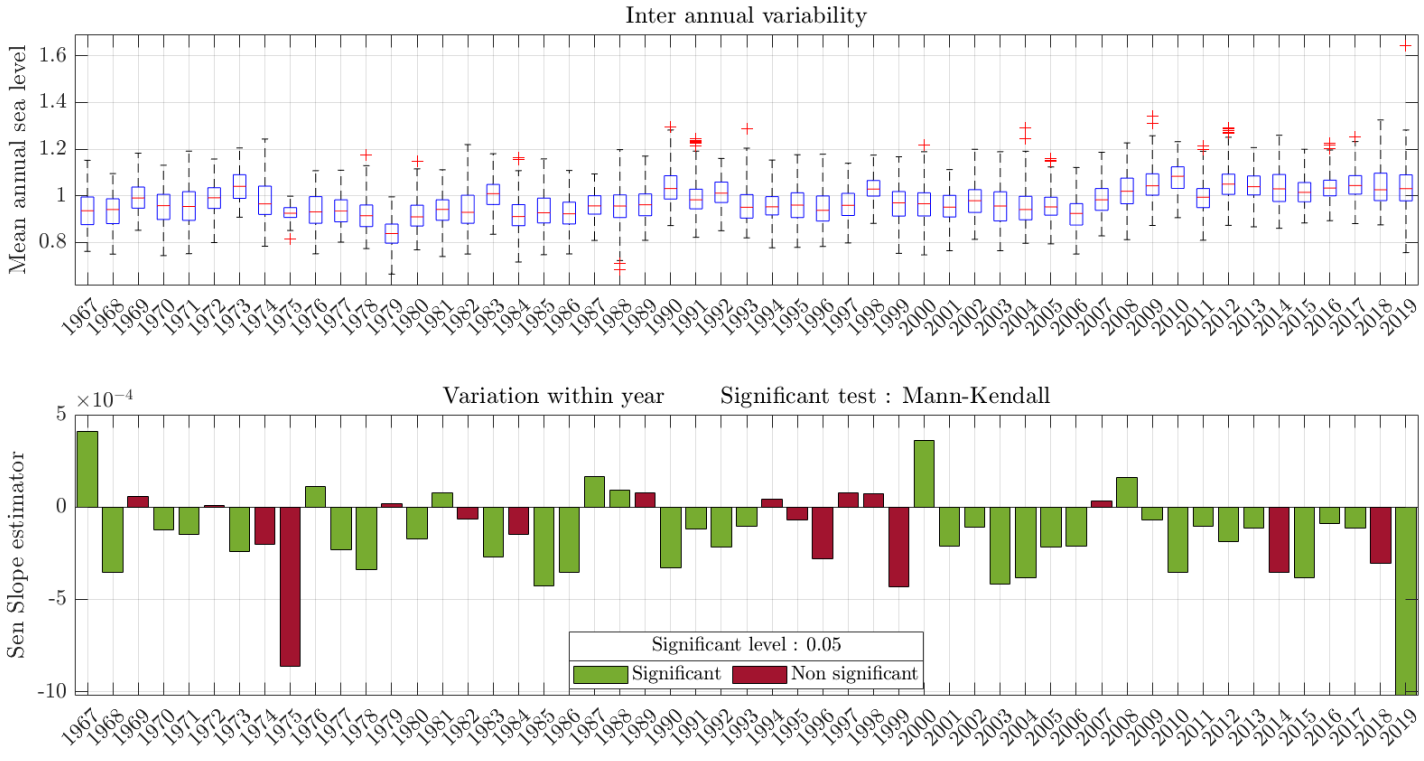

Figure B.7 Sen's slope estimation for the inter annual variability of mean annual sea level

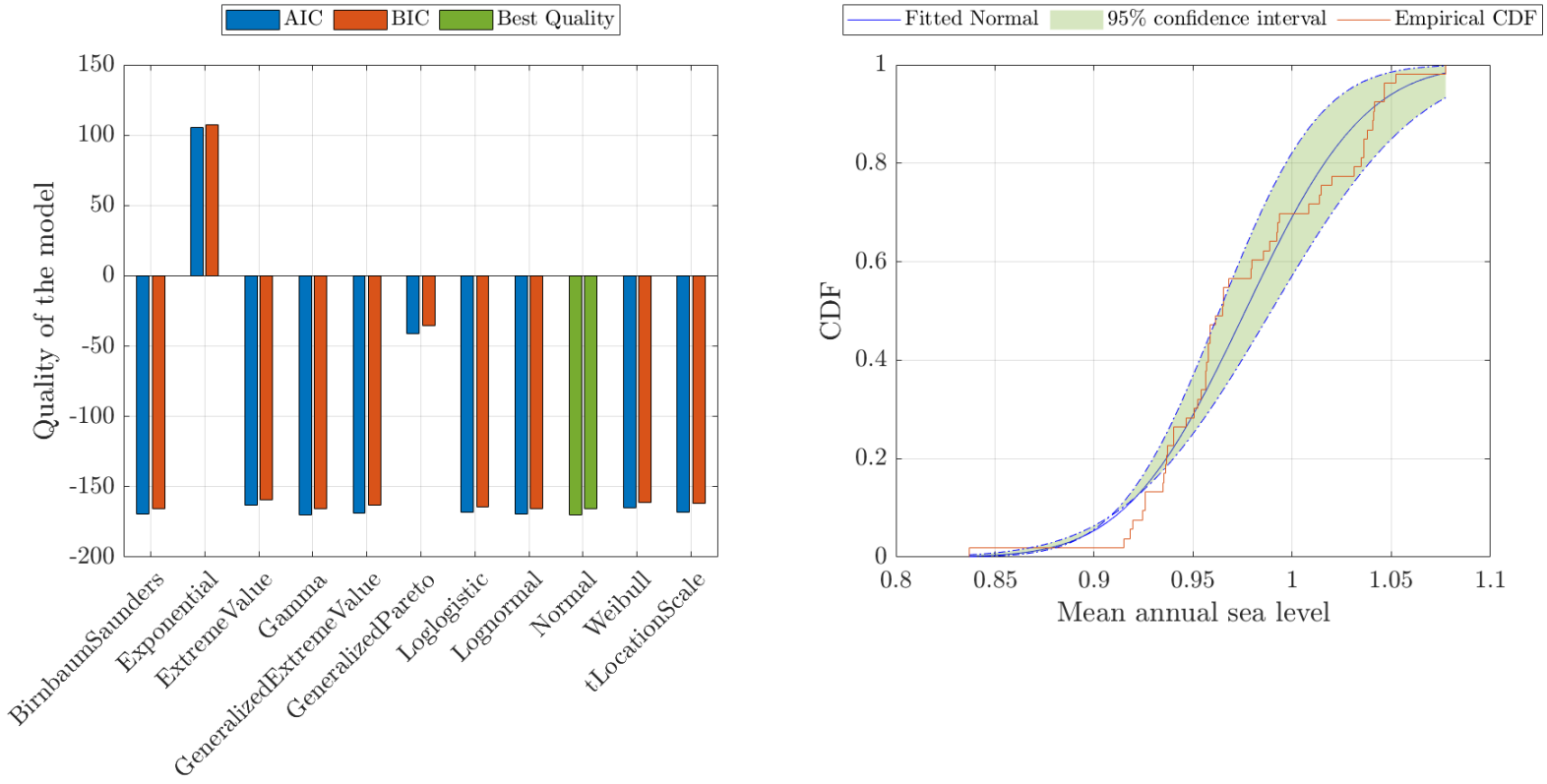

Figure B.8 Goodness-of-fit estimation (left panel) and visualization of the CDF with confidence interval overlaid to the empirical CDF (right graph) for the mean annual sea level

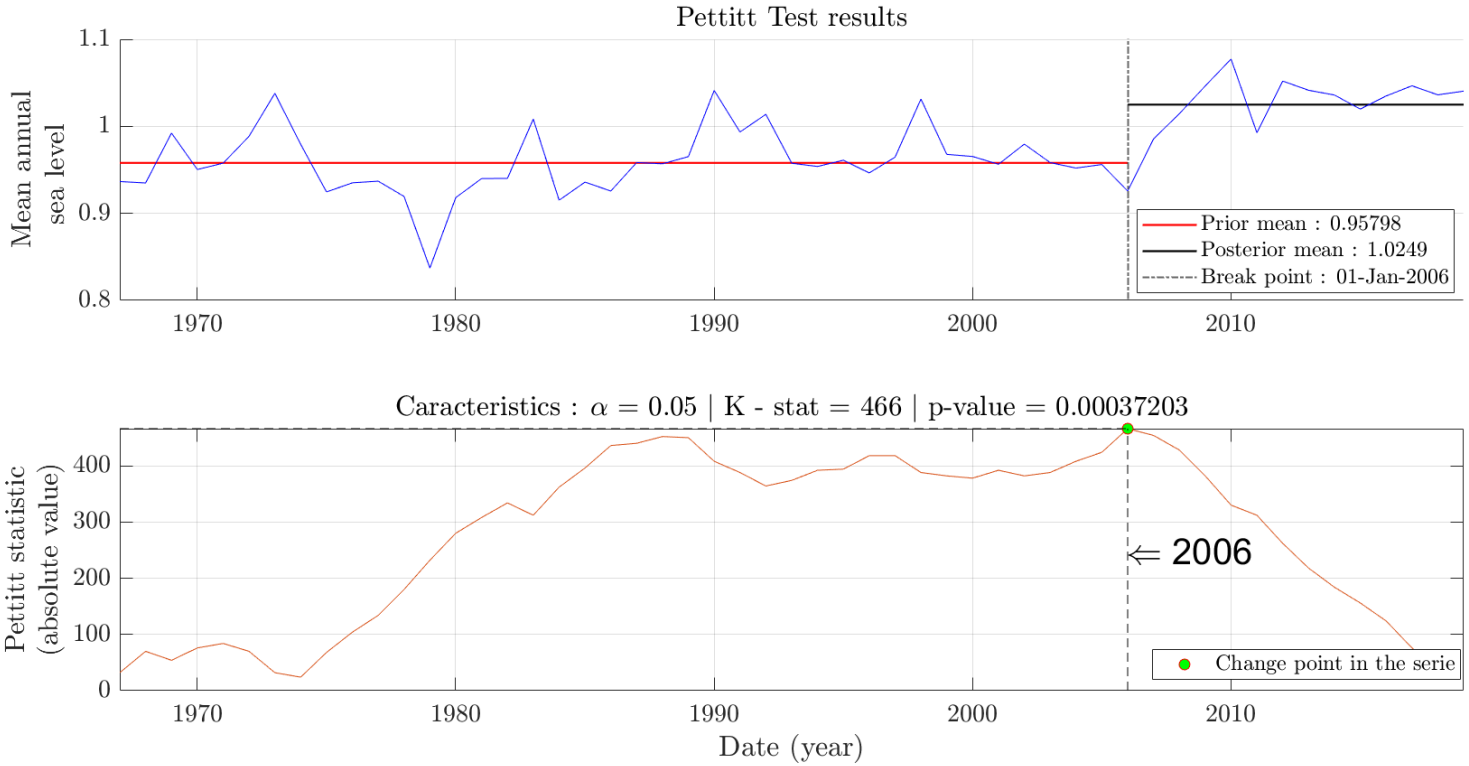

Figure B.9 Pettitt test results including visualization of the prior and posterior mean (top row) and location of the maximum value for the test statistic (bottom row)

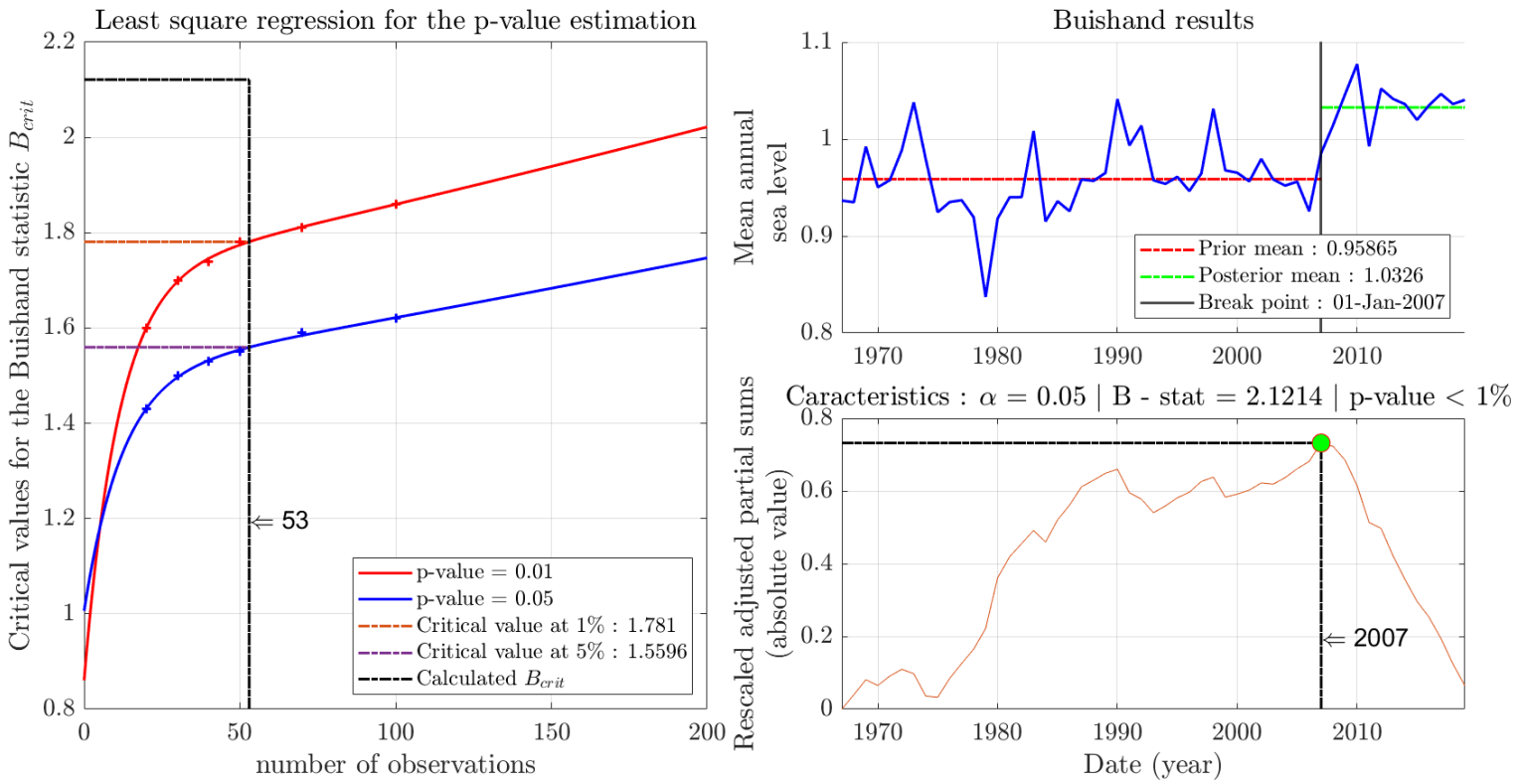

Figure B.10 Buishand test results including p-value estimation (left graph), break point location (bottom right) and prior and posterior mean (top right)

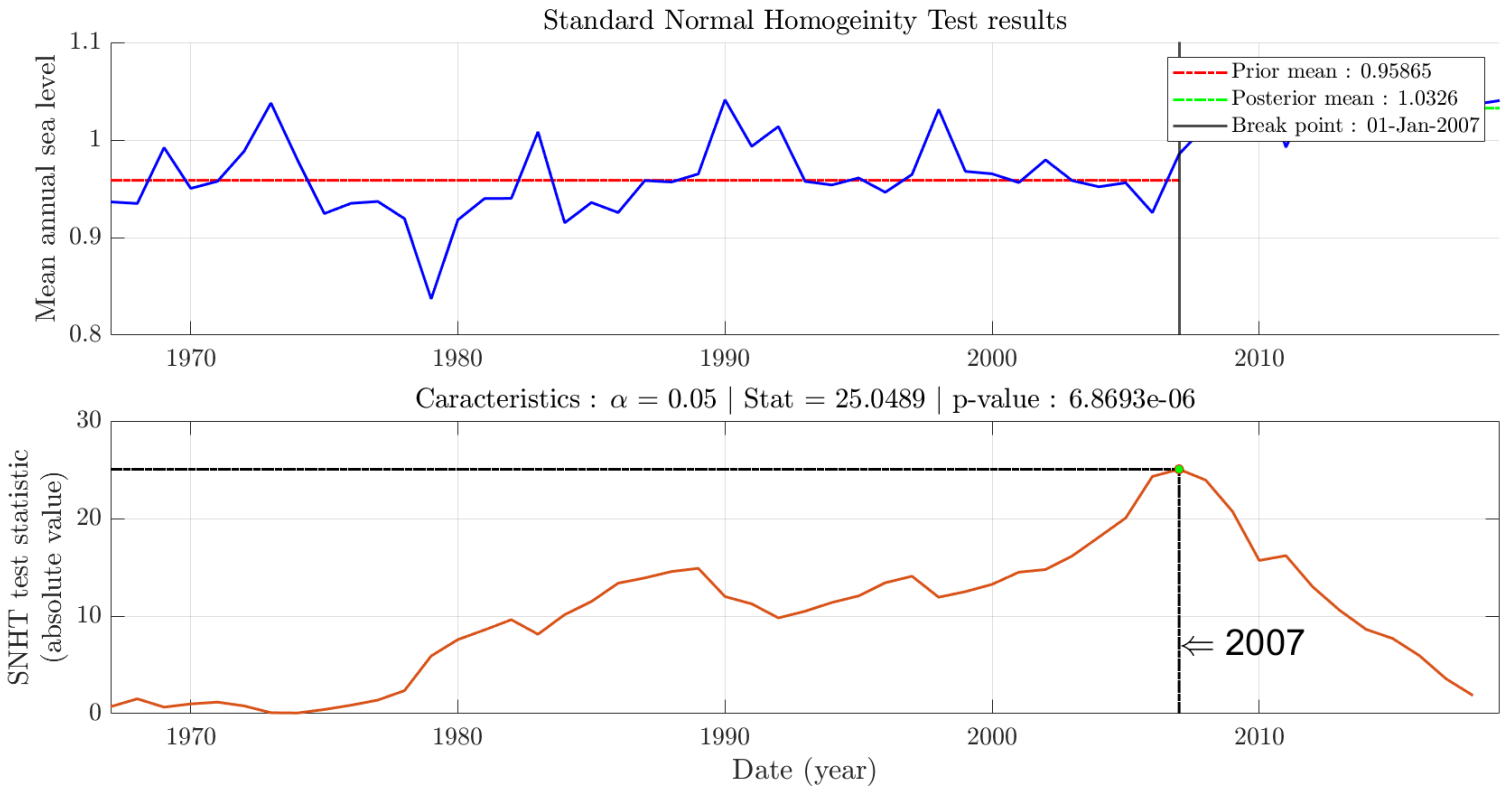

Figure B.11 SNHT results with break point location (bottom) and prior and posterior mean (top)

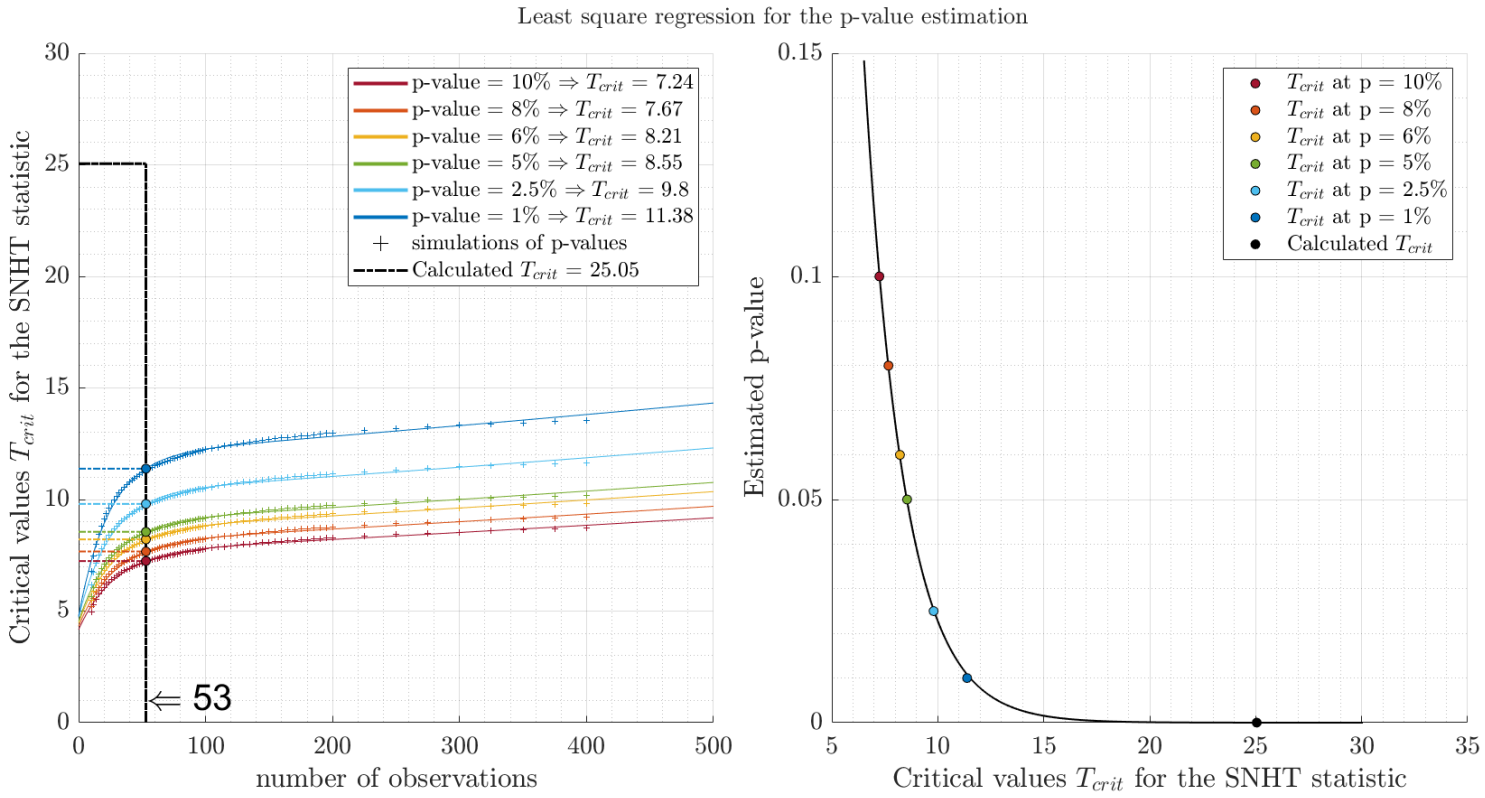

Figure B.12 SNHT p-value estimation using the table provided by Khaliq and Ouarda, (2007)

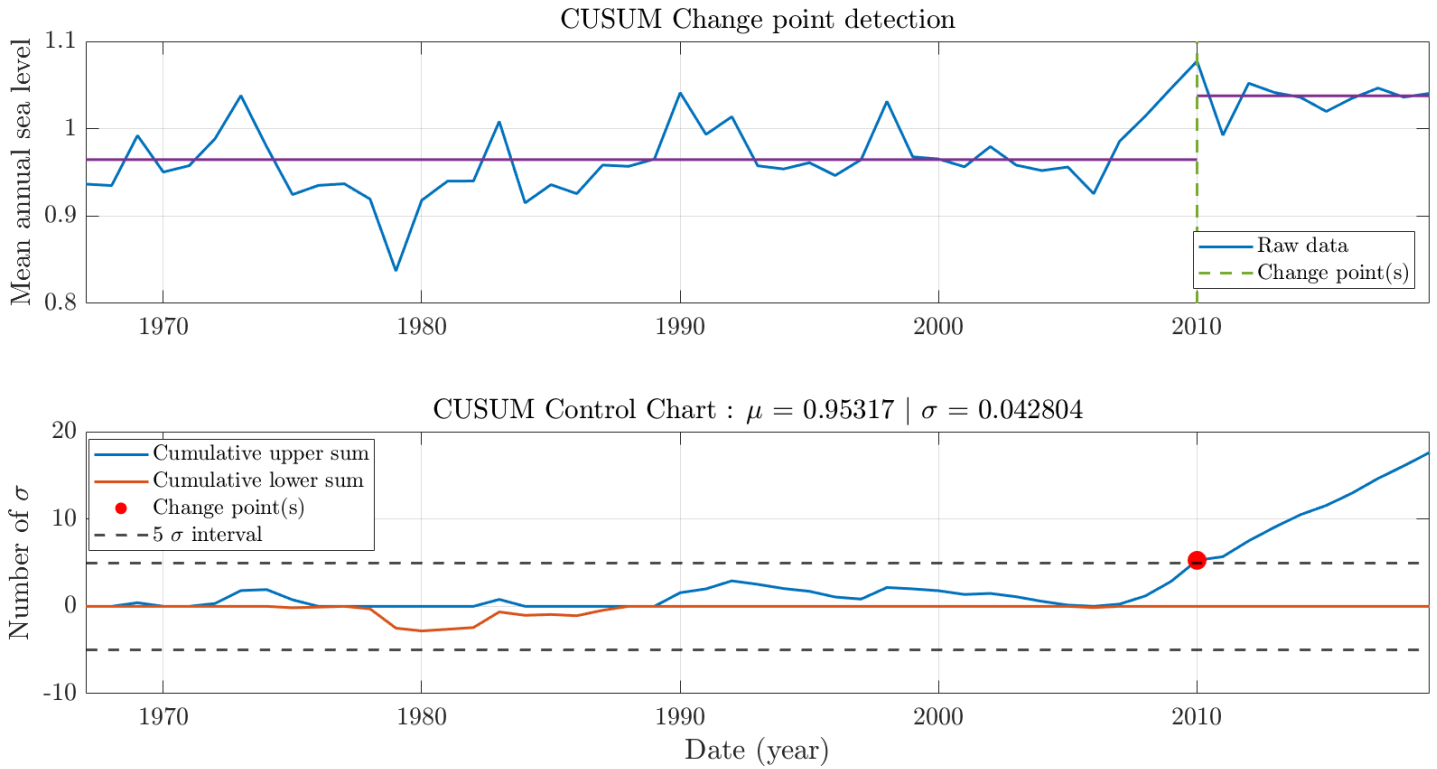

Figure B.13 CUSUM test results with location of change point (top row) and evolution of cumulative upper and lower sums U_i and L_i

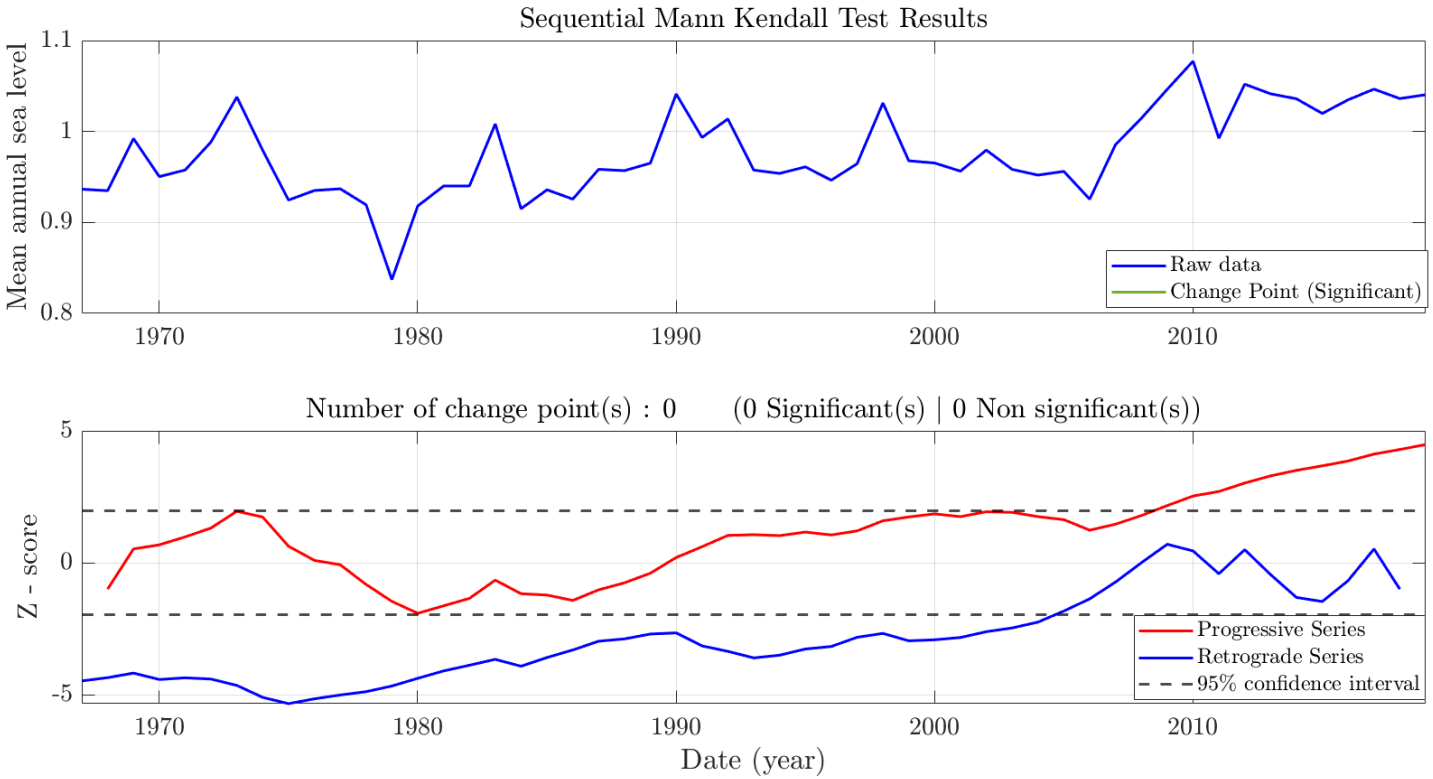

Figure B.14 Sequential Mann-Kendall test results with change point location (top row) and evolution of u and u' over time

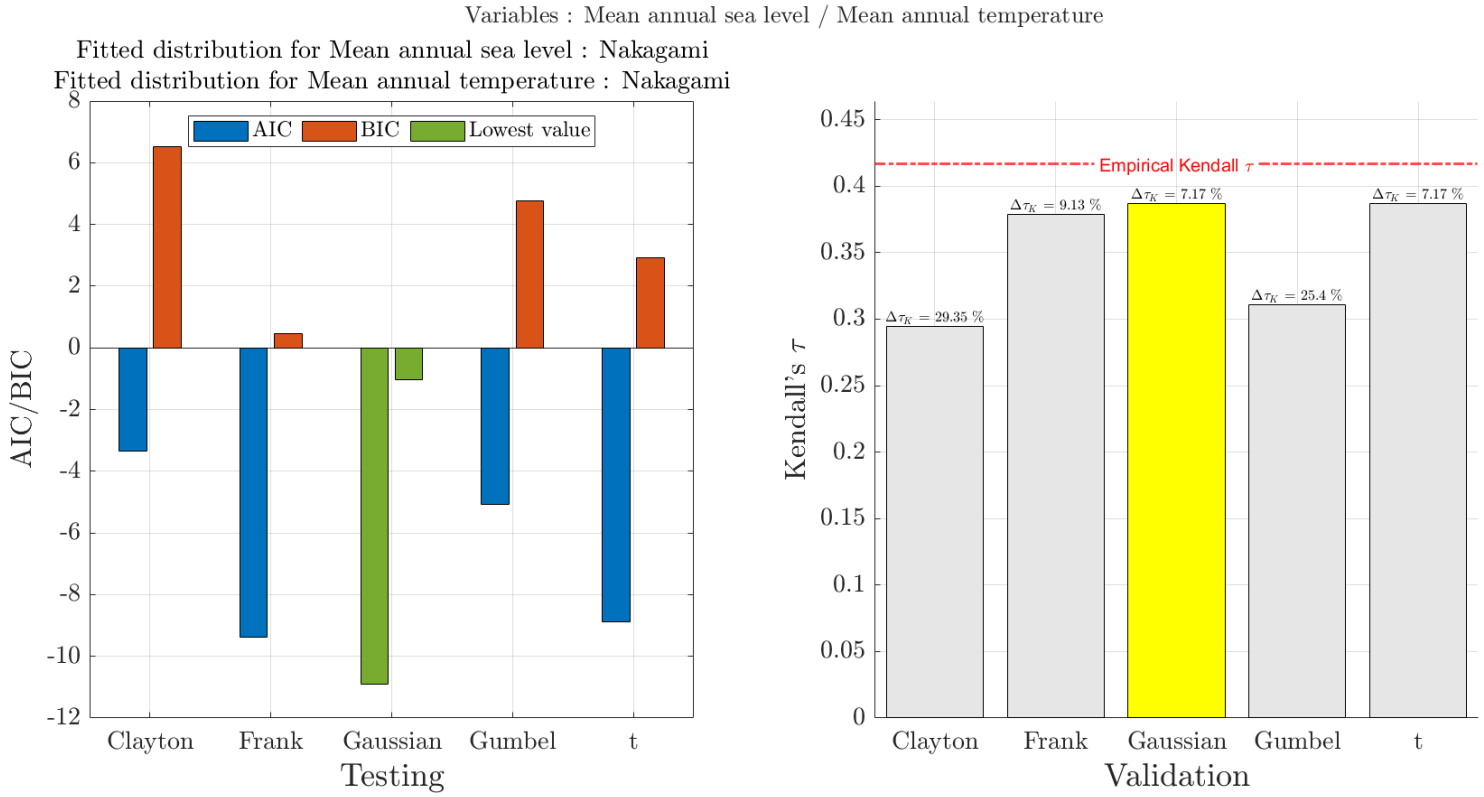

Figure B.15 Comparison of copula models considering ML method to estimate θ_{cop}

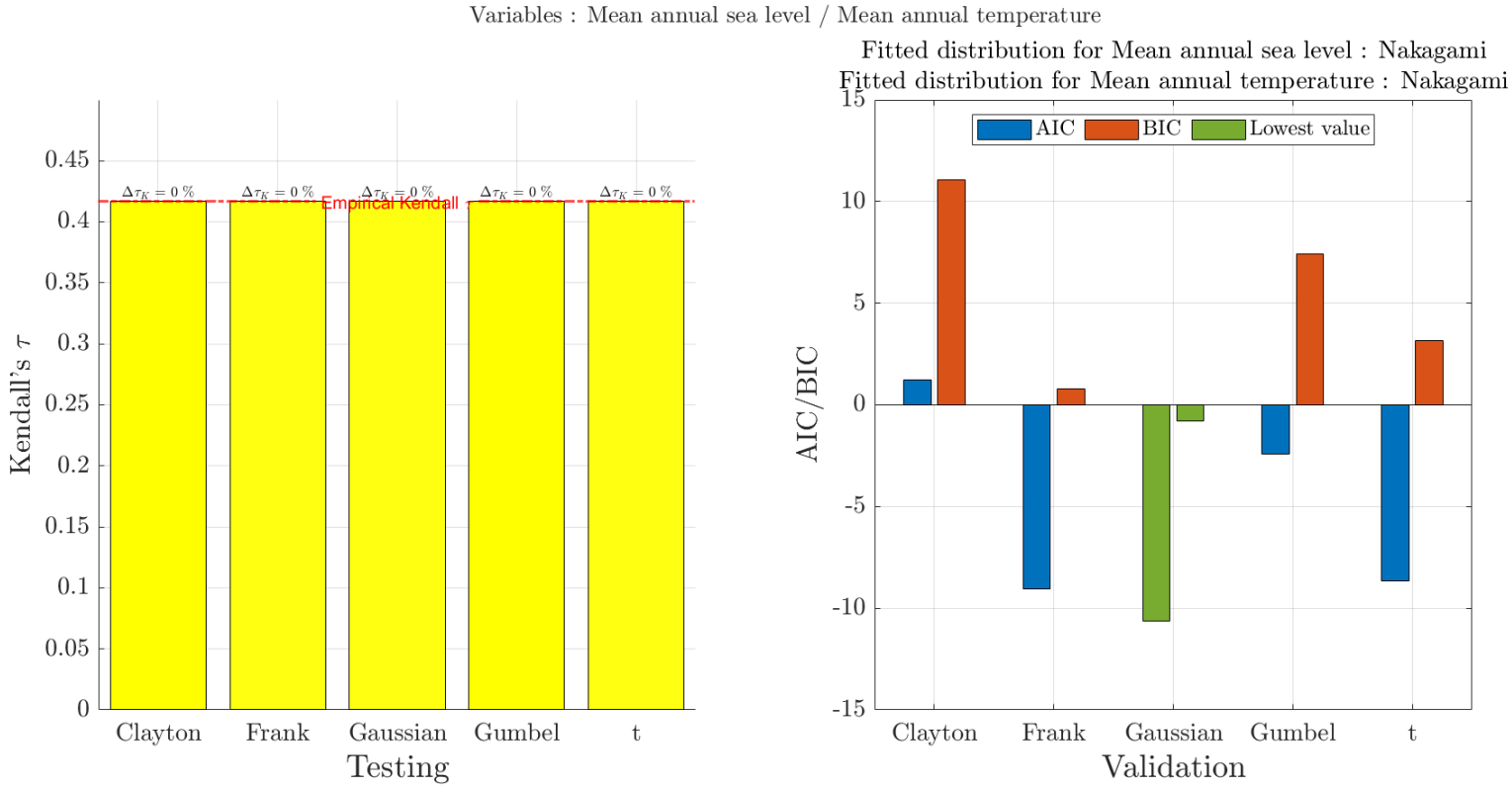

Figure B.16 Comparison of copula models considering MoM method to estimate θ_{cop}

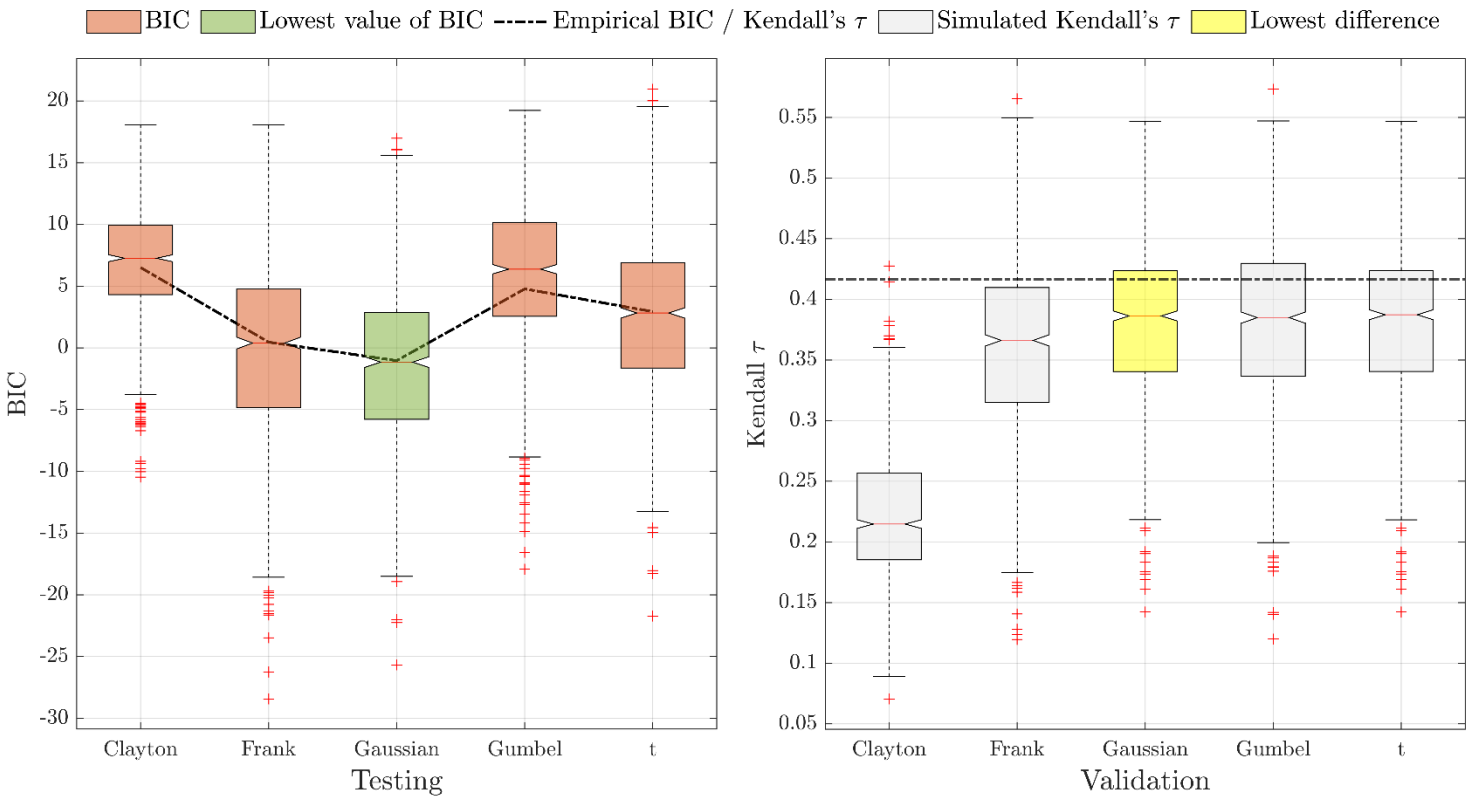

Figure B.17 Bootstrap results to select the best copula considering ML method to estimate copula's parameter

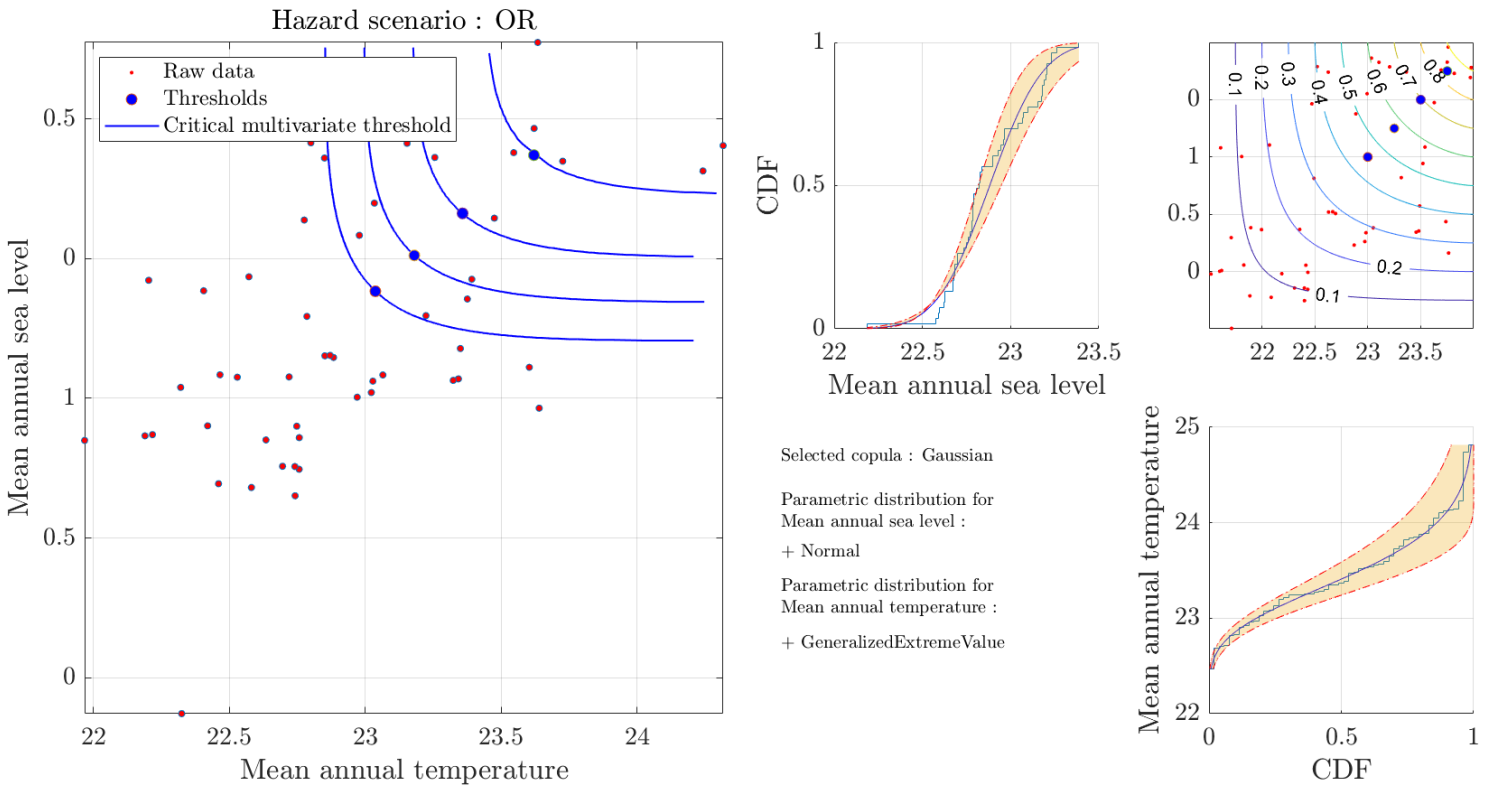

Figure B.18 Isoline of joint probability for thresholds of 0.6,0.7,0.8, and 0.9 on both marginal CDFs u and v

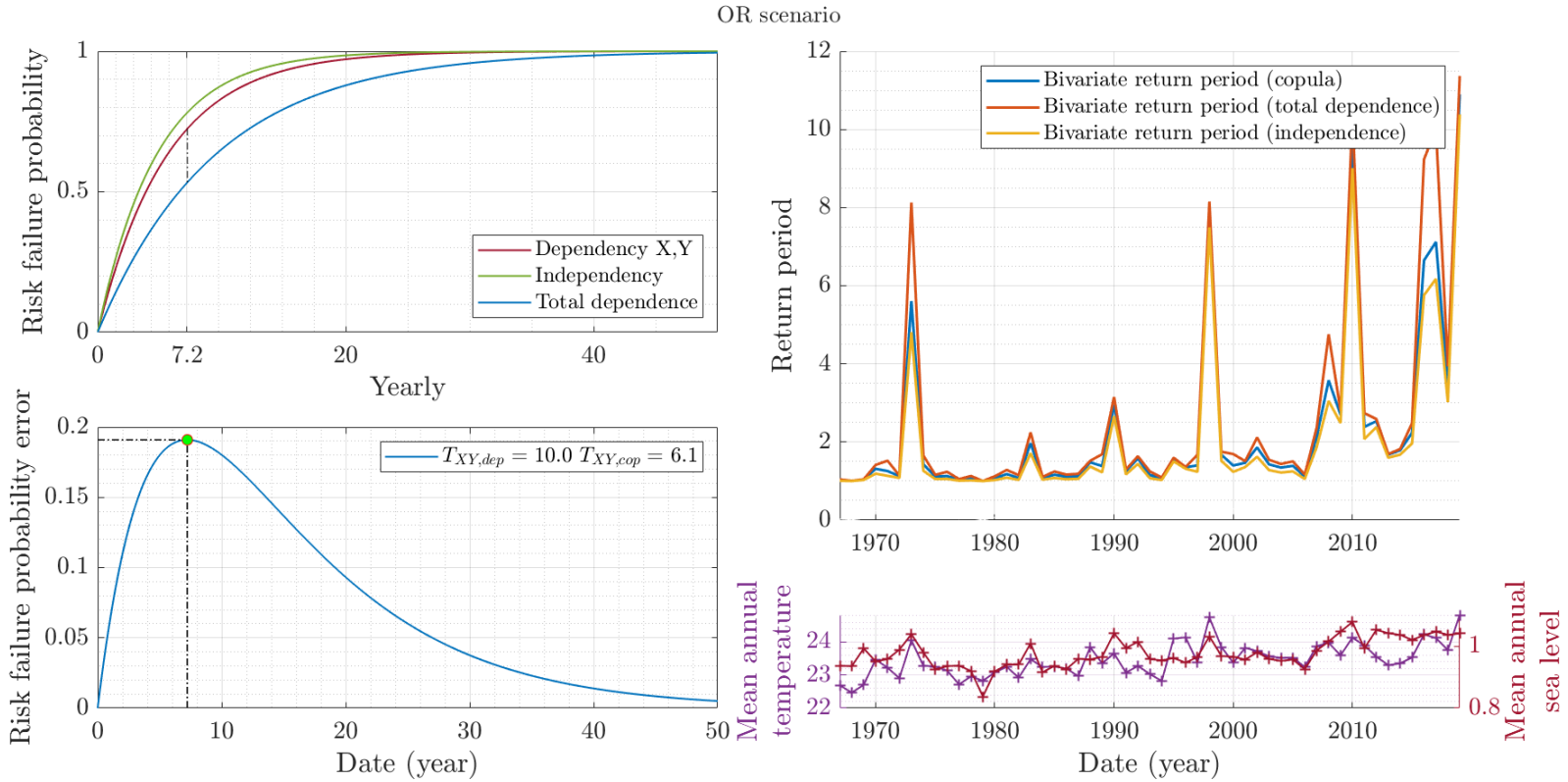

Figure B.19 Risk failure probability (left panels) and evolution of the return period for each considered events (right panels)

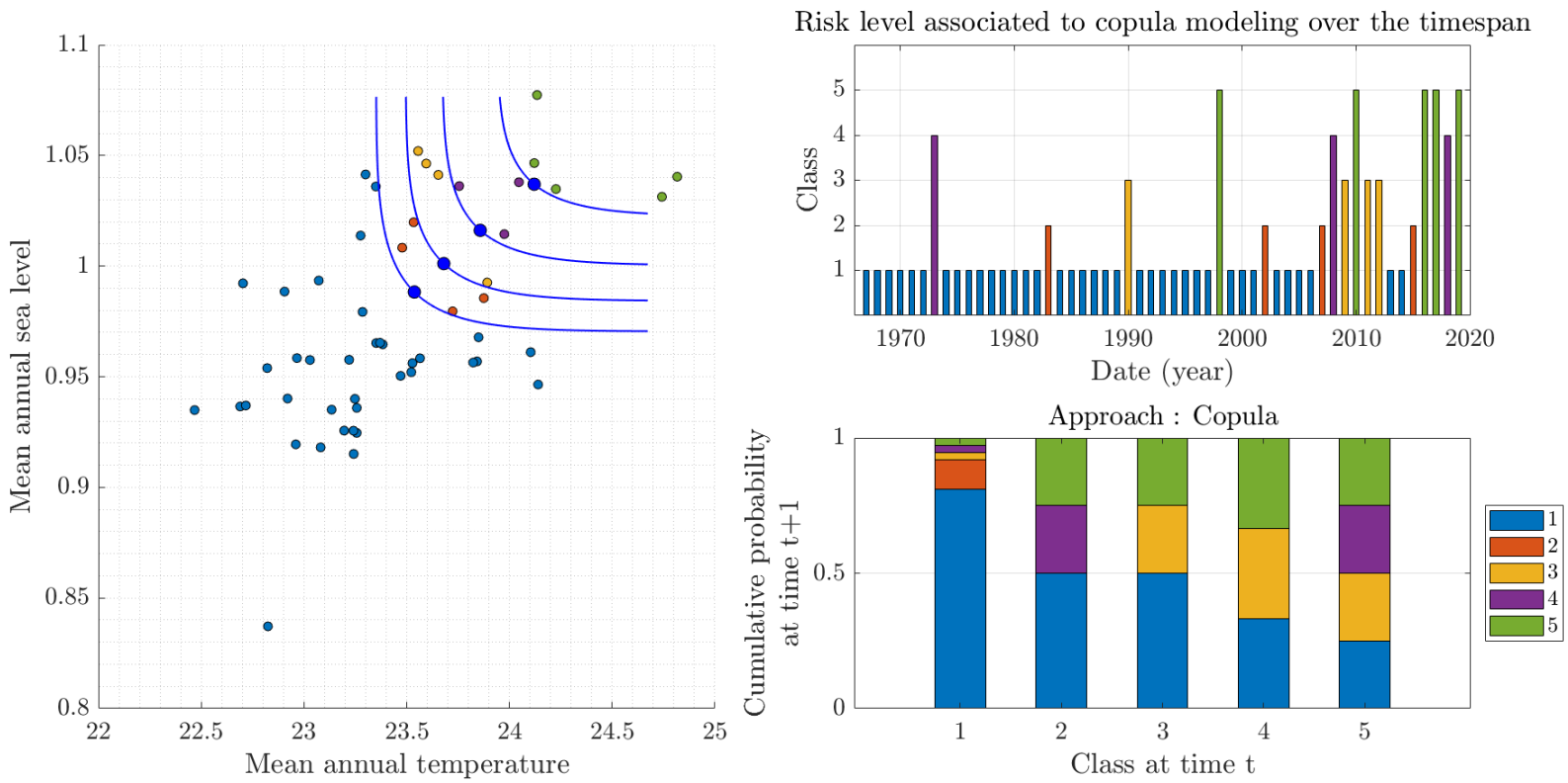

Figure B.20 Risk classification results with evolution of hazard levels (top graph) and conditional probability of risk evolution (bottom graphs) under the OR scenario

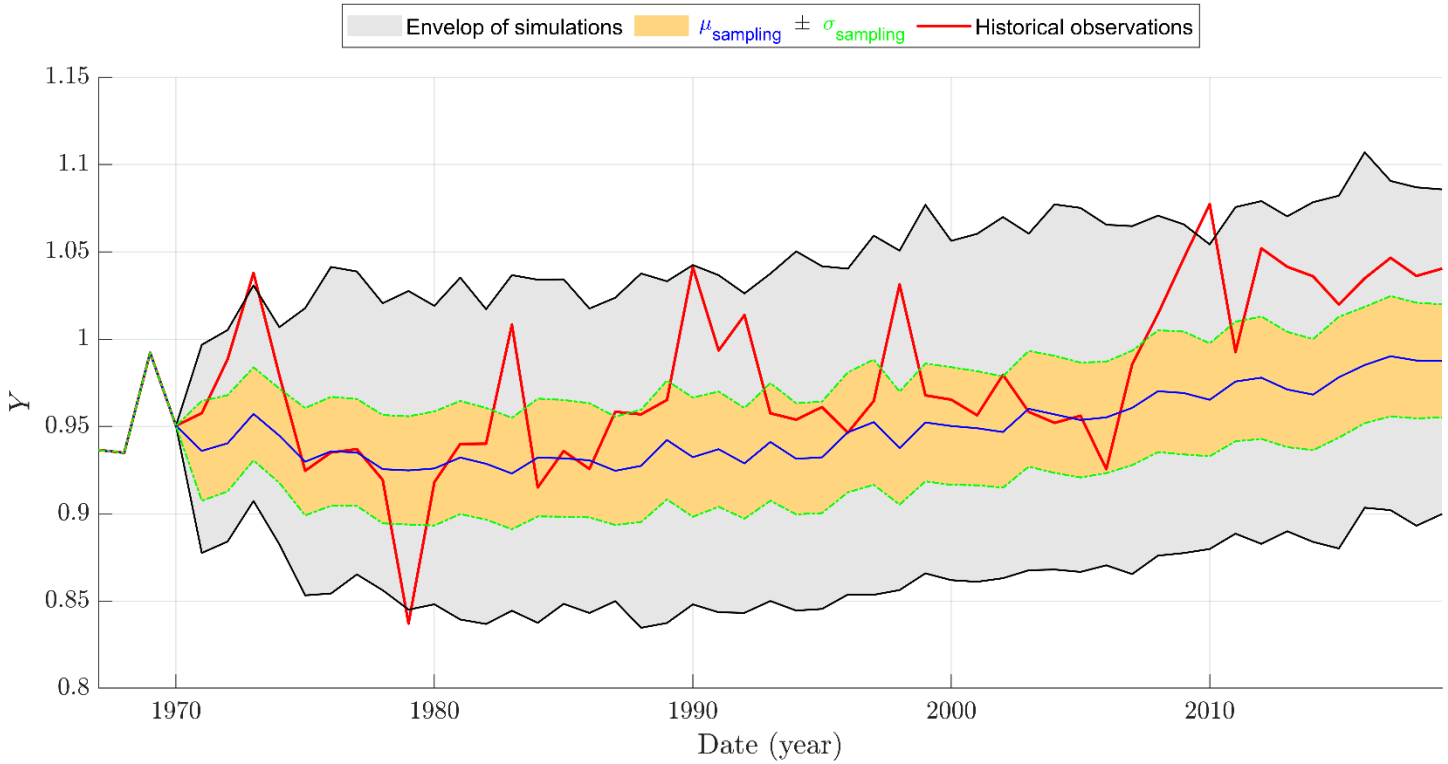

Figure B.21 Range of simulations for the predictand variable Y using the conditional sampling procedure

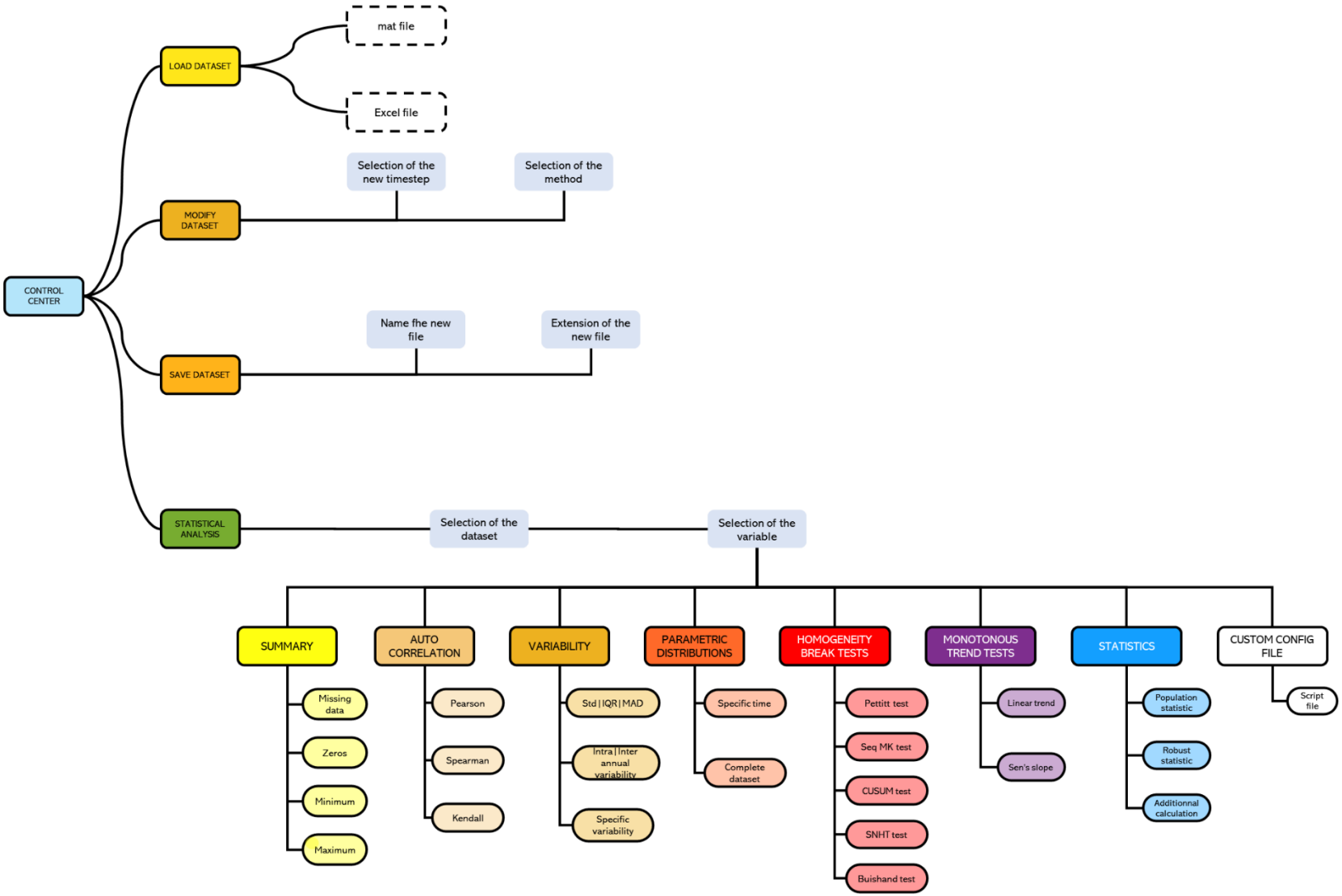

Figure B.22 Flowchart related to the Statistical Analysis Software

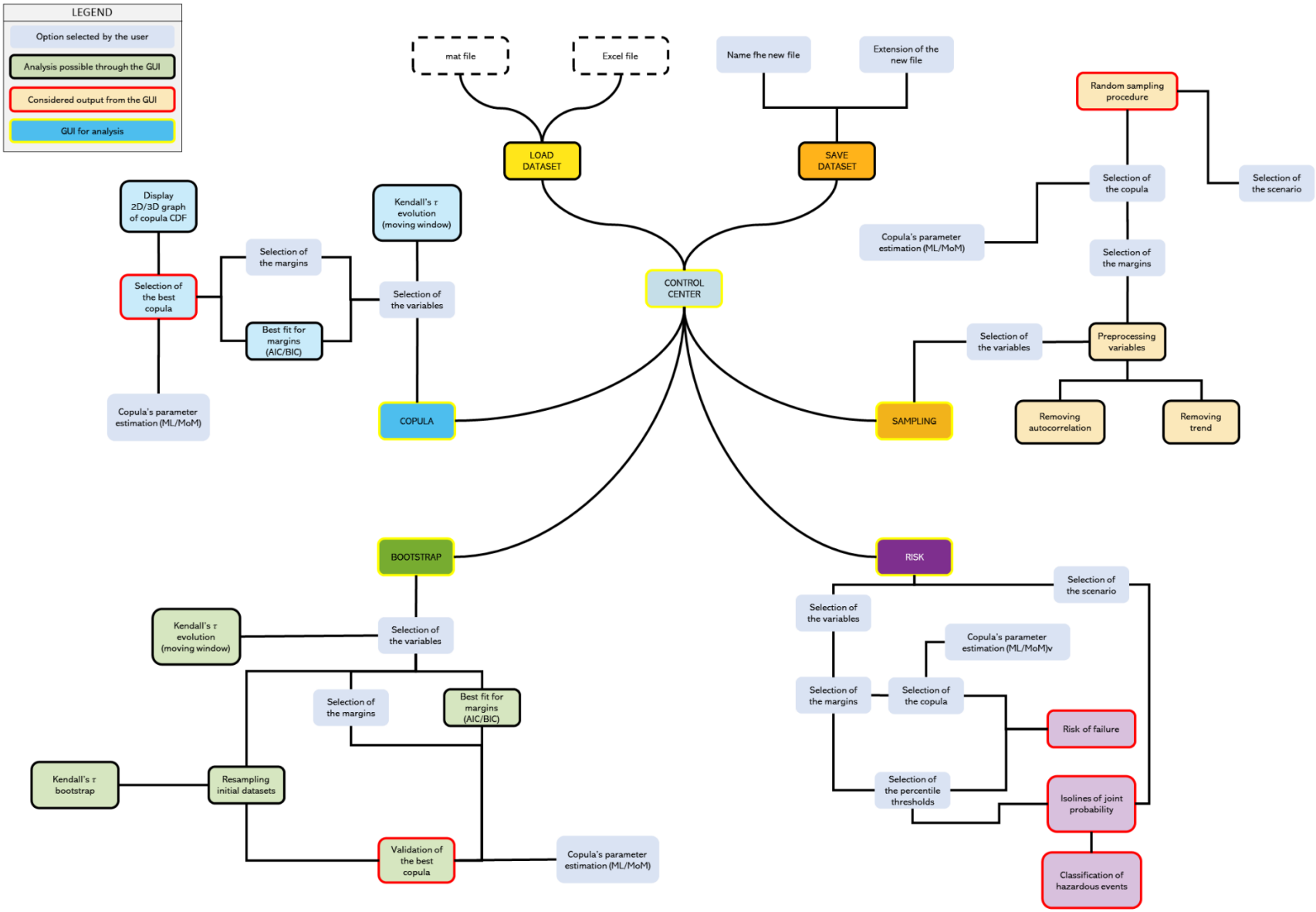

Figure B.23 Flowchart related to the Copula Analysis Software