TÉCNICO LISBOA

Bivariate analysis of intense rainfall on the island of Madeira and its relationship with alluvium flood events

Lucas Rogério Gomes

Dissertação para obtenção do Grau de Mestre em

Engenheira Civil

Orientador:

Doutora Maria Manuela Portela Correia dos Santos Ramos da Silva

Júri:

Presidente: Doutor Rodrigo De Almada Cardoso Proença de Oliveira

Orientador: Doutora Maria Manuela Portela Correia dos Santos Ramos da Silva

Vogais: Doutor António Pedro Nobre Carmona Rodrigues

Outubro 2021

1 Introduction

Intense rainfalls and alluvium flood events are common occurrences on the island of Madeira, some of them being deadly. However, there is a limited understanding of how intense rainfall events relate to alluvium flooding events. This research seeks to analyse rainfall measured at the Funchal gauge station and study its relationship with Funchal's alluvium flooding record, which was collected by Sepúlveda (2011). Particularly, calculating, through bivariate statistical analyses, the return periods of extreme rainfall events. The bivariate nature of this analysis allows for an understanding of how the exceptionality of a rainfall event changes with its progression through time. Then, by associating each extreme rainfall to a destructive alluvium flood event that is geographically and temporally related to the rainfall, conclusions can be made regarding the relationship between the exceptionality of particular rainfall events and their associated alluvium flooding events, paying special attention to the deadly late February 2010 event. Therefore, whilst other studies take into consideration other hydrological factors and use a univariate or categorised approach to rainfall data, this study solely focuses on extreme rainfall and its relationship with particular alluvium flood events and uses the copula approach as the method of arriving at more exact return periods values.

The most important literature relating to this study is the work done by Lopes *et al.,* (2020), Fragoso, *et al.* (2012) and Levizzani, *et al.* (2013). They studied various hydrological and climatological factors that could be related to and could have caused the alluvium floods in Madeira with a focus on the late February 2010 (18-20/02/2010) event. A description of this disastrous event can be found in Table 1 presented section *2 Data*. Conversely to the previously mentioned studies, this study uses a multivariate approach to the analysis of the rainfall data series, whilst those studies merely used a univariate or categorised approach. The bivariate copula method will provide more accurate return periods and permit different perspectives to be taken, namely that of joint and conditional return periods. The multivariate approach also permits for an analysis of the rainfall prior and/or posterior to the annual maximum and how it relates to the annual maximum rainfall. Thus, it was admitted that for one same occurrence there could be different consequences, depending on the previous or antecedent humidity conditions of the hydrological basin on a given date.

To classify extreme rainfall events this study uses the Annual Maximum Series (AMS) technique to classify hourly and daily time-series rainfall as extreme rainfall over a 34-year and 80-year period respectively. The AMS technique, as applied in this thesis, selects the maximum rainfall of each hydrological year (12-month period from October 1 to September 30 of the following year) available in the time-series data, and compiles those values into one series for hourly rainfall and one for daily rainfall. In the research summarize in this paper only hourly rainfall w, all calculations were performed for hourly rainfall data, therefore the hourly AMS is thirty-four values long and the daily is eighty, which represents the number of years collected for each dataset. As previously stated, each AMS rainfall value represents the extreme rainfall event of that hydrological year. Notably, the AMS for the daily data does not contain the late February 2010 rainfall event. The use of the hydrological year is essential to keep the values random since for each hydrological year, hydrological factors are re-set. Then, with the objective of understanding the relationship of the exceptionality of the rainfall of preceding and succeeding hours (to that annual maximum) and the associated alluvium flood events, the notion of a cumulative series was introduced (further explained in section *3 Datasets and modelling approach*).

The statistical methodology used in this thesis revolved around using the AMS data and the cumulative series as two variables (rainfalls) in a bivariate copula analysis. The main purpose was to be able to compute joint and conditional return periods of coupled rainfall events. This methodology first requires the fitting of several types of marginal distribution functions to the hourly and daily AMS and cumulative series that were calculated from the hourly and daily rainfall data. Then, after testing the relative quality of the fitting of each series, the formation of bivariate copulas was performed. With the association of extreme rainfall events, now represented by copulas, with alluvium flood events, it is also possible to understand the exceptionality of specific alluvium events.

2 Data

There were two main sets of data that were used to conceive the implemented approach: rainfall data and alluvium data.

The first set included hourly and daily rainfall records at Funchal rain-gage, respectively, from the 1st of October 1980 to the 30th of September 2014 and from the 1st of October of 1937 to the 30th of September 2017. These datasets were obtained separately. The daily rainfalls comprehended both measurements and data resulting from the filling-gap procedure developed by Espinosa *et al.*, (2021), whereas the hourly rainfalls referred only to measurements provided by the IPMA – Instituto Português do Mar e da Atmosfera, I. P.

The second main set of data is related to recorded alluviums from 1601 to 2010 and was collected from Sepúlveda (2011). For each alluvium event, this author provides some information on the weather conditions and location of the occurrence.

Criteria were defined with the objective of associating extreme rainfall events represented in the hourly AMS and the historical alluviums systematised by Sepúlveda (2011). The purpose was to connect the two main datasets aiming at understanding the association of alluviums and extreme rainfall events on Madeira Island.

An important methodological note is that all the extreme rainfall events were initially defined by the AMS approach and analysed with the objective of Copula study, and not only the ones that were related to alluviums. This means that the computation of return periods was made based on all the extreme rainfall events independently of if they were associated with alluviums or not.

The criteria of selection of the extreme rainfall events that assumable could be connected to alluvium events were **temporal**, **spatial** and **substantive**. The temporal criterion couples the extreme rainfall-alluvium event if the annual maximum rainfall occurred within the previous 6 days of the identified alluviums. The spatial criterion indicates that the alluvium must have occurred specifically in Funchal or was said to have impacted the southern slope of the island or even all over the island. Finally, the substantive criterion was defined as having caused either floods or landslides or damaging impacts to civil infrastructure and human life.

Table 1 presents the coupled (meet the three criteria) alluvium-rainfall events' hydrological year, annual maximum rainfall date, alluvium date and a description of the alluvium event.

Table 1 – Hydrological year of associated alluvium, dates of the maximum hourly rainfall and of the alluvium. Then, the description of each alluvium (adapted from Sepúlveda, 2011).

Hydrological year	AMS date	Alluvium date	Alluvium description
1989/1990	18/09/1990	18/09/1990	"Falling of blocks in Curral das Freiras: "landslide happened after the strong rainfall which happened between 14h and 15h" (DNM, 1990). Floods also took place in Funchal"
1991/1992	29/10/1991	29/10/1991	"In Funchal the rain caused floods and damage to the sewege systems. Also, in Câmera de Lobos floods were registered in the residences and anomalies in the sewer systems."
1993/1994	29/10/1993	29/10/1993	"Funchal was woken up starteled. The intensive rain and streams filled with rubble caused a catastrophe. (…) The trajedy struck various points of the island"
1994/1995	07/10/1994	07/10/1994	"Great rainfall registered during all of the day and provoked some floods a landlides in diverse areas of the island."
1995/1996	22/03/1996	22,23/03/1996	Strong storm with great discharge of water in all the island. Landslides, falling of trees and the obstruction of roads happened.
1997/1998	01/02/1998	07/02/1998	"A bit everywhere, with land giving waybecause of the wight of the rainfall water" (DNM, 8 Fev. 1998).
1999/2000	10/10/1999	10/10/1999	"Strong rainfall in Funchal, followed by landslides."
2001/2002	18/11/2001	18,19/11/2001	"Storm mainly on the south side of the island provoked floods, landslides and the falling of trees."
2002/2003	24/11/2002	24/11/2002	"Storm over all the island, mainly in the south and west, provoked landslides, floods and obstructions of roads."
2006/2007	07/04/2007	7,8,10,11/04/2007	"Intensive rainfall provoked floods in Funchal." "Intensive rain provoked loss of stones in access roads to Curral das Freiras, and also floods."
2009/2010	20/02/2010	18-20/02/2010	"All of the south side of the island was affected by the by the storm. The final official balance indicates that 43 people died, 8 remain lost, 120 were injured and 800 habitations suffered damages, 400 of which there was a total loss or are needing a deep intervention, with a loss of 36 million Euros. (...) The Comissaão Partiaária Mista defined the value of loss at 1080 million Euros,"

3 Datasets and modelling approach

As referenced in the *1 Introduction* and *2 Data* sections, a hypothesis was drawn up that states that an alluvium event is associated with the rainfall event that took place prior to or adjacent to the alluvium date. Thus, instead of analysing the extreme rainfall events as univariate statistical models, the in-time internal relationship of said rainfall events was addressed. Therefore, in the bivariate analysis, a second variable related to cumulative rainfall was considered. A simple usage of an AMS is insufficient for a good understanding of the exceptionality of a rainfall event, because annual maximum rainfall with a given duration may not account for the rainfall conditions during which the alluvium event occurred. Since alluviums are related to rainfall that occurs along time, a more meticulous understanding of the rainfall event prior to the alluvium and its change in time is necessary. In this definition, each rainfall event is identified by an annual maximum rainfall and by an associated cumulative rainfall prior and/or posterior to that maximum. This bivariate model must also have the ability to enjoin the two variables into one distribution so that in the qualitative hydrological sense the two variables can be looked at as one coupled rainfall event that can then potentially be associated with an alluvium flood event. For this purpose, the bivariate copula model was used. The bivariate copula is in essence a bivariate distribution from which joint or conditional probabilities can be calculated and allows for an understanding of a possible non-linear relationship between two variables. Therefore, using the AMS values as the defining characteristic of the rainfall event and as variable number 1, the cumulative rainfall prior and/or posterior to each annual maximum is defined as variable number 2 of the bivariate analysis.

To associate the cumulative rainfall series with the annual maximum three different scenarios were considered: cumulative rainfall in hours or days before the annual maximum, the same for hours or days after the annual maximum and a mix of the two previous scenarios, i.e., cumulative rainfall in hours or days surrounding (before and after) each annual maximum. For each of the three previous scenarios, six hourly and six daily series were created.

These datasets can be described in terms of mathematical formulation. Let X_0 designate the annual maximum series and X_n^B the cumulative hourly or daily rainfall before the yearly max, where the superscript "B" indicates that the cumulative series is composed of hourly or daily rainfall measurements before the annual maximum and the subscript "n" refers to the number of hours or days considered when computing the cumulative rainfall before the annual maximum. The cumulative hourly or daily rainfall series after each yearly maximum uses an equivalent representation, i.e., X_n^A, where "A" indicates that the cumulative series is composed of hourly or daily rainfall measurements after the annual maximum and the subscript "n" refers to the number of hours or days of the rainfall accumulation period. Finally, the cumulative rainfall for n hours/days surrounding the yearly maximum was designated as X_n^{BA}.

Mathematically these series can be defined by the equations (1), (2) and (3), where the index i refers to the rainfalls in consecutive time steps each with duration Δt, with Δt equal to 1 hour or 1 day, respectively for hourly and daily AMS series and $n \leq 6$.

$$X_n^B = X_0 + \sum_{i=1}^{n} Prior\ rainfall_i \qquad (1)$$

$$X_n^A = X_0 + \sum_{i=1}^{n} Posterior\ rainfall_i \qquad (2)$$

$$X_n^{BA} = X_0 + \sum_{i=1}^{n} (Prior\ rainfall_i + Posterior\ rainfall_i) \qquad (3)$$

Based on these hourly and daily rainfall series, eighteen copulas were calculated for the hourly and eighteen for the daily analysis. These sets of eighteen copulas can be subdivided with respect to the before, after and before and after analyses: six copulas containing the six series of cumulative rainfall before the annual maximum, six copulas containing the six series of cumulative rainfall after the annual maximum and six copulas containing the six of series cumulative rainfall before and after the annual maximum. This sums up to eighteen copulas for the hourly rainfall, and eighteen for the daily. The use of copulas also requires that the variables being associated should not be independent and should possess some correlation.

With all the datasets created, the next step in the statistical analysis is to fit all the series with various marginal distribution types using the Maximum Likelihood Estimation method, the Moment Matching Estimation method, Quantile Matching Method with the quantiles set at 0.25 and 0.75 and finally, the Maximum Goodness-Of-Fit Estimation method were used. For each of the estimation methods used, the various distributions were tested and the relative fitting quality was compared using Log-Likelihood Function (LLF), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Then, by ordering these different distributions based on the referred criteria, the best marginal distribution was selected for use in the copulas. The term "relative fitting quality" refers to when a distribution is better fitted to the series than another distribution that was also tested. But, it does not demonstrate that that distribution is sufficiently well fitted in an absolute perspective. The different fitting methods used in the analysis were as follows:

- Maximum Likelihood Estimation (MLE): Normal (nor); Gamma (gam); Weibull (wei); Exponential (exp); Cauchy (cau); Logistic (log); and Lognormal (lnor).
- Moment Matching Estimation (MME): Normal (nor); Exponential (exp) and Logistic (log).
- Quantile Matching Estimation (QME): Normal (nor) and Logistic (log).
- Moment Goodness-of-fit Estimation (MGE) with Cramer-Von Mises (cvm), Kolmogorov-Smirnov (ks) and Anderson-Darling (ad) distances: Normal (nor), Exponential (exp) and Logistic (log).

After all the marginal models were fitted to the series, tested and the best selected, each series has its values reduced according to the distribution that best fitted it. These reduced series are what constitute the two variables of the copula. Whilst the rainfall series were given in millimetres and were symbolised with the variable X, the reduced rainfall series are dimensionless and are represented with the variable U. For example, the hourly and daily X_0 was reduced to U_0, according to the best-fitted distribution and its estimated parameters. For the U notation, the subscripts and superscripts remain the same as the X notation

Once all the reduced series were calculated, the copulas were modelled. For the copula analysis, different copula types also had to be compared, tested and selected. The selected copulas were then studied for nonlinear correlations and return periods.

For the bivariate analysis, the “VineCopula” (R package) copula nomenclature and numbering are presented in Table 2. The table identifies the family and identification number of the copulas used in this study, as provided by the R package.

Table 2 – Copula family name and assigned identification number.

Family name	Designated number
Independence copula	0
Gaussian copula	1
Student t copula	2
Clayton copula	3
Gumbel copula	4
Frank copula	5
Joe copula	6
BB1 copula	7
BB6 copula	8
BB7 copula	9
BB8 copula	10
rotated Clayton copula (180 degrees; survival Clayton")	13
rotated Gumbel copula (180 degrees; survival Gumbel”)	14
rotated Joe copula (180 degrees; survival Joe")	16
rotated BB1 copula (180 degrees; survival BB1”)	17
rotated BB6 copula (180 degrees; survival BB6")	18
rotated BB7 copula (180 degrees; survival BB7”)	19
rotated BB8 copula (180 degrees; “survival BB8”)	20
Tawn type 1 copula	104
rotated Tawn type 1 copula (180 degrees)	114
Tawn type 2 copula	204
rotated Tawn type 2 copula (180 degrees)	214

Once all the copulas, totalling thirty-six for each of the twenty-two families, were calculated, three estimators (LLF, AIC and BIC) were applied to compare the relative quality of the fitted family in much the same way as for the analyses of the marginal distributions. This was performed for each bivariate combination. The best-fitting copula family is selected for further study.

5 Results

With the computation of eighteen copulas modelling eighteen hourly rainfall events and an additional eighteen copulas modelling eighteen daily rainfall events, it was possible to analyse some descriptive information obtained about the association of the two variables in the bivariate analysis. Because of the non-linear nature of copulas, the correlation between the variables that constitute a certain copula can be also analysed with a non-linear perspective. The next step is the calculation of joint and conditional return periods, which can be done directly from the copula cumulative distribution functions. An analysis was done for the bivariate return periods as exemplify in Figure 1 which helps to illustrate the exceptionality of coupled rainfall-alluvium events. The variable combinations selected to be presented in these figures are representative of the other combinations. In this figure, the white circles represent rainfall events that are not associated with an alluvium event, the red circles represent rainfall events that are associated with an alluvium event and the orange circles represent the late February 2010 rainfall-alluvium event. The contour lines depict the return periods that are calculated from the graph's respective bivariate copula where the two variables are the ones represented by the x and y axes. The x-axis is always the hourly AMS series (X_0), and the y-axis is always a cumulative series of rainfall from 1h to 6h surrounding the annual maximum. This analysis is not done for the daily data because the corresponding AMS series does not contain the late February 2010 event.

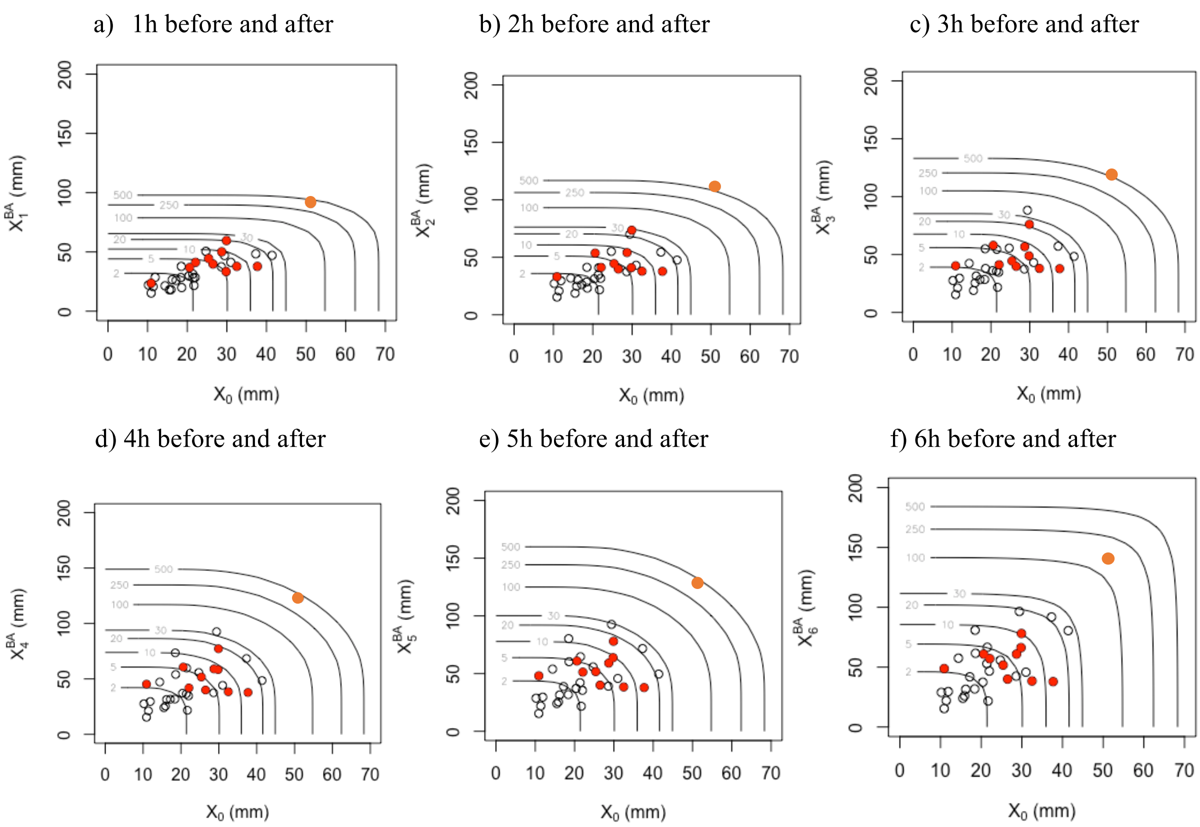

Figure 1 - Contour lines of return periods for coupled hourly annual maximum rainfalls (X_0) and the cumulative hourly rainfalls in 1 to 6 h before and after the annual maximum, X_1^{BA} to X_6^{BA}. All are from the joint "and" copula analysis.

Figure 1 shows the exceptionality of the 2010 late February rainfall that is associated with the deadly alluviums. This event is always set apart from the other thirty-three rainfall events analysed in this study, which shows its true exceptionality. There is also a noticeable tendency for rainfall events that were associated with alluviums (red dots) to have higher return periods. Notably, there is an exception for three rainfall events that aren't coupled with alluvium events. This figure also illustrates how with the increase of hours that are considered for the cumulative rainfall series, the return periods of the events differ, and the values become less clustered.

As part of this study, conditional return periods were also calculated from the bivariate copulas that were modelled. The conditional return periods were calculated in two ways:

1. The return period of AMS given the cumulative series.
2. The return period of the cumulative series given the AMS.

The conditional return periods for the data and series utilised in the copula analysis also prove the 2010 late February rainfall event to be exceptional. However, because of the nature of conditional probabilities, this analysis may not provide correct values for return periods, not even in the 10 to the nth power. This is especially notable for the 2010 late February event where the return periods of over ten thousand or even over one hundred thousand. For example, the return period of the AMS given the cumulative rainfall series of 2h before and after the annual maximum is $T_{X_0|X_2^{BA}} = \frac{T_{X_2^{BA}}}{P(X_0 \geq x_0, X_2^{BA} \geq x_2^{BA})} = 208395.70$ years. All other conditional return periods are in general much larger than the joint analyses. It is concluded that the conditional return period analysis does not result in accurate values.

7 Discussion and conclusions

This study concludes that the methodology of copula analysis is adequate for the purpose of understanding how the temporal distribution of the rainfall during a rainfall event constraints its exceptionality, either in with the use of joint or conditional return periods.

The analysis also showed that the joint and conditional return periods confirmed the exceptionality of the late February 2010 rainfall event. However, these joint and conditional probabilities do not result in similar return period values, even in the x10 to the nth power. Conditional return periods were sometimes unreasonable high, suggesting that they may not be adequate to characterize the rainfall events. The "and" return periods values were larger than its "or" counterpart, which is conceptually understandable. However, the "or" combination might not provide the most exact probability values when trying to relate extreme rainfalls to another event, in this case, alluvium flooding. This is because the "or" combination expresses the return period of either the annual maximum (or greater) happening or the cumulative, which also includes the maximum (or greater) happening. This research concludes that the "and" joint return period values might be the best to estimate the actual and real values of the return periods of rainfall events and their associated alluvium events. Further study could be performed on this point which would allow for a clearer and more definite understanding of this issue. For the 2010 alluvium events, Figure 2 was created to further compare the hourly "or" and "and" bivariate analysis, and to compare these two types of bivariate models to the univariate approach. The figure also is aimed at providing additional insights on the characteristics and exceptionality of the rainfall that triggered the late February 2010 alluvium flood event.

a) Cumulative rainfall before b) Cumulative rainfall after c) Cumulative rainfall before and after

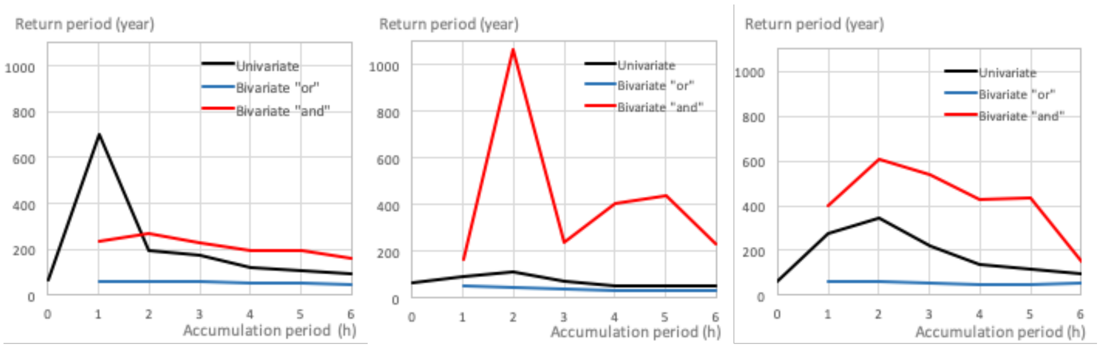

Figure 2 - Rainfall event of the 20th of February 2010. Univariate and bivariate return periods for the annual maximum rainfall and the cumulative rainfall, from 1 to 6 h: a) before; b) after; and c) before and after.

From the previous figures, a conclusion can be made that the univariate approach also results in high values for return periods. It shows particularly high return periods for the cumulative rainfall of one hour before the annual maximum. Whilst the bivariate "and" approach presents higher values for the cumulative rainfalls of two hours before and after the annual maximum. What is seemingly apparent is how the joint bivariate approach results differ from the univariate approach results, and how within the joint approach the "or" and "and" analyses substantially differ in their return period values.

Another conclusion that can be made from this thesis is that the use of the Annual Maximum Series technique, AMS, was useful because of its simplicity in the composition of the series, but, possibly, it could have not captured the fullness of the original data and its intense rainfalls. A possible improvement could be to use another method for selecting extreme rainfall events. For example, one possible technique is the threshold technique (or partial duration series) used by Liu *et al.* (2013) or Mase (1996). The threshold technique allows the researcher to set a limit, whereby any value that is above it is considered an extreme value provided some pre-requisites are met. This allows each series to possibly have more than one value per year, which results in longer series, which in turn generally results in more accurate fittings for marginal distributions and higher quality copula analyses.

Another discussion researchers could have on this topic is the use of a multivariate analysis of more than two variables. There could be an attempt to quantify the "degree of destruction" or "intensity" of the alluvium flood events that were coupled with extreme rainfall events. Thereby creating time-series data with quantified measurements/analysis of the alluvium events. Then, by using copulas where one of the variables would represent this series of intensity values for alluvium flood events and the other variable represents the rainfall measurements for the extreme rainfall events, it would be possible to analyse the correlations between the intense rainfalls and their coupled alluviums. In such a scenario, it would be possible to compare if more destructive alluviums were correlated or not with more extreme rainfalls.

Further developments of the research done could be its extension to other areas of the island of Madeira and not just the capital, Funchal. There were many alluvium events that Sepúlveda (2011) collected and are presented in Annex I that didn't get associated with any of the rainfall events simply because of the spatial criterion described in chapter *2 Data*. Using rainfall data from different locations in Madeira will allow for a more complete understanding of the rainfall and alluvium interactions.

References

Embrechts P., Copulas: A Personal View, *The Journal of Risk and Insurance*, Vol. 76, No. 3, 639-650, 2009, https://doi.org/10.1111/j.1539-6975.2009.01310.x.

Espinosa, L.A., Portela M.M., Pontes Filho, J.D., Studart, T.M.dC, Santos, J.F. and Rodrigues R., Jointly Modeling Drought Characteristics with Smoothed Regionalized SPI Series for a Small Island, *Water*, 11(12):2489, 2019, https://doi.org/10.3390/w11122489.

Fragoso, M., Trigo, R.M., Pinto, J.G., Lopes, S., Lopes, A., Ulbrich, S. and Magro, C., The 20 February 2010 Madeira flash-floods: synoptic analysis and extreme rainfall assessment. *Natural Hazards and Earth System Sciences*, *12*(3), pp.715-730, 2012, https://doi.org/10.5194/nhess-12-715-2012.

Levizzani, V., Laviola, S., Cattani, E. and Costa, M.J., Extreme precipitation on the Island of Madeira on 20 February 2010 as seen by satellite passive microwave sounders. *European Journal of Remote Sensing*, *46*(1), pp.475-489, 2013, https://doi.org/10.5721/EuJRS20134628.

Liu, B., Chen, J., Chen, X., Lian, Y. and Wu, L., Uncertainty in determining extreme precipitation thresholds. *Journal of hydrology*, *503*, pp.233-245, 2013, https://doi.org/10.1016/j.jhydrol.2013.09.002.

Lopes, S., Fragoso, M. and Lopes, A., Heavy Rainfall Events and Mass Movements in the Funchal Area (Madeira, Portugal): Spatial Analysis and Susceptibility Assessment. *Atmosphere*, *11*(1), p.104, 2020, https://doi.org/10.3390/atmos11010104

Mase, S., The threshold method for estimating total rainfall. *Annals of the Institute of Statistical Mathematics*, *48*(2), pp.201-213, 1996, https://doi.org/10.1007/BF00054785.

Yamaka, W., Gupta, R., Thongkairat, S. and Maneejuk, P., *Structural and Predictive Analyses with a Mixed Copula-Based Vector Autoregression Model* (No. 202108), University of Pretoria Department of Economics Working Paper Series, 2021, https://doi.org/10.1142/S0218488520400103.

Zeng, X., Ren, J., Wang, Z., Marshall, S. and Durrani, T., Copulas for statistical signal processing (Part I): Extensions and generalization. *Signal processing*, *94*, pp.691-702, 2014, https://doi.org/10.1016/j.sigpro.2013.07.009.