Precipitation and its extremes

The hydro-climatological controls of the statistical properties of extreme precipitation distributions.

I. E. van der Veer

Delft University of Technology

TU Delft

Precipitation and its extremes

The hydro-climatological controls of the statistical properties of extreme precipitation distributions.

by

I. E. van der Veer

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on June 15, 2022.

Student number:	4482891	
Project duration:	August 30, 2021 – June 15, 2022	
Thesis committee:	dr. ir. R. J. van der Ent	TU Delft, supervisor
	ir. G. J. Gründemann	TU Delft. supervisor
	prof. dr. ir. R. Uijlenhoet	TU Delft
	prof. dr. A. P. Siebesma	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.
Cover image: raindrops, photo taken May 2022.

Preface

Hereby, I present to you my master thesis. This thesis is written as final report to finish my master in Water Management and Environmental Engineering. These two tracks are belonging to the master Civil Engineering at the TU Delft. My research is about extreme precipitation. I chose this subject because extreme precipitation has become more extreme in recent years. Floods all over the world are more and more in the news. I did see the consequences of a flood myself, when I was cycling around the Netherlands. When I reached Limburg, in the summer of 2021, the rivers had overflowed their banks. I had never seen a flood before, let alone a flood this extreme. I was surprised that something like this could happen in the Netherlands, a country with a great infrastructure and a long history of water defense. This event made me even more interested in the causes of extreme precipitation. The second reason for me to choose this topic is that extreme precipitation is likely to change even more in the future due to climate change. How it will change, requires more information about the variables determining the behavior of extreme precipitation. I hope that this research gives a first insight into these variables.

I would like to thank my supervisor Ruud. We had meetings every two weeks, where we had nice and fruitful discussions about my progress and next steps. I left every meeting with many new ideas to improve my research and extract new information from the data. I also would like to thank my supervisor Gaby, who also attended these meetings in the beginning until she went on maternity leave. Remko and Pier, I am also grateful for your contribution to my research. You provided feedback from a different point of view, which was very helpful and improved the quality of my research.

I really enjoyed doing my thesis. I learned many new things about handling large datasets, gaining more insight into the information they provide and presenting it in a comprehensive and clear way. I am very proud of the final results and hope you enjoy reading my thesis and learn something (more) about the controls of extreme precipitation.

Irene van der Veer
Delft, June 6, 2022

Abstract

Extreme precipitation can be characterized by the tail behavior of precipitation probability distributions. The tail contains the most extreme precipitation events and tells something about both the magnitude and frequency of these events. Here the heaviness amplification factor is used to represent the tail behavior. This research determines the most important hydro-climatological controls of the heaviness amplification factor. The study area is the entire world, where Antarctica and the oceans are not taken into account. A pre-defined set of 17 controls divided into four groups is used. The four groups are: general climate condition, geography, land cover and climate variability. The relation between the heaviness and the controls is determined with multiple linear regression, where the standardized controls are used. Since multiple linear regression is applied, multicollinearity might be a problem. This term refers to the occasion where independent variables are correlated to another independent variable or a linear combination of variables. This could lead to erroneous and unreliable results. To deal with this problem, three elimination methods are applied. This first method consists of a correlation analysis, where two controls are highly correlated if their correlation coefficient is higher than 0.9. The control with the lowest regression coefficient is eliminated. This is followed by another regression analysis with all the remaining controls, where controls with a low regression coefficient are also eliminated. The second method applies an iterative regression procedure, where in each iteration the control with the lowest coefficient is eliminated until three controls remain. The last method calculates the variance inflation factor, which is a measure for multicollinearity. It removes the controls with the highest variance inflation factor if higher than 10, until none of the controls has a score above this limit. The described analysis is done for the entire world, but also for 44 IPCC climate reference regions. The conclusion of the analysis is that the most important controls for the heaviness belong to the general climate condition and the climate variability. In total, 13 of the 17 controls occur in a top three for at least one region in one or more of the elimination methods. The most important control is precipitation variability which is in 57% of the regions in the top three based on the results of methods 1, 2 and 3. Each region gives a different top three or at least a different order of controls. The heaviness is calculated based on the outcomes of the multiple linear regression. The R-squared ranges from 0.2 to 0.8, with an average value around 0.3. This suggests that the controls explain about 30% of the variance. The average error for the world between the data and the calculation of the heaviness is 0.11. When the heaviness of a single region is calculated, the absolute error between the data and the calculation of the heaviness reduces from 0.12 to 0.10 when using regional coefficients. This result leads to the conclusion that regional processes do play a role in the behavior of extreme precipitation. This makes it less appropriate to define a global set of most important controls. None of the three elimination methods is able to solve the problem of multicollinearity completely. On average, 35% of the controls have a opposite sign between the correlation and regression coefficient.

Contents

1 Introduction

Precipitation is very important for life on Earth. It is one of the most abundant and easily accessible sources of fresh water. However, extreme amounts of precipitation can have a large (negative) impact on the human society and its ecosystems. One example of the consequences of extreme rainfall is the flooding in the summer of 2021 in parts of Germany, Belgium and the Netherlands. Extreme amounts of unexpected precipitation in the days just before the flooding were a major cause of this event, which led to fatalities in Belgium and Germany and to economic losses in the entire affected area (Jonkman et al., 2021). Kreienkamp et al. (2021) investigated the causes of this event. They found that the observed rainfall in the affected catchment broke existing records. Moreover, they stated that climate change increased the intensity of rainfall and that events like this are likely to occur more frequently in the future.

Extreme precipitation can, among other things, also lead to landslides, damaged crop production and degradation of water quality due to sewer overflow. To limit its effect on the human society now and in the future, more insight into this phenomenon is needed. This knowledge can be used for infrastructure planning and design which can help to keep our feet dry in a changing climate consisting of more extreme weather events.

Extreme precipitation is described by the tail of precipitation distributions. The tail tells something about both the magnitude and frequency of extreme precipitation events. The tails can be divided into two types: heavier than exponential (heavy) and less heavy than exponential (thin). A region with a heavy tail experiences larger precipitation extremes and more variability between extremes. Regions with thin tails have lower extremes and less variability in the amount of precipitation (Papalexiou et al., 2013).

Research has been done to explore the tail behavior of extreme precipitation, using different probability distributions. Papalexiou et al. (2013) used more than 15 000 station records from all over the world to assess the performance of different tails based on the Pareto, Lognormal, Weibull and Gamma tail distributions. The data covers time periods from 50 to 172 years. They found that heavy tailed distributions are representing the reality more accurately than light-tailed distributions. Ragulina & Reitan (2017) stated that, based on station data, there are regional differences in the generalized extreme value (GEV) shape parameter for extreme precipitation. They also found that the shape parameter depends on elevation. If the elevation increases, the shape parameter decreases. The relationship found between the shape parameter and the elevation is linear. The effect of latitude remains unclear. Their results are based on two sets of stations. The first one has 169 stations covering parts of the United States and Great Britain and 71 stations in Norway. The second set includes the stations in the first set, but is extended with stations in Europe, Australia and South Africa. The total number of stations in this set is 1495.

The tail can be characterized by various variables. In this research it is represented by the heaviness amplification factor h. This is a variable developed by Gründemann et al. (2020) and is based on return periods. The heaviness amplification factor expresses the additional in- or decrease between selected return periods, compared to an exponential tail. In this case, h is based on the 2-year (T2), 20-year (T20) and 200-year (T200) return periods. Using T1, T10 and T100, resulted in a negative average heaviness value, which is not realistic (tested but not shown). Using larger return periods is computationally more expensive. T2, T20 and T200 resulted in expected heaviness amplification factors and are therefore chosen. b expresses the difference between T200 and T20 and a is the additional in- or decrease between T200 and T20 compared to an exponential tail, shown in Figure 1.1.

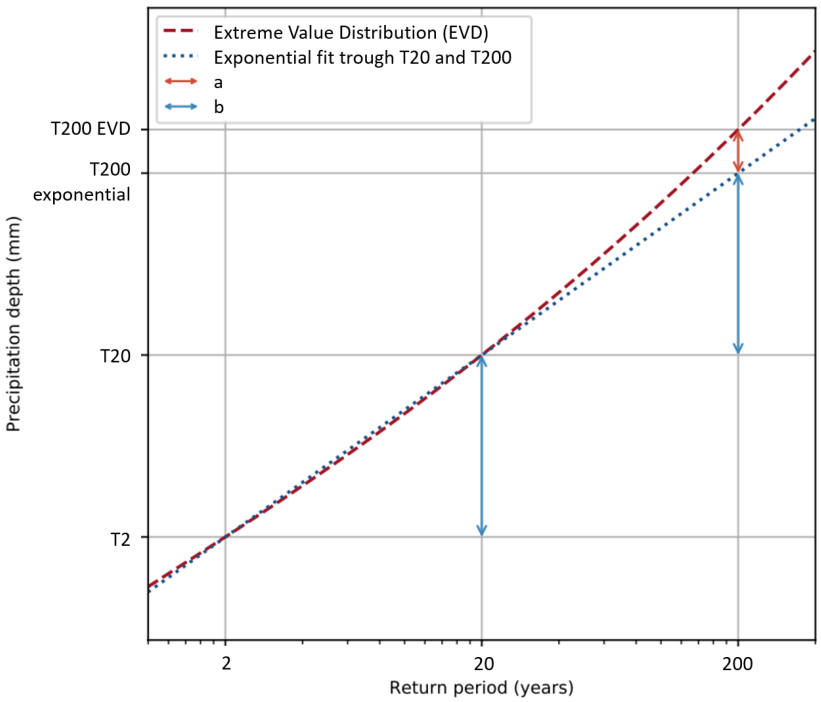

Figure 1.1: Visual explanation of the heaviness amplification factor, based on T2, T20 and T200. Where a represents the additional in- or decrease compared to an exponential tail and b the increase between T20 and T200 (modified from Gründemann et al. (2020)).

To obtain h, Equation 1.1 is used. This is a modified version of Equation 5 from Gründemann et al. (2020).

$$h = \frac{a}{b} = \frac{\text{T200} - 2 \cdot \text{T20} + \text{T2}}{\text{T20} - \text{T2}} = \frac{\text{T200} - \text{T20}}{\text{T20} - \text{T2}} - 1 \tag{1.1}$$

where T200, T20 and T2 are the precipitation return-levels, in this case belonging to a 1-day extreme precipitation event. If $h < 0$, the distribution is considered to have a thin tail and if $h > 0$, the distribution is considered heavy-tailed. The heaviness amplification factor can also be calculated for other precipitation durations. The results in Gründemann et al. (2020) show that the heaviness depends on the climatic zone and the orography. Generally speaking, the heaviness is higher in arid regions and lower in mountain regions. Moreover, as found by Ragulina & Reitan (2017), the heaviness is lower at higher elevations. Gründemann et al. (2020) state that more in-depth analysis is needed to draw conclusions on the variables that control the heaviness. The presented overview shows that the tail behavior of extreme precipitation changes with location. It probably depends on the elevation/orography and the climatic zone. However, no global analysis on the factors that influence the tail behavior has been done yet.

The main objective of this thesis is to identify the dominant factors that control the tail behavior of extreme precipitation. The research question that this thesis addresses is:

> What are the dominant hydro-climatological controls of the tail behavior of extreme precipitation distributions?

This research looks at the accumulated 1-day precipitation, where calendar days are used. The tail behavior is characterized by the heaviness amplification factor h and the return periods are calculated with the Metastatistical Extreme Value method. ERA5 precipitation data is used for the calculation of h. This research looks at extreme precipitation on land, excluding Antarctica. The intended outcome is a list of controls that affect the tail behavior of extreme precipitation and a quantification of the relation obtained with multiple linear regression. This research only looks at predefined controls and therefore does not provide an extensive list of all the controls which influence the tail behavior of extreme precipitation.

2
Materials and Methods

2.1. Tail behavior

The return periods needed to calculate the heaviness amplification factor are determined with the Metastatistical Extreme Value (MEV) method, a new method developed by Zorzetto et al. (2016). The main difference with the Generalized Extreme Value (GEV) and Peak-Over-Threshold (POT) method is that MEV makes use of all the measurements and not only of the extreme ones. Gründemann et al. (2020) uses the GEV, POT and MEV method to calculate the tail behavior of extreme precipitation using satellite and reanalysis-based global precipitation datasets. They conclude that the MEV method leads to a smoother and more coherent spatial pattern of the extremes. Other advantages are that it requires less years of data to obtain reliable results and is less sensitive to precipitation duration, compared to GEV and POT. One disadvantage is that MEV tends to overestimate the extremes. The discrete MEV-Weibull cumulative distribution function is given by Equation 2.1 (Zorzetto et al., 2016).

$$\zeta_M(x) = \frac{1}{M}\sum_{j=1}^{M}\left\{1-\exp\left[-\left(\frac{x}{C_j}\right)^{w_j}\right]\right\}^{n_j} \tag{2.1}$$

where j is the hydrological year, C_j the Weibull scale parameter, w_j the Weibull shape parameter and n_j the number of wet events in year j, all being stochastic variables. With these parameters and the `mevpy` package provided by Zorzetto (2018), the required return periods are calculated.

2.2. Data

For the calculation of h (Equation 1.1) and most of the controls, the ERA5 dataset is used. ERA5 is a reanalysis product from 1979 to near real time. It provides many observations both at the near surface and the upper air. It is selected because it has a high spatial and temporal resolution, covers the entire globe and has no gaps in the data due to the reanalysis (Hersbach et al., 2020). For h, the return periods are needed. These return periods are based on hourly precipitation data obtained from the ERA5 HRES dataset with a resolution of 0.25 x 0.25°. The data covers the period from January 1, 1979 until December 31, 2019. The hourly precipitation data is grouped into daily values. A dataset was made available which contained yearly C, w and n values used to calculate T2, T20 and T200 (Gründemann, 2021). The parameters of the MEV Weibull are based on all events above a threshold of 1 mm/day and estimated for each calendar year based on the PWN method (Greenwood et al., 1979). The final result is one heaviness value for each grid cell.

For most of the controls, the ERA5 monthly reanalysis product is used. Data from January 1, 1979 until December 31, 2019 is selected. If the required control was not obtained from ERA5, another dataset was used which is explained in Section 2.4. Linear interpolation is performed to transform the data to the ERA5 grid. Therefore, the CDO package `remapbil` is used, which applies bilinear interpolation from one grid to, in this case, the ERA5 grid (Schulzweida et al., 2006). The ERA5 grid is used as standard grid, with a resolution of 0.25 x 0.25°. It is a regular grid which results in different grid cell sizes for different latitudes.

2.3. Study area

The study area is based on the ERA5 grid. The oceans are filtered out and Antarctica is excluded. This means that only the grid cells from 90°N to 60°S are taken into account. It was not possible to determine h for every grid cell in the initial study area due to insufficient amount of data. Insufficient means that there were not enough precipitation events in a year. When less than 5 events occurred in a year, this year was excluded from the dataset. If this resulted in a total number of years less than 30, the heaviness result would not be reliable and the grid cell is excluded (Gründemann et al., 2020). This leads for example to the exclusion of the eastern part of the Sahara. Moreover, only the IPCC land regions (explained in more detail in Section 2.3.1) are used in the study. Therefore, any grid cell outside these areas is excluded from the grid as well. The resulting study area is shown in Figure 2.1, where the green region is the study area and the grey area the excluded land area.

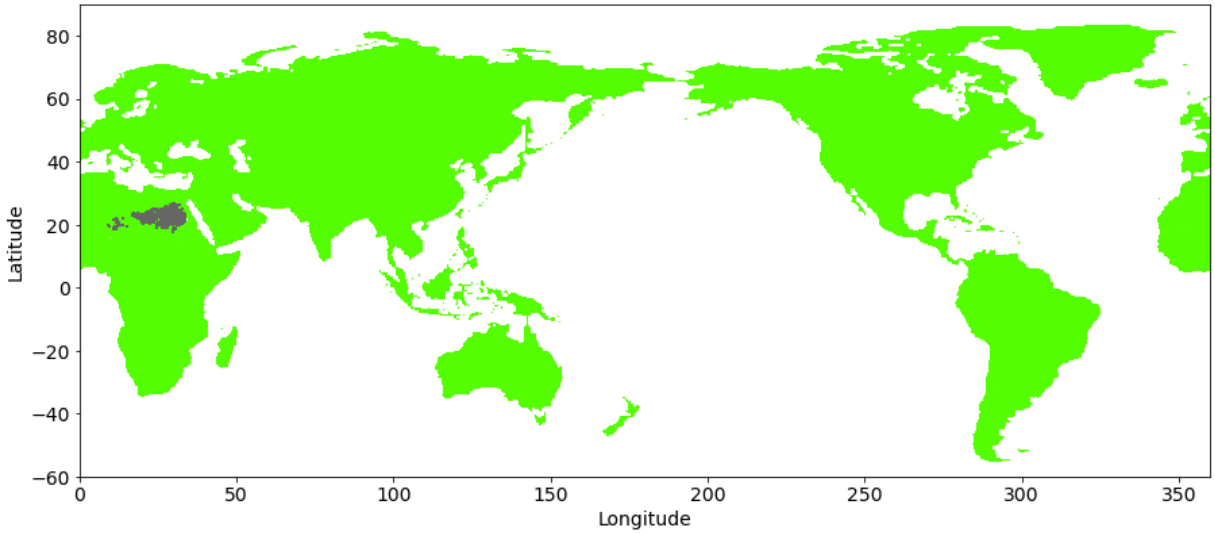

Figure 2.1: Area of study based on the ERA5 grid (Hersbach et al., 2019) and the IPCC land regions (Iturbide et al., 2020). The green region represents the area that is taken into account in this research and the grey region the excluded land area due to insufficient amount of data.

This study area is used to filter all the controls after linear interpolation. This means that all variables have the same number of grid cells with a value. The total number of filled grid cells is equal to 277 378. The grid has 601 latitude values and 1440 longitude values. Due to the grid choice, the grid cells around the Equator have larger areas than the grid cells near the poles. To investigate the influence of the grid cell area on the result of the heaviness amplification factor, a simple study into this phenomenon is done. The method and result are given in Appendix A. Based on this outcome, the conclusion can be drawn that the change in heaviness is small if present. Therefore, the use of the original ERA5 grid is maintained.

2.3.1. Climate Reference Regions

To gain more insight into the regional differences, the analysis is not only performed at global scale, but also at regional scale. The regions defined by the IPCC are used. The IPCC divides the world into 46 land and 15 ocean regions, which show large similarities with the Köppen-Geiger climate classification. Each region has similar climatological conditions based on precipitation and temperature data (Iturbide et al., 2020). In this case, only the land regions are used, where the two regions covering Antarctica (WAN and EAN) are not taken into account. This results in 44 regions for this research, shown in Figure 2.2.

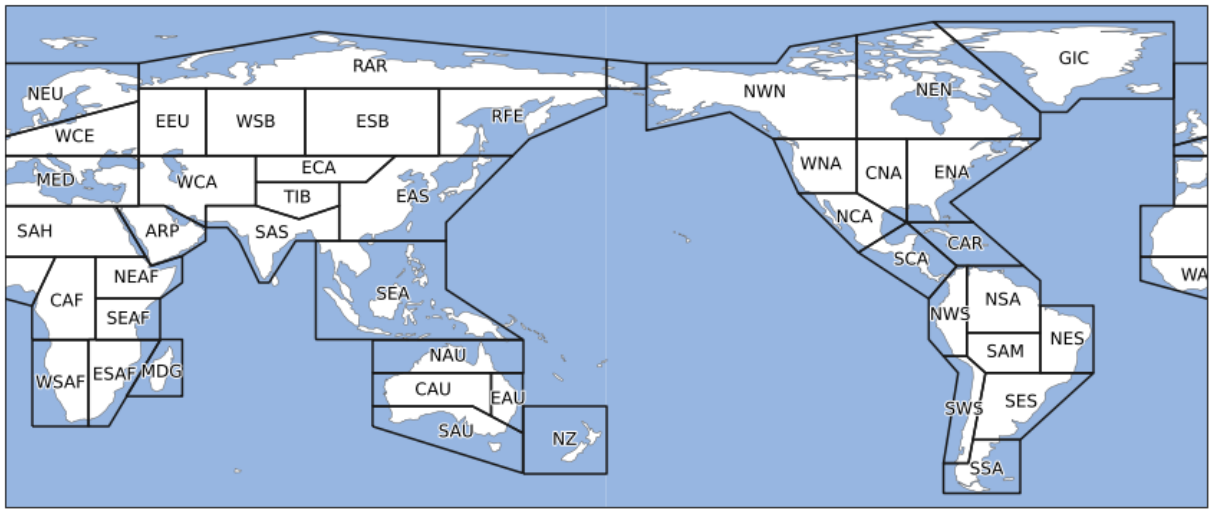

Figure 2.2: IPCC climatic reference regions, only showing the land regions that are taken into account in this research (Iturbide et al., 2020).

To determine if the regions have a different heaviness behavior than the world, a two-sided T-test for the means is performed (Dekking et al., 2005). The variances are unequal in this case. The variable of interest is the p-value, which is a number from 0 to 1, indicating how likely it is that means of two samples occurred by chance. If the p-value is low, it is likely that the differences are not caused by chance but that the datasets differ from each other. In this case, a significance level of 0.05 is chosen, meaning that regions with a p-value lower than 0.05 have a different behavior than the world.

2.4. Controls

The controls that are taken into account are divided into four groups: controls representing the general climate condition, geography, land cover and climate variability. The equations, python packages and datasets to obtain each control are given in more detail in this section. The ERA5 dataset (Hersbach et al., 2019) used in this research for the controls gives monthly averages. When the control contains a time dimension, this dimension is days. For example, the average monthly precipitation is given in mm/day. For all the controls, the mean of the yearly averages is calculated based on 41 years of data. This mean is used in further analysis. Thus, each grid cell has 17 different values, one for each control.

2.4.1. General climate condition

The controls belonging to this category are given below, including their SI quantity. The total precipitation per wet day and the relative humidity were not directly given by ERA5 but calculated from other ERA5 products.

$$\text{2m temperature [T]: } T_{i,j} = \frac{1}{n}\sum_{k=1}^{n} T_{k,i,j} \tag{2.2}$$

$$\text{Total yearly precipitation [L/t]: } P_{i,j} = \frac{12}{n}\sum_{k=1}^{n} P_{k,i,j}\cdot D_k \tag{2.3}$$

$$\text{Evaporation [L/t]: } \mathrm{E}_{i,j} = \frac{12}{n}\sum_{k=1}^{n} \mathrm{E}_{k,i,j}\cdot D_k \tag{2.4}$$

$$\text{Number of wet days [t/t]: } W_{\text{days},i,j} = \frac{12}{n} \sum_{k=1}^{n} W_{\text{days},k,i,j} \tag{2.5}$$

$$\text{Total precipitation per wet day [L/t]: } \mathrm{P}_{\text{Wdays},i,j} = \frac{P_{i,j}}{W_{\text{days},i,j}} \tag{2.6}$$

$$\text{Total cloud cover [-]: } C_{t,i,j} = \frac{1}{n} \sum_{k=1}^{n} C_{t,k,i,j} \tag{2.7}$$

$$\text{Relative humidity [\%]: } \mathrm{RH}_{i,j} = \exp\left(\frac{17.625 \cdot T_{d,i,j}}{T_{d,i,j} + 243.04} - \frac{17.625 \cdot T_{i,j}}{T_{i,j} + 243.04} \right) \tag{2.8}$$

where i and j refer to the latitude and longitude, D_k the number of days in month k, T_d the dew point temperature in Kelvin and n the total number of months within the period of interest. In this case, 41 years are taken into account which results in a n value of 492.

The main reason to take these controls into account is because they can be related to the climatic zone. The climatic zone, in turn, influences the heaviness. Moreover, temperature and extreme precipitation influence each other. This relationship is location dependent and can be both positive and negative. It strongly depends on the duration of extreme precipitation (Utsumi et al., 2011). The number of wet days and amount of precipitation per wet day are taken into account since it can be used as an indicator for precipitation mechanisms. Dry and wet regions might have the same total amount of precipitation, but the previous mentioned controls will be different in these regions. Clouds play an important role in the current climate. They generate precipitation, are a measure for atmospheric stability and also interfere with sunlight and infrared radiation (Dastrup et al., 2020). Moreover, the updraughts of clouds can transport air from the surface to higher parts of the atmosphere in a short amount of time, relevant for generating precipitation (Boucher et al., 2013). In this research, the total cloud cover product is used. The relative humidity is taken into account because when the amount of water vapor in the air increases, thus the relative humidity increases, it is more likely that a (heavy) precipitation event will occur (Kreienkamp et al., 2021).

2.4.2. Geography

Geography, or surface orography, influences the precipitation patterns (Roe, 2005). To account for the geography, the elevation is used. The elevation is selected because the amount of precipitation and also the intensity and frequency of precipitation increases with elevation (Chu, 2012). Moreover, both Ragulina & Reitan (2017) and Gründemann et al. (2020) state that the shape of extreme precipitation distributions depends on elevation. The Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) is used. This dataset is based on 11 raster-based elevation sources. The most important source is the SRTM Digital Terrain Elevation Data. Additional sources were used to cover the areas outside the SRTM coverage and to fill gaps. This dataset has a resolution of approximately 250 meters and covers the entire globe, excluding the oceans (Danielson & Gesch, 2011). The data is linearly interpolated to the ERA5 grid.

To further account for the orography, slope aspect (S_a), slope riserun (S_r), wind speed (V) and wind direction vs slope (ρ) are taken into account. The slope aspect determines the direction of the slope in degrees, where 0° refers to the North. The slope riserun is a measure for the steepness, defined as rise over run. Where rise is the distance covered in vertical direction and run the distance in horizontal direction. They are calculated using the elevation data and the python package `richDEM`. The aspect represents the direction of the maximum slope of the cell. The riserun for a cell is determined by applying a central difference estimation of the surface of that cell and of its neighbours. The riserun is then the maximum of the calculated surface (Horn, 1981). Wind speed is taken into account since research (e.g. Back & Bretherton (2005))

suggests that higher precipitation amounts are related to faster winds. The wind speed is given by:

$$\text{Wind speed [L/t]: } V_{i,j} = \frac{1}{n}\sum_{k=1}^{n}\sqrt{u_{k,i,j}^2 + v_{k,i,j}^2} \tag{2.9}$$

here $u_{i,j}$ is the wind speed in m/s in the u direction and $v_{i,j}$ the wind speed in v direction in m/s .

Not only the wind speed, but also the wind direction can be relevant for precipitation, especially in mountain regions. If the dominant wind direction and the slope aspect have the same direction, larger amounts of precipitation are expected (Beullens et al., 2014). Thus, most precipitation appears at the side where the slope aspect is parallel to the dominant wind direction. Both factors are combined in one variable called α, given in Equation 2.11. Since both variables can have a similar direction with varying steepness of the slope, α is multiplied with the slope riserun to account for this. The control of interest is called ρ, given in Equation 2.10, representing the wind direction vs the slope.

$$\text{Wind direction vs slope [-]: } \rho_{i,j} = S_{r,i,j} \cdot \alpha_{i,j} \tag{2.10}$$

where $S_{r,i,j}$ is the slope riserun and α given by:

$$\alpha_{i,j} = \begin{cases} \left[1 - \dfrac{|\phi_{i,j} - S_{a,i,j}|}{90}\right] & \text{if } |\phi - S_{a,i,j}| \leq 180 \\ \left[1 - \dfrac{360 - |\phi_{i,j} - S_{a,i,j}|}{90}\right] & \text{if } |\phi - S_{a,i,j}| > 180 \end{cases} \tag{2.11}$$

where $S_{a,i,j}$ is the slope aspect and $\phi_{i,j}$ the dominant wind direction given by:

$$\phi_{i,j} = \text{mod}\left[180 + \frac{180}{\pi}\arctan 2\left(\frac{1}{n}\sum_{k=1}^{n} v_{k,i,j}, \frac{1}{n}\sum_{k=1}^{n} u_{k,i,j}\right), 360\right] \tag{2.12}$$

where $u_{i,j}$ is the wind speed in m/s in the u direction, $v_{i,j}$ the wind speed in v direction in m/s and $\arctan 2$ the arc tangent of $\frac{1}{n}\sum_{k=1}^{n} v_{k,i,j}$ divided by $\frac{1}{n}\sum_{k=1}^{n} u_{k,i,j}$.

2.4.3. Land cover

Land cover can influence the daily precipitation extremes. Regions where the land cover changed, mainly to urban areas, experience more extreme precipitation and a larger increase in extremes compared to rural areas (Rahimpour Golroudbary et al., 2016). Deforestation on the other hand can lead to a decrease in precipitation (Bagley et al., 2014). Three controls representing the land cover are taken into account in this study. The first one is soil type $\text{ST}_{i,j}$, which is time invariant and therefore no additional processing of data was needed. It is a categorical variable where 1 is a coarse soil, 2 a medium soil, 3 a medium fine, 4 a fine, 5 a very fine, 6 an organic and 7 a tropical organic soil (Copernicus Climate Change Service, 2019). The other two controls of this category are:

$$\text{Albedo [-]: } A_{i,j} = \frac{1}{n}\sum_{k=1}^{n} A_{k,i,j} \tag{2.13}$$

$$\text{Leaf area index [L}^2\text{/L}^2\text{]: LAI}_{i,j} = \frac{1}{n}\sum_{k=1}^{n}\text{LAI}_{\text{low},k,i,j} + \frac{1}{n}\sum_{k=1}^{n}\text{LAI}_{\text{high},k,i,j} \tag{2.14}$$

where LAI_{low} and LAI_{high} are two LAI products, where the first one refers to low vegetation like crops and grass and the second one to high vegetation, for example trees.

2.4.4. Climate variability

Precipitation mechanisms can differ from location to location. Therefore, the underlying patterns leading to extreme amounts of precipitation are different per region (Liu et al., 2020). To account for this, the variability of precipitation in time is taken into account as well. This control tells something about the spread among precipitation events for each location. Temperature variability is also treated as a control.

$$\text{Precipitation variability [-]: } P_{\text{var},i,j} = \frac{\sigma_{P_{i,j}}}{\overline{P_{i,j}}} \tag{2.15}$$

$$\text{Temperature variability [-]: } T_{\text{var},i,j} = \frac{\sigma_{T_{i,j}}}{\overline{T_{i,j}}} \tag{2.16}$$

where $\sigma_{P_{i,j}}$ is the standard deviation for location i,j and $\overline{P_{i,j}}$ the average precipitation, both in mm/day based on monthly averages. The same holds for temperature, which is given in Kelvin.

2.5. Multiple linear regression

To determine the controls that are most dominant for the heaviness, various methods can be applied. The first idea was to use Principal Component Regression (PCR) which consists of a principle component analysis followed by a multiple linear regression (MLR) with the principle components. PCR is used when the independent variables are correlated to each other, which is the case for the selected controls, shown in Appendix B. Correlation between the independent variables can lead to multicollinearity. Multicollinearity refers to the linear relation between two or more independent variables. The existence of this relationship can lead to unreliable estimates of the MLR (Alin, 2010). PCR solves this problem by determining principle components that are orthogonal to each other. Noori et al. (2010) for example, applies PCR to predict river flows and concludes that PCR improves the outcomes compared to MLR. However, when PCR is applied to the controls in this research, the component explaining the least amount of variance turned out to be the most important one in the MLR. Therefore, applying PCR is less appropriate in this case. Moreover, interpretation of the final results and the link to the original controls is complicated. This makes it difficult to determine the most important controls.

Based on the outcome of this process, the choice was made to apply MLR to assess the linear relationship between the heaviness and the controls. MLR is a method often used to determine the factors that influence (extreme) precipitation (e.g. Um et al. (2011)). In this research, MLR is applied to the standardized controls, meaning that the controls have zero mean and unit variance. For each control, the following equation is applied to scale the data:

$$x_{\text{scaled},i,j} = \frac{x_{i,j} - \mu}{\sigma} \tag{2.17}$$

where $x_{\text{scaled},i,j}$ is the scaled value for each grid cell for a certain control, $x_{i,j}$ is the original value, μ is the mean of the control and σ is the standard deviation. The resulting regression coefficients are, in this case, a measure for the dominance of each control in relation to the heaviness. The coefficients also provide the direction of the dependency. The main problem with MLR is that it is prone to multicollinearity, as stated before (Alin, 2010). Therefore, three different elimination procedures are applied to the data to cope with this problem. These procedures are described in the Section 2.6. The MLR is applied to all the regions including the entire world. This means that there is a global set of final controls as well as 44 regional sets of final controls for all three elimination methods.

2.6. Elimination of controls

The multiple linear regression is done with a subset of controls. This subset is determined based on an elimination procedure. There are two reasons to apply an elimination method. The first one is to eliminate multicollinearity in the controls. When this is present, the coefficients of the MLR can have opposite signs compared to the correlation coefficient or become insignificant and the result will be unstable (Falk & Miller, 1992). Generally speaking, there are three different options to deal with multicollinearity: excluding variables, transform variables or adding data (Kumari, 2008). In this case, exclusion of variables is chosen. Transforming variables is already done by standardizing them (Equation 2.17). Another option for transformation is to combine variables, which would make it difficult to assess the dominance of each of the original variables. Adding more data is considered not feasible at this point. The second reason for applying the elimination procedure is that the goal of this research is to determine the most important controls. Therefore, controls that do not have a large influence on the heaviness amplification factor are also eliminated from the dataset. The result of the elimination method is a set of final controls for each region, which is used in the multiple linear regression. Since there is not one best solution to solve the problem of multicollinearity (Kumari, 2008), three different elimination methods are used and described below.

2.6.1. Method 1 - Unlimited number of controls

The first step of this method is a correlation analysis between the controls. Correlation is a measure of how strongly two variables are linearly related to each other. Two controls are considered highly correlated if they have a correlation coefficient higher than 0.9 for that region, indicating that multicollinearity might be a problem (Kumari, 2008). If this is the case, these two controls are used in a regression analysis with heaviness. The control with the lowest regression coefficient is eliminated. It is possible that a control has multiple combinations with a high correlation. The control is then eliminated if in at least one of these combinations the other control is considered more relevant.

After the correlation analysis, a regression analysis is done with the selection of controls. For each region and the world, the regression coefficients are determined. The relative importance of each control is calculated. Controls that are not relevant in any situation, are also filtered out. In this case, not relevant indicates that the relative importance is lower than 1%. The relative importance of a control, c_{rel}, is given by:

$$c_{\text{rel},x} = \frac{c_{\text{control},x}}{\sum_{q=1}^{p} c_{\text{control},q}} \tag{2.18}$$

where c_{control} is the coefficient of the controls and p the total number of controls used in the regression analysis.

2.6.2. Method 2 - Three most important controls

The second method selects the three most important controls based on the following iterative regression procedure. First, the regression coefficients are determined for all the controls in the dataset. The second step is to determine the control with the lowest absolute coefficient and exclude this control from the dataset. With the remaining controls, the regression analysis is performed again. This process is repeated until three controls are left.

2.6.3. Method 3 - Variance Inflation Factor

The third and last method makes use of the Variance Inflation Factor (VIF). This variable is a measure for the inflation of the variance due to correlation of the independent variables (Shrestha, 2020). It is different from the correlation coefficient (method 1) since it also described the multicollinearity among a linear combination of multiple controls. The correlation coefficient only described the correlation between two controls (Alin, 2010). The equation for the VIF is (Shrestha, 2020):

$$\text{VIF} = \frac{1}{1-R^2} \tag{2.19}$$

where R^2 is the coefficient of determination based on linear regression. The minimum value of the VIF is 1, indicating no correlation between the independent variables. A VIF higher than 10 indicates strong multicollinearity between variables and severe impact on the results of the MLR. The variable with the highest VIF value, if larger than 10, is eliminiated from the dataset. The VIF is calculated again with the reduced set of controls. If there are still variables with a VIF higher than 10, the largest is again eliminated. This is repeated until all the variables have a VIF value smaller than 10. To eliminate controls that are considered not relevant for the heaviness, the same step is applied as in method 1 (Equation 2.18).

2.7. Calculation of the heaviness

With the results of the MLR, the heaviness is calculated. In this case, the same dataset is used to calculate the coefficients and to determine the heaviness. No split sample test is applied. This is done because the goal of the research is not to be able to give a predication of the heaviness, but to be able to determine the most important controls. The determination of the heaviness based on the MLR gives and indication of the performance of the MLR. Two different types of calculations are done for both methods. The first type is the calculation of the heaviness for the world based on the coefficients of the world. The second type is the calculation of the heaviness for the regions based on the coefficients of the world and the coefficients obtained when only using regional data. The performance of the regression is expressed by the R-squared and Mean Squared Error (MSE), given in Equations 2.20 and 2.21, respectively (Pham, 2019).

$$\text{R}-\text{squared} = 1 - \frac{\sum(h_{i,j} - \hat{h_{i,j}})^2}{\sum(h_{i,j} - \overline{h})^2} \tag{2.20}$$

$$\text{MSE} = \frac{1}{p}\sum(h_{i,j} - \hat{h_{i,j}})^2 \tag{2.21}$$

where $h_{i,j}$ the measured heaviness, $\hat{h_{i,j}}$ the predicted heaviness value both for location i,j, $\overline{h}$ the average heaviness and p the number of data points.

2.8. Uncertainty and stationarity in the heaviness data

To get an idea about the uncertainty in the calculations of the heaviness, a bootstrap analysis is done. This method was first introduced by Efron (1992). It is a resampling technique used to estimate various properties about the data. The main advantage of this method is that it is simple to implement. It is a widely used technique to assess the uncertainty (Li et al., 2010).

In this research, 200 samples of 30 years are determined. 30 years are selected since it turns out that 30 years is enough to give a good heaviness result and resembles one climatic period (Arguez & Vose, 2011). The 30 years are selected at random, with replacement. For each set of 30 years, the return periods and thus the heaviness is calculated. The corresponding set of controls is calculated as well. The 30 years are used to

calculate the mean of the yearly averages. Thus, for each sample there is one heaviness value and one value per control for each location. To all the samples, elimination methods 1, 2 and 3 are applied. This is used to calculate the width of the 90% confidence interval.

The uncertainty analysis is done for the world, where all the data is used, for the world where 1000 random points are selected and for region MDG. MDG is one IPCC region and covers Madagascar, located at the east side of the African coast. The 1000 random points of the world are used to assess if the result of region MDG is in fact different from the global results or that this might be due to a smaller sample size. The 1000 random points are selected once and the analysis is repeated with the same 1000 points.

The same analysis is used to determine if the heaviness data is stationary. The 200 samples are used, but two additional samples are created. The first sample covers the period from 1979 until 2009 and the second sample the period from 1989 until 2019. For each set of 30 years, the return periods are calculated and the average heaviness value and standard deviation are determined. These values are compared to the results of the 200 samples and the time series of 41 years to see if there is a significant trend in the data.

3
Results

3.1. Tail behavior

The heaviness amplification factor based on the described precipitation data is shown in Figure 3.1. The heaviness is largest over the Sahara, the Arabian Peninsula and Australia. It is lower in the Northern part of South America, over Indonesia, central Africa and the northern part of Europe. Both thin and heavy tailed regions are present. Some orographic features of the Earth like the Himalayas in India and the Rocky Mountains in North-America are visible. The average heaviness value is 0.154 indicating an average heavy tail. The standard deviation is 0.246

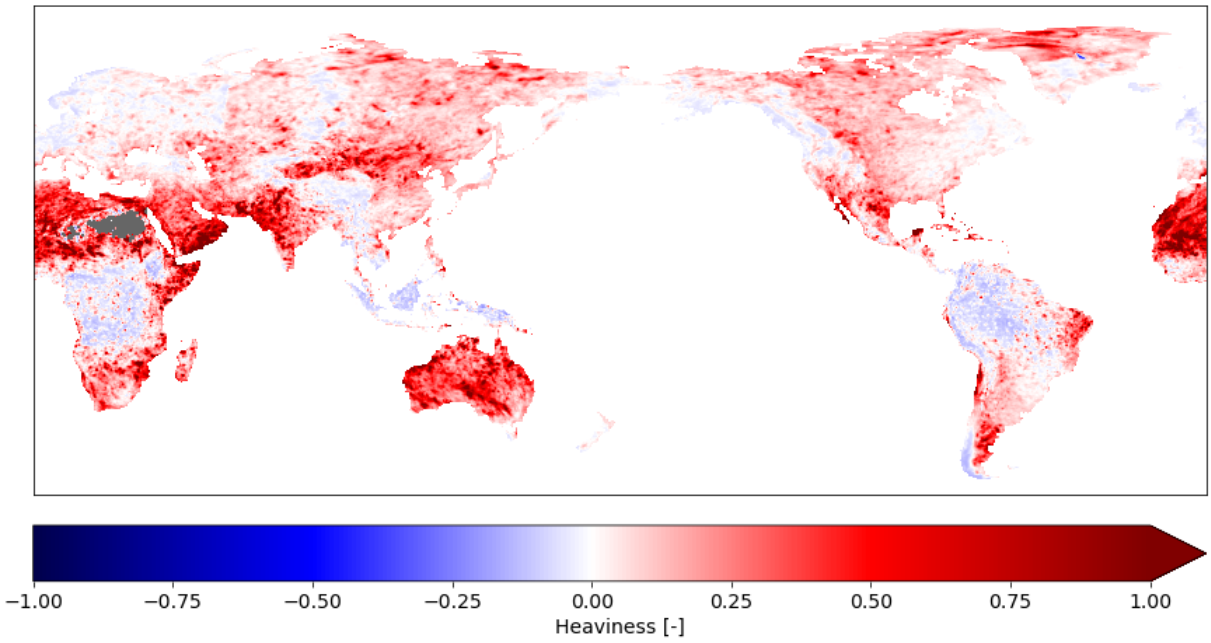

Figure 3.1: Heaviness amplification factor based on T2, T20 and T200 belonging to daily extremes calculated with Equation 1.1. The heaviness is based on precipitation data from 1979 until 2019 obtained from ERA5 HRES (Gründemann, 2021).

3.1.1. Climate Reference Regions

Based on the regional results, the boxplot shown in Figure 3.2 is produced. The range in heaviness values differs per region, where most of the spread is to higher heaviness values. SAH and NEAF have the highest heaviness values and largest spread. SAH and ARP have the highest mean heaviness values and are both dry desert regions, which also holds for NEAF. Most of the regions with lower average heaviness values and a smaller spread are located in coastal areas, like NSA, NEU and NZ. The range of the heaviness values is not clearly related to the average. Some regions with lower average heaviness values have a larger spread and vice versa. The location of the regions can be found in Figure 2.2.

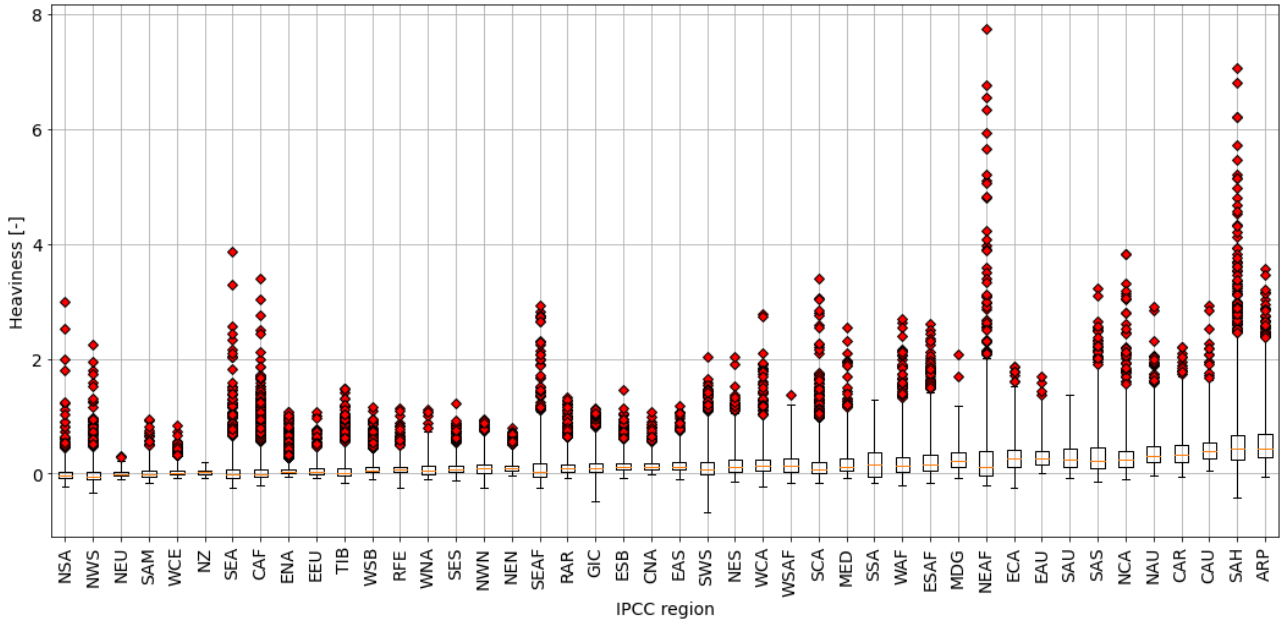

Figure 3.2: Box-and-whiskers plot of the heaviness amplification factor for each IPCC land region excluding Antarctica. The line in the box shows the median and the borders the lower (Q1) and upper (Q3) quartile values. The whiskers show the range of the data, where they extend from Q1 and Q3 a maximum of 4 times the distance from Q1 to Q3. The red squares show the outliers. The values are sorted based on the median heaviness value. The heaviness is calculated with Equation 1.1 and the location of the regions can be found in Figure 2.2.

The p-values for all the regions, when compared to the world, is given in Figure 3.3. The p-value is based on the two-sided T-test, where 0.05 is chosen as the significance level. This figure shows that only two regions have a p-value higher than 0.05 and are thus similar to the world. The other regions have a different heaviness behavior than the world, based on a comparison of the means.

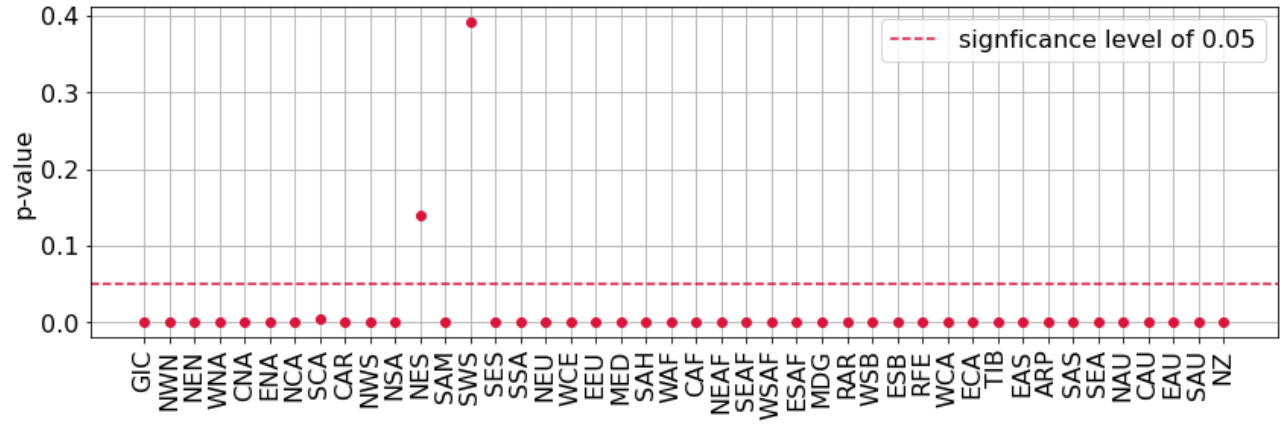

Figure 3.3: P-values of the regions, when the mean of the region is compared to the mean of the world. The p-value is based on the two-sided T-test and a significance level of 0.05 is chosen (Dekking et al., 2005). The location of the regions can be found in Figure 2.2.

3.2. Controls

3.2.1. Global results

The result of the amount of precipitation per wet day (PW_{days}) is shown in Figure 3.4. In this figure, orographic features, like mountain ranges, are visible. Generally speaking, coastal regions experience more precipitation per wet day than more inland regions. The leaf area index (LAI) based on the ERA5 data is given in Figure 3.5. This figure shows a less smooth behavior than PW_{days}. Regions with forests, have higher LAI values. Dry, desert regions show LAI values around zero. High values occur in the Amazon and Indonesia. Orographic features, mainly mountain ranges, are visible as well.

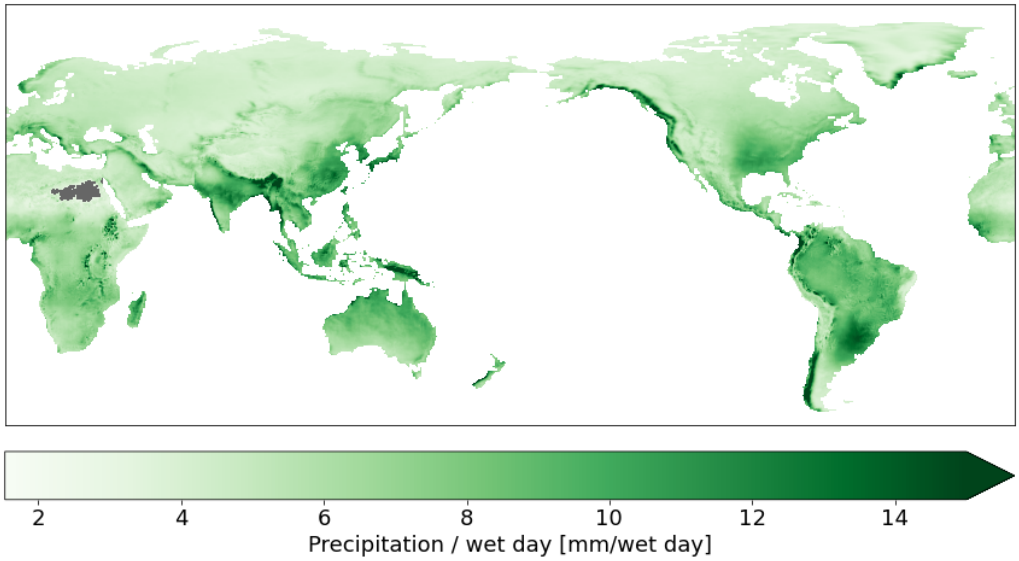

Figure 3.4: Average precipitation per wet day, calculated with Equation 2.6, based on ERA5 monthly reanalysis data from 1979 until 2019 (Hersbach et al., 2019).

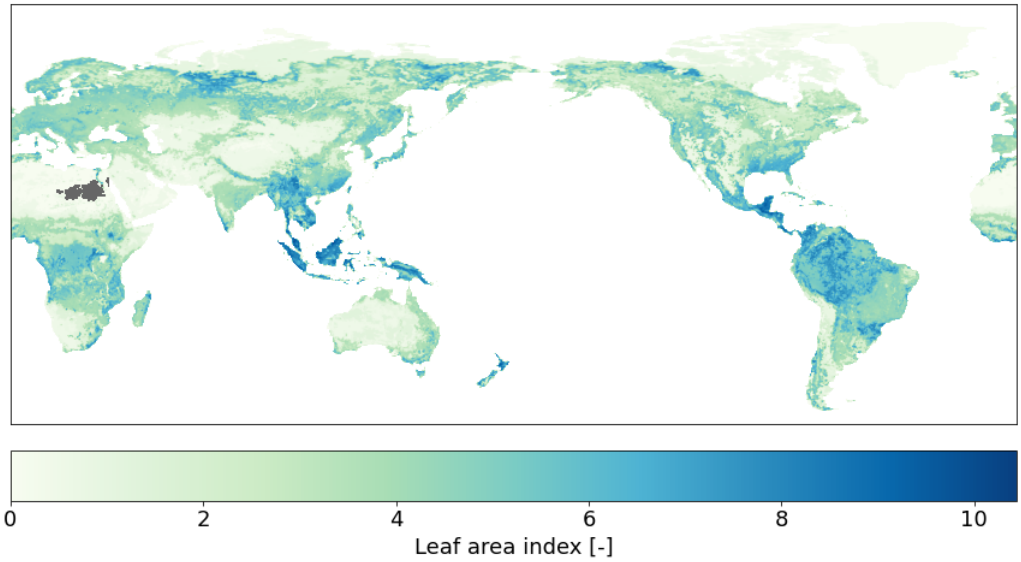

Figure 3.5: Average leaf area index, calculated with Equation 2.14, based on ERA5 monthly reanalysis data from 1979 until 2019 (Hersbach et al., 2019).

The controls belonging to the general climate condition (Appendix C.1) also show orographic features. Moreover, they all show similar patterns except for temperature. Temperature shows high similarities with elevation as expected. The other controls report low values in dry regions like the Sahara and high values around Patagonia, in Indonesia, over central Africa and in the Himalayas. The controls related to geography (Appendix C.2) look different from the general climate condition controls and from each other. The patterns are less gradual except for elevation. The other controls show a more diverse and rapid changing pattern, where the mountain ranges are still visible, but less pronounced. The slope riserun shows the mountains clearly. For the land cover (Appendix C.3), the albedo shows a gradual pattern, with the highest values in the north. Also the mountains and the Sahara are clearly visible. The soil type shows a pattern where larger areas have the same type. The precipitation variability (Appendix C.4) is highest in the Sahara. For the temperature variability, the northern part of the world shows the highest variability. Large mountain ranges are visible in both controls.

3.2.2. Controls versus heaviness

For the controls, a scatter density plot is produced. The result for PW_{days} and LAI is shown in Figures 3.6a and 3.6b. They show that most data points have a heaviness value between 0 and 0.5. The points belonging to PW_{days} cluster between a value of 0.005 and 0.010 mm/wet day. These points show a decaying trend indicating that larger amounts of precipitation per wet day are associated with lower heaviness values. This plot also shows that if the amount of precipitation per wet day is high, the heaviness is low and vice versa. For the LAI, there are a couple of spots where there is a densification of points. The general trend is slightly decaying, indicating a less heavy tail behavior of precipitation with increasing LAI values.

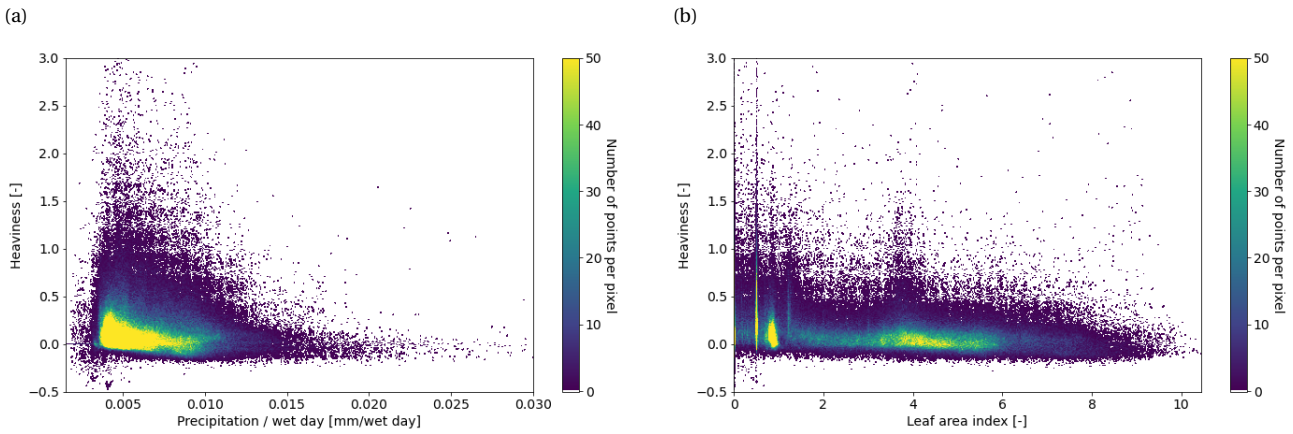

Figure 3.6: a) Scatter density plot of precipitation per wet day (Equation 2.6) vs heaviness (Equation 1.1). b) Scatter density plot of the leaf area index (Equation 2.14) vs heaviness (Equation 1.1). The more yellow the color, the larger the number of points in this grid cell. Both controls are obtained or calculated from ERA5 monthly reanalysis data covering the period from 1979 until 2019 (Hersbach et al., 2019).

The results of the density scatter plots of the other controls are given in Appendix D. Most of the controls show one spot where a densification of points is visible. Most of the times, there is also an in- or decrease visible within this higher density space. Temperature, albedo and temperature variability show multiple spots with a higher density. The slope aspect gives a denser pattern along the entire x-axis. The density only varies in the y-direction. The control wind direction vs slope shows a more or less symmetric pattern around the line y = 0. Soil type shows a discrete pattern and multiple spots with a higher density of points.

3.3. Elimination of controls

3.3.1. Method 1 - Unlimited number of controls

The final result of both steps of method 1 is shown in Figure 3.7. This figure shows the controls remaining after applying the correlation and regression analysis. The number in the grid cell indicates the relative importance of the control for that region based on the MLR with the final set of controls. If a grid cell is black, the control for that specific region is eliminated in the correlation step. If it is is grey, it means that it is eliminated in the regression step. The total number of controls taken into account for each region as well as the total number of times a control is used are shown as well.

The controls with the highest relative importance are mainly located in the upper part of the graph. These controls belong to the general climate condition. The variability in precipitation and temperature are also considered relevant and have relatively many regions where they are relevant. The slope aspect, slope riserun, wind speed, wind direction vs slope, albedo, soil type and leaf area index are relevant in a couple of regions and have a low relative importance. It is interesting to see that none of the controls is considered relevant in all the regions. Moreover, based on the MLR with the final controls, some controls have a relative importance of 0%.

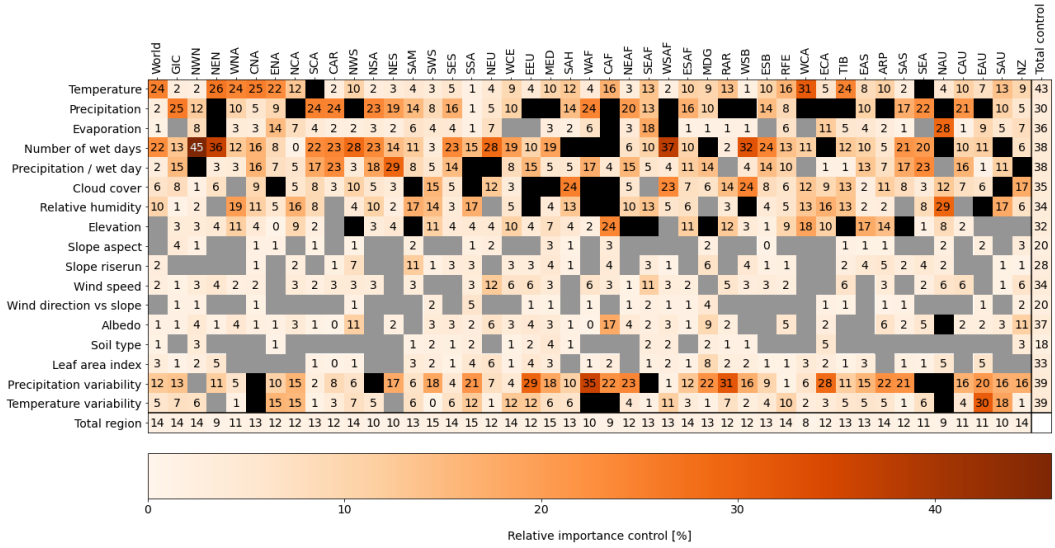

Figure 3.7: Result of the elimination procedure for method 1 where the controls used in the MLR for each region are shown as well as their relative importance based on the MLR with the final set of controls. If a grid cell is black, this control was eliminated in the correlation step. If a grid cell is grey, this control has a relative importance less than 1% and is eliminated. The total number of controls used for each region is given by the column Total region, and the total number of times a control is used in a region is given by Total control. The location of the regions can be found in Figure 2.2.

3.3.2. Method 2 - Three most important controls

For method 2, the final result is shown in Figure 3.8. This figure also shows that the controls belonging to the general climate condition and the climate variability are more important than the controls belonging to the geography and land cover. Many of the controls belonging to these last two categories are not taken into account at all, represented by the white grid cells.

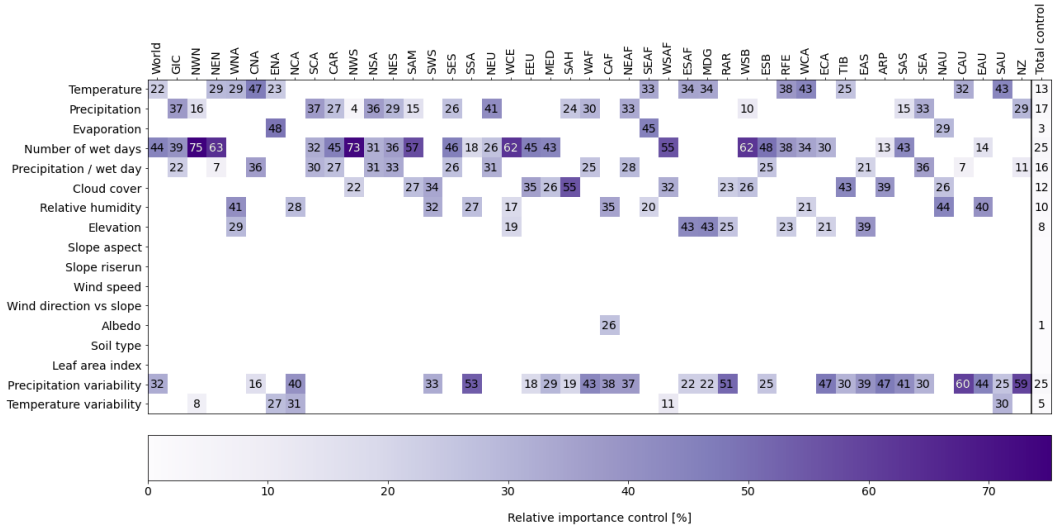

Figure 3.8: Result of the elimination procedure for method 2 where the controls used in the MLR for each region are shown as well as their relative importance based on the MLR with the final set of controls. If a grid cell is white, the control was eliminated. The total number of times a control is used is given by Total control. The location of the regions can be found in Figure 2.2.

3.3.3. Method 3 - Variance Inflation Factor

The result of method 3, given in Figure 3.9, shows that the temperature and precipitation are more often eliminated compared to the other two methods. This is due to the high VIF scores these controls have, indicating that they are highly correlated with other controls. The remaining part the figure is more or less similar, where the most important controls belong to the general climate condition and climate variability categories. This indicates that the exclusion of temperature and precipitation makes other controls in the general climate condition or climate variability categories more important. For example, cloud cover is reported important much more often for method 3 than for method 1 and 2.

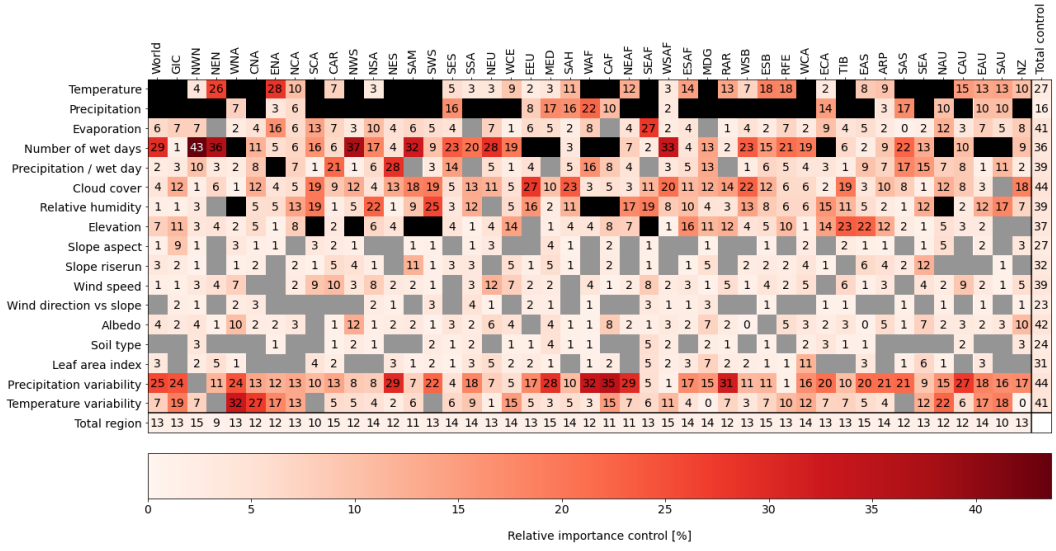

Figure 3.9: Result of the elimination procedure for method 3 where the controls used in the MLR for each region are shown as well as their relative importance based on the MLR with the final set of controls. If a grid cell is black, this control was eliminated in the VIF step (Equation 2.19). If a grid cell is grey, this control has a relative importance less than 1% and is eliminated. The total number of controls used for each region is given by Total region and the total number of times a control is used is given by Total control. The location of the regions can be found in Figure 2.2.

3.4. Multiple linear regression

3.4.1. Correlation controls and heaviness

The correlation between heaviness and the controls for each region is determined. Part of this result is given in Table 3.1. The entire table is given in Appendix E. The table shows the high diversity of correlation coefficients among both the controls and regions. Temperature for example, has both negative and positive correlation values with the heaviness. This pattern is visible for almost all controls. Moreover, controls also show a large range of correlation coefficients, where differences can be as large as a factor 10. This is for example the case for cloud cover (C_t).

Table 3.1: Correlation between the heaviness and the controls for a number of regions. The heaviness is calculated with Equation 1.1. The meaning of the abbreviations of the controls and their equations can be found in Section 2.4. The location of the regions can be seen in Figure 2.2. The entire table is given in Appendix E.

	World	GIC	NWN	NEN	WNA	CNA	ENA	NCA	SCA
T	0.225	−0.307	−0.393	−0.361	0.257	0.287	0.539	0.340	0.188
P	−0.347	−0.371	−0.493	−0.557	−0.412	−0.288	0.028	−0.364	−0.387
E	0.176	0.178	0.259	0.292	0.289	0.100	−0.634	−0.039	0.081
W_{days}	−0.488	−0.437	−0.644	−0.587	−0.531	−0.393	−0.152	−0.399	−0.372
P_{Wdays}	−0.092	−0.295	−0.392	−0.400	−0.211	−0.153	0.151	−0.029	−0.336
C_t	−0.484	−0.162	0.042	0.001	−0.367	−0.445	−0.461	−0.276	−0.398
RH	−0.459	0.019	−0.081	0.120	−0.426	−0.452	−0.066	0.080	−0.115
ELV	0.176	−0.007	−0.245	−0.162	−0.076	0.291	−0.309	−0.432	−0.212
S_a	−0.030	0.203	−0.099	−0.071	−0.234	−0.144	−0.097	−0.067	0.118
S_r	−0.060	−0.035	−0.256	0.036	−0.100	0.222	−0.178	−0.170	−0.236
V	0.178	0.046	−0.013	−0.077	0.073	0.443	−0.205	0.296	0.279
ρ	−0.001	−0.067	−0.011	−0.032	−0.086	−0.238	−0.032	−0.112	−0.047
A	0.007	−0.468	0.222	0.271	0.137	−0.086	−0.460	−0.138	−0.004
ST	−0.066	−0.154	0.159	−0.038	0.222	0.006	−0.043	−0.165	−0.011
LAI	−0.299	−0.154	−0.116	−0.303	−0.284	−0.075	−0.003	−0.291	0.017
P_{var}	0.555	0.439	0.468	0.461	0.347	0.465	0.614	0.463	0.174
T_{var}	−0.062	0.487	0.475	0.198	0.301	−0.324	−0.552	0.052	0.261

3.4.2. Most important controls

The regression coefficients of the used controls for all three methods is given in Appendix F. To determine the most important controls and make comparison between methods 1, 2 and 3 possible, the top three for methods 1 and 3 is selected as well based on the largest absolute regression coefficients. The number of times a control is present in a top three of a region is shown in Figure 3.10. The figure also shows the sign of the regression coefficient, where a bar to the bottom refers to a negative coefficient and a bar to the top to a positive one. If a control is not given in the figure, this control is not present in any top three.

The figure shows that, based on methods 1 and 3, the variability in precipitation is most important and that the regression coefficient is always positive. Based on method 2, the most important control is either the variability in precipitation or the number of wet days since both are present in the top three of 25 regions. The three methods show quite some similarities, but method 3 seems to be a bit different. This method shows for example that precipitation per wet day is less important than predicted by methods 1 and 2. The

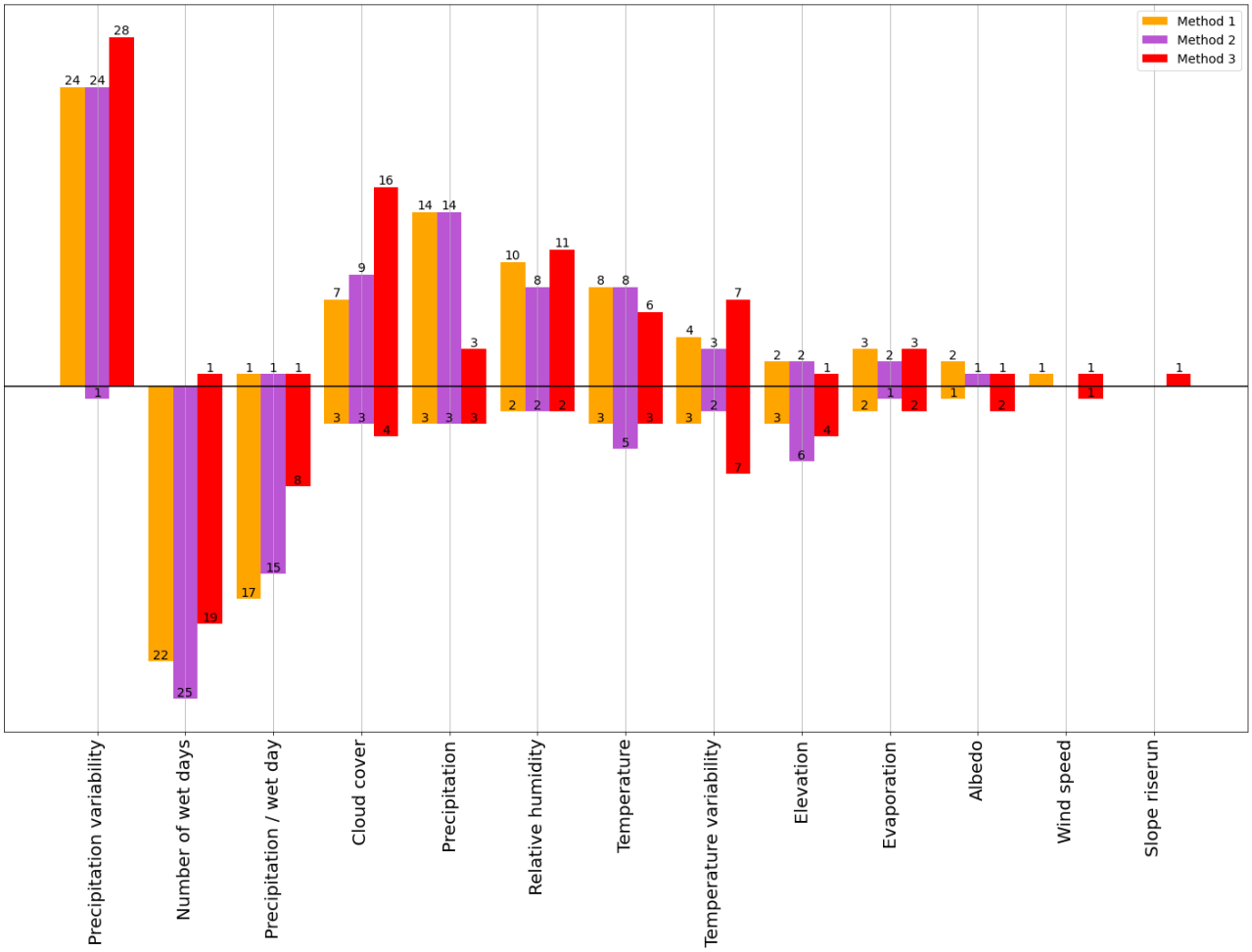

Figure 3.10: Number of times a control is considered most important, i.e. it is in the top three of controls which have the highest absolute regression coefficient. The figure also shows how many times the coefficient is negative or positive. If a bar extends to the top, it is positive and if it extends to the bottom, it is negative. The results are given for elimination methods 1, 2 and 3.

reverse holds for cloud cover. The wind speed is not in the top three of method 2 and the slope riserun is only once in a top three, predicted by method 3. The coefficients for the controls can, most of the times, be both positive and negative.

To get more insight into the spatial pattern of the important controls, they are spatially shown in Figure 3.11 for method 1. The text shows the entire top three for that specific region. The colors are chosen in such a way that similar looking colors are referring to similar controls. For example, precipitation related controls have a blue shade. This figure shows that sometimes several adjacent regions share the exact same control. Much more regions share controls that belong to similar groups. The figure also shows that there is a large variation between the regions. The same result for methods 2 and 3 is given in Appendix G.

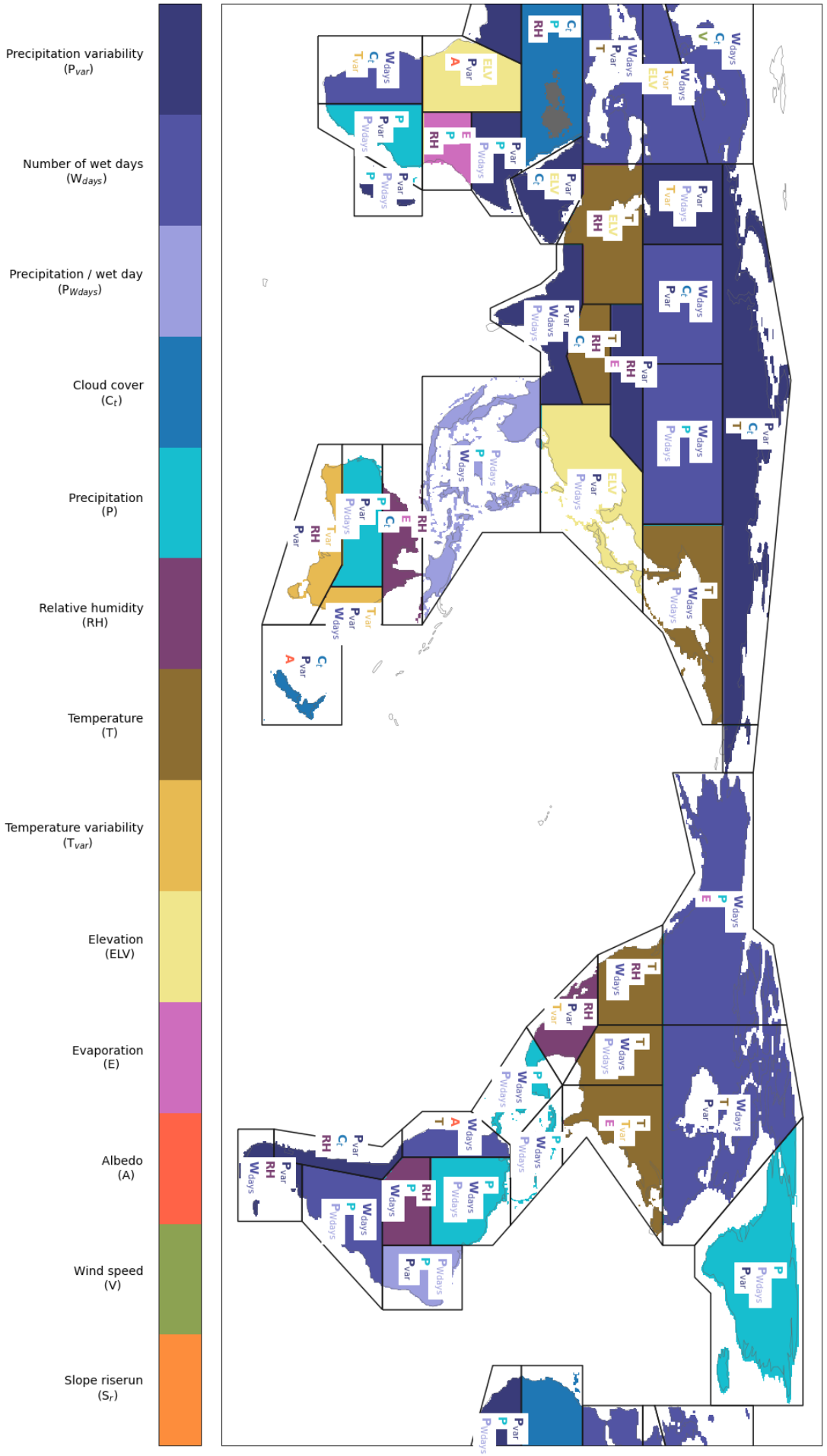

Figure 3.11: Top 3 most important controls based on elimination method 1. The color of the region is determined based on the most important control. The list shows the entire top three.

3.5. Calculation of the heaviness based on the results of the MLR

3.5.1. Calculation of the heaviness for the world

The result of the calculation of the heaviness based on the coefficients of the controls for the world obtained with method 1 is given in Figure 3.12. This figure shows a more smooth behavior compared to Figure 3.1, where the heaviness data is shown. The underlying pattern seems to be coherent between the two heaviness products, but the dotted behavior is not reproduced using the MLR. The error map is given in Figure 3.13, where the error is defined as the absolute difference between the calculated heaviness and the data. This figure shows that regions with a high heaviness value have, generally speaking, the largest error. Moreover, the not reproduced dotted behavior can be seen in the error map. Elevation is eliminated for the world based on method 1 due to a low relative importance. However, it is interesting to see that the error map also shows some orographic features. Both the heaviness calculation as well as the error maps for methods 2 and 3 are given in Appendix H. Method 2 shows an even smoother behavior, since less controls are taken into account. The error map shows larger errors and has the same structure as the error map of method 1. Method 3 is similar to method 1 despite the exclusion of both temperature and precipitation. The standard deviation and average error for all three methods is given in Table 3.2. The standard deviation of the calculations is lower than for the data since the heaviness data has a standard deviation of 0.246.

Table 3.2: Standard deviation of the heaviness calculations for the three elimination methods and the average error with the heaviness data shown in Figure 3.1. The standard deviation of the heaviness data is 0.246

	Method 1	Method 2	Method 3
Standard deviation	0.154	0.148	0.151
Average error	0.107	0.109	0.108

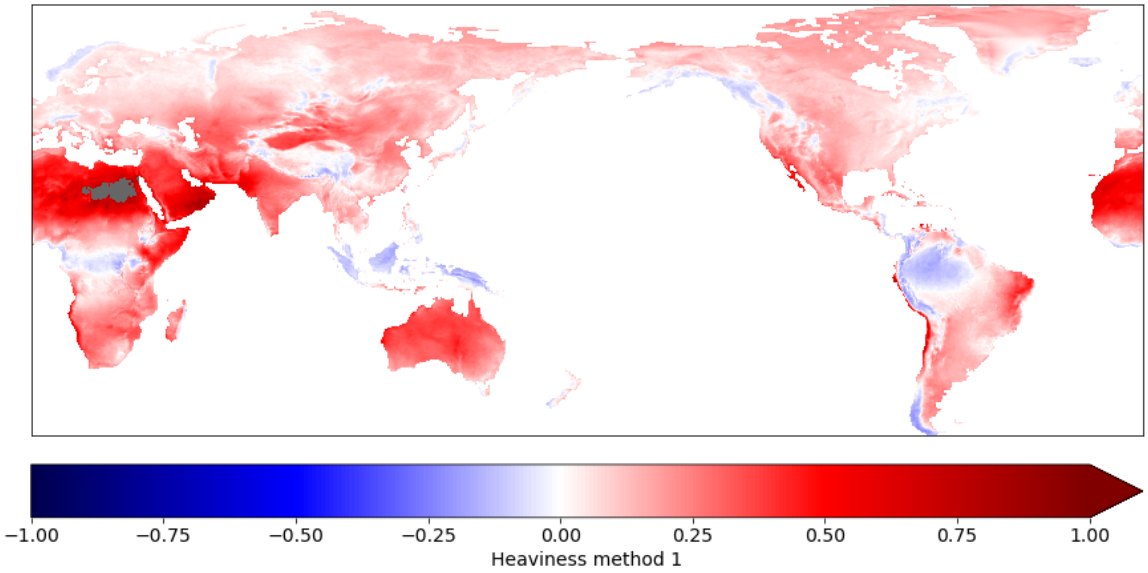

Figure 3.12: Calculation of the heaviness amplification factor based on the coefficients of the MLR and the remaining controls determined with elimination method 1. The used controls can be obtained from Figure 3.7.

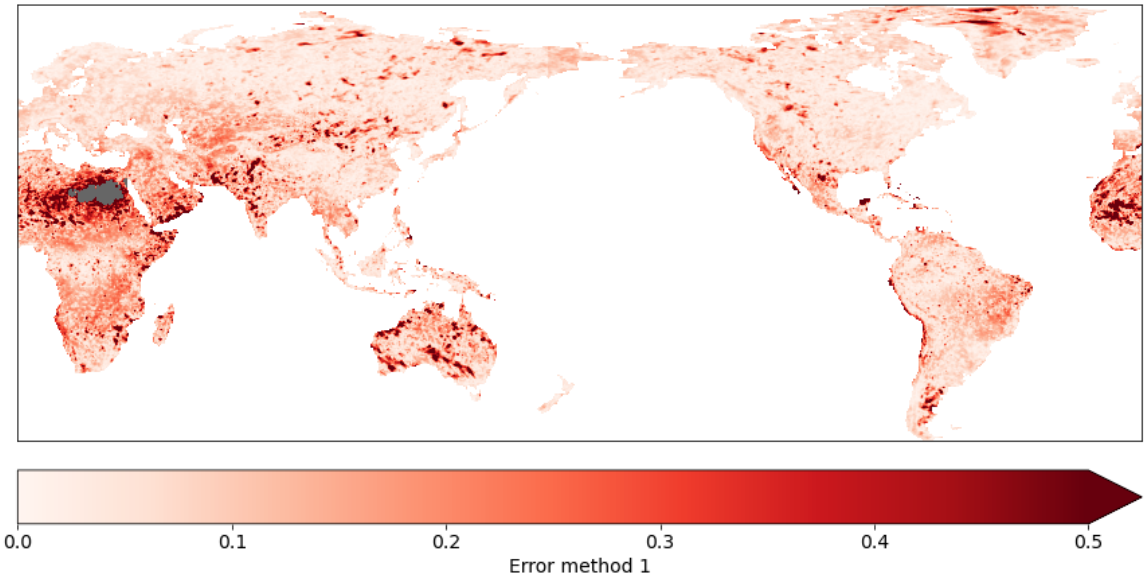

Figure 3.13: Absolute error between the heaviness data and calculation based on coefficients of the remaining controls determined with elimination method 1. The average error is 0.107. The used controls can be foud in Figure 3.7.

3.5.2. Calculation of the heaviness for regions

For all regions, the heaviness is calculated using regional and global coefficients. The average absolute error between these calculations and the heaviness data for all regions and elimination methods is given in Appendix I.1. For region MDG, the results are given in Table 3.3. This table also shows the mean of the average regional errors when using regional and global coefficients.

Table 3.3: Average absolute difference between the use of regional and global coefficients to calculate the heaviness of region MDG, covering Madagascar. The average error using the results of all regions for all three elimination methods is given as well.

	MDG		All regions	
	Regional	Global	Regional	Global
Method 1	0.124	0.150	0.100	0.122
Method 2	0.131	0.170	0.108	0.127
Method 3	0.123	0.153	0.101	0.124

The average difference is larger for the global analysis than for the regional analysis. Moreover, the smallest average difference is obtained for method 3 in region MDG. When all regions are used, method 1 reports the lowest error. Figure 3.14a shows the heaviness calculation based on the regional analysis and 3.14b shows the error between this calculation and the data for method 1. Figures 3.14c and 3.14d show the same for the use of global coefficients. The same results for methods 2 and 3 are given in Appendix I.2.

All calculations of the heaviness, for all three methods and sets of coefficients, are not representing the scattered pattern of the heaviness data, but give a more smooth result. Generally speaking, the calculations of the heaviness are not good at reproducing the original regional data. For the regional analysis, the error shows a more scattered pattern. For the world coefficients, the error is largest at the east coast. The regional coefficients are able to predict a larger range of heaviness values. Similar results are obtained with methods 2 and 3.

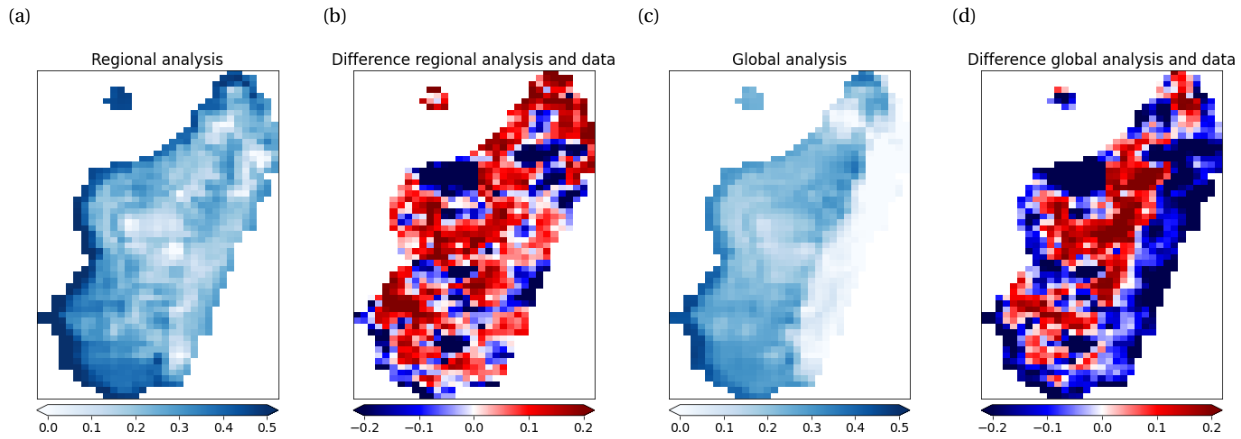

Figure 3.14: a) Calculated heaviness based on regional coefficients. b) Error between the calculated heaviness and the heaviness data for the regional analysis. c) Calculated heaviness based on global coefficients. c) Difference between the calculated heaviness and the data for the global analysis. Results are given for region MDG based on the outcomes of method 1. The used controls can be found in Figure 3.7.

3.5.3. Coefficient of determination and mean squared error

The performance of the MLR is expressed in terms of R-squared and Mean squared error (MSE). These scores are shown in Figure 3.15a and 3.15b for all three elimination methods using global coefficients for the world and regional coefficients for the regions. Keep in mind that the total number of controls used in method 1 and 3 is larger than the number of controls used in method 2.

The R-squared between methods 1 and 3 is quite similar. Sometimes it decreases a bit for method 3. The R-squared is lower for method 2, which is expected since only three controls are taken into account. The same, but reverse, pattern is given by the MSE. The region with the largest and smallest R-squared and MSE are given Table 3.4. Dry regions have less good scores, as both NAU and SAH are dry desert regions with a low R-squared or high MSE. NEU scores not really good on the MSE either, but is not a dry region. SSA and NZ have, based on all three methods, the best R-squared and MSE score.

Table 3.4: Regions with the highest and lowest R-squared (Equation 2.20) and MSE (Equation 2.21) based on the MLR for each of the elimination methods. The location of the regions can be seen in Figure 2.2.

	R-squared		MSE	
	highest	lowest	highest	lowest
Method 1	SSA	NAU	SAH	NZ
Method 2	SSA	NAU	NEU	NZ
Method 3	SSA	NAU	NEU	NZ

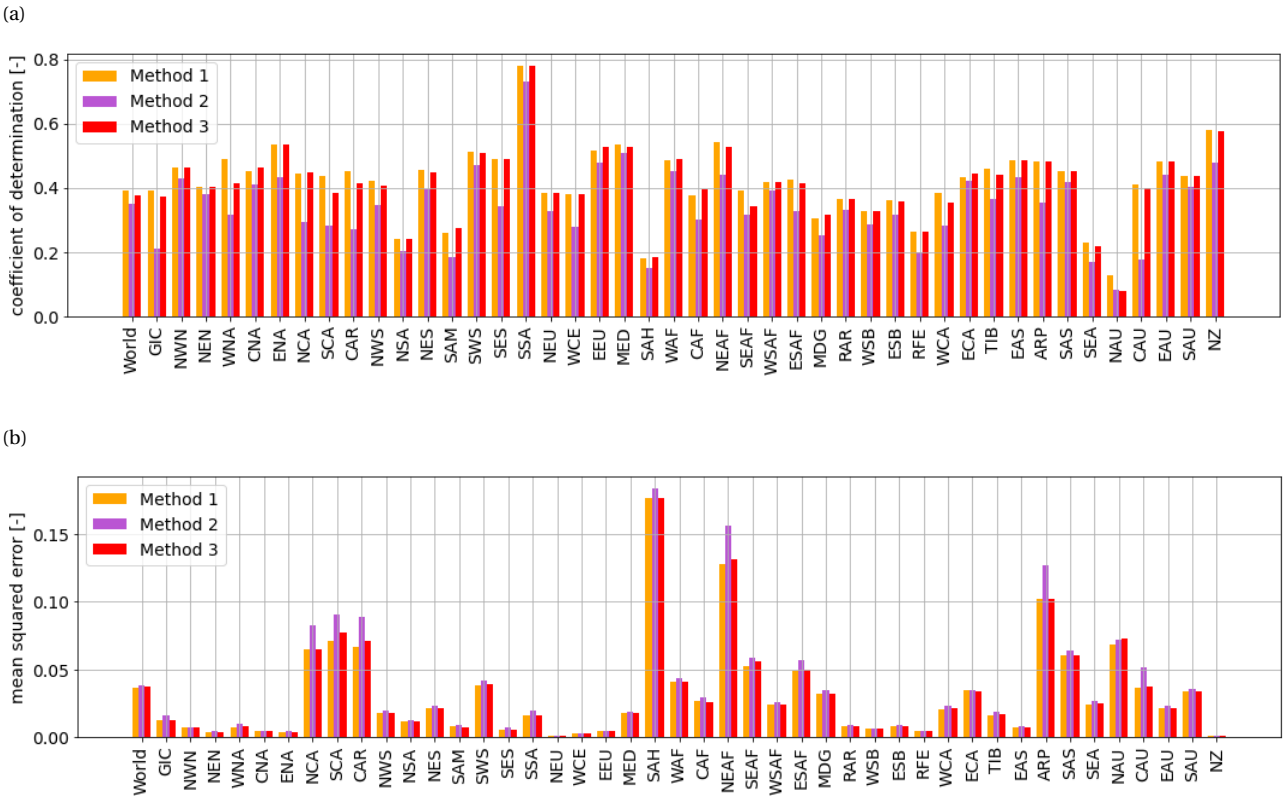

Figure 3.15: a) R-squared of the multiple linear regression, The R-squared is calculated with Equation 2.20. b) Mean squared error of the multiple linear regression, calculated with Equation 2.21. For both parameters, the results are given for all three elimination methods and regions including the entire world. The location of the regions can be seen in Figure 2.2.

3.6. Uncertainty and stationarity

Based on the bootstrap analysis of 200 samples with 30 random years of data, the stationarity of the results is determined. For each sample, the heaviness is determine and the average for the controls based on the 30 years of data is determined. The average heaviness value of the samples, 41 years of data and the first and second 30 years of data is calculated. The results, visualized in a histogram, are shown in Figure 3.16a for the world and in Figure 3.16b for MDG.

When looking at the different means, the mean of the samples is always the highest. Moreover, for the world, the mean of the second 30 years is lower than the mean of the entire 41 years. They overlap for region MDG. This indicates that the mean of the second 30 years is not significantly higher than the mean of the first 30 years, which suggests that there is not a significant trend in the heaviness data. The absence of the trend is supported by similar standard deviations in space of the first and second 30 years. This suggests that the mean but also the spread between these two time periods is similar. The standard deviations for the 41 years and the two time periods of 30 years is given in Table 3.5.

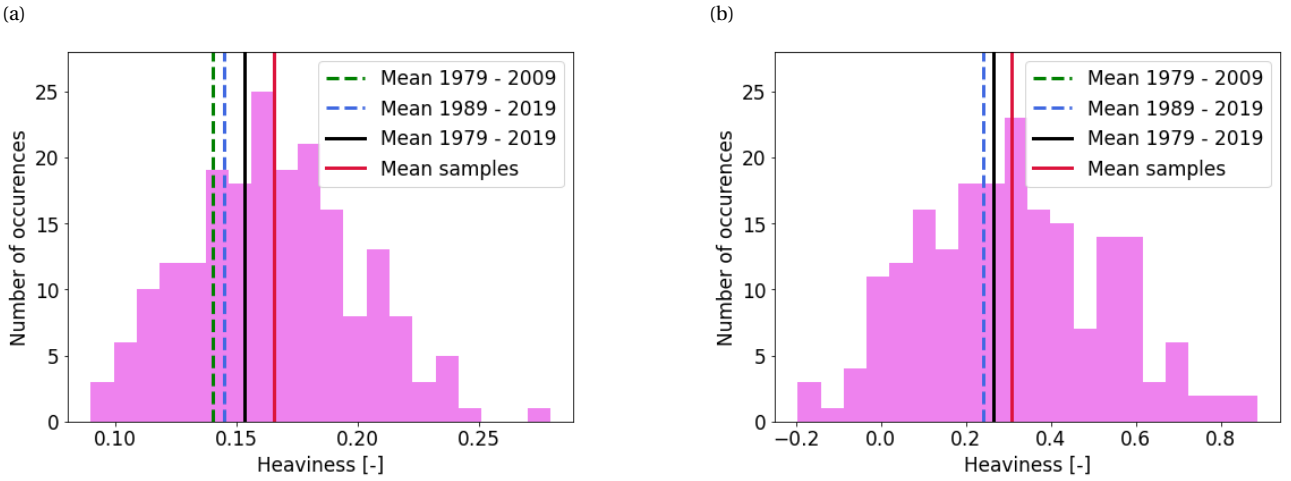

Figure 3.16: a) Histogram of the average heaviness values of 200 samples for the world. b) Histogram of the average heaviness value for MDG. The mean of the sample means and the mean based on 41 years of data are also shown. The 200 samples have 30 random years of data, which is used to calculate the heaviness with Equation 1.1.

Table 3.5: Standard deviation of heaviness in space for three different time periods. The heaviness is obtained with Equation 1.1 and the results are shown for the entire world and region MDG, covering Madagascar.

	World	MDG
1979 - 2019	0.246	0.215
1979 - 2009	0.239	0.269
1989 - 2019	0.244	0.213

When looking at at all the samples for the world, the mean heaviness values are positive, suggesting an average heavy tail for all the 200 samples. This also holds for the 1000 random points which is given in Figure 3.17. For MDG, some of the samples in this case have a heaviness value smaller than 0. This indicates a change from a heavy to a thin tail and would result in a different behavior for extreme precipitation. The spread of the means of the samples is larger for MDG than for the world. This is reflected by the standard deviation of the means of the samples, which is 0.034 for the world and 0.219 for region MDG. For the 1000 random samples this is 0.039, similar to the result of the world. From the histograms, it can also be concluded that MDG indeed shows a different behavior than the world, since the 1000 random points of the world are similar to the results of the entire world.

To determine the uncertainty in the results, the heaviness is calculated 200 times based on the coefficients determined with elimination methods 1, 2 and 3. The mean, minimum and maximum average calculated heaviness value with the multiple linear regression based on the 200 samples is given in Table 3.6. When comparing the values, the range of heaviness values for the world is smaller than for MDG. The random points give similar results as the entire world dataset. Method 3 shows large similarities with method 1. Method 2 reports a smaller range of values that can be predicted.

Figures 3.18a, 3.18b and 3.18c show how often a control is in the top three of the 200 samples, for the world, 1000 random points and MDG based on elimination method 1. For the 1000 random points, an uncertainty range is given. This uncertainty range is based on 50 samples of 1000 random points. The outcome gives the

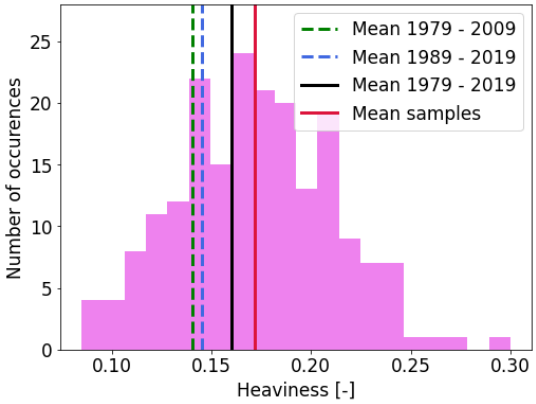

Figure 3.17: Histogram of the average heaviness values of 200 samples for 1000 random points of the world. The mean of the sample means and the mean based on 41 years of data are also shown. The 200 samples have 30 random years of data, which is used to calculate the heaviness with Equation 1.1.

Table 3.6: Mean, minimum and maximum average calculated heaviness values with the multiple linear regression based on 200 samples. The result is shown for all three elimination methods and the world, 1000 random points and region MDG, where the 1000 random points are drawn once.

	World			World - 1000 grid cells			MDG		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
Mean	0.233	0.166	0.238	0.166	0.172	0.167	0.375	0.307	0.363
Min	0.048	0.090	0.046	0.002	0.085	0.009	−0.659	−0.196	−0.693
Max	0.437	0.280	0.439	0.355	0.300	0.357	1.732	0.886	1.815

lower and upper limit of how often each control is used. The same figures are produced based on methods 2 and 3 and given in Appendix J. Methods 1 and 2 show similar results, in the sense that they show a more robust result for the world compared to the region. The 1000 random points show similar results as the world, but there is a bit more variation in the third most important control. For the region, much more controls occur sometimes in the top three. Method 3 is much more consistent for all three groups and shows three dominant controls for both the world and the region.

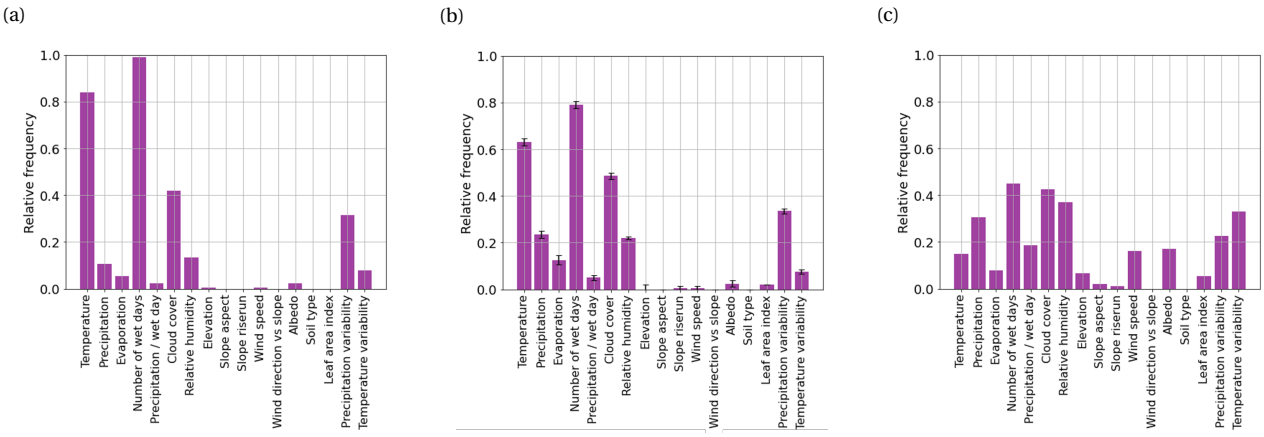

Figure 3.18: a) Relative frequency of how often a control is the top three for the 200 samples for the world. b) Relative frequency of how often a control is the top three for 200 samples and 1000 random points. The lower and upper limit are also shown, based on 50 random samples of 1000 grid points. c) Relative frequency of how often a control is in the top three for the 200 samples for MDG. The results belong to elimination method 1.

The width of the 90% confidence interval is calculated and shown for the world and MDG in Figures 3.19a and 3.19b for elimination method 1. The results for methods 2 and 3 are given in Appendix J. These figures show that the confidence interval for the region is much larger than for the world. When looking at the world, the confidence interval is larger in dry regions and over Indonesia and Greenland. The width of the interval is in the same order as the heaviness value. The confidence interval for method 2 is a bit smaller.

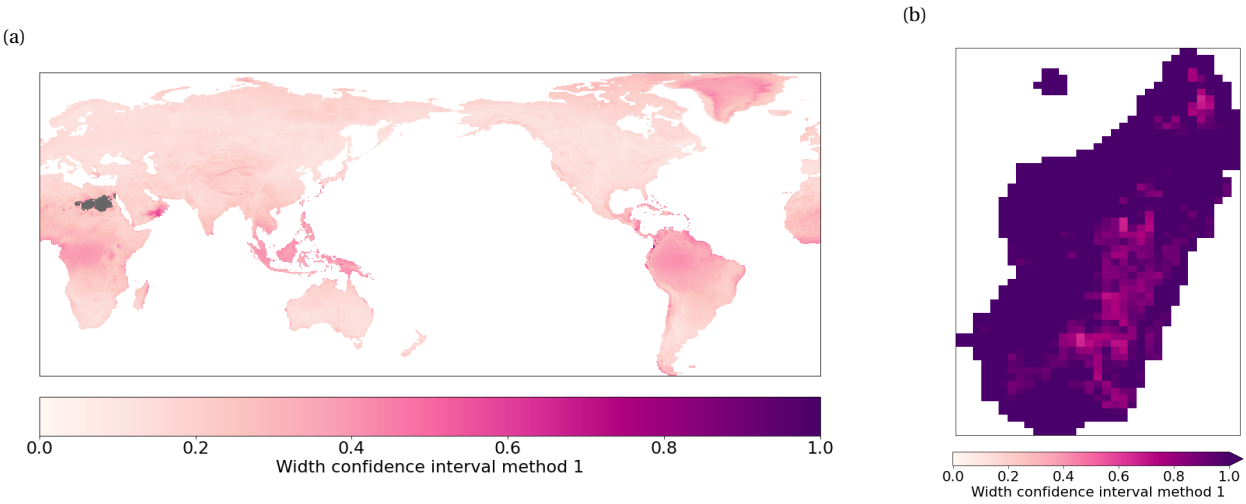

Figure 3.19: a) 90% confidence interval for the world. b) 90% confidence interval for region MDG, where only the data belonging to this region is used. Results are obtained with elimination method 1 and the confidence interval is based on the heaviness value (Equation 1.1) calculated based on 200 samples with 30 random years of data.

4
Discussion

4.1. Interpretation of the results

4.1.1. Heaviness

The average heaviness value based on the ERA5 precipitation data is 0.154, suggesting an average heavy tail. This is coherent with Papalexiou et al. (2013), who stated that heavy-tailed distributions represent the reality more accurately than light-tailed distributions. The results of the heaviness showed in Figure 3.1 can be compared to the results obtained by Gründemann et al. (2020) (their Figure 6c). The figures show similar patterns of the heaviness, but the calculated heaviness in this research is lower than in Gründemann et al. (2020), where negative values of the heaviness are hardly present. The main reason for this difference is the data source used. Gründemann et al. (2020) used MSWEP data, which is based on satellite and reanalysis products, corrected with daily rain gauge observations. ERA5 is a reanalysis product with a coarser spatial resolution. Extremes can be very sensitive to resolution, which can lead to differences in the heaviness amplification factor (Herold et al., 2017). As suggested by both Ragulina & Reitan (2017) and Gründemann et al. (2020), the heaviness shows lower values at higher elevations. As they also suggested, a clear relation with latitude is not visible in the results. The climatic zone seems to influence the results of the heaviness as the IPCC regions show different average values and differences in spread among the heaviness values.

4.1.2. Most important controls

When the top three of the controls are spatially plotted for each region (Figure 3.11 for method 1), it can be concluded that precipitation related controls are dominant over Europe and Canada, where temperature is most important over the United States and the east coast of Asia. There is a more divers pattern on the southern hemisphere. Regions do show different top threes from each other, indicating that regional processes do play a role and the climatic zone is a relevant control for the heaviness.

For the world, the most important controls are: temperature, number of wet days and precipitation variability. Where temperature and precipitation variability are positively correlated with heaviness and the number of wet days has a negative relation with heaviness. Number of wet days (Appendix Figure 6) and precipitation variability (Appendix Figure 16) show similar patterns with heaviness. If the number of wet days is high, the heaviness is low. If the precipitation variability is high, the heaviness is high as well. This can be seen in the dry regions for example. For temperature (Appendix Figure 3), the similarity with heaviness is less clear since temperature shows more variation in the latitudinal direction. However, temperature also shows large similarities with elevation. As Ragulina & Reitan (2017) and Gründemann et al. (2020) suggest, elevation is important for the heaviness. In this research, temperature might substitute elevation due to the high correlation between these two controls, which leads to the exclusion of elevation.

4.1.3. Relation of most important controls with climatic zones

To investigate if the patterns of the most important controls show similarities within regions with similar climatic characteristics, the world is divided into four climatic zones: polar, temperate, subtropical and tropical (based on Bickel et al. (2006)). The regions which belong to which zone is given in Figure 4.1. The polar zone is not taken into account, since only two regions belong to this zone. The percentages given are obtained by adding the results of the top threes of the three elimination methods and dividing them by the total number of regions in this climatic zone multiplied with three. In the temperate zone, the number of wet days is the most important control. It is the top three in 54% of the regions, similar to results of the entire world where 51% of the regions have this control in the top three. For the subtropical zone, the variability in precipitation is the most important control since it is in 78% of the regions in the top three, which is larger than the 55% of the entire world. For the tropical zone, precipitation per wet day, number of wet days and variability in precipitation are all relevant in 54% of the regions. Based on these numbers, the conclusion can be drawn that in different climatic zones, different controls are considered most important. Moreover, there is more coherence between the regions in the tropical zone since three controls occur in more than 50% of the regions. This similarity can be caused by the high amount of yearly precipitation in the tropical zone, where the other zones have less and more varying precipitation amounts.

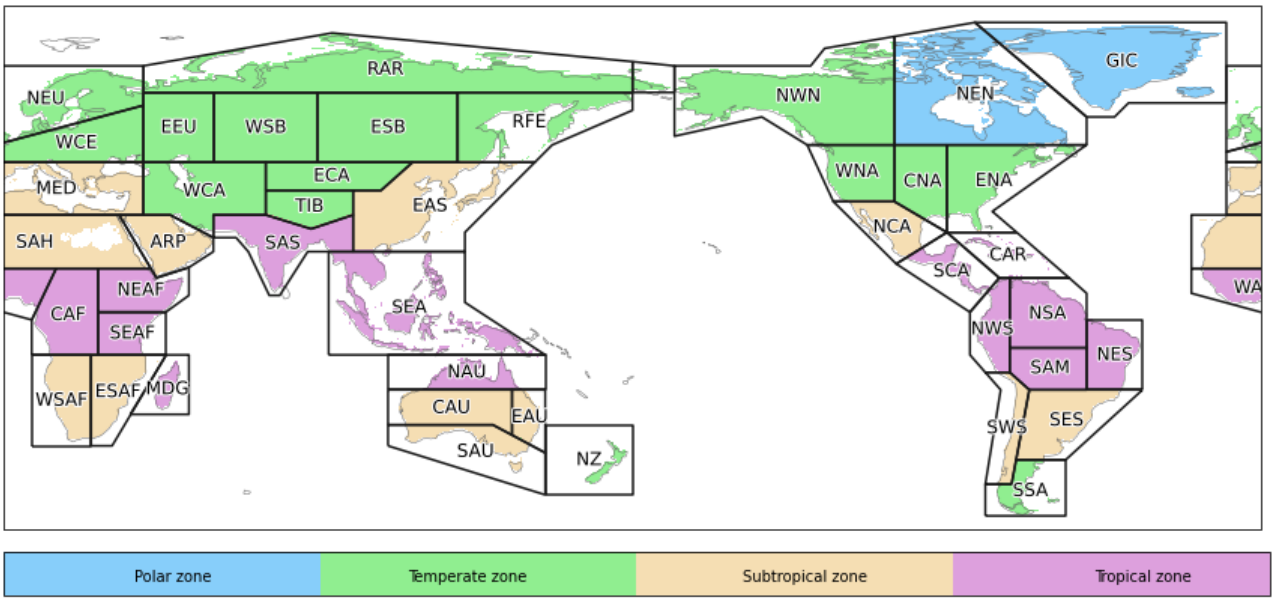

Figure 4.1: Climatic zone of the IPCC climate reference regions. The climatic zones are based on Bickel et al. (2006).

4.1.4. Relation of most important controls to precipitation systems

There are various precipitation generating mechanism on Earth. Some of them are discussed here and the regions where they occur are determined. The most important controls in these regions are obtained to see if there is coherence between these controls and the precipitation mechanisms. The location of the regions can be found in Figure 4.1. The first precipitation generating mechanism is the atmospheric river. These rivers occur in the mid-latitudes and transport water vapor through narrow corridors (Gimeno et al., 2014). The regions under influence of atmospheric rivers are: SEAF, MDG, EAS, CAU, SAU, WNA, NCA, SES, NEU and WCE. The most important control for these regions is the precipitation variability. It occurs in 67% of the top threes. This is higher compared to the result of all the regions, where it occurs in a top three in 55% of the regions. The second mechanisms discussed here is the monsoon. This occurs in the following regions:

CAF, NEAF, SEAF, SAS, EAS, SEA and NAU, located around the Equator in Africa and southeast Asia. Monsoons are caused by the reversal of large scale winds. This leads to extreme precipitation (Krishnamurti et al., 2013). Precipitation variability is import in these regions as it occurs in 76% of the regions. When all the regions are included, this is only 55%. Regions under the influence of the intertropical convergence zone (ITCZ) experience intense rainfall. This zone is centered around 6 degrees north from the Equator and has a yearly cycle (Ortiz-Royero et al., 2013). The ITCZ crosses SAH, ARP, NEAF, CAF, SEAF, SAS, EAS, SEA, NWS, NSA, SAM, NES and WAF. These regions show variation among the top three of most important controls. 12 controls occur in the top three at least once. However, precipitation variability is the most important control for these regions and occurs in the top three in 62% of the regions. The El Nino/Southern Oscillation (ENSO) has also links with large scale precipitation patterns (Ropelewski & Halpert, 1987). The main regions influenced by the ENSO are: ESAF, MDG, SAS, SEA, NAU, EAU and NSA. These regions do have variability in precipitation, number of wet days and precipitation wet day as most important controls. All the precipitation generating mechanisms discussed so far result in periods with heavy precipitation and periods with almost no precipitation. The most important control in the regions under influence of these mechanisms is precipitation variability. The variability can be linked to the cyclic behavior of these mechanisms.

Orographic precipitation leads to precipitation in WCA, ECA, TIB, EAS, RFE, NZ, WNA, NCA, GIC, NWS and SWS (Roe, 2005). The most important control in these regions is the precipitation variability, occurring in 60% of the top threes. It is interesting to see that elevation, a measure for orography, is relatively speaking more important in these regions since it occurs in 25% of regions in the top three, while this is 13% when all the regions are taken into account. Temperature is also more important compared to the world (30% vs 13% globally) and is also related to orography. The last precipitation mechanisms discussed here are the stratiform and convective systems. These systems cause precipitation mainly in the mid-latitudes (Rulfová & Kyselỳ, 2013). The mid-latitudes (temperate zone, Figure 4.1) are dominated by number of wet days (in 54% of the top threes), but also the temperature (38% vs 23% globally) and the cloud cover (42% vs 33% globally) are important.

4.1.5. Heaviness calculation based on the coefficients of the multiple linear regression

When the heaviness is calculated based on the coefficients determined with the MLR, a smooth pattern is the result. This smooth pattern is similar to the heaviness data, but the scattered behavior is not reproduced by this calculation. This can have a couple of reasons. The first reason is that not all the relevant controls are taken into account. This research looks at 17 predefined controls, but it is possible that more controls influence the heaviness. This is also indicated by the R-squared and MSE, where the R-squared is on average 0.3 and the MSE around 0.05 for most of the regions. This means that, generally speaking, only 30% of the variance is explained by the controls used in the MLR. A reason for this could be that local processes play a role in extreme weather events. These local processes are not captured in the averaged, large scale controls used in this research. Moreover, the grid can also be too coarse to capture these local processes.

The argument of regional processes is strengthened by the calculation of the heaviness for the regions based on regional and global coefficients. The average error for the regional analysis decrease with more than 0.2 compared to the use of global coefficients. This indicates that a regional analysis works better for a specific region. The calculation of the heaviness also shows that regions with a high heaviness value have a larger error. They are less well predictable due to the limited availability of precipitation data. Moreover, the predictions of the heaviness have a smaller range than the original data, which is reflected by the decrease in standard deviation from 0.25 to 0.15. This increases the error in the regions with a high heaviness.

4.1.6. Uncertainty analysis

The uncertainty analysis shows that the width of the confidence interval is in the same range as the heaviness itself. Moreover, the width of the confidence interval increases for the regional analysis. One explanation for this is that the top three of controls for the world shows less variation than the top three of region MDG. This can lead to less variation in the final set of controls and therefore more similar results. Another option is that on a global scale, some controls counteract the behavior of other controls, where the result remains the same. Since the regions are smaller, this is less likely to occur. The larger confidence interval for region MDG cannot be caused by the lower number of grid points used since the analysis is also done for 1000 random points of the world. These 1000 random points give results similar to the world itself. Method 2 shows a confidence interval which is a bit smaller than for methods 1 and 3. This is due to the fact that only the top three controls are selected in method 2, which leads to less variation in the results. Method 3 is the most robust in the determination of the top 3, also for a regional analysis. This is probably caused by the set up of this elimination method. In contrast to methods 1 and 2, method 3 eliminates the controls with a highest score. In this case, the highest VIF score. It turns out that controls with a high VIF score are controls which are among the most important ones in methods 1 and 2. Due to the limited choice of controls which are more correlated to heaviness, the same controls are more often in the top three for method 3. The controls determined by method 3 are also different than the controls determined by method 1 and 2.

4.2. Limitations of the methodology and possibilities for further research

4.2.1. MEV-Weibull

The MEV-Weibull equation, given in Equation 2.1, is used to calculate the return periods. As Zorzetto et al. (2016) explains, shorter timeseries are sufficient since all precipitation events above a small threshold are taken into account. However, the number of years that are needed to get reliable results is not given. In this research, 41 years are used. When 20 years were tried, incorrect values of the heaviness were reported where small spots had unreasonable high values (not shown). Therefore, more research into the optimal number of years needed to calculate the heaviness is necessary. This is especially relevant for dry, arid regions where less precipitation events occur.

4.2.2. Precipitation duration

The heaviness is based on 1-day precipitation extremes and the corresponding 2, 20 and 200-year return periods. The heaviness can also be calculated for other precipitation durations and return periods. This could change the heaviness amplification factor and the dependency on the controls. It is suggested to do more research into the dependency of heaviness on the duration of extreme events and the chosen return periods.

4.2.3. Stationarity

The MEV method assumes stationarity of the data. There is no significant trend in the heaviness data. Nevertheless, more research into the non-stationarity in the parameters of the MEV-Weibull method and its implications should be performed. Examples of non-stationarity in extreme events are for example given by Salas & Obeysekera (2014). Moreover, the trend in the controls is not taken into account. However, it is likely there there is a trend in some of the controls. It is interesting to investigate this trend further and explore similarities between the trend in the controls and the MEV parameters. This could give more information about the behavior and possible changes of the heaviness in the future.

4.2.4. Area of the grid cell

A small study into the impact of the area of the grid cell on the results is performed. The conclusion was that the area did not impact the heaviness of most of the grid cells significantly. The fit to the original and grid-adjusted heaviness values has a slope of 0.990 and an intercept of 0.016. This is almost a perfect linear fit. There are some grid cells that have a slightly different heaviness value. The maximum difference in heaviness value between the two grid is 1.0, but it occurs for only a few points. However, more elaborate research into the impact of the different grid cell sizes to the heaviness amplification factor is advised. The ideal situation would be to transfer the data to a grid with equal area sizes and repeat the analysis described in this research. Moreover, the effect of increasing or decreasing the grid cell size is also an interesting topic for further research. As discussed above, the results of the heaviness of Gründemann et al. (2020) give higher values than the heaviness in this research, where a larger grid cell size is used.

4.2.5. Multiple linear regression

About 70% of the variation in heaviness is not explained by the controls taken into account in this research. Therefore, it is worthwhile to take more and/or other controls into account and assess the importance on the heaviness. Moreover, this research only considered a linear relationship between heaviness and the controls using MLR. Other possibilities of assessing the linear dependency like PLSR (Abdi & Williams, 2010; Gebremicael et al., 2019) and ANOVA could be performed to see if this leads to other or more robust results. It is also interesting to investigate the possible non-linear relationships that might exist. This could be done by making (non-)linear combinations of controls and adding non-linear terms of individual controls, e.g. square roots or polynomials, to the dataset. However, this would complicate the process of determining the most important controls.

4.2.6. Multicollinearity

Three elimination methods were compared in this research to deal with the problem of multicollinearity and to reduce the set of controls. To see if the methods tackle the problem of multicollinearity, the difference in sign between the correlation and regression coefficient is determined. If there is a difference, this suggests that the original relation between the heaviness and the control is suppressed in the MLR. This can be caused by the fact that the relation is not strong and the sign difference is a random variation around zero. The other possibility is that more controls carry the same information and are thus redundant in the MLR. This leads to erroneous signs (Falk & Miller, 1992). The difference in sign between the correlation and regression coefficient for method 1 is given in Figure 4.2. The same figures for methods 2 and 3 are given in Appendix K. These figures show that there are quite some controls that show different signs for some of the regions where they are used. Based on the outcomes of the three elimination methods, 35% of the controls have a different sign. The difference in signs in the controls related to land cover and geography are probably caused by the first reason mentioned by Falk & Miller (1992), since the regression coefficients are around zero for these controls. For the other controls, however, it is likely that there are still redundant sets of information in the dataset. None of the procedures was capable of removing all these dependencies. This makes the results less robust and more unstable. Therefore, it is suggested that more research be conducted into the best solution to solve the problem of multicollinearity in this specific case.

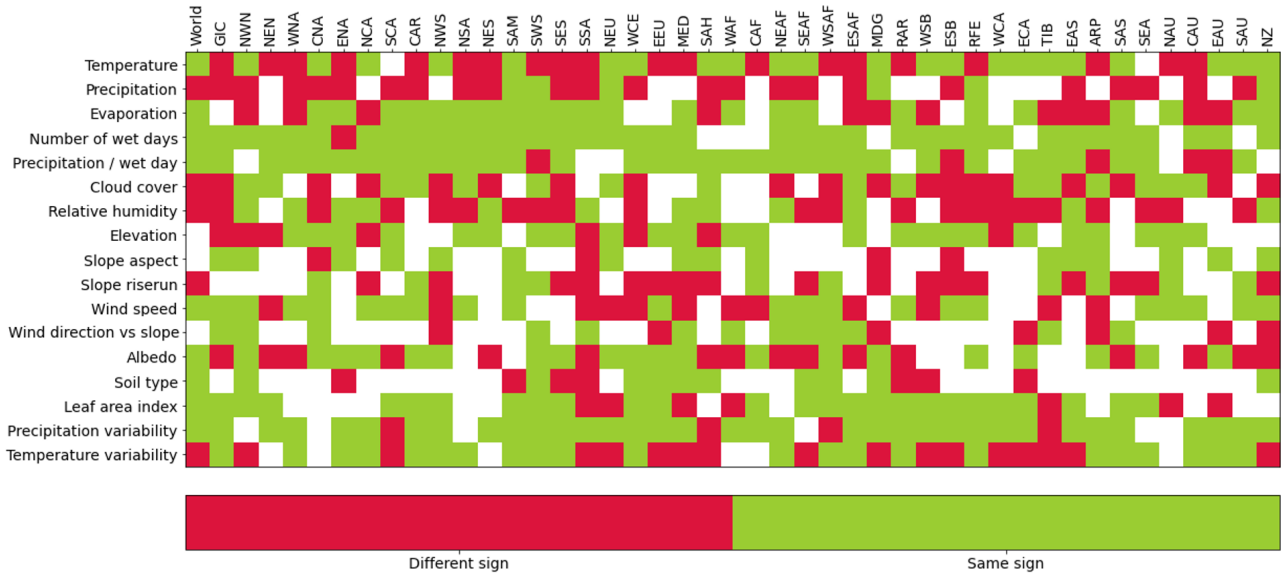

Figure 4.2: Difference in sign between the correlation and regression coefficients. If a grid cell is red, the signs are different. If it is green, the signs are the same. The difference in signs can be an indication for multicollinearity. The location of the regions can be found in Figure 2.2. The results for methods 2 and 3 can be found in Appendix K.

4.2.7. Regional analysis

The regional analysis is based on the regions classified by the IPCC. These regions are still quite large. It is interesting to see how the results would change if the regions were different. The Köppen Geiger classification could be used for example, were regions across the world can have the same climatic zone. Other possibilities are for example using catchments of large rivers, countries as regions or dividing the world in an equal amount of equally sized shapes. K-means is a clustering technique that can also be used to define clusters with similar characteristics (Likas et al., 2003).

4.2.8. Uncertainty analysis

The uncertainty analysis shows that the width of the confidence interval is of the same order of magnitude as the heaviness values. The confidence interval increases when only considering a single region. Due to limitations in computational power, only 200 samples were constructed with 30 random years of data. It is worthwhile to create more samples, at least 1000, to get more reliable results (Hesterberg, 2011). Moreover, the number of 30 years was chosen since it leads to proper results of h and resembles one climate period. However, including more data would increase the robustness of the results.

5
Conclusion

This thesis identified the dominant hydro-climatological controls of the tail behavior of extreme precipitation distributions. The tail behavior of extreme precipitation was characterized by the heaviness amplification factor. The heaviness is determined based on the 2, 20 and 200 year return periods for 1-day precipitation events, calculated with 41 years of data ranging from 1979-2019. The return periods were calculated with the MEV method. Based on expert judgement I pre-defined a set of 17 hydro-climatological controls. Multiple linear regression was applied to determine the dependency of the heaviness on each of the standardized controls. To eliminate the problem of multicollinearity, the multiple linear regression was preceded by an elimination procedure. Three different elimination procedures were applied. The analysis was done for the entire world, but also for 44 IPCC climate reference regions. Only land cells were taken into account.

Based on the results presented in this report, 13 of the 17 included controls are in the top three of at least one region based on one or more elimination methods. These controls, including there average percentage of top threes they are in, are: precipitation variability (57%), number of wet days (50%), precipitation per wet day (32%), cloud cover (31%), precipitation (30%), relative humidity (26%), temperature (24%), temperature variability (19%), elevation (13%), evaporation (10%), albedo (5%), wind speed (2%) and slope riserun (1%). The R-squared ranges from 0.2 to 0.8 where most of the regions have a R-squared around 0.3, indicating that 30% of the variance is explained by the controls. This indicates that not all the variability in the heaviness is explained by the controls used in this research. This is reflected by the calculation of the heaviness, which is much more smooth than the original data and has an average error of 0.11. Regions have different top threes from each other and the world. When using regional coefficients to calculate the heaviness instead of global coefficients, the average error between the data and the heaviness calculation reduces from 0.12 to 0.10. This suggests that regional processes play a role in the behavior of the heaviness and that makes it difficult to define a generic set of controls. Furthermore, the three elimination methods give similar controls that are most important. They do not solve the problem of multicollinearity. In, on average, 35% of the controls, there are sign differences.

To conclude, a first idea about the dominant controls for the heaviness is presented in this research. These controls belong mainly to the general climate condition and climate variability categories. The land cover and geography controls are of less importance for the heaviness, which is in contrast to suggestions of previous research. There are many options for further research, including using more controls, applying non-linear regression techniques, adjusting the area, considering regional processes and using different precipitation durations. These are advised paths that could increase the knowledge about the processes that drive the behavior of the heaviness amplification factor.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433–459.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 370–374.

Arguez, A., & Vose, R. S. (2011). The definition of the standard WMO climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society*, *92*(6), 699–704.

Back, L. E., & Bretherton, C. S. (2005). The relationship between wind speed and precipitation in the Pacific ITCZ. *Journal of climate*, *18*(20), 4317–4328.

Bagley, J. E., Desai, A. R., Harding, K. J., Snyder, P. K., & Foley, J. A. (2014). Drought and deforestation: has land cover change influenced recent precipitation extremes in the Amazon? *Journal of Climate*, *27*(1), 345–361.

Beullens, J., Van de Velde, D., & Nyssen, J. (2014). Impact of slope aspect on hydrological rainfall and on the magnitude of rill erosion in Belgium and northern France. *Catena*, *114*, 129–139.

Bickel, K., Gary, R., Köh, M., & Rodrigue, V. (2006). Consistent representation of lands. In *2006 IPCC Guidelines for National Greenhouse Gas Inventories* (Vol. 4). Agriculture, Forestry and Other Land Use.

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., ... others (2013). Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *K., Tignor, M., Allen, SK, Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, PM, Cambridge University Press, Cambridge, UK*.

Chu, H.-J. (2012). Assessing the relationships between elevation and extreme precipitation with various durations in southern Taiwan using spatial regression models. *Hydrological Processes*, *26*(21), 3174–3181.

Copernicus Climate Change Service. (2019). ERA5 Monthly Averaged Data on Single Levels from 1979 to Present.

Danielson, J. J., & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. US Department of the Interior, US Geological Survey.

Dastrup, R. A., et al. (2020). 8.3 Controls of weather and climate. *Physical Geography and Natural Disasters*.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how* (Vol. 488). Springer.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569–593). Springer.

Falk, R. F., & Miller, N. B. (1992). *A primer for soft modeling*. University of Akron Press.

Gebremicael, T., Mohamed, Y., & Van der Zaag, P. (2019). Attributing the hydrological impact of different land use types and their long-term dynamics through combining parsimonious hydrological modelling, alteration analysis and PLSR analysis. *Science of the Total Environment*, *660*, 1155–1167.

Gimeno, L., Nieto, R., Vázquez, M., & Lavers, D. A. (2014). Atmospheric rivers: A mini-review. *Frontiers in Earth Science*, *2*, 2.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: definition and relation to parameters of several distributions expressable in inverse form. *Water resources research*, *15*(5), 1049–1054.

Gründemann, G. J. (2021). Personal communication.

Gründemann, G. J., Zorzetto, E., Beck, H. E., Schleiss, M., Van de Giesen, N., Marani, M., & van der Ent, R. J. (2020). Extreme Precipitation Return Levels for Multiple Durations on a Global Scale. *Earth and Space Science Open Archive ESSOAr*.

Herold, N., Behrangi, A., & Alexander, L. V. (2017). Large uncertainties in observed daily precipitation extremes over land. *Journal of Geophysical Research: Atmospheres*, *122*(2), 668–681.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., ... others (2019). *ERA5 monthly averaged data on pressure levels from 1979 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(6), 497–526.

Horn, B. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, *69*(1), 14-47. doi: 10.1109/PROC.1981.11918

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., ... others (2020). An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. *Earth System Science Data*, *12*(4), 2959–2970.

Jonkman, S. N., Slager, K., van den Hurk, B., Rongen, G., Asselman, N., Strijker, B., ... others (2021). Hoogwater 2021: Feiten en Duiding.

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., ... others (2021). Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021.

Krishnamurti, T., Stefanova, L., & Misra, V. (2013). Monsoons. In *Tropical meteorology* (pp. 75–119). Springer.

Kumari, S. (2008). Multicollinearity: Estimation and elimination. *Journal of Contemporary research in Management*, *3*(1), 87–95.

Li, Z., Shao, Q., Xu, Z., & Cai, X. (2010). Analysis of parameter uncertainty in semi-distributed hydrological models using bootstrap method: A case study of SWAT model applied to Yingluoxia watershed in northwest China. *Journal of Hydrology*, *385*(1-4), 76–83.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, *36*(2), 451–461.

Liu, B., Tan, X., Gan, T. Y., Chen, X., Lin, K., Lu, M., & Liu, Z. (2020). Global atmospheric moisture transport associated with precipitation extremes: Mechanisms and climate change impacts. *Wiley Interdisciplinary Reviews: Water*, *7*(2), e1412.

Noori, R., Khakpour, A., Omidvar, B., & Farokhnia, A. (2010). Comparison of ANN and principal component analysis-multivariate linear regression models for predicting the river flow based on developed discrepancy ratio statistic. *Expert Systems with Applications*, *37*(8), 5856–5862.

Ortiz-Royero, J., Otero, L., Restrepo, J., Ruiz, J., & Cadena, M. (2013). Cold fronts in the Colombian Caribbean Sea and their relationship to extreme wave events. *Natural Hazards and Earth System Sciences*, *13*(11), 2797–2804.

Overeem, A., Buishand, T., Holleman, I., & Uijlenhoet, R. (2010). Extreme value modeling of areal rainfall from weather radar. *Water Resources Research*, *46*(9).

Papalexiou, S., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, *17*(2), 851–862.

Pham, H. (2019). A new criterion for model selection. *Mathematics*, *7*(12), 1215.

Ragulina, G., & Reitan, T. (2017). Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach. *Hydrological sciences journal*, *62*(6), 863–879.

Rahimpour Golroudbary, V., Zeng, Y., Mannaerts, C. M., & Su, Z. B. (2016). Land cover effects on extreme precipitation in the Netherlands. In *Living planet symposium* (Vol. 740, p. 175).

Roe, G. H. (2005). Orographic precipitation. *Annu. Rev. Earth Planet. Sci.*, *33*, 645–671.

Ropelewski, C. F., & Halpert, M. S. (1987). Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly weather review*, *115*(8), 1606–1626.

Rulfová, Z., & Kyselỳ, J. (2013). Disaggregating convective and stratiform precipitation from station weather data. *Atmospheric research*, *134*, 100–115.

Salas, J. D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, *19*(3), 554–568.

Schulzweida, U., Kornblueh, L., & Quast, R. (2006). CDO user's guide. *Climate data operators, Version*, *1*(6), 205–209.

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics*, *8*(2), 39–42.

Um, M.-J., Yun, H., Jeong, C.-S., & Heo, J.-H. (2011). Factor analysis and multiple regression between topography and precipitation on Jeju Island, Korea. *Journal of Hydrology*, *410*(3-4), 189–203.

Utsumi, N., Seto, S., Kanae, S., Maeda, E. E., & Oki, T. (2011). Does higher surface temperature intensify extreme precipitation? *Geophysical research letters*, *38*(16).

Zorzetto, E. (2018). *Metatstatistical Extreme Value Analysis in Python.* `https://github.com/EnricoZorzetto/mevpy`. GitHub.

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters, 43*(15), 8076–8082.

Appendices

A. Area

In this research, a regular grid from the ERA5 dataset is used. This implies that the areas of the grid cells decrease with increasing latitude. This can have an effect on the estimated return periods and therefore on the calculation of the heaviness amplification factor. Overeem et al. (2010) suggests that, based on radar data, the areal reduction factor (ARF) decreases with increasing area sizes for maximum annual precipitation. To investigate the impact of the area on the heaviness amplification factor a simple procedure is applied. The three variables of the MEV: shape, scale and number of wet days parameters, are averaged for a certain number of grid cells in the latitudinal direction. This is done according to the scheme in Table 1. The grid cells are combined to compensate for the decrease in area sizes in latitudinal direction. This small research gives a first idea about the dependency of heaviness on the grid cell size.

Table 1: Number of grid cells combined for a given absolute latitude range. For the combined grid cells, the average parameters of the MEV Weibull (Equation 2.1) are calculated. These are used to obtain the new heaviness value (Equation 1.1) for the combined grid cells.

Latitude range	Combined grid cells
0-30	0
30-45	2
45-60	4
60-75	6
75-90	8

For this new set of data, the return periods are calculated. These return periods are used to calculate the heaviness amplification factor. To be able to compare the original and new heaviness data, the new heaviness data is converted back to the original grid, where multiple grid cells have the same value. The result is given in Figure 1.

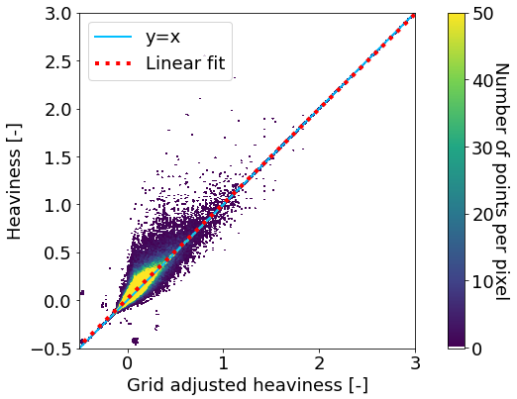

Figure 1: Grid adjusted heaviness versus the original heaviness, where also the linear fit to the data and the line $y = x$ is given. The more yellow the color, the more points are located in that area.

This figure also shows the linear fit of the data and the line $y = x$, which would indicate no change in the heaviness due to area regridding. The fit is very close to the line $y = x$ with a slope of 0.990 and an intercept of 0.016. For some locations, there is a different between the two heaviness values. This holds mainly for grid cells with a heaviness value between 0 and 1. Moat of the values that are a bit off from the fit, get a lower heaviness value after the grid is adjusted. Table 2 shows the slope and intercept for the fit to each of the sets of latitudinal ranges.

Table 2: Slope and intercept for the fitted line between the original heaviness values and the grid-adjusted heaviness values. The result is show for each of the groups defined in Table 1 including the entire dataset.

Latitude range	Slope	Intercept
0-90	0.990	0.016
0-30	1.000	0.000
30-45	1.013	0.012
45-60	1.032	0.015
60-75	1.017	0.026
75-90	1.105	0.049

B. Variance Inflation Factor of the controls

The investigate the multicollinearity between the controls, the VIF score for each control and region, including the entire world is calculated. The VIF score is a measure for multicollinearity, calculated with Equation 2.19. The VIF calculates the dependency of a control to another control or a linear combination of controls (Alin, 2010). A control is highly correlated if the VIF score is higher than 10. This is visually shown in Figure 2. If a grid cell has a dark color, it means that the VIF score is higher than 10 for that region. Mainly the controls belonging the general climate condition and the climate variability have high VIF scores. The conclusion based on this figure is that there is indeed multicollinearity present in the dataset.

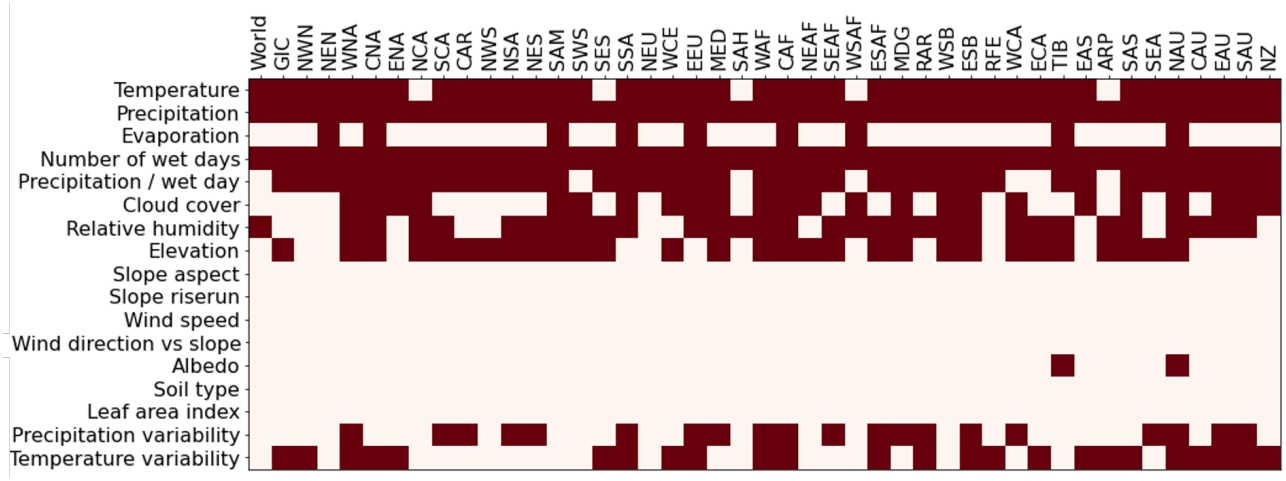

Figure 2: Dark grid cells represent the VIF scores that are higher than 10. The result is given for the entire world and all the IPCC climate reference regions. The VIF is calculated with Equation 2.19. A high VIF score indicates multicollinearity in the dataset, where a variable can be correlated to another variable or a linear combination of multiple variables (Alin, 2010).

C. Result controls

C.1. General climate condition

The results of the controls belonging to the general climate condition are shown here. These controls are: temperature (Figure 3), precipitation (Figure 4), evaporation (Figure 5), number of wet days (Figure 6), precipitation per wet day (not shown here but given in Figure 3.4), cloud cover (Figure 7) and relative humidity (Figure 8).

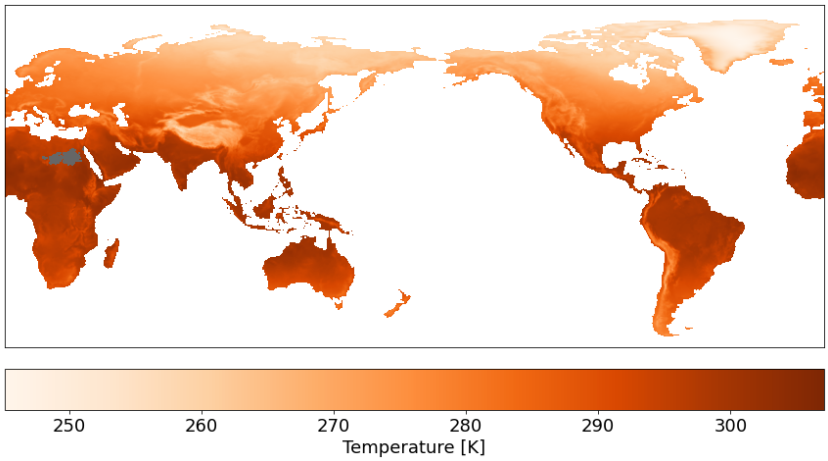

Figure 3: Average temperature in Kelvin calculated with Equation 2.2 where the temperature is obtained from ERA5 which reports monthly averaged data. The data is obtained from 1979 until 2019 (Hersbach et al., 2019).

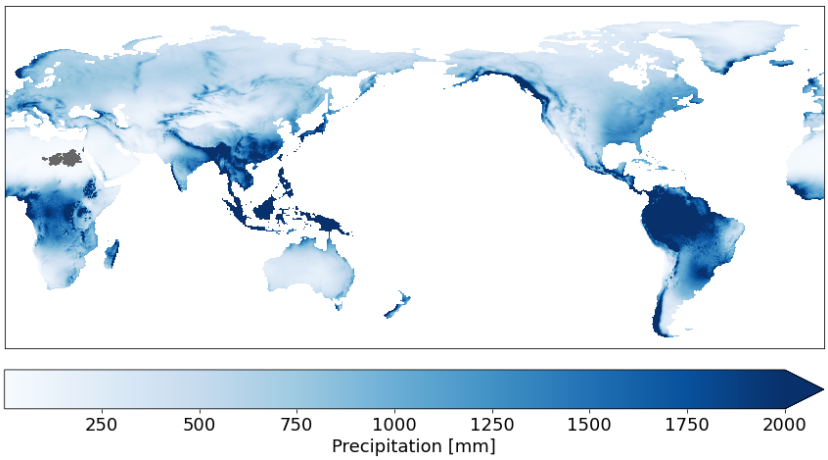

Figure 4: Average yearly precipitation in mm calculated with Equation 2.3 where the precipitation is obtained from ERA5 which reports monthly averaged daily precipitation data. The data is obtained from 1979 until 2019 (Hersbach et al., 2019).

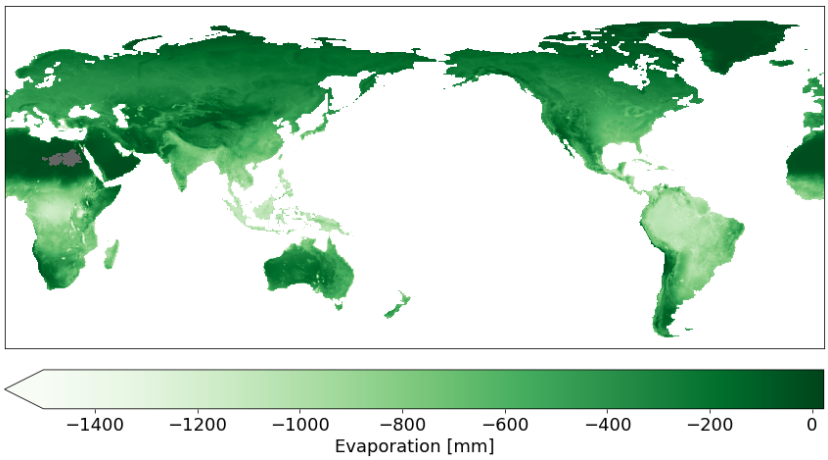

Figure 5: Average yearly evaporation in mm calculated with Equation 2.4 where the evaporation is obtained from ERA5 which reports monthly averaged daily evaporation data. The data is obtained from 1979 until 2019 (Hersbach et al., 2019). Negative values correspond to a net upward evaporation.

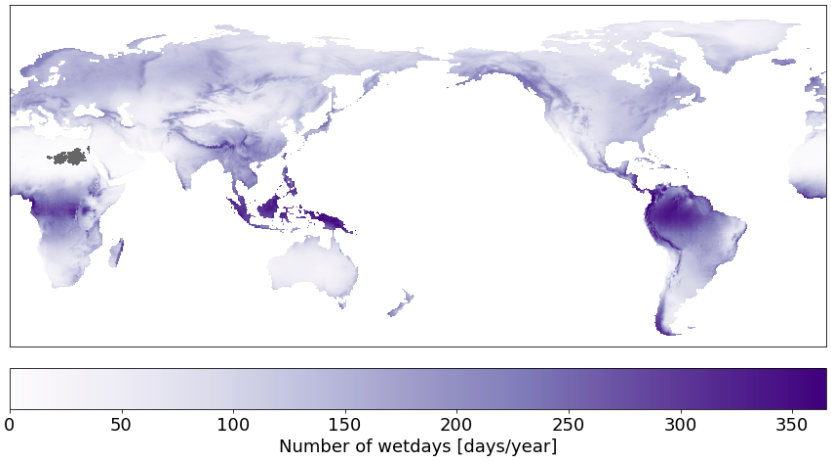

Figure 6: Average number of wet days per year calculated with Equation 2.5 where the number of wet days are obtained from ERA5 which reports monthly averaged data. The data is obtained from 1979 until 2019 (Hersbach et al., 2019).

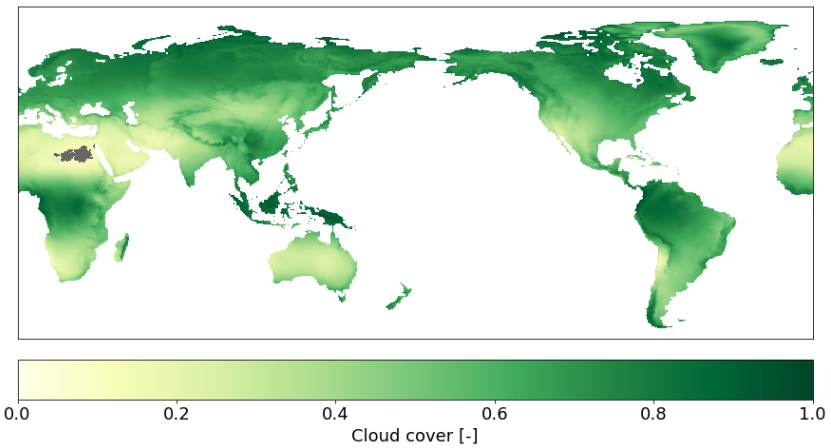

Figure 7: Average cloud cover calculated with Equation 2.7 where the cloud cover is obtained from ERA5 which reports monthly averaged data. The data is obtained from 1979 until 2019 (Hersbach et al., 2019).

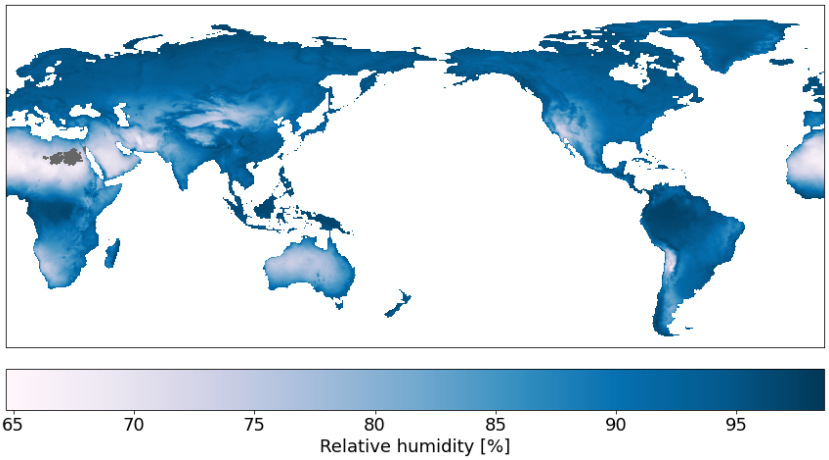

Figure 8: Average relative humidity calculated with Equation 2.8 where relative humidity is obtained from the temperature and dew point temperature in Kelvin given by ERA5 which reports monthly averaged data. The data is obtained from 1979 until 2019 (Hersbach et al., 2019).

C.2. Geography

The results of the controls belonging to the geography are shown here. These controls are: elevation (Figure 9), slope aspect (Figure 10), slope riserun (Figure 11), wind speed (Figure 12) and the combined control of wind direction and slope aspect (Figure 13).

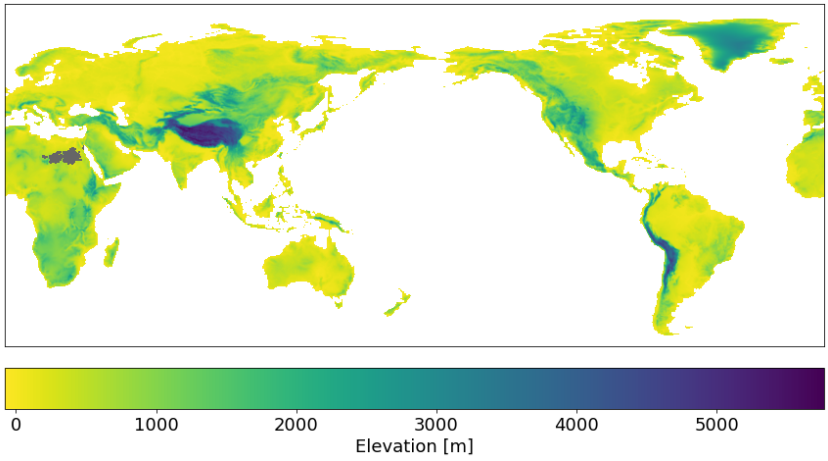

Figure 9: Elevation given by GMTED2010 and linear interpolated to the ERA5 grid (Danielson & Gesch, 2011).

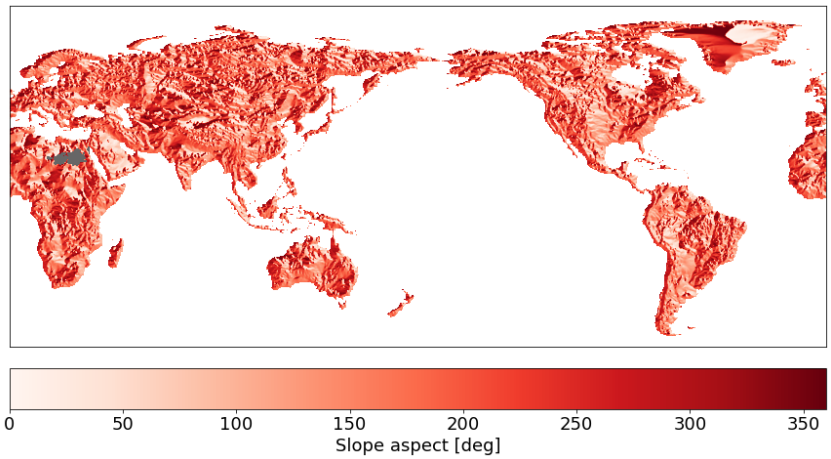

Figure 10: Slope aspect given in degrees from the North calculated from the elevation dataset with the `richDEM` package. The slope represents the maximum slope of the cell (Horn, 1981).

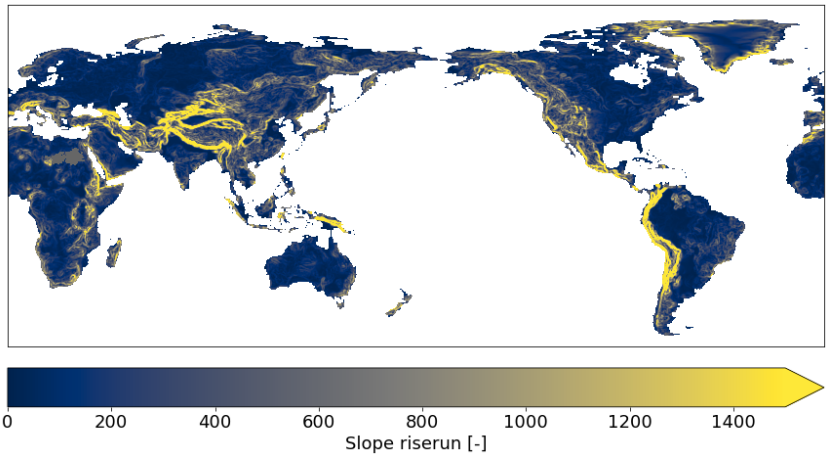

Figure 11: Slope riserun representing the rise over run, where rise is the vertical distance and run the horizontal distance. The slope riserun is calculated from the elevation dataset with the `richDEM` package (Horn, 1981).

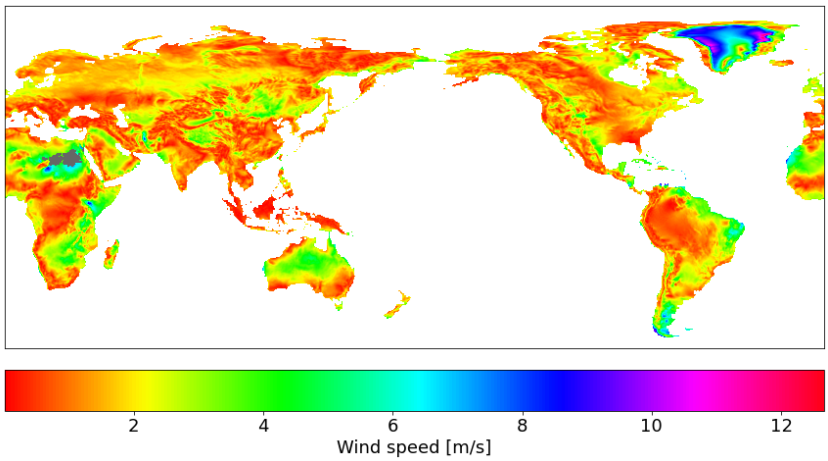

Figure 12: Average wind speed in m/s calculated with Equation 2.9 using data from ERA5 covering the time period of 1979 until 2019 (Hersbach et al., 2019).

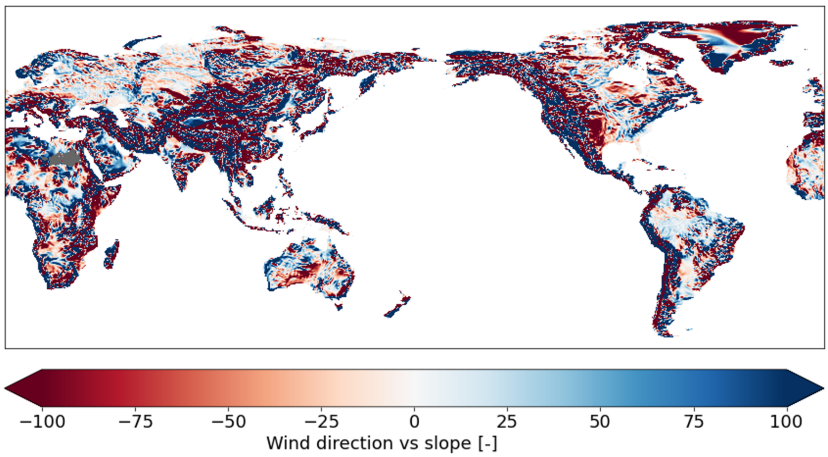

Figure 13: Wind direction vs slope ρ calculated with Equation 2.10. ERA5 monhtly reanalysis data from 1979 until 2019 is used for the wind direction (Hersbach et al., 2019). The slope aspect and slope riserun are calculated with the `richDEM` package using elevation data (Danielson & Gesch, 2011).

C.3. Land cover

The results of the controls belonging to the land cover category are shown here. These controls are: albedo (Figure 14), soil type (Figure 15) and leaf area index (not shown here but given in Figure 3.5).

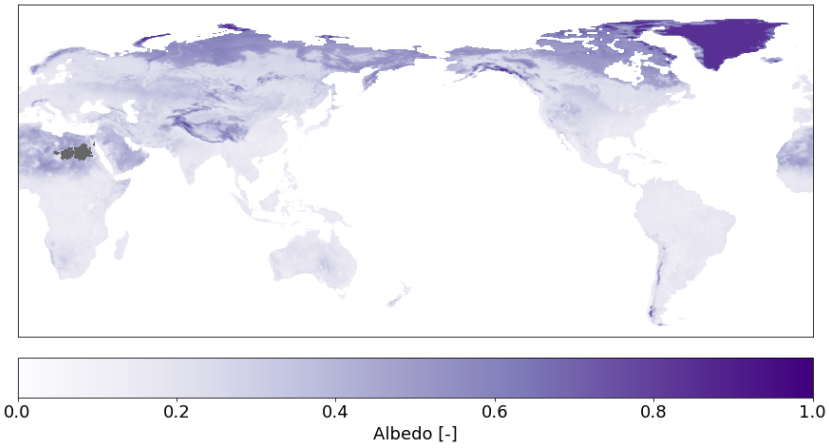

Figure 14: Average albedo calculated with Equation 2.13 where ERA5 monthly reanalysis data from 1979 until 2019 is used (Hersbach et al., 2019).

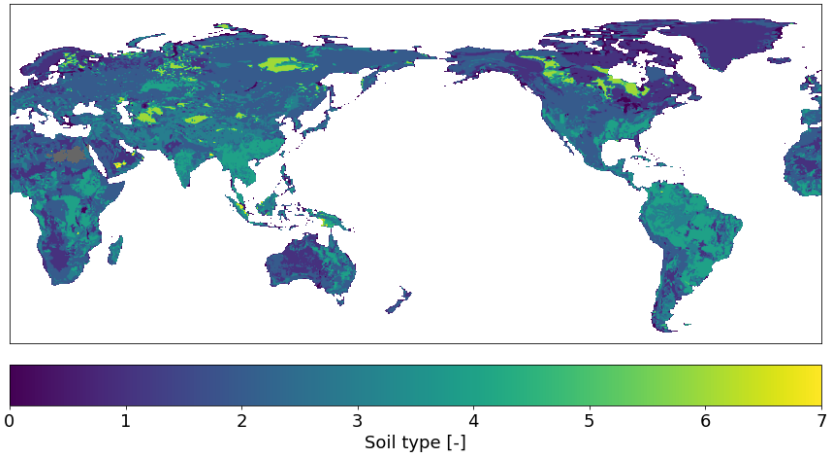

Figure 15: Soil type given by ERA5 monthly reanalysis data, where the soil type is fixed and does not change over time (Hersbach et al., 2019).

C.4. Variability

The results of the controls belonging to the variability category are shown here. These controls are: precipitation variability (Figure 16) and temperature variability (Figure 17).

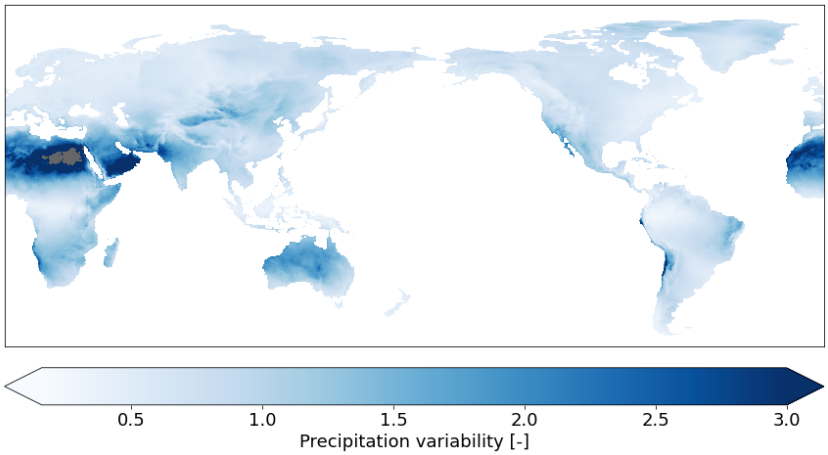

Figure 16: Variability in precipitation calculated with Equation 2.15, where precipitation data from ERA5 monthly reanalysis is used covering the time period from 1979 until 2019 (Hersbach et al., 2019).

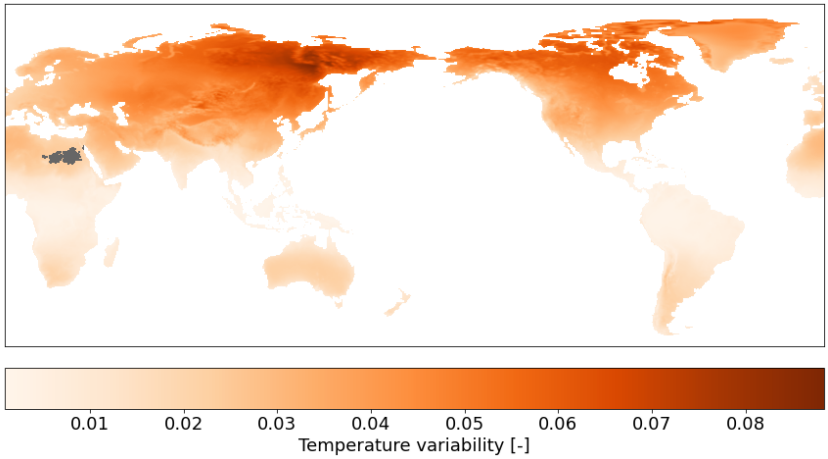

Figure 17: Variability in temperature calculated with Equation 2.16, where temperature data from ERA5 monthly reanalysis is used covering the time period from 1979 until 2019 (Hersbach et al., 2019).

D. Scatter density plots

D.1. General climate condition

The results of the scatter density plots with the controls belonging to the general climate condition are shown here. The controls are: temperature (Figure 18a), precipitation (Figure 18b), evaporation (Figure 18c), number of wet days (Figure 18d), precipitation per wet day (not shown here but given in Figure 3.6a), cloud cover (Figure 18e) and relative humidity (Figure 18f).

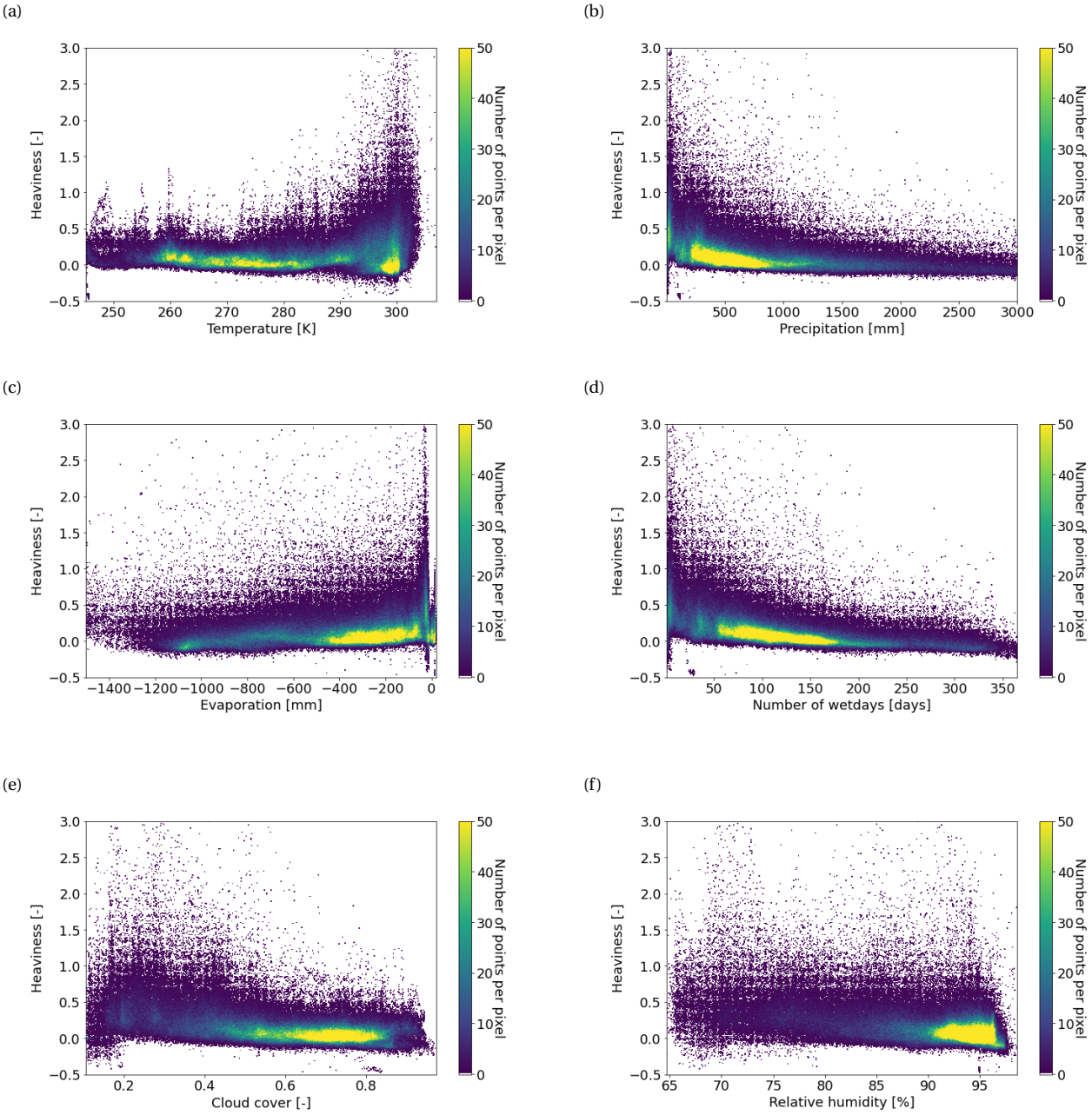

Figure 18: Scatter density plots of controls versus the heaviness, where the following controls are plotted: a) Temperature obtained with Equation 2.2. b) Precipitation obtained with Equation 2.3. c) Evaporation obtained with Equation 2.4. d) Number of wet days obtained with Equation 2.5. e) Cloud cover obtained with Equation 2.7. f) Relative humidity obtained with Equation 2.8. The heaviness is calculated with Equation 1.1, where precipitation from ERA5 HRES is used (Gründemann, 2021). For the controls, the ERA5 monthly reanalysis is used (Hersbach et al., 2019). All the data covers the period from 1979 until 2019.

D.2. Geography

The results of the scatter density plots with the controls belonging to the geography are shown here. These controls are: elevation (Figure 19a), slope aspect (Figure 19b), slope riserun (Figure 19c), wind speed (Figure 19d) and the combined control of wind direction and slope (Figure 19e).

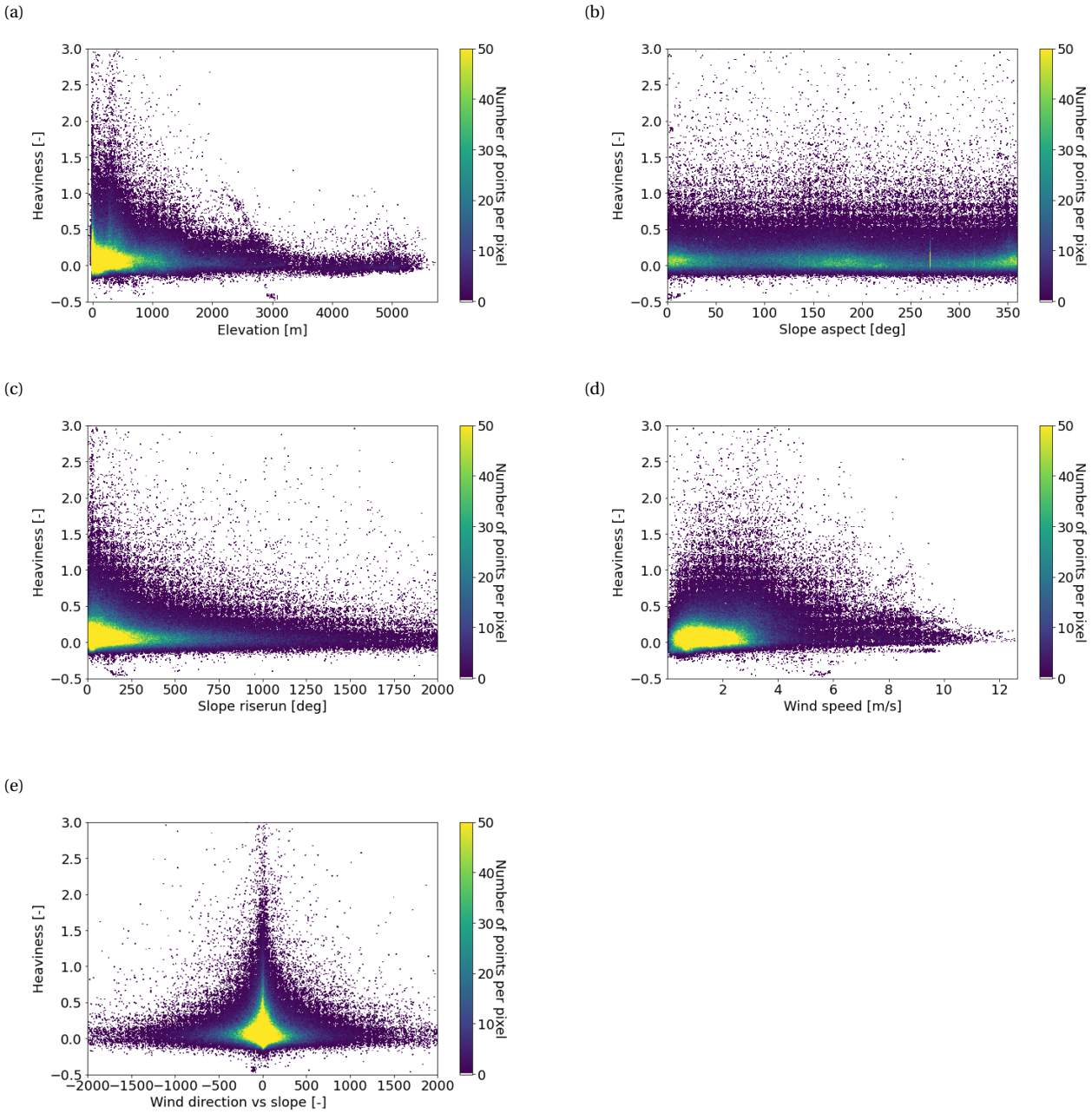

Figure 19: Scatter density plots of controls versus the heaviness, where the following controls are plotted: a) Elevation obtained from GMTED2010 (Danielson & Gesch, 2011). b) Aspect of slope obtained from elevation using python package `richDEM` (Horn, 1981). c) Riserun of the slope obtained from elevation using python package `richDEM` (Horn, 1981). d) Average wind speed obtained with Equation 2.9. e) Wind direction vs slope ρ obtained with Equation 2.10. The wind speed and wind direction are obtained from the monthly reanalysis product from ERA5 (Hersbach et al., 2019). The heaviness is calculated with Equation 1.1, where precipitation data from ERA5 HRES is used (Gründemann, 2021). All data covers the period from 1979 until 2019 if applicable.

D.3. Land cover

The results of the density scatter plots of albedo (Figure 20a), soil type (Figure 20b) and leaf area index (not shown here but given in Figure 3.6b) are given below. For soil type, a box-and-whiskers plot is made since this one is more appropriate due to the discrete values of the soil type.

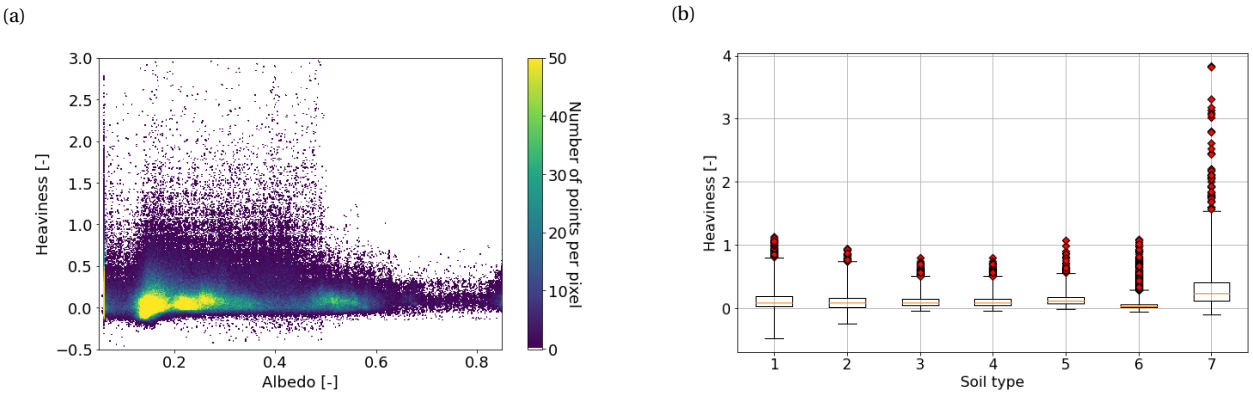

Figure 20: Scatter density plots of controls versus the heaviness, where the following controls are plotted: a) Albedo obtained with Equation 2.13. b) Soil type. The heaviness is calculated with Equation 1.1, where precipitation from ERA5 HRES is used (Gründemann, 2021). For the controls, the ERA5 monthly reanalysis is used (Hersbach et al., 2019). All the data covers the period from 1979 until 2019.

D.4. Variability

The results of the density scatter plots with the controls belonging to the variability category are shown here. These controls are: precipitation variability (Figure 21a) and temperature variability (Figure 21a).

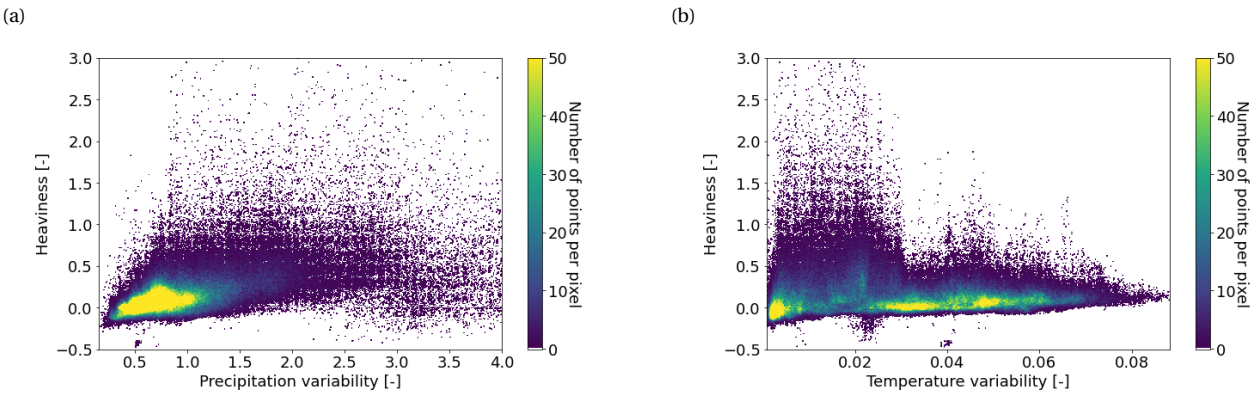

Figure 21: Scatter density plots of controls versus the heaviness, where the following controls are plotted: a) Precipitation variability obtained with Equation 2.15. b) Temperature variability obtained with Equation 2.16.. The heaviness is calculated with Equation 1.1, where precipitation from ERA5 HRES is used (Gründemann, 2021). For the controls, the ERA5 monthly reanalysis is used (Hersbach et al., 2019). All the data covers the period from 1979 until 2019.

E. Correlation coefficients

Correlation coefficients of the controls with heaviness for all regions are given in Table 3.

Table 3: Correlation coefficients of the controls with heaviness. The meaning of the abbreviations of the controls can be found in Section 2.4. The location of the regions is given in Figure 2.2.

	T	P	E	W_{days}	P_{Wdays}	C_t	RH	ELV	S_a	S_r	V	ρ	A	ST	LAI	Pvar	Tvar
World	0.225	−0.347	0.176	−0.488	−0.092	−0.484	−0.459	−0.085	−0.030	−0.060	0.178	−0.001	0.007	−0.066	−0.299	0.555	−0.062
GIC	−0.307	−0.371	0.178	−0.437	−0.295	−0.162	0.019	−0.007	0.203	−0.035	0.046	−0.067	0.100	0.021	−0.154	0.439	0.487
NWN	−0.393	−0.493	0.259	−0.644	−0.392	0.042	−0.081	−0.245	−0.099	−0.256	−0.013	−0.011	0.222	0.159	−0.116	0.468	0.475
NEN	−0.361	−0.557	0.293	−0.587	−0.400	0.001	0.120	−0.162	−0.071	0.036	−0.077	−0.032	0.271	−0.038	−0.303	0.461	0.198
WNA	0.257	−0.412	0.289	−0.531	−0.211	−0.367	−0.426	−0.076	−0.234	−0.100	0.073	−0.086	0.137	0.222	−0.284	0.347	0.301
CNA	0.287	−0.288	0.100	−0.393	−0.153	−0.445	−0.452	0.291	−0.144	0.222	0.443	−0.238	−0.086	0.006	−0.075	0.465	−0.324
ENA	0.539	0.028	−0.634	−0.152	0.151	−0.461	−0.066	−0.309	−0.097	−0.178	−0.205	−0.032	−0.46	−0.043	−0.003	0.614	−0.552
NCA	0.340	−0.364	−0.039	−0.399	−0.029	−0.276	0.080	−0.432	−0.067	−0.170	0.296	−0.112	−0.138	−0.165	−0.291	0.463	0.052
SCA	0.188	−0.387	0.081	−0.372	−0.336	−0.398	−0.115	−0.212	0.118	−0.236	0.279	−0.047	−0.004	−0.011	0.017	0.174	0.261
CAR	0.211	−0.486	−0.167	−0.395	−0.154	−0.471	−0.160	−0.182	0.064	−0.137	−0.005	−0.203	−0.288	−0.291	−0.302	0.318	0.034
NWS	0.128	−0.256	0.409	−0.567	−0.136	−0.341	−0.099	−0.238	0.133	−0.009	0.472	0.080	−0.304	−0.274	−0.342	0.481	0.484
NSA	0.124	−0.406	0.279	−0.406	−0.334	−0.405	−0.254	−0.033	0.017	0.016	0.403	0.033	−0.084	−0.106	−0.223	0.381	0.219
NES	0.075	−0.534	0.470	−0.377	−0.434	−0.162	−0.181	−0.011	−0.119	0.138	0.443	0.074	0.100	0.012	−0.210	0.380	0.100
SAM	0.079	−0.213	0.114	−0.352	0.013	−0.234	−0.123	−0.075	−0.055	0.061	0.266	0.034	−0.066	−0.020	0.018	0.156	0.367
SWS	0.321	−0.463	0.329	−0.590	−0.108	−0.511	−0.151	−0.058	0.170	0.164	0.041	−0.067	0.006	−0.059	−0.197	0.585	−0.065
SES	−0.098	−0.553	0.489	−0.539	−0.348	−0.342	−0.324	−0.160	−0.197	−0.091	0.044	−0.008	−0.191	−0.26	−0.298	0.315	0.498
SSA	0.714	−0.649	0.367	−0.799	−0.498	−0.812	−0.811	0.103	−0.292	−0.006	−0.165	0.251	−0.221	0.249	−0.108	0.841	0.682
NEU	−0.315	−0.505	0.374	−0.522	−0.442	0.046	0.015	−0.017	−0.221	−0.195	−0.180	−0.141	0.197	0.067	0.033	0.174	0.411
WCE	0.145	−0.326	−0.198	−0.488	−0.091	−0.308	−0.315	−0.140	−0.103	−0.079	−0.179	0.028	−0.114	−0.137	−0.282	0.475	0.236
EEU	0.521	−0.664	0.038	−0.644	−0.474	−0.571	−0.589	−0.354	−0.054	−0.142	−0.395	−0.116	−0.138	0.177	−0.526	0.629	0.239
MED	0.605	−0.621	0.385	−0.687	−0.391	−0.615	−0.532	−0.161	−0.117	−0.230	0.246	0.129	0.449	−0.111	−0.397	0.681	0.015
SAH	0.179	0.069	−0.053	0.063	0.223	0.288	0.002	−0.170	−0.002	−0.106	−0.013	−0.049	−0.003	−0.073	0.006	−0.097	−0.139
WAF	0.475	−0.455	0.591	−0.501	−0.271	−0.560	−0.543	−0.101	−0.032	−0.284	−0.013	−0.100	0.515	−0.173	−0.316	0.594	0.557
CAF	0.427	−0.420	0.430	−0.409	−0.235	−0.400	−0.389	−0.291	0.122	−0.136	0.275	−0.005	0.495	−0.021	−0.204	0.465	0.448
NEAF	0.206	−0.401	0.210	−0.509	−0.284	−0.430	0.053	−0.318	−0.116	−0.007	0.367	0.010	0.153	−0.404	−0.491	0.627	0.058
SEAF	0.436	−0.324	0.436	−0.411	−0.280	−0.437	−0.223	−0.417	−0.217	−0.097	0.494	−0.121	0.287	−0.147	−0.250	0.396	0.177
WSAF	−0.197	−0.596	0.526	−0.591	−0.457	−0.456	−0.126	−0.409	0.083	0.131	0.167	0.130	0.243	−0.212	−0.496	0.295	0.384
ESAF	0.286	−0.296	−0.017	−0.230	−0.271	−0.220	0.199	−0.455	−0.173	−0.099	0.028	−0.088	−0.042	−0.214	−0.092	0.199	−0.032
MDG	0.295	−0.400	−0.158	−0.284	−0.282	−0.340	−0.083	−0.334	0.093	−0.324	0.111	0.009	−0.288	−0.360	−0.345	0.318	−0.098
RAR	−0.347	−0.509	0.184	−0.517	−0.148	0.056	−0.025	−0.120	−0.074	−0.126	−0.341	−0.012	−0.041	−0.046	−0.034	0.335	0.208
WSB	0.380	−0.491	0.162	−0.495	−0.331	−0.309	−0.374	−0.101	−0.089	−0.093	−0.005	−0.043	−0.075	0.020	−0.310	0.477	0.005
ESB	0.275	−0.503	0.144	−0.547	0.001	−0.393	−0.476	0.101	−0.064	−0.018	0.152	0.024	−0.177	0.125	−0.281	0.398	−0.037
RFE	−0.026	−0.164	−0.197	−0.314	0.089	−0.312	−0.345	0.068	0.001	−0.033	−0.126	−0.040	−0.218	0.103	0.086	0.359	0.265
WCA	0.491	−0.347	0.003	−0.442	0.040	−0.437	−0.305	−0.291	−0.056	−0.255	0.150	0.027	−0.306	−0.079	−0.368	0.499	−0.240
ECA	0.493	−0.556	0.526	−0.604	−0.132	−0.560	−0.576	−0.451	−0.017	−0.310	0.272	0.090	−0.450	0.021	−0.387	0.599	0.440
TIB	0.462	0.026	−0.325	−0.175	0.229	−0.436	−0.011	−0.507	0.121	0.092	−0.218	−0.058	−0.279	−0.145	0.142	0.201	−0.162
EAS	−0.063	−0.435	0.072	−0.550	−0.046	−0.502	−0.397	−0.345	−0.038	−0.134	0.164	0.073	−0.068	−0.251	−0.292	0.538	0.335
ARP	0.203	−0.108	−0.086	−0.129	0.363	0.182	0.301	−0.204	−0.011	0.092	0.130	−0.052	−0.085	0.004	−0.108	0.473	−0.433
SAS	0.272	−0.547	0.453	−0.634	−0.363	−0.568	−0.556	−0.163	0.002	−0.139	0.315	−0.035	0.310	−0.183	−0.496	0.609	0.323
SEA	0.086	−0.284	0.050	−0.323	−0.184	−0.325	−0.179	−0.101	−0.046	−0.061	0.264	−0.008	−0.188	−0.162	−0.225	0.324	0.233
NAU	0.027	−0.120	0.110	−0.112	−0.017	−0.123	−0.037	−0.066	−0.063	−0.040	−0.119	0.091	0.114	0.050	0.002	0.118	0.057
CAU	0.083	−0.134	−0.121	−0.225	0.073	−0.235	0.144	−0.216	0.005	0.137	−0.108	0.046	0.000	−0.058	−0.007	0.328	−0.193
EAU	0.527	0.069	−0.047	−0.029	0.405	−0.070	0.090	−0.225	−0.272	−0.120	0.554	0.038	−0.120	−0.051	−0.102	0.533	−0.411
SAU	0.584	−0.561	0.434	−0.601	−0.099	−0.609	−0.602	−0.041	−0.070	−0.287	−0.193	−0.042	0.430	0.154	−0.193	0.589	0.479
NZ	0.397	−0.433	−0.264	−0.610	−0.266	−0.470	−0.424	−0.140	−0.316	−0.226	−0.329	0.146	−0.053	−0.035	−0.074	0.674	0.064

F. Regression coefficients

Coefficients of the controls obtained with the multiple linear regression for elimination methods 1, 2 and 3 are shown in Tables 4, 5 and 6, respectively.

Table 4: Regression coefficients for the controls obtained with elimination method 1. If a cell is blank, this control is not taken into account for the multiple linear regression for that specific region. The meaning of the abbreviations of the controls can be found in Section 2.4 and the locations of the regions is shown in Figure 2.2.

	T	P	E	W_{days}	P_{Wdays}	C_t	RH	ELV	S_a	S_r	V	ρ	A	ST	LAI	P_{var}	T_{var}
World	0.147	0.016	0.006	−0.134	−0.016	0.037	0.061			0.016	0.015		0.010	−0.007	−0.018	0.073	0.033
GIC	0.018	0.160		−0.083	−0.094	0.050	−0.007	0.023	0.025		0.007	−0.009	−0.010		−0.006	0.084	0.047
NWN	−0.006	0.031	−0.021	−0.117		0.004	−0.007	0.009	−0.004		−0.009	−0.003	0.011	0.009	−0.006		−0.017
NEN	0.047			−0.065	−0.006	0.012		0.008			0.009		−0.002		−0.009	0.021	
WNA	−0.191	0.080	−0.028	−0.099	−0.029		−0.158	−0.090			0.018		−0.038			0.046	0.013
CNA	0.121	0.027	0.016	−0.079	−0.078	0.046	0.053	0.021	0.007	0.008	0.012	−0.005	−0.010				
ENA	−0.083	−0.034	−0.054	0.032	0.026		−0.020	−0.003	−0.005				−0.006	0.004		0.039	−0.058
NCA	0.202		0.123	−0.015	−0.093	0.086	0.258	0.148		0.042	0.055		−0.061			0.253	0.245
SCA		0.452	0.076	−0.422	−0.330	−0.163	0.164	−0.042	0.022		0.041		0.022		0.031	−0.048	−0.023
CAR	−0.068	0.617	−0.060	−0.602	−0.582	−0.089				−0.046	−0.099		−0.020		−0.025	0.219	0.090
NWS	0.033		0.008	−0.088	−0.012	0.032	0.013		0.004	0.023	−0.012	−0.004	−0.035		−0.004	0.020	0.023
NSA	−0.010	0.081	0.013	−0.079	−0.065	−0.018	0.036	−0.011			0.011						0.017
NES	−0.021	0.120	0.014	−0.088	−0.176	0.021	−0.015	−0.029					−0.013			0.108	
SAM	0.011	−0.034	0.014	−0.027	0.019		0.038		−0.006	0.027	0.007			0.004	0.008	0.015	0.016
SWS	−0.022	−0.052	0.029	−0.020	0.031	−0.094	0.088	−0.067		0.009		−0.017	0.022	−0.014	−0.016	0.110	−0.005
SES	0.021	0.063	0.018	−0.089	−0.056	0.020	0.013	−0.016		0.012			−0.015	0.007	−0.004	0.018	0.026
SSA	−0.011	0.010	0.009	−0.101			−0.113	−0.031	0.013	0.024	0.021	0.039	0.018	−0.020	0.033	0.142	−0.081
NEU	−0.003	−0.004	0.005	−0.021		0.010		−0.003	−0.002		0.009		0.005		−0.005	0.006	−0.001
WCE	0.026	0.031		−0.058	−0.024	0.010	0.016	0.031		0.009	0.019		−0.011	−0.003	−0.004	0.013	0.037
EEU	−0.006			−0.015	−0.023			−0.006		0.005	−0.009	0.002	−0.007	0.004	−0.007	0.043	−0.018
MED	−0.041		0.012	−0.078	−0.021		−0.018	−0.030	−0.016	0.020	−0.015	0.005	0.016	−0.019	0.013	0.075	−0.028
SAH	0.160	−0.182	0.027		0.064	0.316	0.166	0.059	−0.015	0.019			0.023	−0.020		0.139	0.078
WAF	0.028	0.166	0.046		−0.117			−0.014			0.042	−0.008	−0.005		0.010	0.245	
CAF	−0.062				−0.019			−0.090	0.013	−0.015	−0.014		0.067		−0.011	0.083	
NEAF	0.062	0.402	0.078	−0.130	−0.310	0.100	0.206				0.022	0.021	−0.083			0.472	0.095
SEAF	0.087	0.087	0.115	−0.069	−0.036		0.087			0.021	0.072	−0.013	−0.013	−0.019	−0.012		−0.008
WSAF	0.010			−0.163	−0.018	0.102	0.025			0.005	0.013	0.007	0.015	−0.012	−0.011	−0.006	0.050
ESAF	−0.159	0.244	0.022	−0.151	−0.172	−0.107	0.100	−0.169			−0.038	−0.020	0.023		−0.021	0.186	−0.048
MDG	0.033	−0.035	0.005		−0.048	0.022			−0.008	0.022		−0.016	−0.031	−0.008	−0.030	0.074	0.005
RAR	0.044		0.005	−0.009		0.047	0.013	−0.04			−0.018		0.008	0.005	−0.009	0.102	0.023
WSB	0.003		−0.003	−0.053	−0.008	0.040		−0.006		0.008	0.005			−0.003	−0.004	0.028	−0.004
ESB	0.054	0.074		−0.120	−0.073	0.040	0.021	0.010	0.005	0.009	0.016				−0.006	0.049	0.022
RFE	0.037	−0.020	−0.014	−0.031	0.024	0.014	0.013	0.023		0.004	−0.006		−0.012		0.003	0.003	0.023
WCA	0.240			−0.088		0.099	0.102	0.145							−0.030	0.048	0.021
ECA	0.023		0.047		−0.006	−0.036	0.066	−0.043				−0.005	−0.011	−0.021	−0.011	0.112	−0.014
TIB	0.144		0.030	−0.072	0.009	−0.079	0.079		0.010	0.014	0.037	−0.007			−0.007	−0.068	0.032
EAS	−0.032	0.042	−0.019	−0.042	−0.052	0.009	−0.010	−0.069	−0.005	0.017					−0.012	0.062	−0.023
ARP	−0.178		0.048	−0.089	−0.119	0.190	−0.049	−0.232	−0.019	0.083	−0.056	0.017	−0.102			0.365	−0.086
SAS	0.021	0.122	0.009	−0.153	−0.126	0.057				0.020		−0.011	−0.014		−0.013	0.156	0.008
SEA		0.124		−0.111	−0.128	−0.018	0.046	−0.009		0.025	0.013		−0.028		−0.009		0.038
NAU	−0.034		0.210			−0.091	0.213	−0.059	−0.016	−0.020	−0.046				−0.037		
CAU	−0.175	0.344	0.021	−0.162	−0.259	−0.116		−0.039			−0.111		−0.042			0.264	−0.077
EAU	0.027		0.034	−0.041	−0.004	0.023			−0.01			−0.007	−0.009		0.022	0.075	−0.111
SAU	0.116	0.088	0.046		−0.093		0.149			−0.009	−0.010		−0.029			0.140	0.157
NZ	0.013	−0.007	−0.011	−0.009		0.024	−0.009		−0.004	0.001	−0.008	−0.003	0.015	−0.005		0.023	−0.003

Table 5: Regression coefficients for the controls obtained with elimination method 2. If a cell is blank, this control is not taken into account for the multiple linear regression for that specific region. Some controls are not present in the table since they are not in the top three of any of the regions. The meaning of the abbreviations of the controls can be found in Section 2.4 and the locations of the regions is shown in Figure 2.2.

	T	P	E	W_{days}	P_{Wdays}	C_t	RH	ELV	A	Pvar	Tvar
World	0.045			−0.089						0.065	
GIC		0.114		−0.120	−0.068						
NWN		0.022		−0.101							−0.011
NEN	0.031			−0.067	−0.008						
WNA	−0.095						−0.130	−0.092			
CNA	0.066				−0.051					0.022	
ENA	−0.028		−0.058								−0.033
NCA							0.124			0.179	0.141
SCA		0.482		−0.417	−0.390						
CAR		0.283		−0.479	−0.286						
NWS		0.008		−0.135		0.041					
NSA		0.111		−0.095	−0.095						
NES		0.189		−0.232	−0.211						
SAM		0.027		−0.103		0.050					
SWS						−0.142	0.136			0.138	
SES		0.049		−0.084	−0.049						
SSA				−0.046			−0.066			0.129	
NEU		0.091		−0.058	−0.069						
WCE				−0.052			0.015	0.016			
EEU				−0.105		0.082				0.042	
MED				−0.127		0.076				0.087	
SAH		−0.142				0.318				0.112	
WAF		0.204			−0.173					0.292	
CAF							0.127		0.097	0.138	
NEAF		0.416			−0.363					0.474	
SEAF	0.106		0.142				0.065				
WSAF				−0.199		0.116					0.041
ESAF	−0.236							−0.292		0.151	
MDG	−0.160							−0.204		0.105	
RAR						0.041		−0.045		0.091	
WSB		0.018		−0.101		0.044					
ESB				−0.047	−0.025					0.025	
RFE	0.046			−0.046				0.028			
WCA	0.074			−0.058			0.036				
ECA				−0.056				−0.040		0.087	
TIB	0.066					−0.112				−0.078	
EAS					−0.030			−0.056		0.057	
ARP				−0.066		0.200				0.238	
SAS		−0.034		−0.099						0.095	
SEA		0.139			−0.152					0.129	
NAU			0.146			−0.129	0.218				
CAU	−0.088				−0.019					0.161	
EAU				−0.047			0.129			0.143	
SAU	0.078									0.045	0.054
NZ		−0.014			0.005					0.027	

Table 6: Regression coefficients for the controls obtained with elimination method 3. If a cell is blank, this control is not taken into account for the multiple linear regression for that specific region. The meaning of the abbreviations of the controls can be found in Section 2.4 and the locations of the regions is shown in Figure 2.2.

	T	P	E	W_{days}	P_{Wdays}	C_t	RH	ELV	S_a	S_r	V	ρ	A	ST	LAI	P_{var}	T_{var}
World			−0.021	−0.10	−0.009	0.015	0.006	−0.026	−0.005	0.012	0.006		−0.015		−0.012	0.088	−0.027
GIC			−0.023	−0.004	0.009	0.038	0.005	0.033	0.027	0.008	0.005	−0.006	−0.009			0.072	0.057
NWN	−0.010		−0.019	−0.105	0.024	0.004	−0.008	0.009	−0.004	−0.002	−0.009	−0.004	0.010	0.009	−0.006		−0.018
NEN	0.047			−0.065	−0.006	0.012		0.008			0.009		−0.002		−0.009	0.021	
WNA		0.022	0.008		−0.008	0.003		−0.006	−0.010	0.004	0.023	−0.007	−0.031		−0.005	0.073	0.097
CNA			0.015	−0.039	−0.029	0.042	0.018	−0.019	0.006	0.007		−0.010	−0.008			0.044	−0.088
ENA	−0.090	−0.011	−0.052	0.018		−0.014	−0.019	−0.003	−0.005				−0.007	0.004		0.039	−0.057
NCA	0.203	0.122	0.117	−0.133	−0.143	0.105	0.267	0.154		0.042	0.052		−0.070			0.257	0.257
SCA			0.129	−0.149	−0.010	−0.185	0.182		0.033	−0.011	0.089				0.041	−0.102	
CAR	−0.073		−0.068	−0.061	−0.213	−0.088	0.018	0.029	−0.022	−0.058	−0.105		−0.014	0.011	−0.021	0.134	0.058
NWS			−0.009	−0.112	−0.004	0.039	0.016		0.005	0.015	−0.011		−0.039	0.008		0.026	0.016
NSA	−0.006		0.020	−0.034	−0.012	−0.008	0.043	−0.013		0.004	0.016	−0.005	−0.003	0.002		0.016	0.009
NES			0.016	−0.016	−0.092	0.043	−0.003	−0.015			0.009	0.004	−0.009		−0.012	0.096	−0.007
SAM			−0.019	−0.092		0.052	0.027		−0.003	0.034	0.006		0.008		0.004	0.022	0.017
SWS			0.031	−0.052	0.016	−0.102	0.136		−0.008	0.007	0.007	−0.020	0.005	−0.015	−0.011	0.116	
SES	0.021	0.063	0.018	−0.089	−0.056	0.02	0.013	−0.016		0.012			−0.015	0.007	−0.004	0.018	0.026
SSA	0.029			−0.166		0.107	−0.105	−0.009	0.014	0.027	0.026	0.039	0.018	−0.023	0.031	0.147	−0.074
NEU	−0.003		0.006	−0.022	−0.004	0.009		−0.003	−0.002		0.009	0.001	0.005		−0.004	0.006	−0.001
WCE	0.020		−0.003	−0.042	−0.004	0.011	0.013	0.031		0.011	0.017		−0.011	−0.004	−0.004	0.012	0.033
EEU	0.009	−0.029	−0.022		−0.018	0.099	−0.057			0.006	−0.011	0.007		0.005	−0.010	0.062	−0.020
MED	0.014	−0.063	0.020			0.036	0.010	−0.007	−0.017	0.020	−0.010	0.007	0.018	−0.018	0.006	0.104	−0.014
SAH	0.153	−0.222	0.027	0.046	0.07	0.314	0.158	0.055	−0.015	0.018			0.022	−0.020		0.142	0.077
WAF		0.166	0.062		−0.124	−0.030		−0.031			0.031	−0.008	−0.008	0.008	0.010	0.248	−0.028
CAF		0.071			−0.057	0.034		−0.055	0.014	−0.017	−0.008		0.055		−0.016	0.231	−0.099
NEAF	0.205		0.074	0.124	−0.075	0.061	0.283	0.127			0.038		−0.034			0.478	0.117
SEAF			0.142	0.013		−0.058	0.098		0.007	−0.008	0.043	−0.016	0.009	−0.029	−0.026	0.028	−0.031
WSAF	0.014	−0.012	0.010	−0.147	−0.015	0.092	0.039	0.005			0.010	0.009	0.014	−0.012	−0.009	−0.005	0.050
ESAF	−0.139		0.04	0.039	−0.056	−0.109	0.100	−0.165		0.011	−0.033	−0.019	0.021		−0.029	0.167	−0.045
MDG				−0.061	−0.058	0.055	−0.019	−0.049	−0.005	0.022	−0.005	−0.016	−0.032	−0.009	−0.032	0.067	0.004
RAR	0.044		0.005	−0.009		0.047	0.013	−0.040			−0.018		0.008	0.005	−0.009	0.102	0.023
WSB	−0.017		−0.012	−0.056	−0.005	0.055	−0.034	−0.011		0.006	0.003		0.002		−0.005	0.027	−0.009
ESB	0.068		0.009	−0.056	−0.025	0.044	0.031	0.020	0.005	0.010	0.015	−0.004		0.005	−0.007	0.042	0.026
RFE	0.037		−0.015	−0.044	0.012	0.013	0.014	0.023		0.004	−0.006		−0.012		0.002	0.003	0.023
WCA			−0.008	−0.058	0.012	0.019	0.020	−0.005	−0.007	−0.012	−0.016		−0.010	0.006	−0.035	0.050	−0.038
ECA	−0.017	−0.083	0.055		0.022	−0.011	0.088	−0.079		0.006		−0.006	−0.011	−0.021		0.116	−0.040
TIB			0.019	−0.028	0.008	−0.086	0.050	−0.103	0.011		0.029	−0.005	−0.017	−0.007		−0.046	0.033
EAS	−0.025		−0.016	−0.006	−0.028	0.011	−0.017	−0.065	−0.005	0.018	0.004		−0.003	−0.004	−0.011	0.060	−0.017
ARP	−0.171	0.069	0.048	−0.158	−0.126	0.186	−0.042	−0.223	−0.020	0.087	−0.057		−0.101			0.367	−0.083
SAS		0.122	0.007	−0.157	−0.126	0.058	−0.010	−0.016		0.020		−0.012	−0.013		−0.012	0.153	
SEA			−0.006	−0.038	−0.044	0.005	0.037	−0.003	−0.004	0.035	0.013		−0.022		−0.019	0.027	0.037
NAU		0.058	0.070		−0.045	−0.069		−0.029	−0.033		−0.013	0.010	0.017		0.009	0.091	−0.126
CAU	−0.160		0.036	0.109	−0.094	−0.093	0.024	−0.032			−0.098		−0.033	0.028		0.292	−0.064
EAU	0.058	−0.046	0.034		0.006	0.014	0.056	0.013	−0.009		−0.013	−0.006	−0.012		0.016	0.079	−0.075
SAU	0.116	0.088	0.046		−0.093		0.149			−0.009	−0.010		−0.029			0.140	0.157
NZ	0.014		−0.011	−0.013	−0.003	0.024	−0.009		−0.005		−0.007	−0.002	0.014	−0.005		0.023	−0.001

G. Top 3 most important controls

The top three most important controls for method 2 are shown in Figure 22.

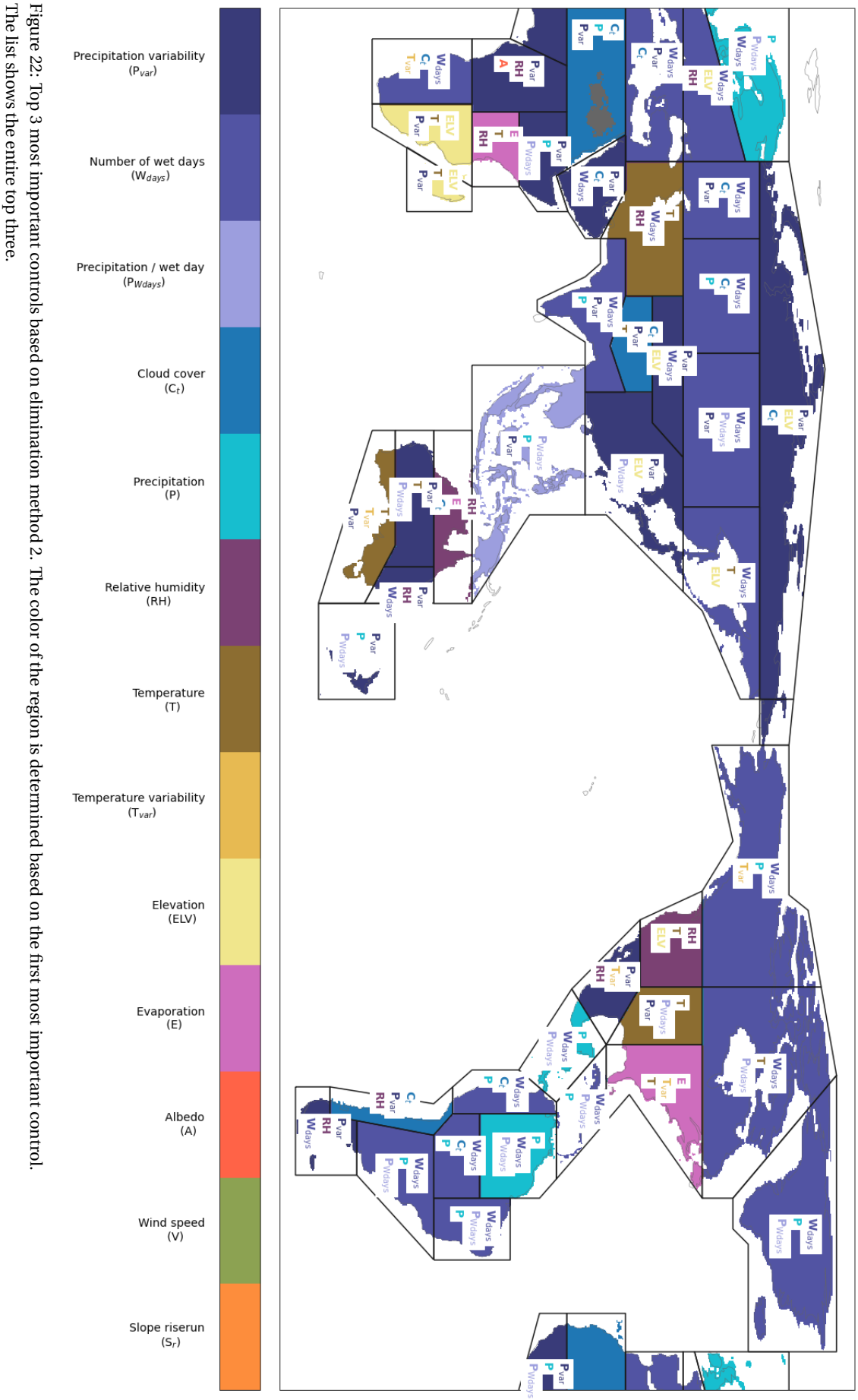

Figure 22: Top 3 most important controls based on elimination method 2. The color of the region is determined based on the first most important control. The list shows the entire top three.

The top three most important controls for method 3 are shown in Figure 23.

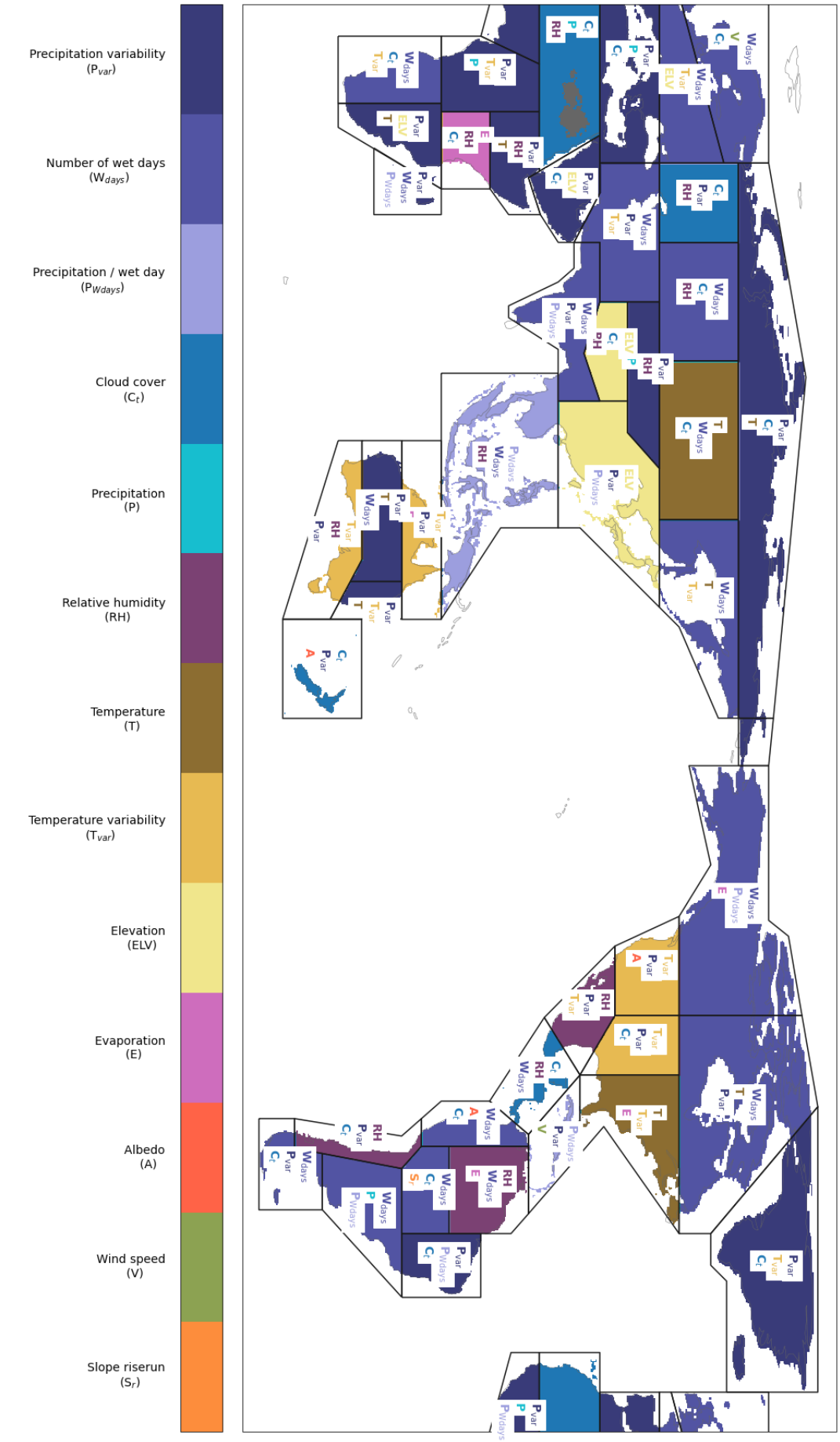

Figure 23: Top 3 most important controls based on elimination method 3. The color of the region is determined based on the first most important control. The list shows the entire top three.

H. Heaviness and error map

The calculation of the heaviness based on the set of controls and coefficients determined with method 2 is shown in Figure 24. The error map is given in Figure 25.

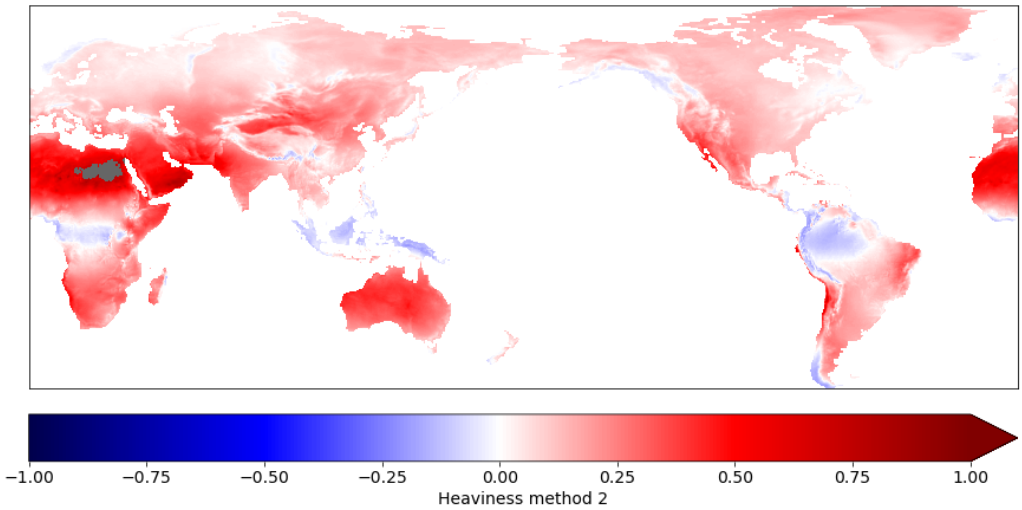

Figure 24: Calculation of the heaviness amplification factor based on the coefficients of the multiple linear regression and the remaining controls determined with elimination method 2. The mean is 0.154 and the standard deviation is 0.148. The used controls can be obtained from Figure 3.8.

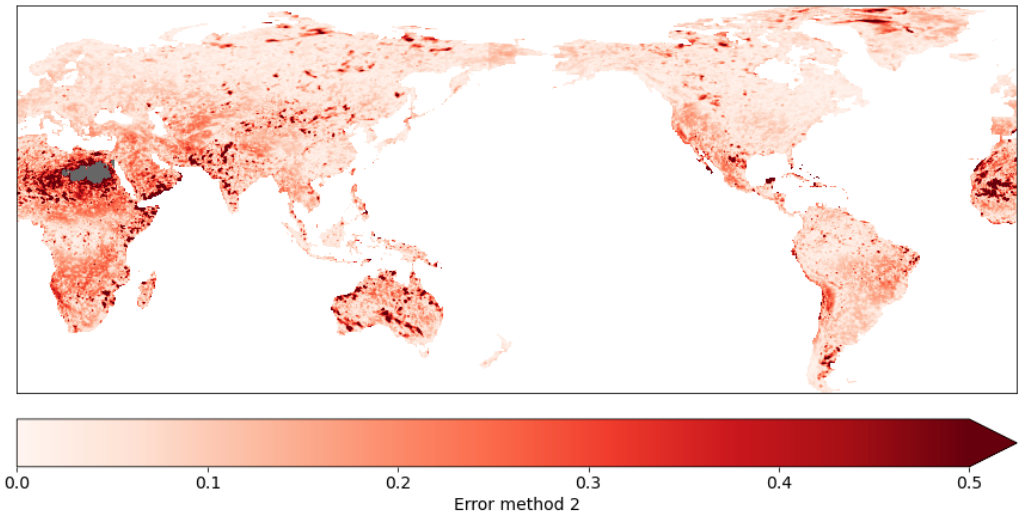

Figure 25: Absolute error between the heaviness data and calculation based on coefficients of the remaining controls determined with elimination method 2. The average error is 0.109. The used controls can be found in Figure 3.8.

The result based on method 3 for the heaviness calculation is given in Figure 26 and Figure 27 shows the errors of this method.

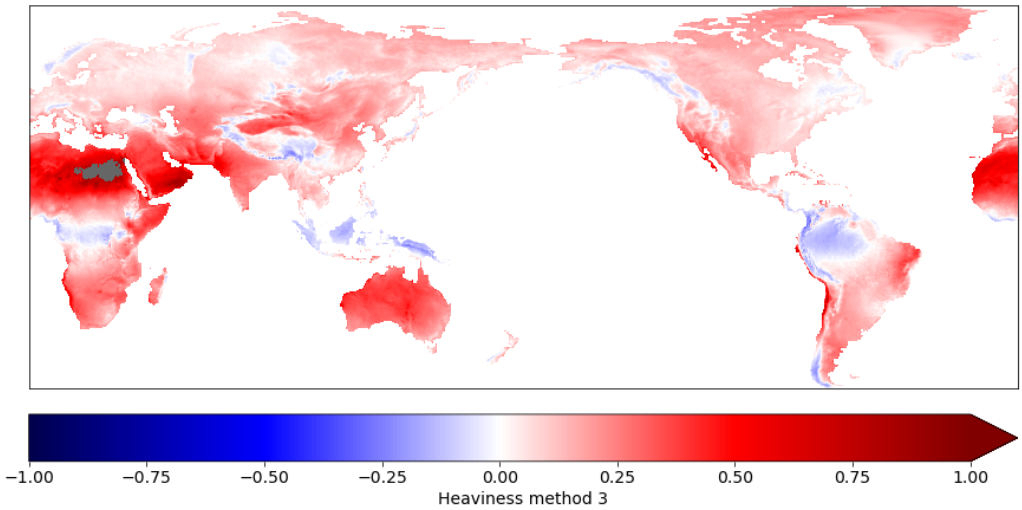

Figure 26: Calculation of the heaviness amplification factor based on the coefficients of the multiple linear regression and the remaining controls determined with elimination method 3. The mean is 0.154 and the standard deviation is 0.151. The used controls can be obtained from Figure 3.9.

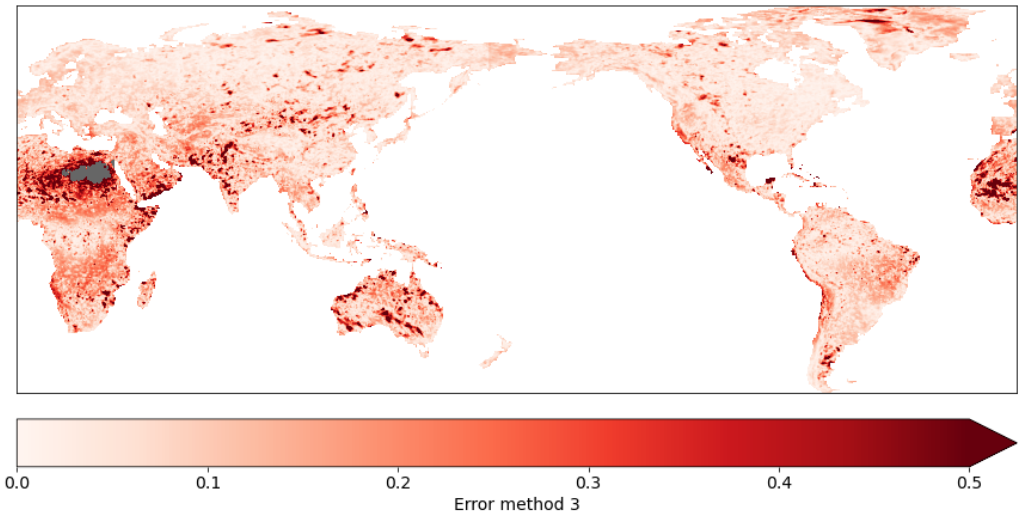

Figure 27: Absolute error between the heaviness data and calculation based on coefficients of the remaining controls determined with elimination method 3. The average error is 0.108. The used controls can be found in Figure 3.9.

I. Regional analysis

I.1. All regions

The absolute average error for each region using global and regional coefficients for all three methods is given in Table 7.

Table 7: Absolute average error for each region using regional and global coefficients for all three elimination methods. The location of the regions can be found in Figure 2.2. The used controls for each region can be found in Figures 3.7, 3.8 and 3.9, for methods 1, 2 and 3 respectively.

	Method 1		Method 2		Method 3	
	Regional	Global	Regional	Global	Regional	Global
GIC	0.079	0.089	0.088	0.086	0.080	0.094
NWN	0.059	0.064	0.060	0.061	0.059	0.066
NEN	0.045	0.051	0.046	0.045	0.045	0.052
WNA	0.061	0.085	0.073	0.110	0.065	0.090
CNA	0.044	0.055	0.048	0.059	0.044	0.053
ENA	0.038	0.046	0.044	0.050	0.038	0.045
NCA	0.157	0.169	0.183	0.188	0.156	0.177
SCA	0.165	0.174	0.179	0.178	0.175	0.175
CAR	0.181	0.244	0.203	0.291	0.190	0.239
NWS	0.074	0.098	0.081	0.089	0.077	0.100
NSA	0.064	0.078	0.066	0.070	0.064	0.073
NES	0.096	0.137	0.103	0.136	0.097	0.134
SAM	0.059	0.090	0.064	0.100	0.058	0.095
SWS	0.125	0.140	0.130	0.153	0.126	0.156
SES	0.048	0.073	0.055	0.086	0.048	0.080
SSA	0.088	0.133	0.090	0.120	0.088	0.119
NEU	0.026	0.051	0.027	0.049	0.026	0.069
WCE	0.036	0.061	0.039	0.063	0.036	0.063
EEU	0.045	0.063	0.047	0.060	0.045	0.055
MED	0.086	0.114	0.089	0.114	0.087	0.115
SAH	0.274	0.329	0.281	0.330	0.275	0.335
WAF	0.131	0.153	0.140	0.156	0.130	0.155
CAF	0.090	0.130	0.098	0.133	0.091	0.131
NEAF	0.216	0.227	0.250	0.250	0.219	0.234
SEAF	0.125	0.163	0.145	0.160	0.134	0.159
WSAF	0.110	0.147	0.114	0.174	0.110	0.160
ESAF	0.140	0.175	0.151	0.188	0.141	0.180
MDG	0.124	0.158	0.131	0.170	0.123	0.153
RAR	0.063	0.066	0.065	0.066	0.063	0.066
WSB	0.050	0.063	0.052	0.062	0.050	0.056
ESB	0.060	0.061	0.062	0.063	0.060	0.063
RFE	0.046	0.053	0.048	0.051	0.046	0.053
WCA	0.092	0.131	0.104	0.139	0.095	0.120
ECA	0.128	0.136	0.130	0.137	0.126	0.135
TIB	0.078	0.084	0.082	0.128	0.081	0.101
EAS	0.060	0.069	0.070	0.069	0.060	0.068
ARP	0.217	0.248	0.248	0.263	0.217	0.247
SAS	0.157	0.164	0.162	0.165	0.157	0.163
SEA	0.090	0.106	0.096	0.105	0.090	0.105
NAU	0.185	0.189	0.194	0.195	0.190	0.191
CAU	0.135	0.165	0.165	0.166	0.137	0.163
EAU	0.100	0.132	0.106	0.140	0.099	0.128
SAU	0.136	0.151	0.138	0.146	0.136	0.158
NZ	0.024	0.050	0.027	0.042	0.024	0.069

I.2. MDG

The calculation of the heaviness for region MDG based on the regional analysis and the entire dataset using the results of method 2 is shown in Figures 28a and 28c. Moreover, the error between both calculations and the data is also plotted and given in Figures 28b and 28d for the regional and global analysis.

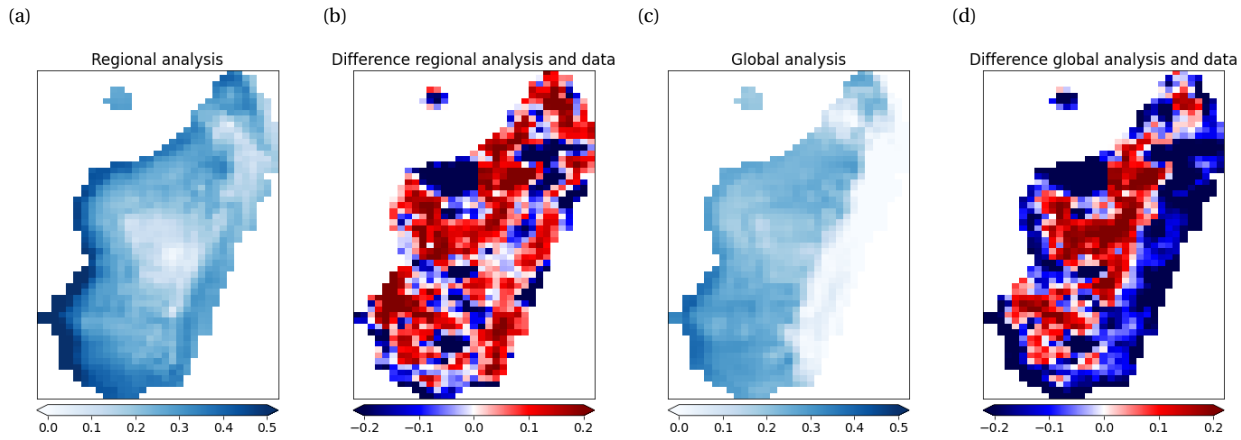

Figure 28: a) Calculated heaviness based on regional coefficients. b) Error between the calculated heaviness and the heaviness data for the regional analysis. c) Calculated heaviness based on global coefficients. c) Difference between the calculated heaviness and the data for the global analysis. Results are given for region MDG based on the outcomes of method 2. The used controls can be found in Figure 3.8.

The calculation of the heaviness for region MDG based on the regional analysis and the entire dataset using the results of method 3 is shown in Figures 29a and 29c. Moreover, the error between both calculations and the data is also plotted and given in Figures 29b and 29d for the regional and global analysis.

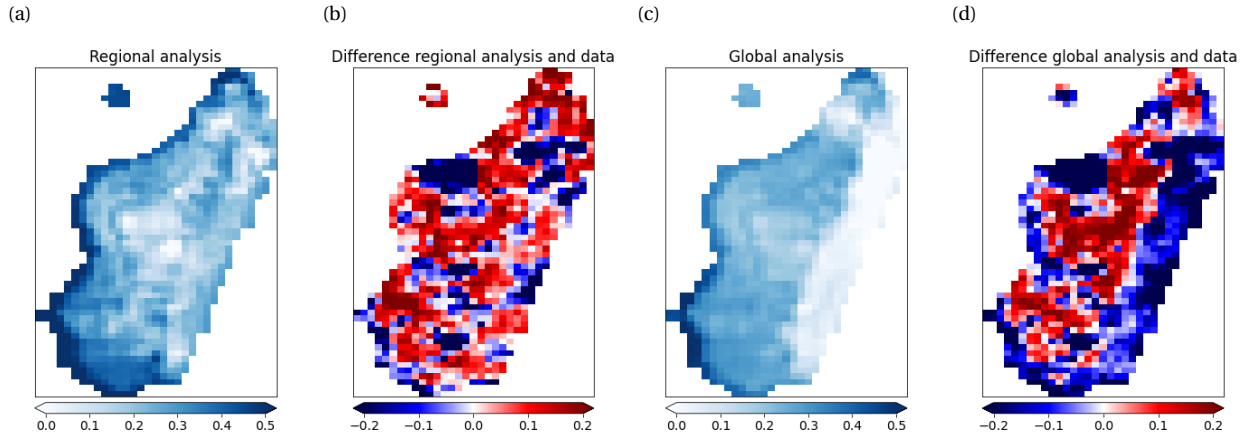

Figure 29: a) Calculated heaviness based on regional coefficients. b) Error between the calculated heaviness and the heaviness data for the regional analysis. c) Calculated heaviness based on global coefficients. c) Difference between the calculated heaviness and the data for the global analysis. Results are given for region MDG based on the outcomes of method 3. The used controls can be found in Figure 3.9.

J. Uncertainty

The relative occurrence of a control in a top three based on method 2 is given in Figures 30a, 30b and 30c, for the world, 1000 random points and region MDG. The length of the confidence intervals is given in Figures 31a and 31b.

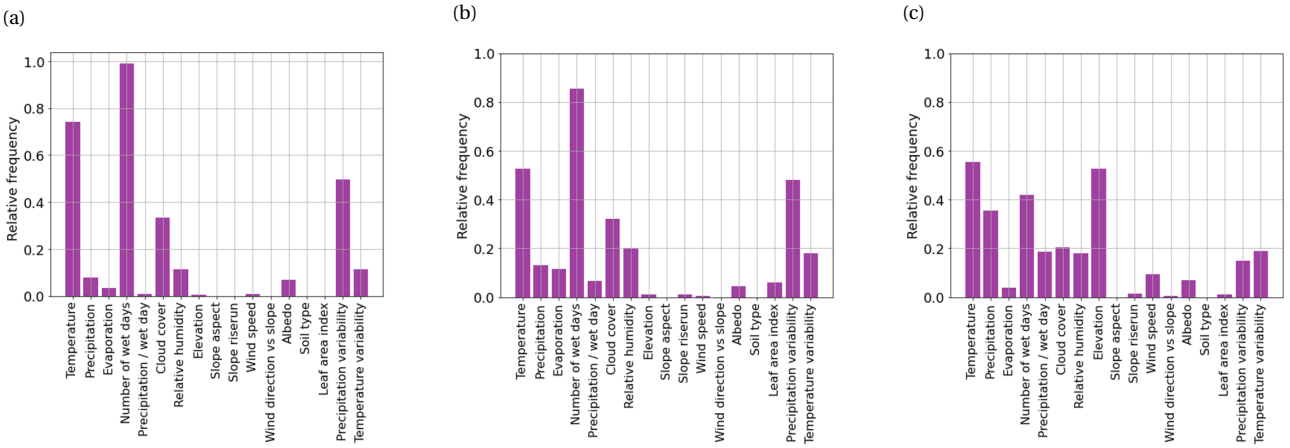

Figure 30: a) Number of times a control is the top three for the 200 samples for the world. b) Number of times a control is the top three for 200 samples and 1000 random points. c) Number of times a control is in the top three for the 200 samples for MDG. The results belong to elimination method 3.

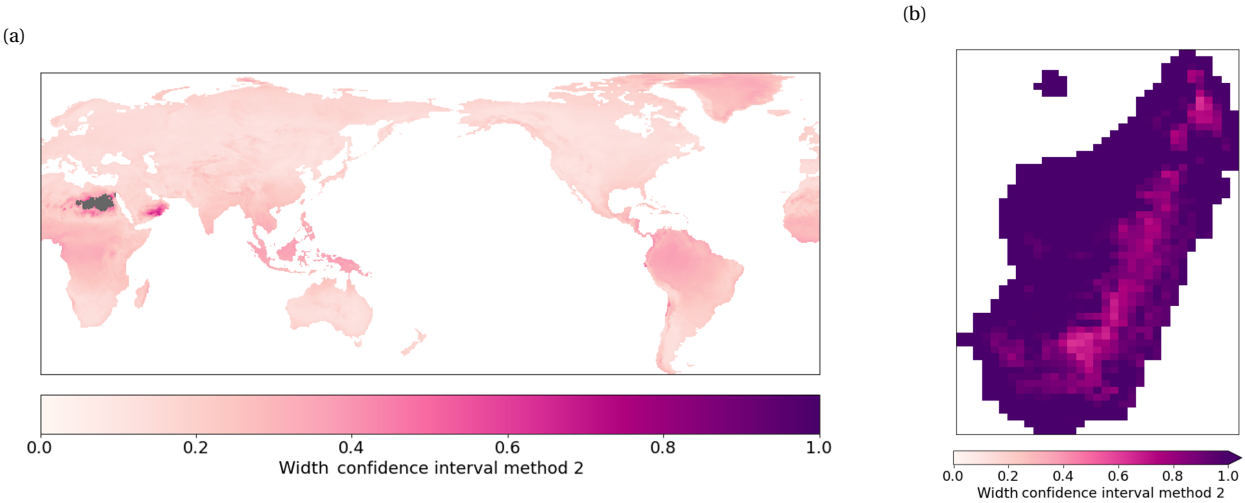

Figure 31: a) Confidence interval for the world. b) Confidence interval for region MDG, where only the data belonging to this region is used. Results are obtained with elimination method 2.

The relative occurrence of the controls in the top three based on method 3 is given in Figures 32a, 32b and 32c, for the world, 1000 random points and region MDG. The length of the confidence intervals is given in Figures 33a and 33b.

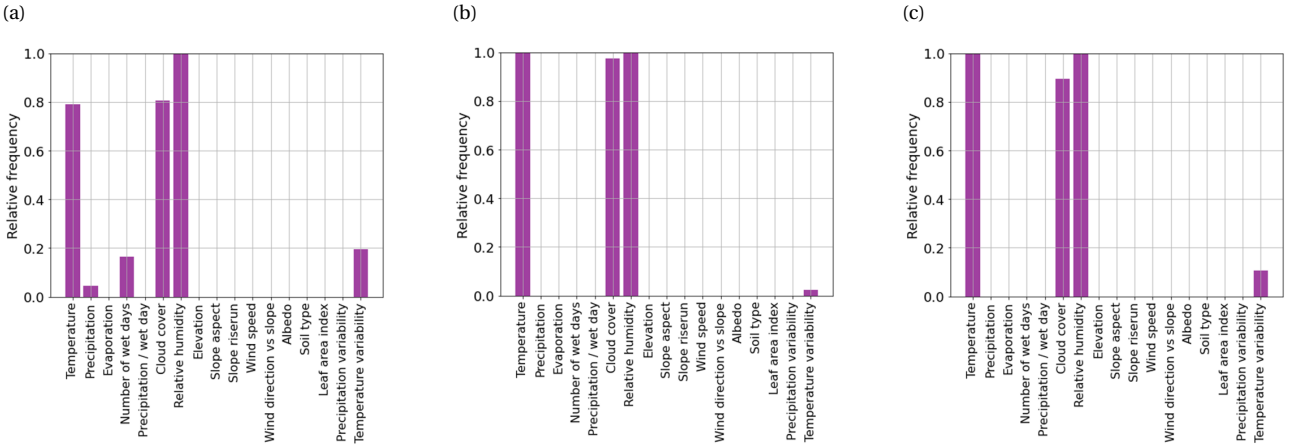

Figure 32: a) Number of times a control is the top three for the 200 samples for the world. b) Number of times a control is the top three for 200 samples and 1000 random points. c) Number of times a control is in the top three for the 200 samples for MDG. The results belong to elimination method 3.

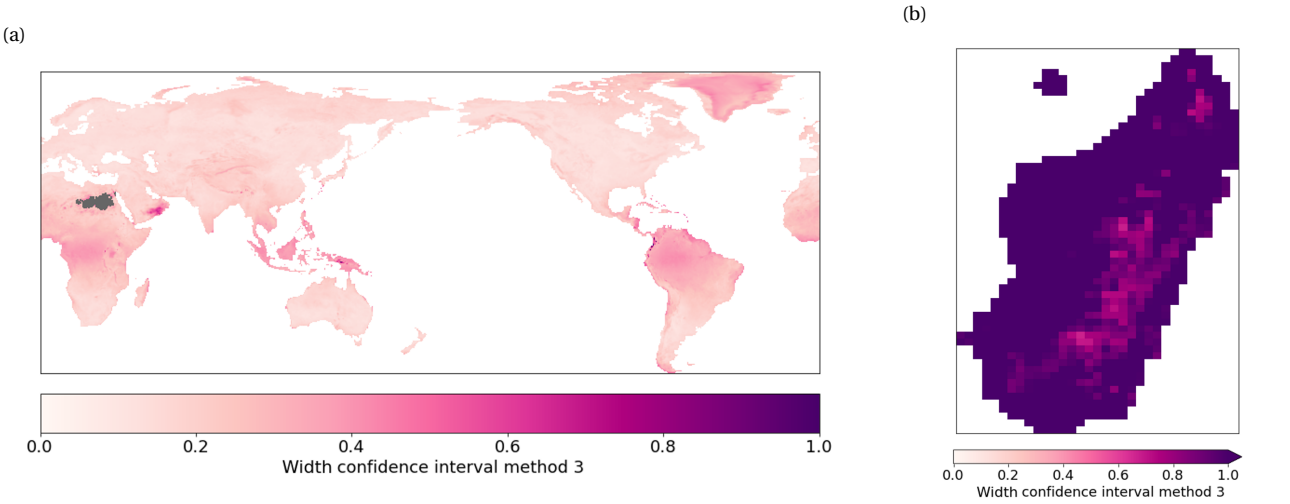

Figure 33: a) Confidence interval for the world. b) Confidence interval for region MDG, where only the data belonging to this region is used. Results belong to elimination method 3.

K. Signs of the correlation and regression coefficients

The sign differences between the correlation and regression coefficients for method 2 is shown in Figure 34.

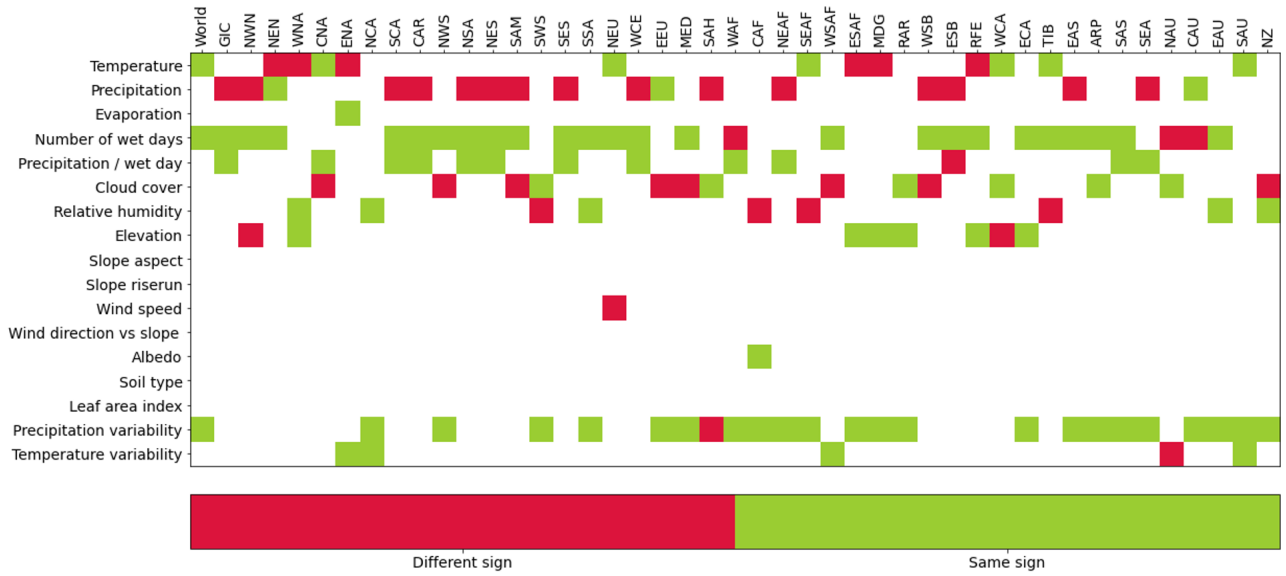

Figure 34: Difference in sign between the correlation and regression coefficients based on elimination method 2. If a grid cell is red, the signs are different. If it is green, the signs are the same. The location of the regions can be found in Figure 2.2.

The sign differences between the correlation and regression coefficients for method 3 is shown in Figure 35.

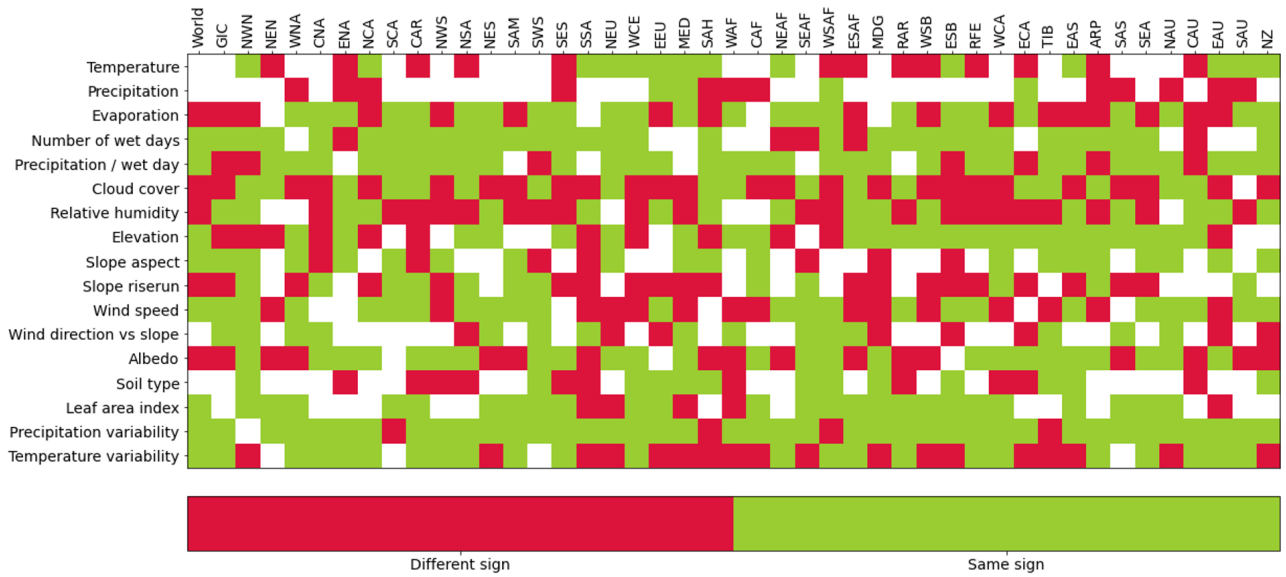

Figure 35: Difference in sign between the correlation and regression coefficients based on elimination method 3. If a grid cell is red, the signs are different. If it is green, the signs are the same. The location of the regions can be found in Figure 2.2.