UNIVERSITAT DE BARCELONA

Intensity-Duration-Frequency of Rainfall in Catalunya

Maximum Expected Precipitation and IDF Relationships at High Temporal and Spatial Resolution

Alba Llabrés Brustenga

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento 4.0. España de Creative Commons***.

This doctoral thesis is licensed under the ***Creative Commons Attribution 4.0. Spain License***.

Intensity - Duration - Frequency of Rainfall in Catalunya

Maximum Expected Precipitation and IDF Relationships at High Temporal and Spatial Resolution

PhD Thesis by Alba Llabrés Brustenga

UNIVERSITAT DE BARCELONA

Intensity - Duration - Frequency of Rainfall in Catalunya

Maximum Expected Precipitation and IDF Relationships at High Temporal and Spatial Resolution

Memòria presentada per optar al grau de doctor per la Universitat de Barcelona

Programa de doctorat en Física

Autor/a: Alba Llabrés Brustenga

Director/a: Raúl Rodríguez Solá
M. Carmen Casas Castillo

Tutor/a: Joan Bech

UNIVERSITAT DE BARCELONA

A day without rain is like
a day without sunshine.

Weathering - A. R. Ammons

Acknowledgements

This project would never have been possible without the program of Doctorats Industrials and the trust deposited in me when I was selected to participate as PhD student.

The good collaboration that always existed with Servei Meteorològic de Catalunya (SMC) was also of great importance to accomplish the task at hand. During the three years of collaboration and way afterwards, people at SMC have offered me their help and friendliness and I sincerely thank them for that. Especially my supervisor, Anna, I owe you for your patience and commitment.

I would also like to thank my advisers, Angel, for your excitement to start this project, that you transmitted fairly well to me, and for not letting go entirely after your retirement. And very especially to Raúl and M. Carmen for taking over the task of guiding me.

Finally, I would like to take the opportunity to thank my family. My parents for their endless support. My two brothers and sister-in-law for providing actual assistance or just much needed inspiration during this long journey. And last but not least, a recognition to my cousins for encouraging me week after week.

Alba Llabrés Brustenga
Barcelona, 22^{nd} September 2019

Contents

Introduction

Intensity - Duration - Frequency relationships of extreme precipitation events are of high interest in hydrology. The building of infrastructure, like urban drainage, or designation of flood zone, to cite two examples, need, as groundwork, the recurrence intervals at which extreme episodes of rainfall can occur, i.e., the maximum intensities that can be expected for a given duration and return period.

The work presented in this manuscript was done in the context of a Doctorats Industrials program. It is a program with financial support of Generalitat de Catalunya that encourages the collaboration between enterprises, universities and PhD students. In this case, the work was developed during three years in collaboration with Servei Meteorològic de Catalunya (SMC). Therefore, the objectives of the project are defined from a consensus between research and practicality. This collaboration is also disclosed in the database of the project, which is data provided exclusively by the SMC, and the tools and software that have been used as well as some of the followed policies, because tools or practices already in use in the SMC have been preferred when possible.

The manuscript is structured in three main parts. The first part (Chapters 1, 2 and 3) is dedicated to present available data and submit it to a strict quality control and homogeneity check. Quality control of the daily precipitation database was one of the requirements of the SMC collaboration as it is of high interest to define tools to tackle the quality control of historical data and the objective was to attain a knowledge of the global state of this massive database. Chapter 3 includes a selection of daily rainfall series that are reliable enough to proceed to the calculation of IDF relationships at high spatial and temporal resolution.

The second part of the manuscript (Chapters 4 and 5) is dedicated to present the methodology used for the obtainment of the IDF relationships and prepare the selected series of daily rainfall to obtain, first, the series of expected maximum rainfall in 24 hours, and, subsequently, the intensity - duration - frequency relationships through the use of a scaling exponent that is also determined from daily data. In this way, in Chapter 4, series of annual maximum rainfall in 24 hours are obtained from daily data after a correction with a multiplicative factor derived in the scope of the project. In Chapter 5 lays the groundwork of the temporal downscaling by obtaining the empirical scaling exponent needed to implement the monofractal downscaling methodology.

Finally, in Chapter 7, the results of IDF relationships are presented and discussed. The obtained results are provided as a set of maps that display the maximum expected

intensity at a selection of durations (daily and subdaily) and for a selection of return periods. The results are of high spatial resolution as they are obtained in a spatial grid of 1 km x 1 km. The results also have high temporal resolution as the intensity - frequency relationship can be calculated at any duration at the locations of the grid nodes using the provided empirical scaling parameter with validity as long as the duration ranges between 1 hour and few days.

To aid in the comprehension of the manuscript, it should be noticed that a list of acronyms is available at the end. Also that Fig[6.14] displays the names of the 42 counties in Catalonia and Fig[1.2] displays the main features of catalan topography; both counties and topographic structures are often mentioned in the text to locate the regions that are being discussed.

Chapter 1

AVAILABLE DATA IN CATALONIA

Daily rainfall data available in Catalonia are the foundation to achieve the main goals of the project. Therefore, available rainfall data have been carefully quality controlled and selected. In order to understand the origin of subsequently used data and the climatic characteristics of the region in which the project applies, the present chapter is dedicated to present the study area as well as the available dataset of daily rainfall data.

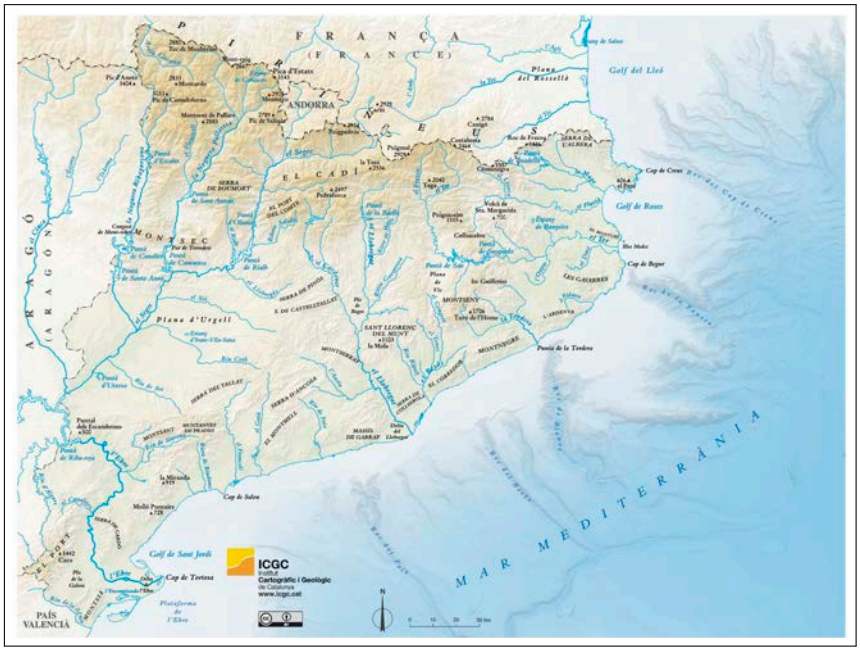
The spatial scope of the study is restricted to the territory of Catalonia because the project is funded by a program of the Government of Catalonia, called "Doctorats Industrials", and it is in collaboration with the SMC, which provides data and has interest in working with the results. However, some measuring sites just outside of Catalonia's borders are also included, in order to minimise boundary condition problems.

1.1 Study area and its climate

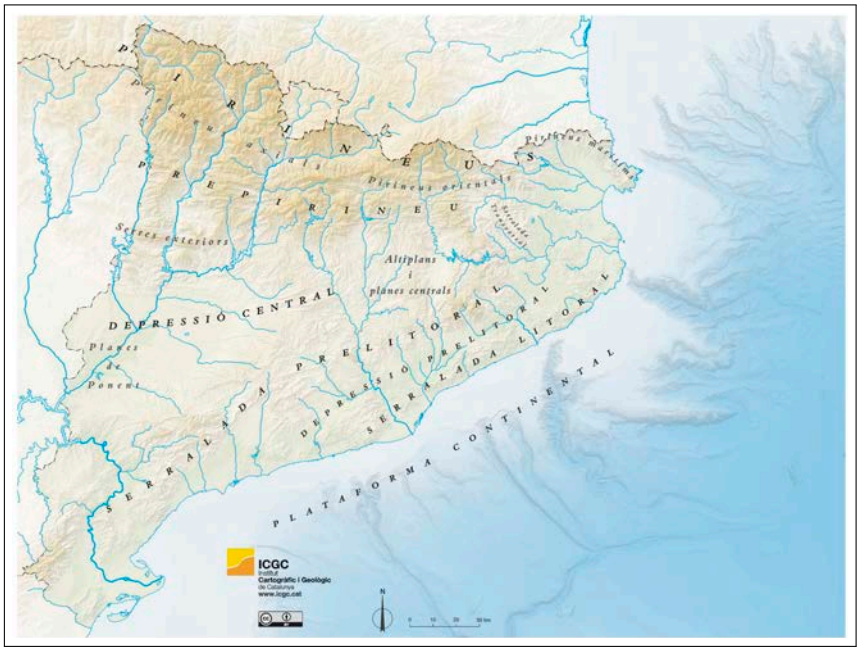
Catalonia ($32 \cdot 10^3$ km^2) is located in the north-eastern corner of the Iberian Peninsula, limited on the east by the Mediterranean Sea (along 580 km of coastline) and the Pyrenees on the north. Fig[1.1] shows the location of Catalonia in a map of Europe.

Figure 1.1: location of Catalonia in Europe (source: Instamaps ICGC).

At Catalonia's latitude (between 40°N and 43°N), confluence of cold polar and warm tropical air masses occurs. This particular location added to a complex relief gives rise to a large variety of climates. The orography of Catalonia is represented in Fig[1.2] which displays the main units of relief and rivers; altitudes over 2000 m a.s.l. are found at distances of less than 100 km from the sea. The most important features are the Pyrenees, the Central Depression, the Ebro Basin and the Coastal Ranges.



(a) Catalonia's topography in detail.

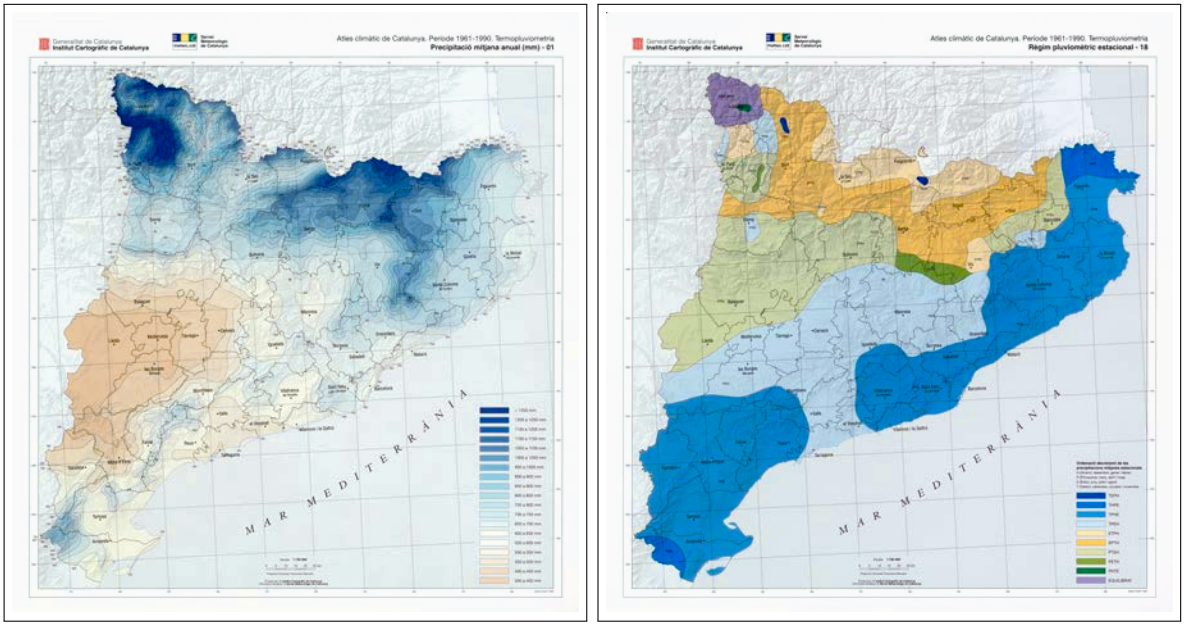


(b) main mountain ranges and plains in Catalonia.

Figure 1.2: digital elevation model of Catalonia with the main features of its topography. Source: ICGC.

Catalonia's climate is, in general, characterised by a very irregular rainfall pattern with mild winters and hot dry summers. The nearby presence of the Mediterranean Sea is a crucial driver of the region's climate, in particular, it smooths air temperature and can give rise to torrential rainfall in autumn, especially in the coastal areas. As distance from the coastline increases, diurnal as well as seasonal temperature variations tend to grow larger and precipitation amounts, in general, get reduced. In mountainous regions, temperature is lower and precipitation more abundant. In the eastern Pyrenees, summer is the rainiest season because of the importance of summer storms. The most north-western area, also in the Pyrenees, has characteristics of an Atlantic climate, i.e., a precipitation with higher amounts and a more regular pattern.

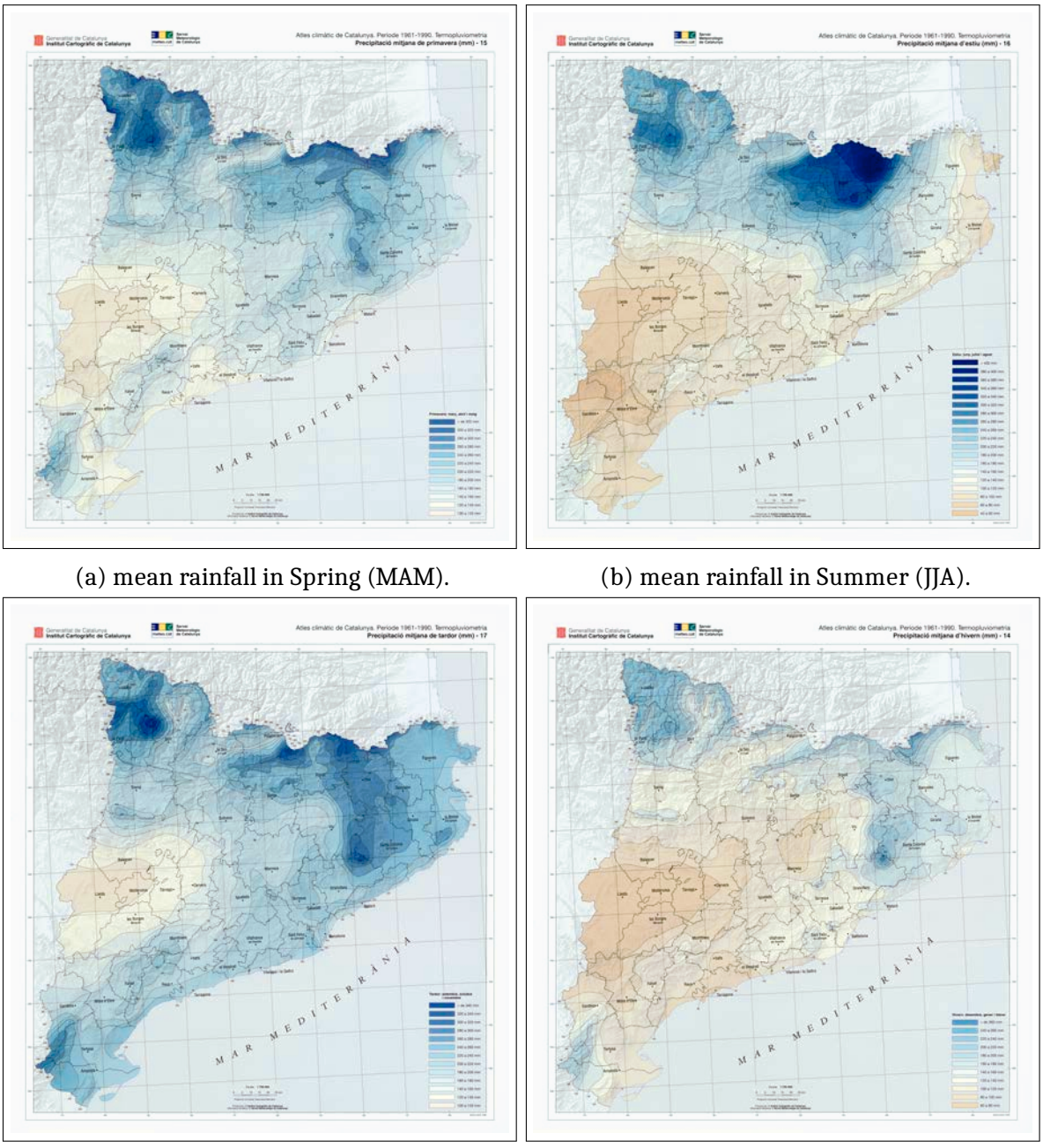
The Climatic Atlas of Catalonia (Martín Vide and Raso Nadal, 2008) is a publication of 32 sheets with temperature and precipitation maps. In Fig[1.3] two slides from the Climatic Atlas of Catalonia are displayed. The first slide (Fig[1.3a]) shows that mean annual accumulated amounts can be lower than 400 mm in the western area and over 1200 mm on the northern mountainous regions. The second slide displayed from the Climatic Atlas (Fig[1.3b]) shows the seasonal rainfall pattern distinguishing the order of the seasons in which the highest rainfall is collected; the rainiest season is most often autumn, but only in the region that is most influenced by the Mediterranean, whereas both spring and summer are the rainiest seasons in some areas. Winter is not a rainy season because the usual synoptic situation of westerlies brings Atlantic low pressure systems with a production of rain that is shielded in Catalonia by the high Iberian lands.



(a) mean annual rainfall (mm) in Catalonia. (b) rainiest climatological season pattern.

Figure 1.3: slides from Atles climàtic de Catalunya 1961-1990.
Source: Martín Vide and Raso Nadal (2008).

The seasonal pattern of the rainfall varies greatly in Catalonia's territory. The following slides from the Climatic Atlas of Catalonia (Martín Vide and Raso Nadal, 2008) show the spatial distribution of mean total amounts of precipitation in each season in the climatic period 1961-1990 (see Fig[1.4]).



(a) mean rainfall in Spring (MAM).

(b) mean rainfall in Summer (JJA).

(c) mean rainfall in Autumn (SON).

(d) mean rainfall in Winter (DJF).

Figure 1.4: seasonal rainfall climatological depths in the period 1961-1990.
Source: Martín Vide and Raso Nadal (2008).

In fact, the whole Iberian Peninsula is a complex region in meteorological terms because of its complex relief and geographical location (Esteban-Parra et al., 1998). That is, between mid-latitudes and subtropical latitudes; between two continents: Europe and Africa; influenced by two seas: Atlantic Ocean and Mediterranean Sea. Furthermore, the Iberian Peninsula is a territory where extreme weather events occur and, especially on the region of Mediterranean influence, both droughts and heavy rain events are usual (Llasat and Puigcerver, 1994), (Lana and Burgueño, 2000). Because of its complexity, rainfall is a meteorological variable of great interest in this territory (Saladié et al., 2007).

1.2 Data of the project

Rainfall data used in this project is provided by the SMC, although some measuring sites are managed by other official Meteorological Services. The initial dataset consists of 2,142 series, among them 1,726 are in Catalonia while the rest are near its borders (in France, Andorra or Spain).

The coordinate projection system used in this project is ETRS89/UTM in the reference zone 31N, which includes the whole study area.

Each available time series contains daily data measured at a specific location. The policy of the SMC demands that in the cases where a weather station is moved, even for a distance of half a kilometre, the name of the station (as well as its unique identification code) changes and data is stored in two different time series. These data come from official weather networks of manual measurements (like the Meteorological Network of Catalonia and the Balearic Islands created in 1894) or automatic weather stations (like the ones managed under the SAIH which in Catalonia started in 1996). In the case of manual measurements, daily collected data is read in the morning (between 7 a.m. and 8 a.m. depending on the protocol of the network) and stored as rainfall of the previous day. For this reason, for use in this project, data originated in automatic weather stations (that might store rainfall information hourly) are accumulated from 8 a.m. until 7:59 a.m. of the following day.

Data stored in the SMC's database comes from measuring sites which have been managed in the past by several different organisations. Nowadays, data from currently official meteorological services is periodically introduced in the database, i.e., measures managed by the SMC itself (which are introduced in real time) and measures managed by AEMet (which are introduced annually).

The network XEMA includes 209 automatic precipitation stations which have been carefully quality controlled under the management of the SMC; in particular, since 2008, data from this network is validated daily by technicians from the SMC, in the period 2008-2016 there are 189 stations which underwent a strict quality control, among them 171 are operational at least until 2016. A number of 150 stations of the network XEMA have been drawn from those which have data after 2008 and cover the study area evenly. This selection of 150 series of confidently good quality in the period 2008-2016 has been

used to study the common differences in daily precipitation at different spatial points (establishing a threshold for the automatic quality control) and to perform a verification of the automatic quality control procedure explained in the next chapter. A different selection of 120 automatic weather stations has been used to obtain and study the empirical correction factor to transform one-day maxima into 24 hours maxima (see Chapter 4). The selection of 120 series is different from the previous selection because a long period of continuous measures is needed in this case; the correction factor is calculated from series with a minimum of 15 years of data in the period 1988-2016 instead of strictly quality controlled series in the period 2008-2016. A different set of 105 AWS with at least 15 years of hourly data is used to construct IDF relationships at high temporal resolution in order to validate the results of the temporal downscaling from daily to hourly maximum expected intensity; these 105 series were selected later in the time progress of the project and cover the period 1990-2017.

1.2.1 Spatial distribution of daily rainfall data

The spatial distribution of the available series is shown in Fig[1.5]. The spatial density of available series is extremely high. Specifically, a nearest neighbour analysis yields an observed mean distance of 1.5 km and nearest neighbour index of 0.46 indicating randomness (which would be the index 1) but with a tendency to clustering (index 0); light tendency to clustering is reasonably expectable because measuring sites tend to be located in places of interest, e.g. near villages and accessible.

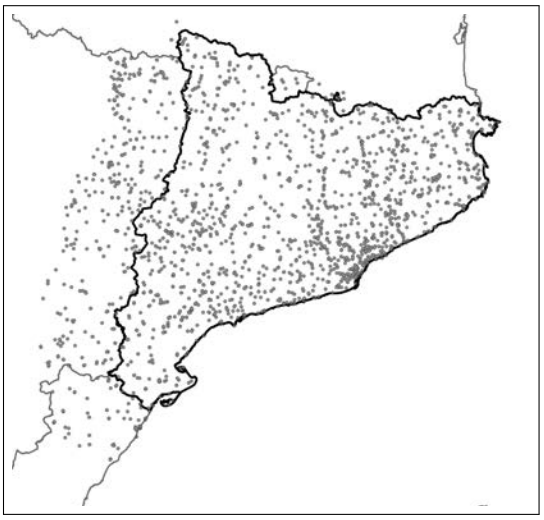


Figure 1.5: spatial distribution of available daily rainfall measuring sites.

The series displayed, however, are the whole set of available series at any time, i.e., not all of them actually have data at the same time. In fact, most of the series are extremely short (50% of them have less than 15 years), hence, not useful for climatic studies. There are a few series, though, which are extremely long (10% have more than 50 years and some of them cover over a century). Regarding completeness of daily data, the majority of series have a very good percentage of available data during their measuring period (50% of the series have over 95% of days with measures).

An important feature of the available set of series is that by choosing a minimum length of the time series, the territory can still be covered in a uniform way (with the drawback of a consequent loss in spatial density). Indeed, choosing a minimum length of 30 years, the remaining time series cover the territory with observed mean distance of 4.8 km and nearest neighbour index of 0.69 (such subset increases the randomness of the spatial

distribution of the measuring points because it filters most of the short series that were relocated in the same areas).

1.2.2 Historic evolution of the number of daily rainfall series

On the analysis of available data, it is also interesting to observe the number of series available per year as well as the evolution of this number in time. To this end, Fig[1.6] displays the number of available measuring sites through the years making a distinction according to the minimum completeness of the series in the given year.

Even if the instrumental history of Catalonia starts in the late 18th century, only sparse measures appear since 1789 until 1855 and they have not been considered for lack of continuity, nor are they digitised in the database of the SMC. On the recent delimiter of the study period, 2016 was the latest whole year that was considered in the present project; it is a project that started in 2015 and the final selection of series and the subsequent study was performed during 2017. Therefore, the considered period of available rainfall data is 1855-2016. Nonetheless, a narrower period needs to be selected in order to have a required minimum of simultaneously available series.

The oldest currently available daily rainfall series is *(BN037) – Universitat de Barcelona – C. del Carme* which began its observations in 1855 in the city of Barcelona; the measuring site was located in the ancient building of the University of Barcelona in the centric street "Carrer del Carme".

Since 1895 more than 10 measuring sites are available simultaneously forming a network instead of isolated measuring sites. The Meteorological Network of Catalonia and the Balearic Islands started in 1894; it was created by the Granja Experimental de la Diputació de Barcelona to fill a need of the agricultural sector (Anduaga, 2012).

During the 1910s, the number of available measuring sites keeps increasing to reach around a hundred series, managed under the Catalan Pluviometric Network, thanks to the work done by the Astronomical Society of Barcelona (1910-1023) (Prohom Duran, 2006) in collaboration with the Central Meteorological Observatory of Madrid.

It is also during the decade of 1910 when weather stations that will keep operative for over a century start taking measures. One example of these stations is *(BB017) – Observatori de l'Ebre* which has data for over 110 years (since January 1905) with only one period of interruption of measures, between April 4th and December 20th 1938, during the Spanish Civil War (actually, the dates of inactivity tightly include the period in which the Battle of Ebro took place). It is the measuring site that has been measuring daily rainfall in a more continuous way and for a longer time in Catalonia. Actually, another available station, *(BA053) – Abadia de Montserrat*, currently operative as well, started measuring before, in October 1st 1901, but has more periods of missing data, giving a completeness of only 68% next to the 99% of *BB017*. Other weather stations currently operative that have been measuring with a high continuity and for over a century are: *(AE052) – Cadaqués, (AP024) – Pantà de Foix, (MI006) – Moià-EEPP* and *(BN007) – Observatori Fabra*. The case of Observatori Fabra is remarkable given the fact that its daily

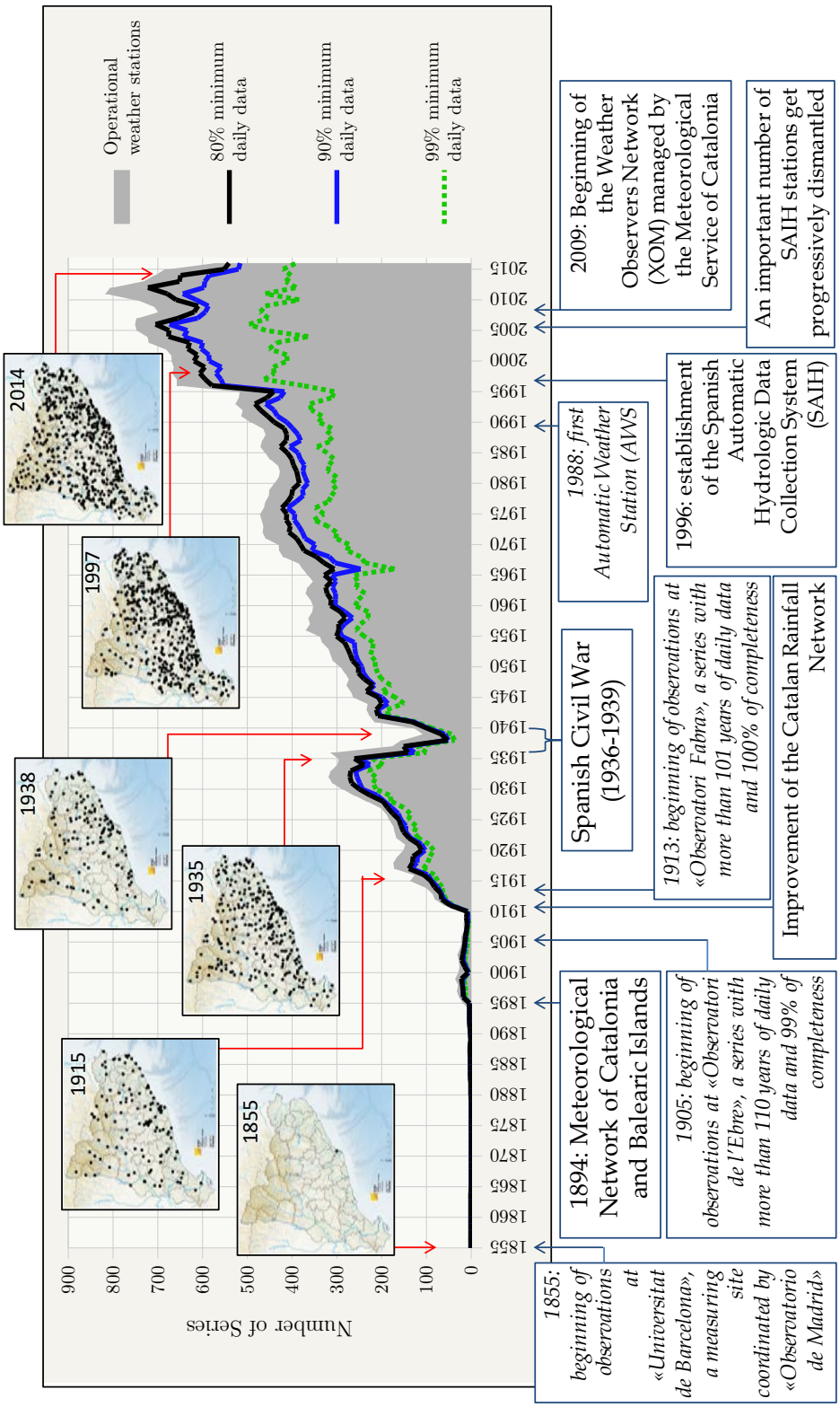


Figure 1.6: temporal evolution of available daily rainfall series, lines show a distinction depending on completeness, the most relevant features of the timeline are pointed out.

rainfall series is continuous (with absolute completeness) since July 1st 1913.

From the 1920s, the number of weather stations increases continuously until 1936 when, due to the Spanish Civil War (1936-1939), the number of available series drops severely (up to the point of finding in 1939 under a 25% of the mean number of series available during the first half of the 1930s, regarding 80% completeness).

The years after the Civil War are a period of recovery for the number of available weather stations, even if, after the initial and steep growth of the first three years, the increase is low and the numbers previous to the war are not reached again until the 1950s.

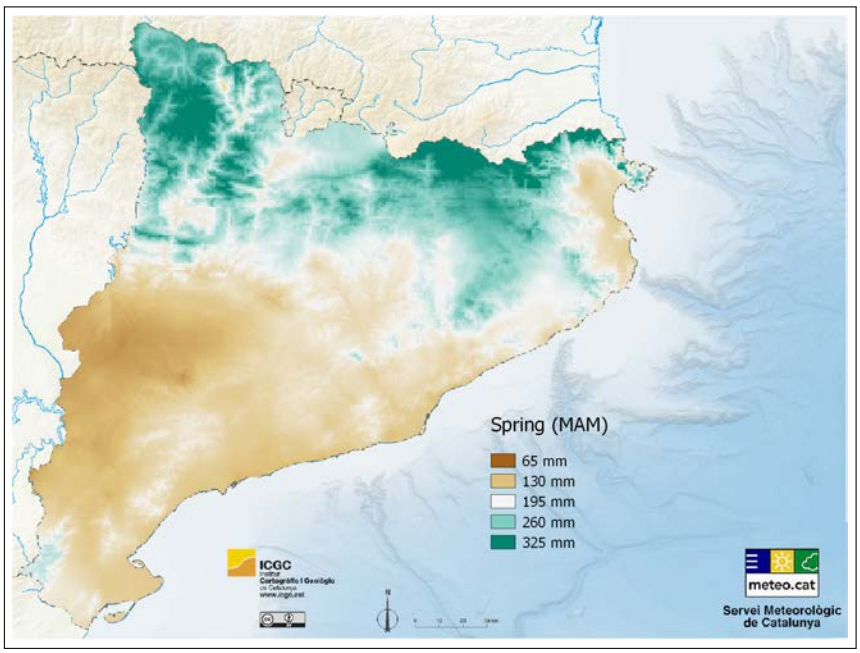
In recent years, a higher number of weather stations are available, after they drastically increased in 1996 with the establishment of SAIH. About ten years later, around 2006, an important part of those SAIH stations got progressively dismantled, but the global number keeps steady thanks to the stations of the XEMA Network (the first one of them started in 1988) and the beginning of XOM (which corresponds to the most recent increase, i.e., in 2009); both XEMA and XOM are networks of, respectively, automatic and manual weather stations, which are currently managed by the SMC.

From 1995 and on, a clear distinction is observed in the number of available series depending on completeness. This fact, which implies a considerable reduction of the more complete series with respect to the total, is a consequence of the increasing number of available measuring sites and the fact that automatic stations at their beginning could have more problems while manual stations had been fading between 2000 and 2010 due to the automation success.

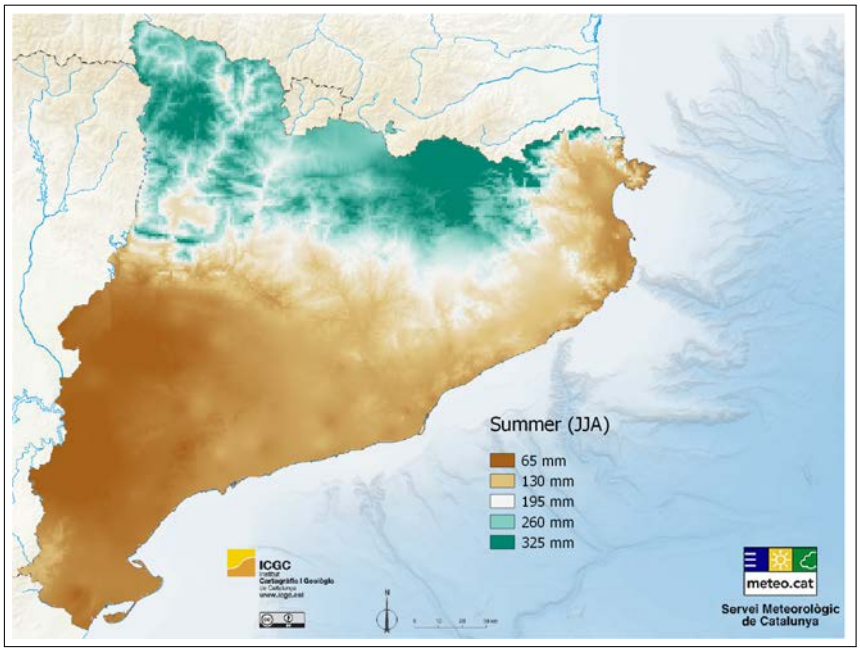
In the analysis of the temporal evolution of daily rainfall available series, apart from the general upward trend, two irregularities in the number of series are noticeable (in 1966 and 2004) with a more important reduction in the number of complete series rather than in the total. The dip in 2004 is probably caused by a change in the communication system of the organisation managing the network at the time (ACA) whereas the causes of the dip in 1966 are unknown.

1.3 Climatology from available data

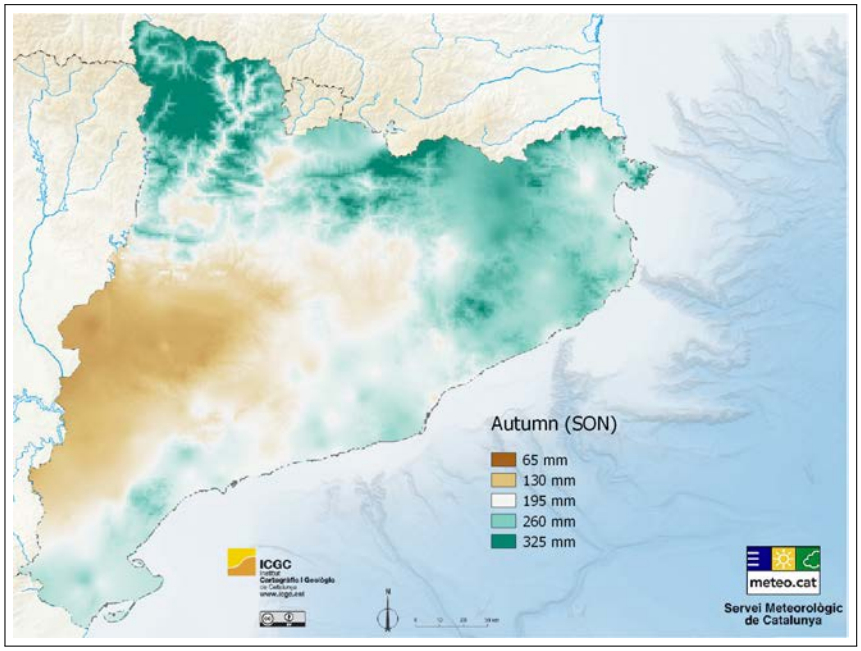
From the available series of daily rainfall which have been operational for at least 20 years and have a maximum of 10% of missing data, climatic rainfall seasonal totals are calculated and displayed in Fig[1.7]. The maps have been obtained with an interpolation technique which is a geographically weighted regression using altitude as an independent variable and kriging to correct residuals. The interpolated precipitation follows the known climatic patterns of the region (compare with Fig[1.4]).



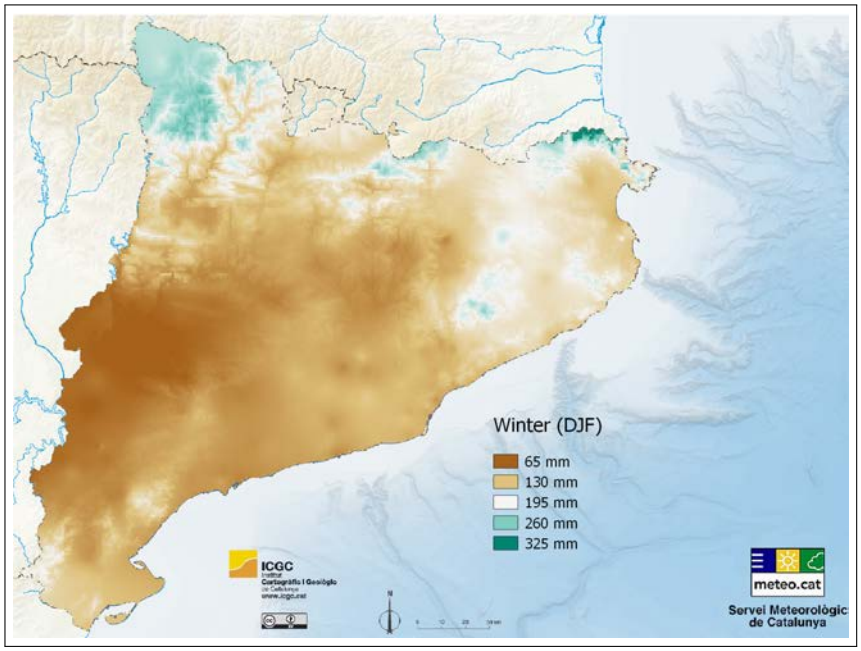
(a) mean rainfall in Spring (MAM).



(b) mean rainfall in Summer (JJA).



(c) mean rainfall in Autumn (SON).



(d) mean rainfall in Winter (DJF).

Figure 1.7: seasonal rainfall climatological depths from series with a minimum length of 20 years of daily data in the period 1855-2016.

Chapter 2

QUALITY CONTROL

The present project uses daily rainfall data provided by the SMC (see previous chapter). The SMC has been working towards collecting and storing all weather data that have been generated in Catalonia during its instrumental history (including the considered period in this project, 1855-2016). During this time period, available weather data have been managed by many different organisations, especially during the 20^{th} century. Consequently, the initial dataset is a mixture of different measuring methodologies (including manual and automatic rain gauges), quality control procedures (if any were applied) and metadata coding.

From 2,142 available measuring sites, a selection of daily rainfall series have been drawn to be used in the present project. The quality control, however, has been applied to the whole dataset. An important objective of the present project, encouraged by the collaboration with the SMC, is to approach a global vision of the database and apply uniform quality control procedures to the whole set of available data in order to single out suitable series for specific future studies as well as organise a more in-depth validation of stored data. In the project, the quality control is needed, as in any study, to avoid introducing non climatic signals to the results. In this particular case, the heterogeneity in the origin of data, the high number of available series and the need to ensure a good quality at daily resolution difficult the application of a quality control procedure.

2.1 Approaches to quality control

The first part of the performed approach is to review quality control procedures for rainfall data found in the literature and be aware of the expected errors that could be found in a daily dataset.

The World Meteorological Organization claims the importance of quality controlling meteorological data and encourages organisations entrusted with data collection and storage to apply such procedures following some guidelines (World Meteorological Organization, 2008).

Nonetheless, there is not an established quality control procedure for precipitation data (Branisavljević et al., 2009). Indeed, several authors as well as different meteorological and climatological organisations use different methods according to the needs and characteristics of their data. A selection of remarkable methods follows, they are commented as examples of different methodologies and the list should not be regarded as exhaustive.

Shearman (1975) presented a quality control routine developed and applied in the Met Office. It was based on an interpolation from surrounding observations that enabled comparison with the value that needed to be checked and could perform a reconstruction in case of accumulations, shifted day or erroneous time measurements.

Feng et al. (2004) used a series of checks in order to flag erroneous data and remove it before reconstruction of missing values and homogeneity controls. The outcome was the building of a daily dataset for 1951-2000 in China; the quality control checks extreme limits, internal consistency, temporal outliers and spatial outliers.

You et al. (2007) explored the performance of three different tests for precipitation based on empirical statistical distributions underlying the observations and decided to adopt the Multiple Intervals Gamma Distribution (MIGD) method for their data.

Chen et al. (2008) described an automated and objective protocol established at NOAA's Climate Prediction Center to quality control, on real time, daily information of gauges in the Global Telecommunications System. The procedure compares historical gauge records, contemporary observations at nearby stations, satellite estimates and numerical forecasts.

Sciuto et al. (2009) developed a procedure to automatically classify daily rainfall data as "validated" or "not validated" before a further manual check. This procedure consists in comparing observational data at one station with data at reference stations making the validation by means of a neural network previously trained with historical data which had been already validated.

Nie et al. (2012) applied several steps of previously established quality control procedures, namely, outlier filtering, duplication station check, internal consistency check, extreme value check, temporal consistency check and spatial consistency check.

Serrano-Notivoli et al. (2017a) developed an automated quality control for daily precipitation data which compares each daily value with a set of two predictions obtained from reference series (one of them is a binomial prediction showing the probability of dry or wet day; the other one is a magnitude prediction that estimates the amount of precipitation for the given day).

Even if there is not an established quality control procedure, most quality checks start by cleaning big errors before facing the doubtful or more difficult cases (as mentioned by Einfalt et al. (2008) it is an approach that yields good results) in order to gain increasing quality. Regardless of the methodology, it should be noticed that the dataset can never be completely free from errors and that rainfall variability plays an important role in the trickiness of quality control procedures, especially in the case of daily data.

2.1.1 Examples in the Iberian Peninsula

Rainfall's nature of spatial and temporal high variability is an added difficulty to the quality control of this variable. In Spain, nonetheless, some endeavours have led to quality controlled databases of rainfall at a daily scale, with enough spatial density and temporal length to be used in climate studies, some of them are also homogenised and others have a gridded format.

Romero et al. (1998) presented a study using a 30 year database of daily rainfall data (for the period 1964-1993) for the Mediterranean regions of Spain. The database was created by selecting 410 stations of the study area (with an average distance of 15 km between them) from a number of 3366 available measuring sites. The series included in the database were selected with a criterion of 90% completeness and a quality control procedure developed by the authors was performed by means of an iterative estimation of the considered value as an interpolation from reference stations. Missing values were filled using the same interpolation method. In this way, the first objective of the authors was fulfilled, namely, obtaining a complete daily rainfall database with the appropriate density to effectively capture spatial variability in the region. The second objective was using the aforementioned database to give a pluviometric characterisation of the region. The authors presented such characterisation as several yearly and seasonal products (distinguishing by decades in order to study possible trends). The obtained products are mean precipitation, number of days with one-day rainfall above a threshold (set to 1 mm to quantify the number of rainy days and to higher quantities to quantify the occurrence of extreme events), mean rainfall per rainy day, mean duration of wet episodes and mean duration of dry episodes, moreover, recurrence intervals were calculated for each decade.

Lana et al. (2004) selected a daily precipitation database for the period 1950-2000 for a detailed analysis of the daily pluviometric regime in Catalonia. The database consists of 75 rain gauges, but not all of them cover the period 1950-2000; 85% of the selected 75 series, have an effective recording period of at least 31 years while the rest of the series have a slightly shorter period. Homogeneity was checked by means of the Von Neumann ratio test applied to monthly totals.

Vicente-Serrano et al. (2010) proposed a model for constructing databases of climate data and developed a database of 828 daily rainfall series for north-east Spain, with varying coverage, for the period 1901-2002. There were 3,106 initial observatories which were combined (in nearby cases) to provide long temporal coverage. Also, gaps were filled and a quality control of daily data was applied (rejecting 0.1% of the values and replacing them with information from the nearest neighbour in case it was at a maximum distance of 15 km and had a Pearson's correlation coefficient over 0.5). Finally, the homogeneity of the series was tested using the AnClim software (Stepanek, 2003) which applies the Standard Normal Homogeneity Test (SNHT) developed by Alexandersson (1986). The authors applied the homogeneity tests successively, rejecting series or periods detected as inhomogeneous, to monthly precipitation amount, monthly average number of rainy days above 1 mm, monthly maximum precipitation and number of days

above the 0.995 percentile.

Herrera et al. (2012) developed a daily precipitation gridded dataset, known as Spain-02, which is publicly available. The dataset has a regular horizontal resolution of 0.2° (approximately 20 km) and covers the period 1950-2003. It was obtained from 2756 series of daily precipitation records that cover peninsular Spain and the Balearic Islands, after selection among a much larger set of available stations; selection criteria involved a minimum of 20 years of data in the final period and good performance in SNHT homogeneity tests. The mean distance between neighbouring selected stations was about 7 km. The grid was finally obtained from a kriging method in a two-step process, after analysis of the most suitable interpolation method. The authors were concerned about the suitability of a gridded dataset to perform studies on extreme events; they analysed the capture of spatial and temporal evolution of extreme events by the developed gridded dataset and concluded that its high quality and resolution made it appropriate for such studies. They caution, however, the use of Spain02 to analyse trends, recommending a selection of long time series instead.

Serrano-Notivoli et al. (2017b) presented a publicly available gridded dataset of daily precipitation in Spain. The dataset was built from 12858 observatories covering the period 1950-2012 in peninsular Spain and 1971-2012 in the Balearic and Canary islands. Raw data was quality controlled and gaps were filled using reference values for each day and location. The authors computed reference values by means of a binomial prediction (wet or dry day) and a magnitude prediction, using generalised linear models, from the 10 nearest neighbours. The final gridded dataset has a 5 km x 5 km spatial resolution and was developed by estimating daily precipitation amounts and their uncertainty at each grid node.

2.1.2 Possible errors in the dataset of the project

Large datasets can contain errors difficult to spot due to the impossibility of manual checking. Nowadays, the SMC is focusing efforts in daily validating new coming in data. However, the validation process becomes massive when looking back at long series.

Errors found in historical records are most often caused by digitalisation mistakes, including misreading of original handwritten data, typographical errors and codification errors (i.e., misusing numbers meant to code no-data or other circumstances).

A common error in a daily rainfall dataset is to have missing values filled with zeros. This can either be a consequence of mistakes in data treatment or a failure at reporting rainfall accumulations of several days. A study of the extension and implication of these errors in Australia can be found at Viney and Bates (2004). Fig[2.1] shows an example of data that reached the project's database as daily rainfall values while measures were actually taken weekly.

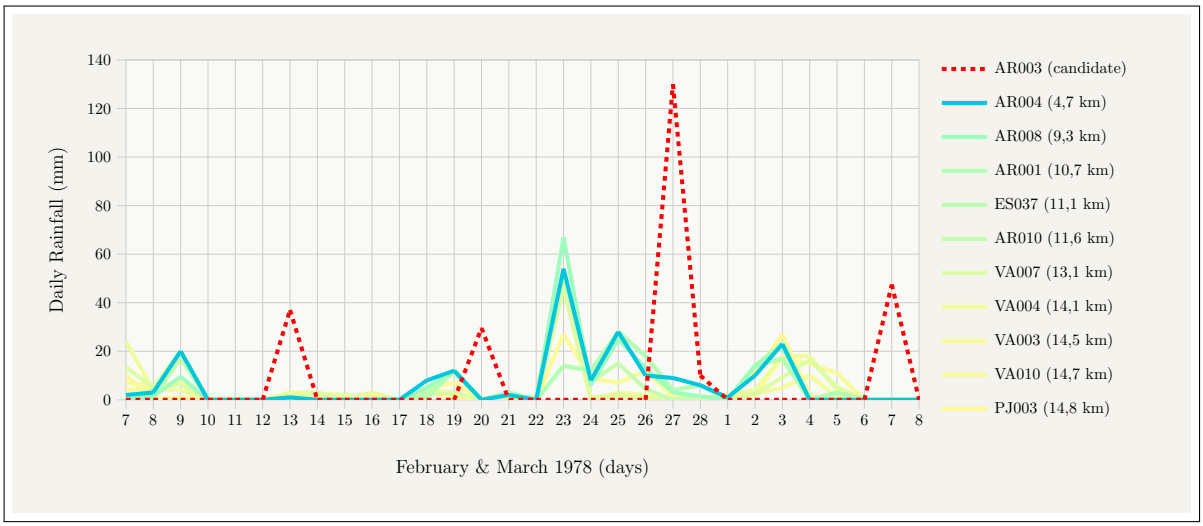
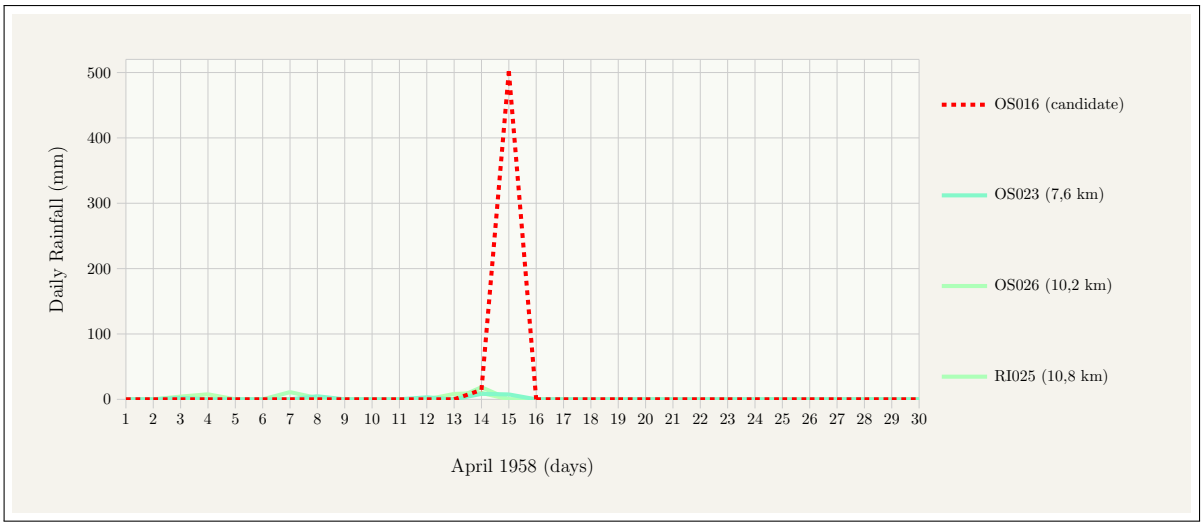
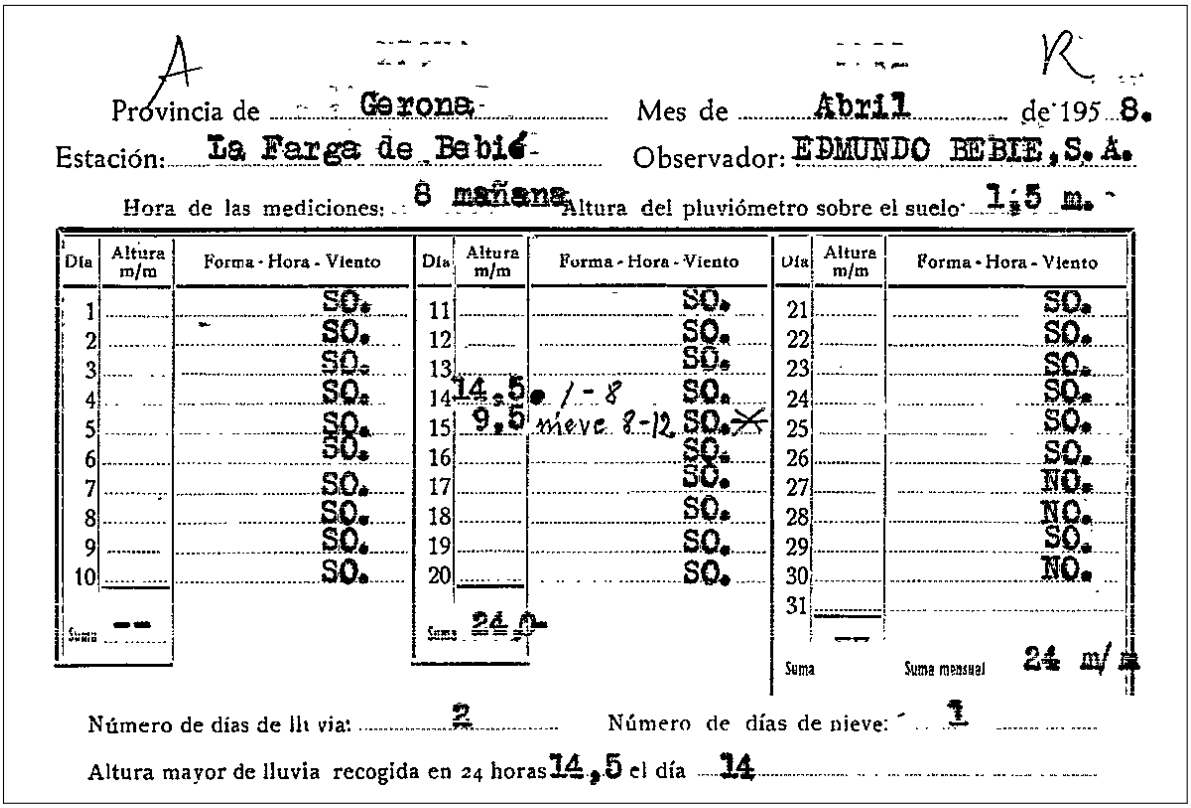


Figure 2.1: example of untagged weekly values.

It is widespread in quality controls to check for outliers which sometimes are removed from the database as a correction method. However, outliers, understood as values widely separated from the main cluster of points in the sample, are statistically present in time series. On the contrary, extremely large values present in a rainfall database are usually digitalisation mistakes or unreported accumulations of several days. Fig[2.2] shows an example of a digitalisation mistake, the value of 502 mm was found during an outlier check and the fact that nearby stations did not present large values made it a suspect value. After checking this case with the original handwritten sheets, it was replaced to its correct value: 9.5 mm (see Fig[2.2b]).



(a) erroneous extremely large value.



Provincia de Gerona — Mes de Abril — de 195 8.

Estación: La Farga de Bebié — Observador: EDMUNDO BEBIE, S. A.

Hora de las mediciones: 8 mañana — Altura del pluviómetro sobre el suelo 1,5 m.

Día	Altura m/m	Forma - Hora - Viento
1		SO.
2		SO.
3		SO.
4		SO.
5		SO.
6		SO.
7		SO.
8		SO.
9		SO.
10		SO.
Suma	--	

Día	Altura m/m	Forma - Hora - Viento
11		SO.
12		SO.
13		SO.
14	14,5	SO.
15	9,5 nieve 8-12	SO.
16		SO.
17		SO.
18		SO.
19		SO.
20		SO.
Suma	24,0	

Día	Altura m/m	Forma - Hora - Viento
21		SO.
22		SO.
23		SO.
24		SO.
25		SO.
26		SO.
27		NO.
28		NO.
29		SO.
30		NO.
31		
Suma	--	

Suma mensual 24 m/m

Número de días de lluvia: 2 — Número de días de nieve: 1

Altura mayor de lluvia recogida en 24 horas 14,5 el día 14

(b) correct value in the original measurement sheet.

Figure 2.2: example of erroneous extremely large value found in the database.

The repetition of identical values is an error that can arise, although rarely, from digitalisation. Another source of this type of error is metadata codification as some protocols used to advise a repetition of the mean value to encode accumulations of several days (a practice which is no longer followed in the SMC database because it can easily lead to confusion). Fig[2.3] shows an example of an identically repeated value for 4 days.

The problem of the shifted day is a common source of errors in daily rainfall data basically caused by the need for a convention regarding the date at which rainfall should be assigned when it is measured in the morning of one day but it has been collected during the previous 24 hours (which mainly correspond to the previous day). On a visual inspection of data, shifted days are easily identified (see Fig[2.4]) but to detect them automatically is complicated.

Other problems encountered in the project's database were inconsistencies between replicated data in the SMC's database (i.e., data that had reached the database by two different ways but had been initially measured with a unique origin) and missing data caused by non introduction of available measures in the database (e.g., 40 series had missing data in the month of December 2012 because data provided annually by AEMet had been introduced before the end of the year in those cases).

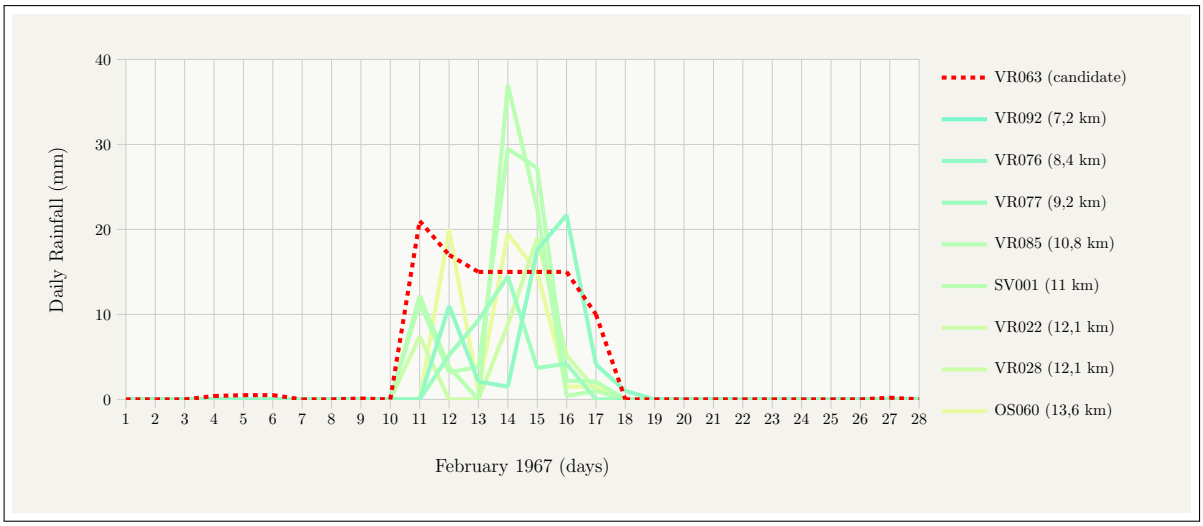


Figure 2.3: example of repeated same value during 4 consecutive days.

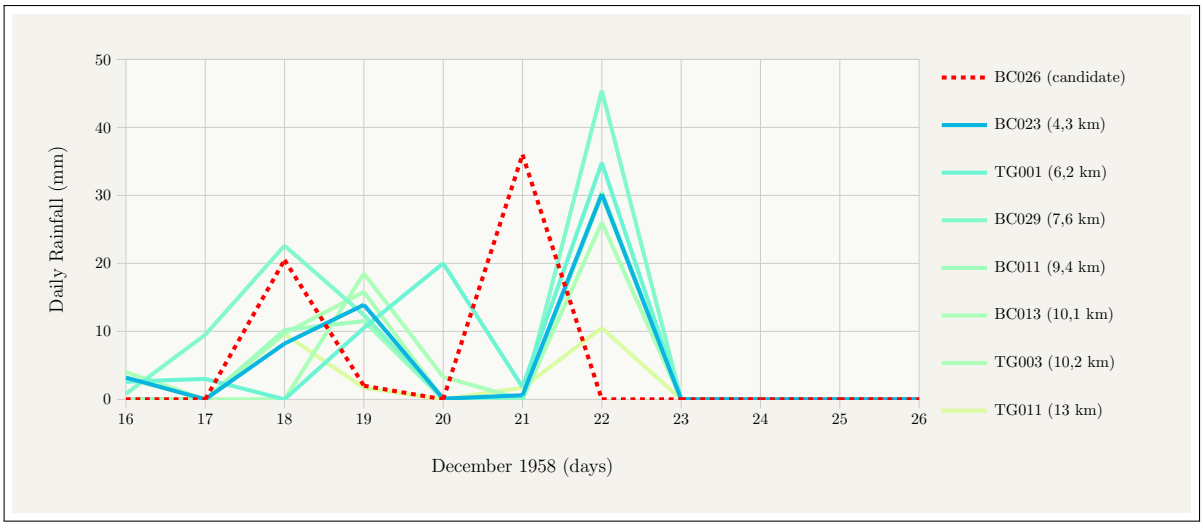


Figure 2.4: example of a station with shifted values, clearly identified against data of nearby stations.

2.2 Applied quality control procedure

In this section an automatic quality control procedure is presented. It has been specially designed for the needs of the database used in the project and it has been applied to the whole set of available rainfall daily data from the SMC. The main problem of the quality control of precipitation, especially for daily data, is a great inherent variability, both spatial and temporal. The procedure that has been developed consists in several successive phases aimed at labelling each series according to its quality and progressively increasing the detail of error detection. On the one hand, doubtful cases are flagged enabling the selection of a dataset with the desired quality. On the other hand, data modification is done at some steps but only when the correct value from the original source can be checked and removal is allowed in clear cases of error (e.g. physically impossible data).

The quality control procedure is mainly automatic, although a manual revision is recommended at some point. It follows different phases which are shown in the scheme of Fig[2.5].

Phase 0 is a necessary initial inventory of available data, it consists in gathering the data which is desired to undergo the quality control, making sure that all measures are comparable (i.e., same time interval and units, in our case, daily precipitation data measured in the morning in millimetres) and identifying each station with its unique coordinates (especially when data from different origins is assimilated, as it is our case, it is important to check if the coordinate system is the same for all the stations); at this point, it is recommended to display available series in a map and identify its spatial density and spots where relative comparison could be difficult.

Phase 1 consists in a basic quality control which flags data which is physically impossible or doubtful; after the automatic check, a manual validation is recommended in order to be aware of doubtful cases and try to find the cause of clear errors (digitalisation errors are the most common at this phase and checking original data could restore the value instead of eventually losing data).

Phase 2 is an automatic labelling of each series according to an annual quality index which has been designed to account for potential errors regarding absolute data of the series (i.e., only data of the individual station, without comparison with other series).

Finally, a relative comparison (i.e., each series with a selection of reference series) is carried out in phase 3 after selecting the best reference series according to the suitability of their location and correlation of data with the the candidate series.

After the quality control is performed, the results will be interpreted with aid from the verification process which has also been carried out in this study.

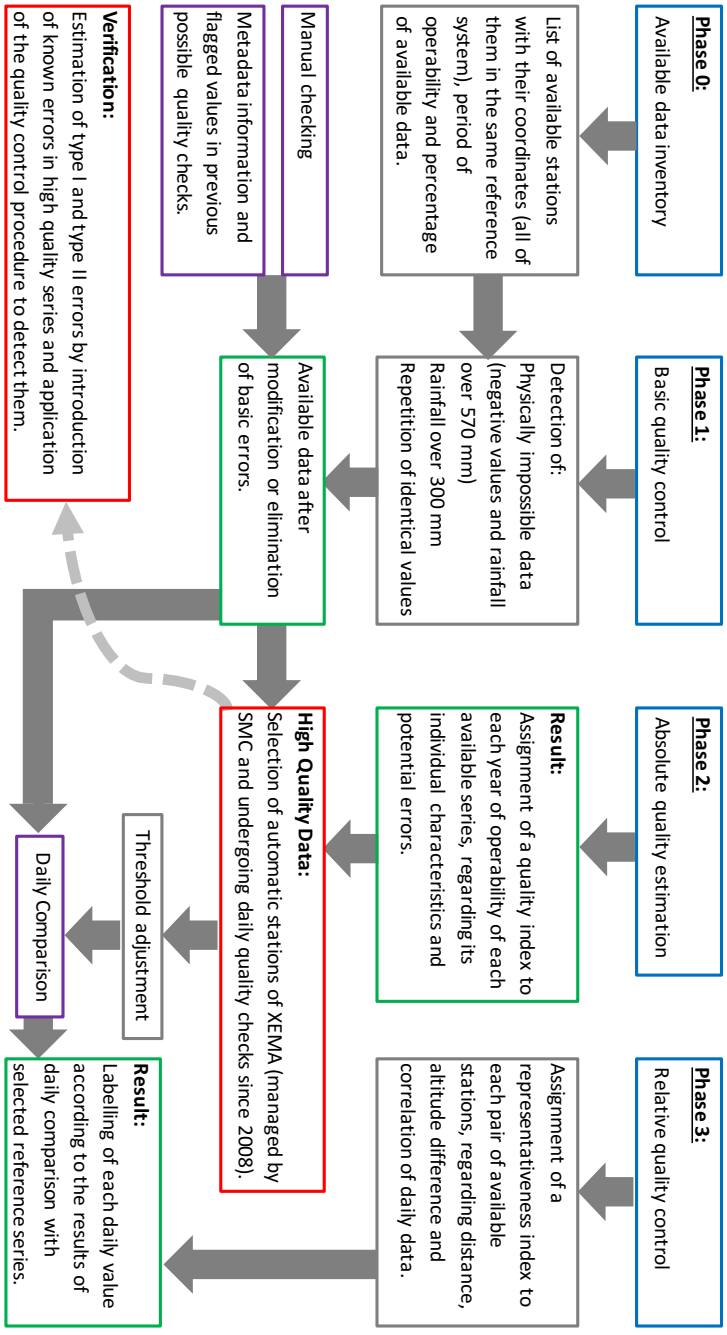


Figure 2.5: scheme of the followed quality control procedure.

2.2.1 Basic quality control

First of all, it is necessary to be sure that the available data are physically possible, in the case of daily precipitation, it is necessary to verify that data do not contain negative values or extremely large values (above 567 mm). The upper threshold for physically impossible values has been set taking into account the climatology of the study area. The maximum value of the probable maximum precipitation in 24 hours determined in Catalonia by Casas et al. (2008) was 567 mm, being 430 mm the highest daily rainfall ever recorded in Catalonia, collected the 13th October 1986 in Cadaqués (in the coast at the northeast) during a flooding episode, comprehensively studied, known as "Aiguat de Sant Eduard" Vigneau (1987). Thus, values above 567 mm incorporated in the database might be errors from the digitalisation process, values intended to encode metadata and erroneously taken as precipitation values or undetected accumulations.

Within the basic controls, it is also recommended to review cases of daily precipitation exceeding a certain threshold (with a compromise between the relevance of the values and the number of cases that exceed it), analysing each case with help of graphic representations and consulting, if possible, the original values or the synoptic situation (from re-analysis). In this project, the threshold has been set to 300 mm.

Finally, the last of the basic controls is to detect identical values repeated for two or more consecutive days, using a filter based on the number of consecutive days and the repeated value, in order to manually review the most serious cases (the highest repeated values or the longer spells, see Table[2.1]). Many of these cases are codification errors because of common old procedure to codify accumulations of several days was to write down the equally distributed value to each one of the days. In these cases, the total precipitation amount can still be used to calculate monthly totals, but it leads to incorrect daily values, hence, it must be removed from the daily dataset.

Precipitation value	Minimum number of days in the spell
Above 10 mm	2
Above 1 mm	8
Above 0 mm	11

Table 2.1: threshold used to manually check repeated identical values during several consecutive days.

2.2.2 Absolute quality control

The absolute quality control consists in analysing the series individually and labelling them, by means of an automatic process, according to their probable quality. In order to

assess the quality of the series, a quality index has been designed, considering the possible problems that are common in daily precipitation series. The quality index, Q, takes values between 0 and 100 to indicate the absolute quality of the series: values above 80 indicate acceptable quality and below 50 very low quality. Q is calculated according to Eq[2.1] and each station has an assigned value for every year of its sample.

$$Q = \frac{1}{5}(P + Q_{gaps} + {Q_{zero}}^{m} + {Q_{zero}}^{w} + Q_{outliers}) \tag{2.1}$$

The terms of Eq[2.1] are the following:

- P: percentage of annual data, calculated as the number n of available daily data divided by 365 (or 366 for leap years).

- Q_{gaps}: index that takes into account the distribution of gaps, penalising mostly hollow interspersed with data, more than cases of unique periods without measurement. If n_{gap} is the number of periods with missing data, L_{gaps}^{max} the length (in days) of the maximum spell with missing data and n the number of days of the year in which the series was operational, Q_{gaps} is calculated as in Eq[2.2].

$$Q_{gaps} = 100 - 100\frac{2n_{gap} + L_{gap}^{max}}{n} \tag{2.2}$$

- ${Q_{zero}}^{m}$: percentage of months with accumulated precipitation over zero with respect to the number of complete months during the given year. It aims to penalise series measuring null precipitation for many months, as it is very likely that, in fact, it corresponds to periods without measurement that have been filled with zeros (false zeros). If m_0 is the number of complete months with total null precipitation and the number of complete months of the considered year of the given series, ${Q_{zero}}^{m}$ is calculated as in Eq[2.3]:

$${Q_{zero}}^{m} = 100 - 100\frac{m_0}{m} \tag{2.3}$$

- ${Q_{zero}}^{w}$: index that indicates the probability of having systematic cases of accumulations of two or more days which had not been properly reported. It is based on the fact that the number of precipitation days in a given year should be independent of the day of the week. In the case of detecting a day of the week with rainfall much less often than the rest throughout the year, the value of this index decreases. However, ${Q_{zero}}^{w}$ is not useful to detect in which moment the accumulation occurred and gives the same treatment to false zeros and to null measures that repeatedly occurred on the same specific day of the week. It is calculated following Eq[2.4], using the coefficient of variation CV (standard deviation divided by the mean value) of the set n_i (where i takes values between 1 and 7, depending on the day of the week) that contains the number of days with precipitation equal or

higher than 1 mm. It is considered that it be calculated as long as the studying year has a minimum of 20 days measuring precipitation equal or higher than 1 mm.

$$Q_{zero}{}^{w} = 100 - 100CV(n_i) \tag{2.4}$$

- $Q_{outliers}$: index related to the proportion of days in which the threshold of outliers is not exceeded with respect to the total number of days of the given year. This threshold is calculated in absolute terms for the months from January to December for each series (i.e., it is not restricted to the given year), using daily precipitation data equal to or greater than 1 mm and three times the interquartile range above the 3$^{\text{rd}}$ quartile (it is considered that it can only be calculated with a minimum of 7 days of rainfall equal or over 1 mm and one complete month of daily data). Given the fact that a 10% of outliers is never reached, the index is scaled from a simple percentage in order to be comparable to the rest of the indexes, the calculation is performed following Eq[2.5] where $n_{outliers}$ is the number of days that exceed the threshold and n is the number of days with data in the given year.

$$Q_{outliers} = 100 - 1000\frac{n_{outliers}}{n} \tag{2.5}$$

2.2.3 Relative quality control

The relative phase of the quality control consists in comparing values measured at a station with those indicated by the available reference nearby stations. Apart from the choice of reference series, a source of difficulties in relative controls at daily level is the comparison between values that are shifted, thus having a correct value but it being assigned to a wrong date (usually shifts of one day). One approach to this problem is to compare daily values with the corresponding day of the reference series and also with the values shifted one day, the drawback is the added complexity in decision making in case of automation. Another approach is to obtain the outcome of the quality check as if no shift was present and then go over the cases of sequentially flagged values in order to determine if they correspond to a shifted period; in this case, the drawback is the unclear interpretation of the results. In the present project, the option of comparing the daily values with the corresponding day and also with shifted days was inspected but the decision to label a value as shifted was complex and could affect the results of the quality control, it also took a computation time longer than it was desired. Therefore, the quality control was executed as if no shift was present and afterwards the distribution of labels was analysed, but the interpretation of those results was not conclusive: the problem is that shifted values usually occur during short periods and the quality control could label shifted values as valid if they are very similar to the correct value.

In order to select auxiliary stations suitable for comparison with the candidate station, an index has been designed to classify their adequacy or representativeness, R, according to their spatial coordinates (considering distance and difference of altitude)

and the correlation between measured data. This representativeness index is assigned to each pair *candidate station – possible auxiliary*. Representativeness is calculated using the expression Eq[2.6], where d is the distance (in km) between the two stations, h is the absolute value of the difference in altitude (in m) and C_{corr} is Pearson's correlation coefficient (r) between the daily data of the candidate station and those of the auxiliary station. The distance term takes values between 0 and 1 linearly decreasing with distance up to the maximum considered radius (50 km). The altitude term takes also values between 0 and 1 but it decreases exponentially with the increase in altitude difference so that there is no upper limit to allowed altitude differences even if small differences are considered more alike than greater differences which are highly penalised with a term approaching zero. Regarding the correlation term, when the value of C_{corr} is negative, the zero value is assigned to this coefficient, as well as if the candidate station and the auxiliary do not have a minimum of 25 days in which both have non-zero precipitation. A part from the correlation term which ensures reference series with a similar rainfall pattern, the terms of distance and difference of altitude have also been considered necessary because of the topography of the territory. Indeed, in the study area, great differences of altitude may be present at short distances while a distinction in the terrain characteristics (such as altitude above sea level or distance to the coastline) might trigger different drivers for meteorological conditions.

$$R = \frac{100}{3}\left(\frac{50-d}{50} + 0.5^{h/500} + C_{corr}\right) \tag{2.6}$$

In order to label daily values of the candidate station after performing a daily comparison with reference series, monthly thresholds are established from high quality data and are applied to determine the outcome of the quality control unless the conditions to some special cases are met. At the end of the procedure, daily values are labelled as valid, "V", doubtful, "D", invalid, "N", or insufficient information, "I".

Any candidate value with less than three auxiliary values fit for comparison will be assigned label "I". More than one auxiliary is necessary in order to judge if a potential error is present in the measure of the candidate or if it is a fault in the auxiliary. More than two auxiliaries are desirable in order to obtain a solid global outcome based on the average of individual labels assigned after comparison of the candidate with each auxiliary.

Some special cases are labelled directly when the necessary requirements are fulfilled. Only in case of having more than 7 auxiliary stations a daily value can be directly labelled according to Table[2.2] or considered valid with more than 15 auxiliary stations if it falls between 1st and 3rd quartile of reference values.

In the general case (i.e., having at least three auxiliaries and not fulfilling the conditions to be directly labelled as a special case following Table[2.2]), candidate values are given an individual label, L, after the comparison with each auxiliary and, afterwards, these individual labels are averaged in order to obtain the global label which is the definitive outcome of the quality control for the considered value. The specifics to obtain the

final outcome label are detailed following Eq[2.7], Eq[2.8], and Eq[2.9].

Candidate daily value	**Reference values of 5 mm or less**		**Reference values under 1 mm**	**Outcome label**
0 mm	under 20%	or	under 1%	N (false zero)
0 mm	over 90%	or	over 75%	V (correct zero)
over 15 mm	over 99%	and	over 90%	N (isolated value)

Table 2.2: outcome labels for special cases.

Individual labels (i.e., for each pair *candidate - auxiliary*) can only be "V" (valid) or "N" (invalid). They are obtained using a previously set threshold, T, calculated following Eq[2.7] which depends on the month of the considered daily measure and the representativeness, R (Eq[2.6]), between the candidate and the auxiliary.

$$T = C_m \cdot \ln(101 - R) \tag{2.7}$$

Where C_m is a coefficient which depends on the month (see Fig[2.6]).

T has a lower limit zero ($\ln(1)$) when representativeness R is 100 (equivalent to candidate and auxiliary being the same station, i.e., at the same location and with perfectly correlated data) and grows logarithmically with increasing distinction between candidate and auxiliary (i.e., increasing distance, increasing altitude difference and reduction in data correlation).

The difference (Dif) used for comparison (calculated with the measured value at the candidate station, PPT_c, and the measured value at the auxiliary station, PPT_{aux}) is, in fact, a relative difference (divided by the average of both values) and scaled multiplying by the maximum, following Eq[2.8].

$$Dif = \frac{|PPT_c - PPT_{aux}|}{Av(PPT_c, PPT_{aux})} Max(PPT_c, PPT_{aux}) \tag{2.8}$$

Where PPT is the daily value of the candidate (c) or the auxiliary (aux).

The function in Eq[2.8] has been chosen for comparison with reference values because it enhances the distinction and does not treat small and large values alike (as would happen with a strict relative difference where the values 1 mm versus 2 mm produce the same comparison (i.e., the double) than 10 mm versus 20 mm).

In case the difference between the measured values of the pair is under the threshold, the outcome of the individual comparison is label "V" (i.e., they measure values similar enough, where the differences can be accounted by the difference in location), while

cases in which the difference exceeds the threshold have an outcome for the individual label of "N".

In order to achieve a final label "V" after relative comparison, the candidate station must obtain more than 50% in a weighted arithmetic mean, W_m, (see Eq[2.9]) of the individual labels of all the pairs *candidate-auxiliary*. In case of obtaining between 20% and 50% of the weighted arithmetic mean, label "D" will be assigned to the candidate value and if less than 20% is obtained, the ultimate label of that value will be "N".

$$W_m = 100 \frac{\sum_{aux}[(R - R_{min})^2 L]}{\sum_{aux}(R - R_{min})^2} \tag{2.9}$$

Where $L = \begin{cases} 1 \textit{ if individual label is V} \\ 0 \textit{ if individual label is N} \end{cases}$, and R_{min} is the minimum allowed representativeness, set to 70.

In the general case, where the outcome label is assigned depending on the results of daily comparison between the candidate station and auxiliaries following Eq[2.7], Eq[2.8], and Eq[2.9], the monthly coefficient C_m is needed to determine the threshold. Coefficient C_m is displayed in Fig[2.6] and has been established after a study based on daily precipitation values of 150 stations of the XEMA during the period 2008-2016 (when these stations have undergone quality checks on a daily basis). The determination of the threshold for each month is performed by organising, according to the representativeness (see Eq[2.6]), the daily scaled relative difference of precipitation (obtained by Eq[2.8]) at pairs of stations. Then, the value of 10 times the interquartile range above the 0.95 percentile of the set of points is found for each R and the obtained values are fitted logarithmically to find Eq[2.7], where the coefficient C_m depends only on the considered month of the year. Fig[2.7] shows the fitting of the thresholds from points obtained at 10 times the interquartile range above the 0.95 percentile as well as all the observed differences.

The empirically obtained coefficients C_m show a seasonal cycle with high values in the summer months and low values in winter (see Fig[2.6] and Fig[2.7]). The highest values correspond to months when summer storms are likely to occur (form midsummer to early autumn, i.e., July to September); these storms are characterised by short episodes of heavy rain which yield an extremely irregular pattern on surface measuring stations with potentially high rainfall amounts at some points and, hence, cause high differences in measures taken even at close locations. On the other hand, winter months (December to February) are characterised by episodes of continuous uniform rain which yield homogeneous spatial patterns with low differences in measures taken at distant locations. In between, we find months corresponding to spring and late autumn which can produce both kinds of rainfall episodes but tend to present widespread rain and usually coincide with the rainiest seasons in the region of our study.

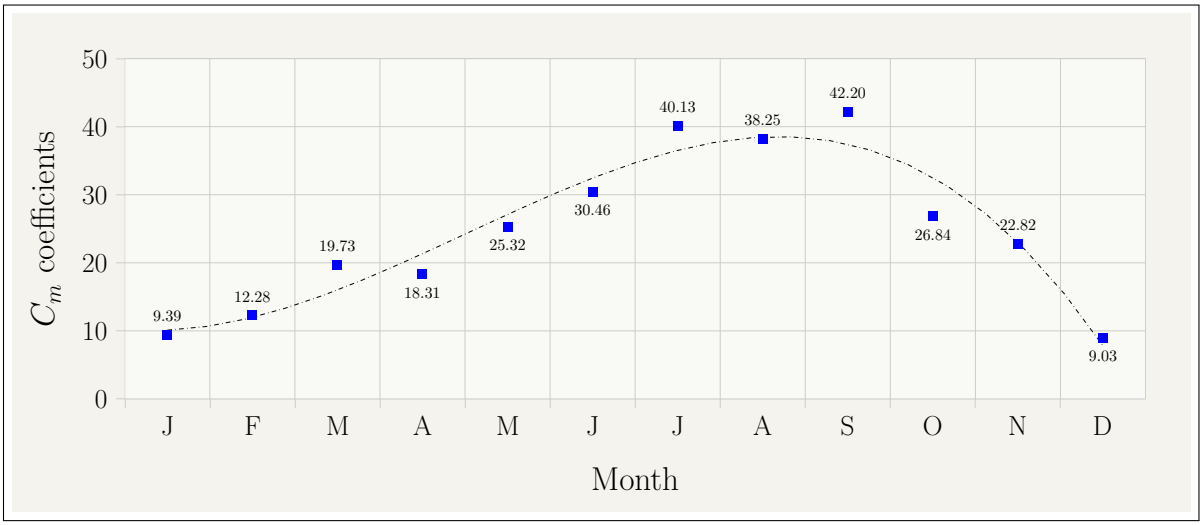


Figure 2.6: coefficients C_m of the threshold's function (depending on the month). Black dashed line is a 3rd order polynomial tendency line displayed to aid visualisation.

2.2.4 Verification of the quality control

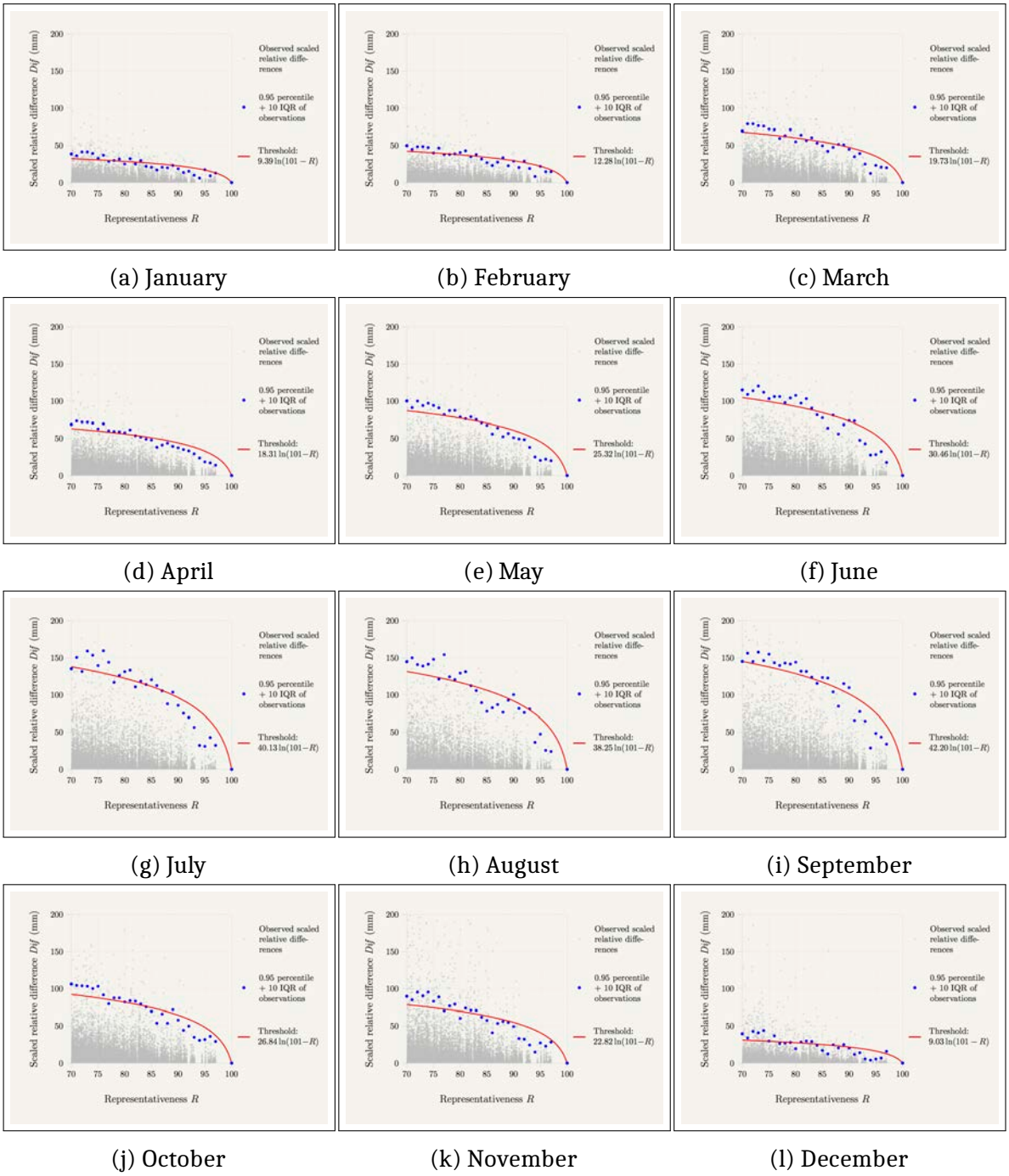
The daily rainfall quality control designed for the set of available stations in the SMC's daily rainfall database has been verified by introducing controlled errors and analysing the obtained results. The quality control method can be verified taking advantage of the availability of data from automatic stations incorporated into the XEMA network managed by the SMC. Data measured at these stations is subject to a daily manual validation by the SMC technicians that has been carried out since 2008 and up to now.

In order to perform a verification of the automatic quality control method, the results obtained in 150 stations across the territory have been analysed after introduction of controlled errors in the data. It is performed by means of a count of false positives (type I error) and false negatives (type II error), see Table[2.3].

		Null hypothesis (H_0): the value is correct	
		TRUE	**FALSE**
Decision about H_0	**Fail to reject (label as "V")**	Correct inference (True Negative)	Type II error (False Negative)
	Reject (label as "N")	Type I error (False Positive)	Correct inference (True Positive)

Table 2.3: error types.

In our case, type I errors are cases in which the method labels a correct value as invalid; it is calculated as the percentage of "N" labelled values over the total number of



(a) January (b) February (c) March

(d) April (e) May (f) June

(g) July (h) August (i) September

(j) October (k) November (l) December

Figure 2.7: threshold values for each month obtained from fitting the values of 10IQR above the 0.95 percentile of observed scaled relative differences in pairs of automatic weather stations from 150 selected XEMA stations in the period 2008-2016.

cases that the method is capable of labelling, considering the control series (which are known to be correct). A total number of 465,551 cases have been considered in order to calculate type I error, it is the number of days that the relative quality control is capable of labelling (considering the 150 control series). The verification result indicates that the type I error committed in the method is on average 0.7%, for values under 50 mm there are a 0.1% of false positives, at most 1.5% for rainfall values under 100 mm and under 10% for rainfall values under 150 mm; greater amounts have not been considered as these quantities have seldom been measured by the analysed stations (in fact, only 63 cases are measures between 100 mm and 150 mm.

Type II errors, in our case, are undetected errors in available data, i.e., cases which the method labels as valid but are actually invalid. Given the fact that actual errors in data are uncontrolled and it is precisely what needs to be flagged, the analysis of this type of error produced by the method is performed by introducing controlled errors in valid data, that is, in the subset of control series that are labelled as valid by the method. These controlled errors have been introduced by multiplying the measured value by a factor and the results have been analysed depending on the rainfall amount indicated by the original value and the magnitude of the introduced error. However, type II error depends on the magnitude of the introduced error (i.e., for errors that yield a value which is very similar to the original measure, failing to reject the error in not a deficiency of the method) and the amount of actual rainfall (i.e., for small amounts, the capacity to distinguish between original measures and values modified by a multiplicative factor is limited by rainfall's inherent variability). In this way, type II error is reduced as the magnitude of the error grows as well as with higher amounts of rain.

Results of type II error obtained on the verification are displayed in Fig[2.8], where it can be seen, for instance, that errors of five times the correct value for rainfall under 20 mm are not detected in 16% of the cases whereas a similar type II error is found for errors of twice the correct values in the case of rainfall up to 100 mm.

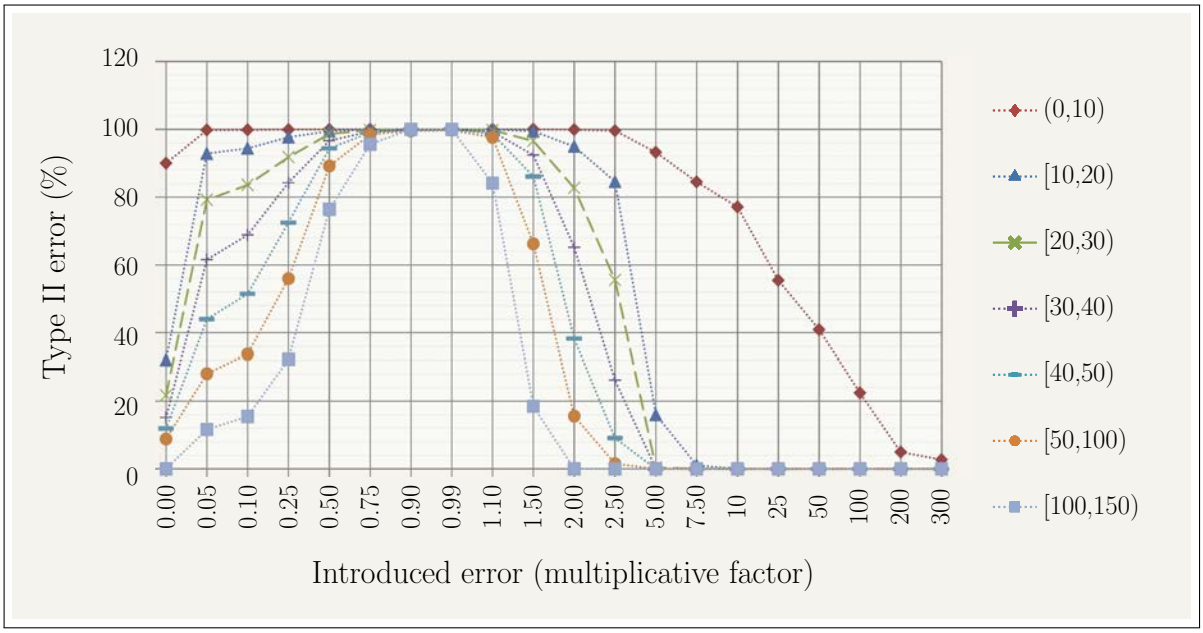


Figure 2.8: type II errors found in the verification of the quality control method, depending on the value of the original data and the multiplicative factor introduced to modify it.

The method becomes optimal when both types of errors are minimised. The performed verification ensures that, in the results obtained through the quality control method, data labelling as "N" is done with great security (since type I error is very low). On the other hand, data labelled as "V" is better accomplished for high amounts rather than for low values. Therefore, this methodology is suitable for quality controlling extreme values (above 50 mm) yielding type II errors (failing to reject invalid data) under 16% when data is twice the correct value.

2.3 Resulting quality controlled series

This section is a discussion of the results after application of each step of the designed quality control. The procedure has been applied to the available dataset of 2,142 series in the period 1855-2016. The outcome at each phase of the quality control is analysed and the final results include a study of the number of doubtful cases detected during the automatic quality control and a way of selecting series with minimum quality.

2.3.1 Basic quality control

The basic quality control consists in extracting a list of physically impossible data (negative values and rainfall above 567 mm) and doubtful cases which should be manually checked (rainfall above 300 mm and consecutively repeated non-zero values). A

total of 12 negative cases were found, among them, 10 cases were in the same station with the value -999; the other 2 were values of -1 and -2. These 12 cases are codification mistakes, for instance, -99.9 is the value used to code missing data (which is interpreted as "missing value" when introduced in the database, hence, it is not read as data when extracting values that have been introduced correctly). On the upper threshold of physically impossible data, only one value was found in the database above 567 mm, an isolated value of 700.5 mm which was removed after checking that it was an erroneous value. However, 28 values above 300 mm were found; they were checked with the information of nearby measuring sites and consulting their original handwritten data when possible. After the check, 4 values were considered to be mistakes: a value of 322.4 was restored to its correct amount (32.4 mm) and the other 3 values (461, 461.1 and 465.1) were removed from the database. The list of consecutively repeated values has a total of 1,058 cases, among them, 9 cases repeated for more than 3 days and 4 cases repeated for more than 10 days; the highest values repeated over the longer spells were checked an removed from the daily database when necessary, although most of them are kept with a flag that only considers them when extracting monthly data from the database as they are not correct at daily resolution but they add to the correct amount of monthly total.

2.3.2 Absolute quality control

Fig[2.9] shows the percentage of stations with a specific value of the global Q index (averaged over the whole period of availability for each series) calculated for all the series of the present project (2,142 in total). A distinction is made between manual stations and automatic stations. As expected, the series coming from automatic stations have shown better global quality indexes than manual stations (which are much more numerous and usually longer, so they have potentially more problems). The average value of this global quality index in the whole set of considered stations is 82.0, the median is 87.1 and the first and third quartile are, respectively, 78.1 and 90.5; these results prove a good working of the index, considering that the value of 80 was set as a threshold of acceptable quality following the design of the index.

Fig[2.10] shows the spatial distribution of the *Q* index (Fig[2.10a]) and the individual indices that form it. It is observed that the indices do not present any pronounced pattern, although there are some peculiarities. The global Quality Index (Fig[2.10a]) has uniformly distributed values, which indicates an even spatial distribution of the problematic measuring sites. Therefore, the index is not dependent on climatology (which is variable in the territory). A zone in the centre and slightly to the north could seem to have a lower presence of low indices but this is related to a lower density of stations.

The two parts of the Quality Index that are related to the percentage of data (Fig[2.10b] and Fig[2.10c]) show a uniform spatial distribution which influences the result of the *Q* index. The majority of low values coincide in both indices but some cases of low percentage of data have medium Q_{gaps} index meaning that the gaps are localised instead of scattered as it is usual.

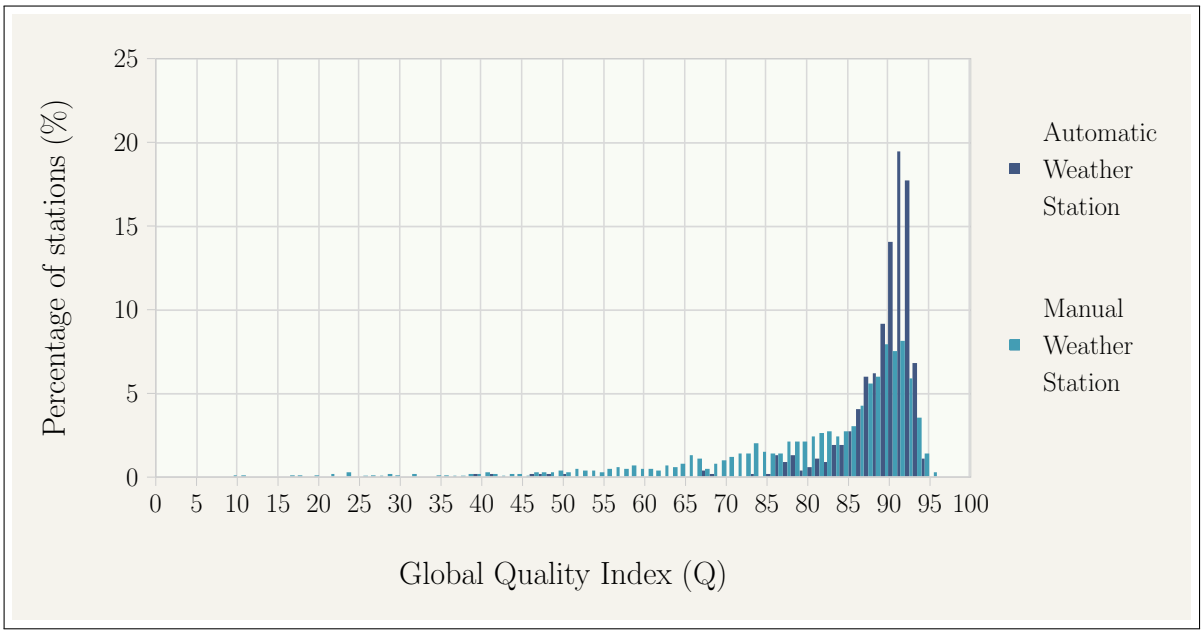
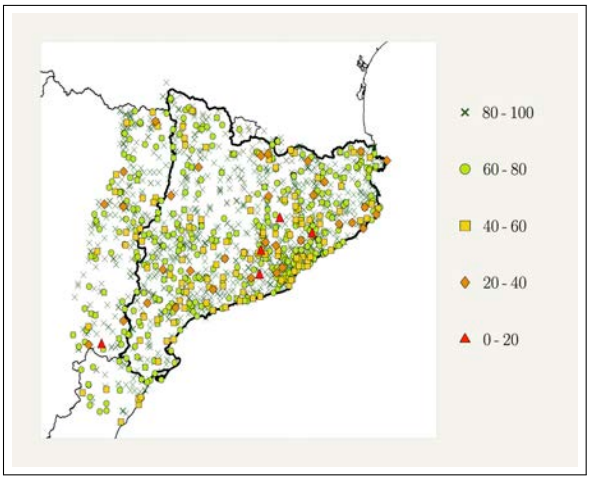


Figure 2.9: percentage of stations for each global *Quality* index value, distinguishing between Automatic and Manual weather stations (averaging the *Quality* index over the whole period with data at each measuring site).

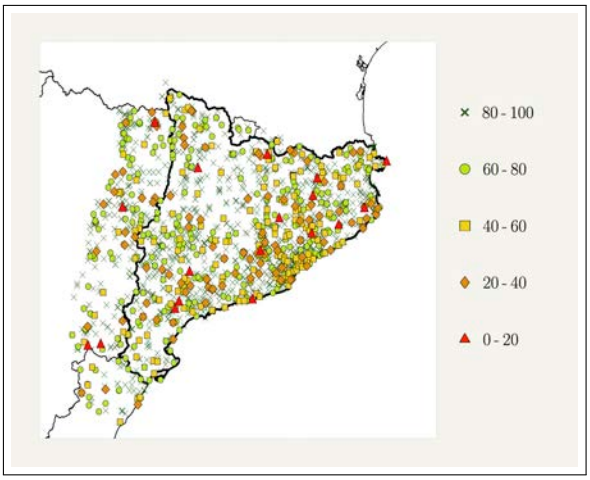
The ${Q_{zero}}^m$ term of the Q index presents high values in general, with only two weather stations under the value of 40 when all the years are averaged. These two cases are *MN - Montornès del Vallès* and *VR101 - Lliçà de Vall - Ca l'Estaper*. The first one is an automatic station which covers a short period (under 10 years) and is currently dismantled; it is suspected to have had poor maintenance. The second case of low values of averaged ${Q_{zero}}^m$ index has allowed to detect a station with a long period of data (from 1992 to 2013, both whole years included) in which no precipitation was registered and each of the days has the value 0 mm. A part from these exceptional cases, the spatial distribution of ${Q_{zero}}^m$ index shows that values between 40 and 80 are concentrated in the coast or, especially, on the southern half of Catalonia (see Fig[2.10f]); these results are consistent with the climatology of those areas, which are more arid and could have longer spells without precipitation, but the three months dry spell is an exceptionally long period in the whole of the study area and, in the majority of the years in which ${Q_{zero}}^m$ presents low values, it will reflect an actual measurement issue in the station.

The ${Q_{zero}}^w$ term of the Q index has its highest values (between 80 and 100) concentrated in the north (in the Pyrenees and eastern pre-Pyrenees) and lower values (between 20 and 40), although there are very few of them (5 cases), in the central coast up to the prelitoral depression (see Fig[2.10e]). This index has proven to be the less satisfactory of the terms of the Q index as it can not detect measurement issues efficiently.

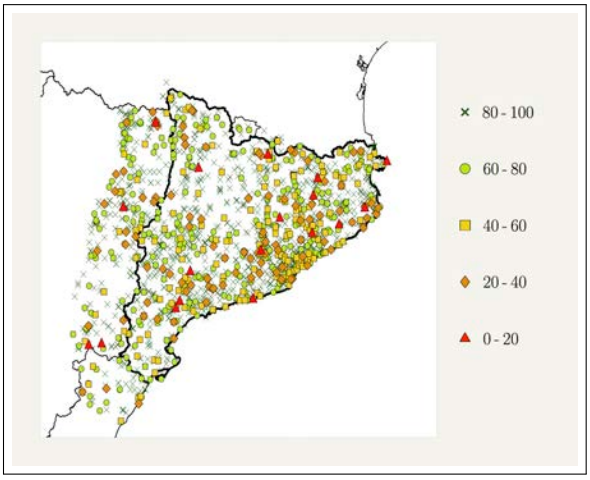
$Q_{outliers}$ is the last terms of Q index and it has a characteristic spatial distribution (see Fig[2.10f]). A remarkable feature is that its values, that range between 75 and 100,



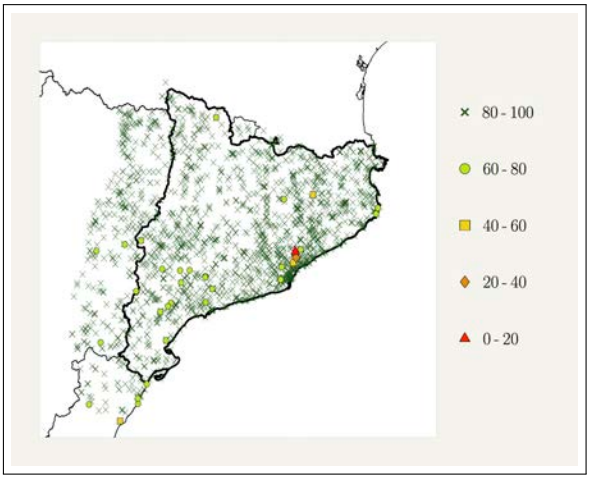
(a) Global Quality Index (Q)



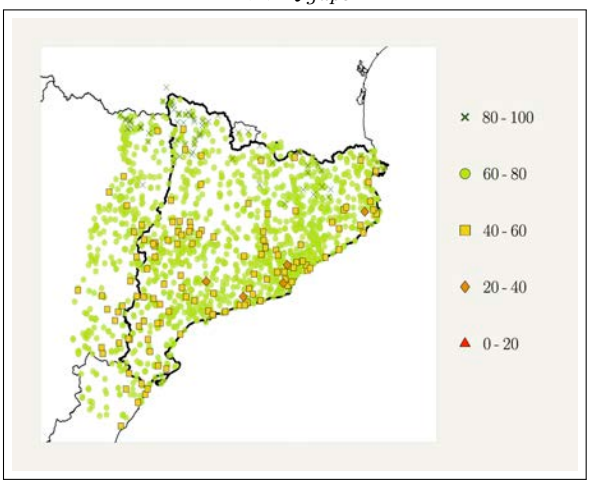
(b) completeness of data



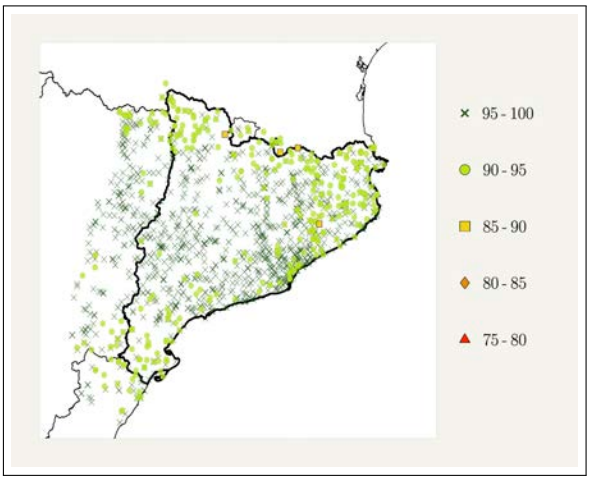
(c) Q_{gaps}



(d) ${Q_{zero}}^{m}$



(e) ${Q_{zero}}^{w}$



(f) $Q_{outliers}$

Figure 2.10: spatial distribution of mean Quality Index (averaged over the whole set of years with data at each measuring site).

are higher than in the other indices; it had already been expected in the design of the index, as outliers are rare cases by definition, but even if the behaviour of outliers was taken into account, the mean results exceed the expected value.

2.3.3 Relative quality control

The outcome of the relative quality control procedure has yielded a label for each daily value of the whole set of available series in the database of the study. Table[2.4] summarises the number of analysed cases and the obtained labels after the application of the quality control procedure.

Daily values	**Valid cases (label "V")**	**Doubtful cases (label "D")**	**Detected errors (label "N")**	**Cases without enough information (label "I")**
14,144,395	10,466,655	15,228	30,521	3,631,991
	(74.0%)	(0.1%)	(0.2%)	(25.7%)

Table 2.4: number of obtained labels for each category.

The results of the quality control show that an important part (25.7%) of the days have not been labelled according to their quality because there is not enough information to reach a decision, and they have been labelled as "I". Regarding the number of series (instead of the total daily values), of the 2,142 initially available series, a 5.7% have a minimum operating period of 15 years over which it has not been possible to perform the relative comparison at any time. The fact that whole series undergo the quality control without obtaining an outcome on any of their days is a weakness of the method triggered by low spatial density of measurements in some temporal periods, because the automatic procedure cannot label values without a minimum number of neighbouring stations that fulfil the necessary requirements. However, certainty in labelling has been prioritised over forcing to produce an outcome.

Regarding the subset of days in which it is possible to perform the comparison, the percentage of data labelled as valid is a 99.6%. These results indicate an outstanding quality of the initial database. Even if a type II error of up to 16% might be committed, the percentage of valid data in the initial database would be between 83.6% and 99.6%.

Apart from the verification process already presented, a manual checking of some cases has been performed in order to further validate the automatic labelling. This analysis has revealed that an important part of correctly detected errors are values with a problem on the day at which the value is assigned. Specifically, shifts of one day. Some of these cases had already been detected at early stages of the planning of the quality control procedure but they were overlooked because of the complication associated with approaching this sort of error.

Furthermore, an analysis of the obtained results is performed based on the number of available series per year which achieve a minimum quality. Three levels of quality have been defined, namely: acceptable, good and excellent quality. The quality of the series at each year has been assigned depending on the completeness of daily data, the quality index previously designed and the performance in the relative quality control (i.e., a maximum percentage of automatically assigned days as "N"). The three levels of quality are described in Table[2.5] where the threshold of $Q = 80$ corresponds to the acceptable quality value according to the design of the index and the number of invalid cases up to 10% is close to the type I error for high amount of collected rainfall (i.e., over 150 mm).

	Acceptable quality	**Good quality**	**Excellent quality**
Completeness of daily data	90%	95%	100%
Minimum quality index	80	85	90
Maximum "invalid" ("N") days	10%	5%	1%

Table 2.5: levels of quality.

Fig[2.11] shows the temporal evolution of the number of series that annually comply with the three aforementioned levels of quality as well as the total available series. The temporal evolution of the availability and quality of the series can be analysed in three broad sections. In the first section (from the beginning of the considered period and until the mid-1970s), the number of available stations as well as the number of high quality series increases progressively, except for the period corresponding to the Civil War, where both the quality of the series and the number of available measurement sites dramatically decreases. In the second section (approximately between 1974 and 1995) the number of available stations is stabilized and so does the number of high quality stations (the series of excellent quality are the more variable with time but there are around 230 stations) so that the percentage of stations that annually exceeds the predetermined excellent quality level is approximately 49%, whereas good quality stations are 73% of total available series. Finally, in the most recent period (as of 1996) there is a sudden increase in the number of available stations caused by the proliferation of automatic stations. The percentage of detected errors in the recent period (since mid-1990s) is progressively reduced even if the number of series that pass the established minimums of quality does not reflect it, as it is strongly influenced by the completeness of the series; the percentile 0.95 of invalid ("N") days has a mean value of 1.4 in the period 1985-1994 and it is reduced to a mean value of 0.6 in the period 2010-2016.

Fig[2.12] displays the number of series regarding the length of years with high quality data. For a shorter required length, the number of series is higher; however, there

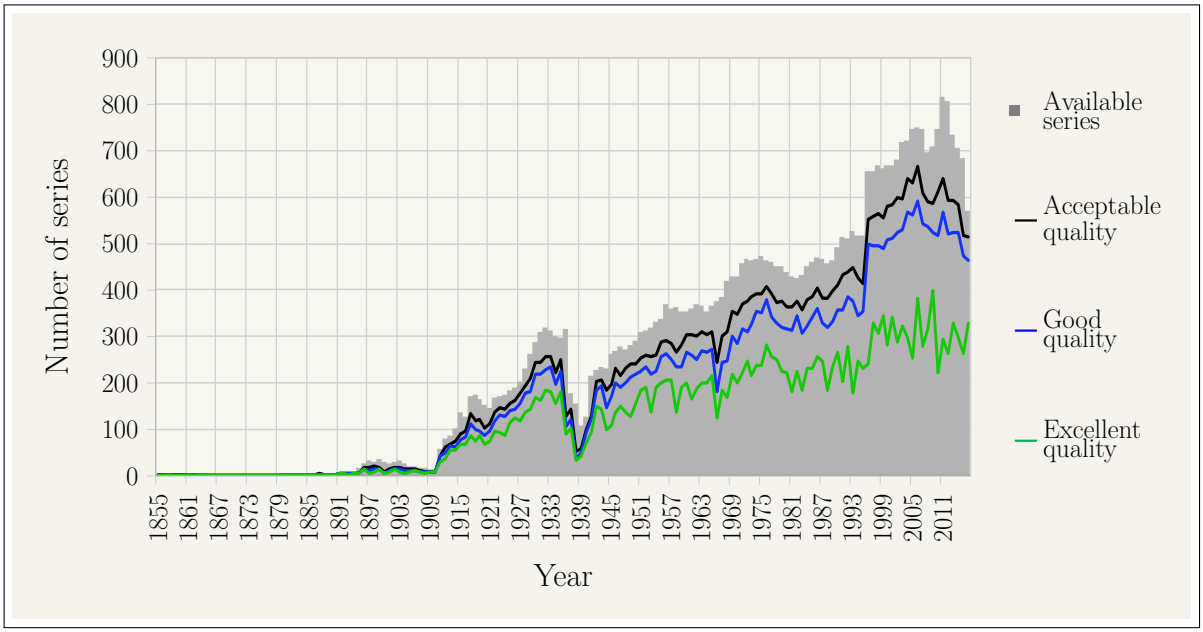


Figure 2.11: temporal evolution of the number of available series at each year and series with a minimum quality: acceptable quality (90% completeness, minimum quality index of 80 and detected errors under 10%), good quality (95% completeness, minimum quality index of 85 and detected errors under 5%) and excellent quality (100% completeness, minimum quality index of 90 and detected errors under 1%).

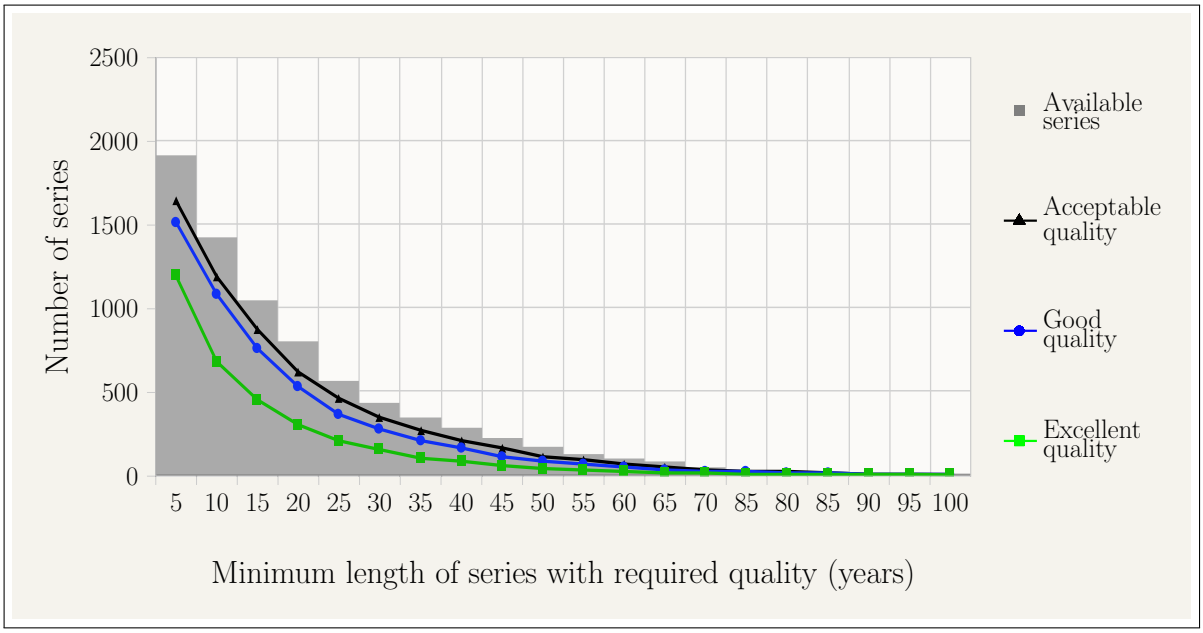


Figure 2.12: number of series with a minimum number of years with acceptable, good or excellent quality.

are 80 series with at least 65 years of data and, among them, 12 series achieve a level of excellent quality in the same minimum number of years. The total number of series with at least 10 years of operability is 1423 and 47.6% of them have excellent quality while 75.8% have good quality. Regarding the series with at least 30 years of data (a total of 434 series), 34.3% have excellent quality while 62.7% have good quality.

Chapter 3

HOMOGENISATION AND SELECTION OF SERIES

In the present chapter a dataset of high quality data is presented. It consists in a subset of available data in the time period 1942-2016. It is a sub-period of the global time period of data presented in previous chapters, in order to select a common period with a minimum coverage of spatial data to be used in the rest of the project. After the exhaustive quality control applied to the whole set of available data (see previous chapter), a subset of high quality series is selected (it is strongly recommended (Aguilar et al., 2003) to quality control data before applying a homogenisation procedure due to the false detection of *change points* caused by erroneous measures) and homogeneity checks are applied in order to select the most suitable series to estimate the maximum expected precipitation at high spatial resolution.

Homogenisation techniques are necessary to prevent non-climatic signals from being introduced to the analysis of data. In the present project, homogeneity of daily rainfall series has been checked at a monthly resolution in order to select homogeneous series from the available dataset. The correction of detected inhomogeneities is beyond the scope of the project and has not been performed. Therefore, the series used in subsequent parts of the project have been chosen after selecting the best series according to their performance in the quality control and homogeneity checks, but the original data of the selected series has been used without attempting to correct daily values nor fill missing data.

3.1 Homogenisation of climatic data

In climatology, a homogeneous series is defined as the temporal series of a meteorological variable with oscillations that are only caused by the atmosphere's behaviour. A non-homogeneous series, however, is a time series with inhomogeneities or artificial *breaks* (also known as *change points*) caused by variations in the measuring con-

ditions (e.g., changes of location or measuring instruments, growth of vegetation). Inhomogeneities can appear as an abrupt change in the mean value of the variable or as an artificial trend. Therefore, the term homogenisation describes the set of techniques used to distinguish the climatic signal from variations caused by artificial factors.

In the process of homogenisation of climatic series it is useful to have available metadata (Aguilar et al., 2003), i.e., stored non-climatic information about the measuring conditions. Therefore, the storage of information related to location, type of instruments and other circumstances of the measuring site and the documentation of any change on these circumstances is an important job of the organisations responsible for the measuring networks. However, it is not always possible to access metadata or they might be incomplete; hence, there are some homogenisation techniques which do not use this information and most of the existing techniques do not need it fundamentally.

Several techniques exist with the objective of detecting and correcting inhomogeneities, however, some of them are more appropriate than others depending on the meteorological variable, the temporal scale or the spatial density of weather stations (in case of relative homogenisation). In fact, different techniques have been developed by different research groups in order to adjust to their own needs and data (Aguilar et al., 2003).

A review of the existing homogenisation techniques and a classification according to their characteristics can be found at Peterson et al. (1998) and Ribeiro et al. (2015) and an analysis of the performance of some methods for precipitation data at Beaulieu et al. (2008) and Venema et al. (2012). The distinction between absolute and relative methods is particularly important. On the one hand, absolute methods use information of exclusively one time series (the candidate series) but in order to distinguish climatic trends from artificial breaks it is necessary to have access to reliable metadata; otherwise, according to Ribeiro et al. (2015), an absolute method could introduce inhomogeneities instead of correcting them. On the other hand, relative methods need one or more reference series in order to distinguish a climatic signal from an artificial inhomogeneity and they might require the reference series to be already homogeneous. Relative methods are clearly recommended instead of absolute methods by Guijarro (2011), and Domonkos and Coll (2016) claim that relative homogenisation can help to remove even relatively small non-climatic biases from data when the network of spatial measurements is sufficiently dense.

Homogenisation procedure is the name given to a set of techniques used to implement a methodology with the objective of homogenising specifically climate data series. The SMC currently works with two homogenisation procedures: HOMER and ACMANT, which are applied to temperature and precipitation series. Both HOMER and ACMANT were developed during the Action COST ES0601 *Advances in homogenisation methods of climate series: an integrated approach (HOME)* which, between 2007 and 2011, analysed the strengths and weaknesses of available homogenisation methods (like MASH, developed by the Hungarian Meteorological Service, Szentimrey (2006); Prodige (Caussinus and Mestre, 2004), AnClim (Stepanek, 2003), Climatol (Guijarro, 2018) or RHtestV3 (Wang and Feng, 2009)) with the objective of achieving a general method for homogenis-

ing climate and environmental data by an improved synthesis of the most effective statistical procedures for detection and correction of inhomogeneities COST Action ES0601 (2011). HOMER was the procedure resulting of the COST Action ES0601 (2011), but ACMANT (also based on the existing PRODIGE Caussinus and Mestre (2004) procedure for the detection of inhomogeneities) was developed in parallel by Peter Domonkos and tested on the benchmark dataset used to analyse existing procedures. ACMANT has the advantage over HOMER of being fully automatic; the implementation of the algorithm by users with different degrees of training was found to be one of the important factors in the results of the homogenisation by the COST Action ES0601 (2011), therefore, automatic procedures, which reduce the user decisions, produce reproducible results and are more suitable for users with low training.

3.1.1 Homogenisation of daily rainfall data

In order to perform climatic studies of precipitation or obtain maximum expected rainfall, it is necessary that the series are homogeneous. The task, however, is complex because of the great spatial and temporal variability inherent in precipitation. There have been recent advances in the development of homogenisation techniques, their application to precipitation and the increase in temporal resolution; for instance, ACMANT (Domonkos, 2011) was developed during the COST Action ES0601 (2011) initially to homogenise temperature series but it was later developed to ACMANT2 that included methods to homogenise precipitation and the option to downscale the correction to the daily scale (Domonkos and Coll, 2016). Nowadays, nonetheless, there is still a need to improve the methods used to homogenise precipitation (Ribeiro et al., 2015).

One of the conclusions of the review made by Ribeiro et al. (2015) is that results of homogenisation at an annual scale are, in general, better than at higher temporal resolution (i.e., shorter duration) because of the increase in the variability of the time series at shorter duration. Therefore, the most usual procedure to homogenise precipitation is to use accumulated monthly or annual rainfall instead of using homogenisation techniques for daily data. A different, but not usual, approach to reduce the complexity added by the high variability is to use derived quantities as the number of rainy days per month, the maximum monthly rainfall in 24 hours or the number of days with rainfall above the 0.995 percentile; monthly homogenisation techniques applied to rainfall derived quantities are applied at Vicente-Serrano et al. (2010) to obtain a daily precipitation database for the north-east of Spain.

The two previously mentioned procedures (HOMER and ACMANT) that are used in the SMC can be used to detected multiple inhomogeneities at a monthly scale in temperature and precipitation time series and they are able to correct data at a daily resolution. Although in case of HOMER, the daily downscaling can only be done with the joint use of SPLIDHOM Mestre et al. (2011). It should be noticed that the detection of multiple breaks in a single temporal series needs mathematical methods that include the joint treatment of mutually dependent factors (Domonkos and Coll, 2016) and it

is not as simple as an iterative correction and new detection of single breaks. Existing comparisons between HOMER and ACMANT are mainly focused on temperature, see for instance: Pérez-Zanón et al. (2015). However, Domonkos and Coll (2016) explains that both HOMER and ACMANT come from a similar theoretical background and, hence, they consider it reasonable to assume both procedures could exhibit a similar efficiency and recommend a choice between HOMER and ACMANT based on the characteristics of the dataset. In particular, Domonkos and Coll (2016) recommend the use of ACMANT for datasets without metadata, dense networks with high correlation between time series and (after the practical advantage of using an automatic procedure) for very large datasets. In summary, relative homogenisation techniques are used to compare the temporal behaviour of each candidate series with that of a reference series (formed by a weighted average of available series in the same climatic area as is the case of ACMANT or as an independent pairwise comparison with each available series in the neighbourhood as is the case of HOMER) to detect periods in which the relative behaviour is different than in the rest of the time period. In the case of precipitation, the comparison is made by means of the ratio of accumulated monthly data. Fig[3.1] shows the mentioned comparison with reference series in the form of a ratio series, as well as the suggested breaks in the mean value of the ratio series; in this case, after HOMER's pairwise detection, the user must choose which suggested breaks are actually considered as non-climatic inhomogeneities (in case they are systematically repeated with different reference series and after checking metadata for the candidate station).

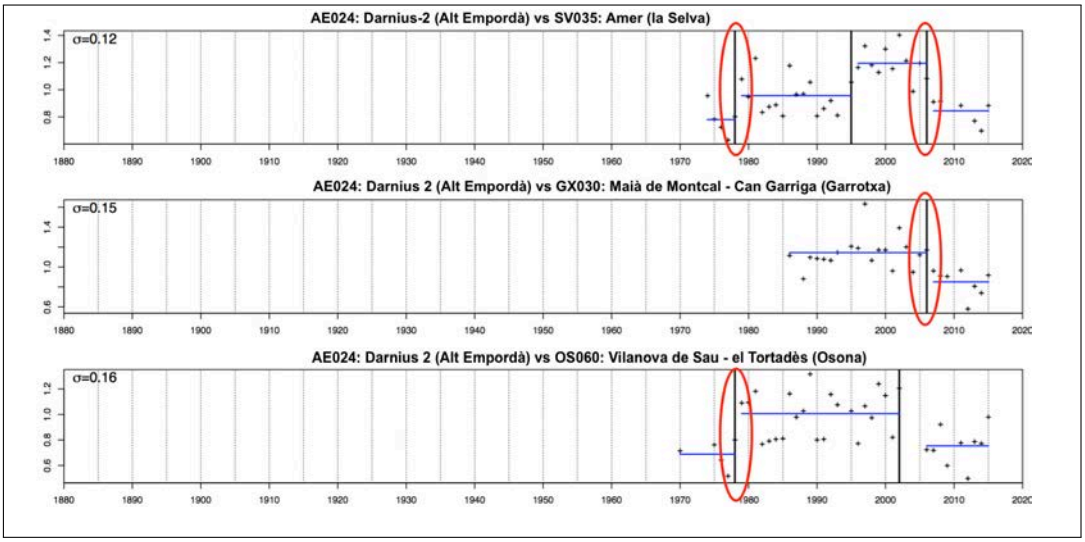


Figure 3.1: sample of the first three pairwise comparisons of HOMER with reference series of a candidate example series. Vertical lines mark automatically suggested breaks; red circles indicate user's choice of actual inhomogeneities due to repetition of the break at the same position with several reference series. The σ value indicates the minimum detectable amplitude, related to the variability of the series.

The station *AE024 - Darnius-2* that is used as an example in Fig[3.1], has breaks shown in Table[3.1] when homogenised automatically with ACMANT. In both cases, there were the same available series, but the two procedures work differently with reference series: HOMER (in the case of pairwise detection, shown in Fig[3.1]) uses all available series at a distance and/or correlation maximum chosen by the user and the comparison is made one by one; ACMANT automatically makes a reference series as a composite of available series with a weight depending on correlation.

	Year	**Month**	**Amplitude (ratio)**
Break 1	1974	9	0.77
Break 2	1979	9	1.46
Break 3	2007	9	0.83

Table 3.1: detected breaks in *AE024 - Darnius-2* by used version of ACMANT in the project.

3.2 Applied homogenisation procedure

In the present project, the 3^{rd} version of the ACMANT procedure was applied to detect homogeneity breaks in monthly precipitation series. As explained by Domonkos and Coll (2016), ACMANT3 is a software package composed by six programs that incorporate 174 sub-routines in total. In particular, the executable used in the present project is ACMANP3month that allows an automatic detection and correction at monthly resolution of precipitation data. In this case, the procedure was chosen to check the homogeneity of preselected series but the proposed correction was not applied. The chosen version of the procedure detects and corrects inhomogeneities in precipitation series, at a monthly or daily resolution, using a logarithmic transformation (which treats low values differently, see Domonkos 2015 and Fig[3.4]) in order to apply similar methods to the ones used in temperature series Domonkos (2015).

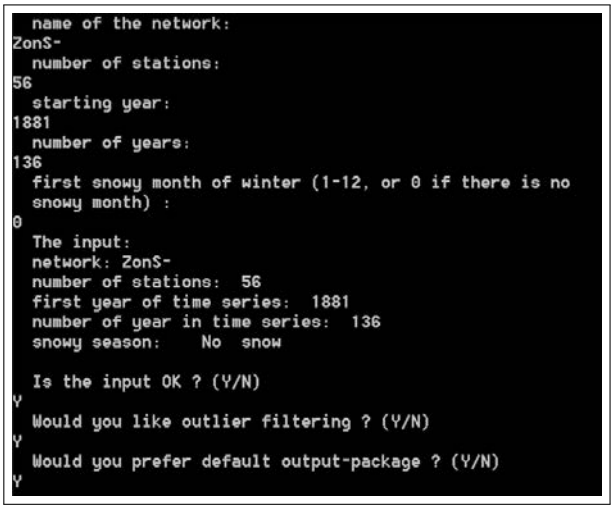


Figure 3.2: screenshot of execution of ACMANT on a selection of 56 series in southern part of the study area.

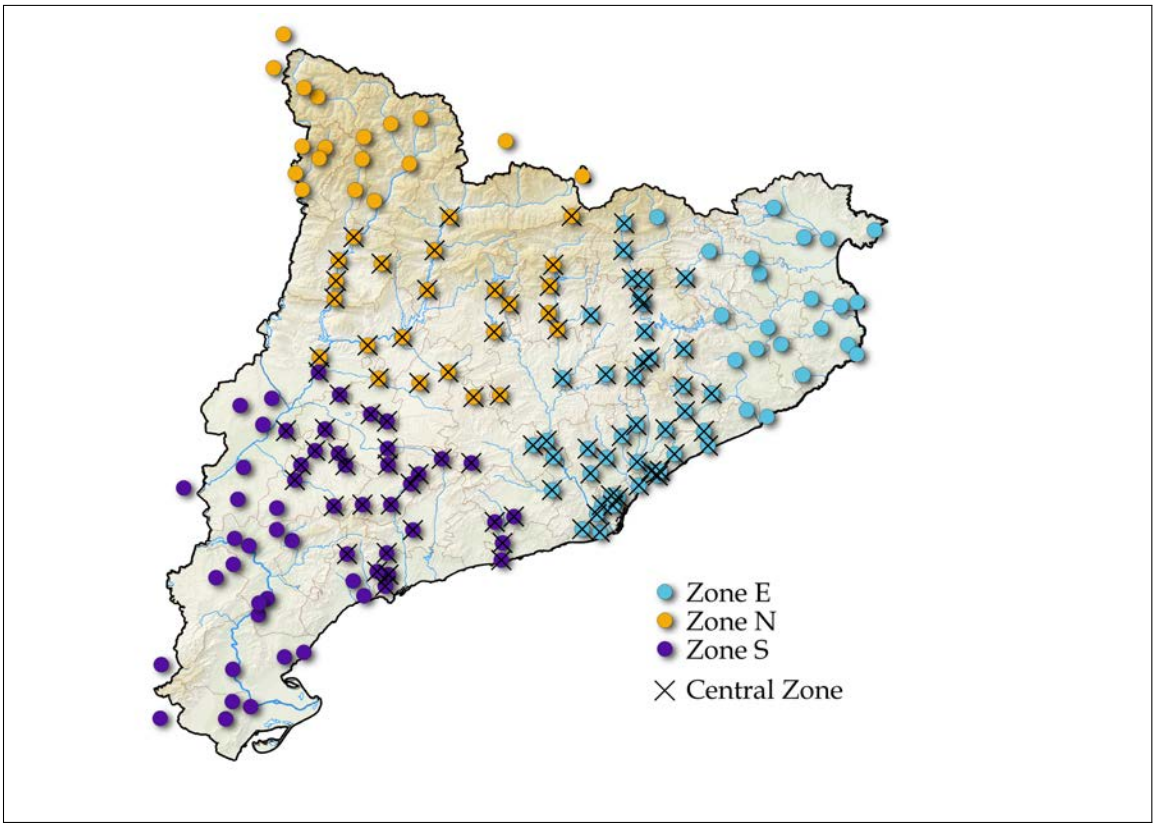


Figure 3.3: map of the 164 pre-selected series that undergo the homogenisation process. Each colour marks one of the zones and the crosses, the overlapping central zone.

The choice of the homogenisation procedure has been made because of the reported good performance of ACMANT (Domonkos and Coll, 2016) and valuing its complete automation and design specific for precipitation. Even if the used version of the program allows the homogenisation of daily precipitation, the procedure has been applied at monthly resolution, giving the fact that only detection was needed (not correction). The procedure also includes the option of dividing the year in a period of rain and another period of snow, but this possibility has been disregarded because few series in the network, with respect to the total, collect snow during the minimum number of required months.

Metadata is not used automatically in ACMANT and it could have been used manually to check the breaks proposed by the homogenisation procedure. However, this check has not been performed because even if metadata is available for the project series, they might be incomplete, especially in former years.

The subset of pre-selected series for homogenisation consists in 164 series in the time period 1881-2016 (because in previous years there is not a network of available series). The pre-selected 164 series have, in the period 1942-2016, a minimum length of 25 years with 99% of daily data and, at each of these years, they perform 80 or above

in the absolute quality control and have a maximum of 1% of suspect not valid values (label "N") in the automatic relative control. These criteria are used to pre-select the series, nonetheless, all the available years in the period 1881-2016 of the pre-selected series have been considered in the process of homogenisation. There were 227 series that fulfilled the pre-selection requirements, however, the final dataset of series to undergo the homogenisation process was trimmed manually from the group of series that complied with the aforementioned criterion; the only objective was to avoid different series with very similar locations to enter the next phase of the project and obtain a uniformly dense network while avoiding nearby points with different information. Therefore, the 164 pre-selected series in the pre-homogenisation stage are a uniform network of carefully selected long series with good quality. Indeed, Pearson's correlation coefficient calculated by ACMANT from monthly data indicates that each of the 164 series has a minimum of 5 auxiliaries with $r > 0.50$ and a minimum of 3 auxiliaries with $r > 0.70$. The two series from the pre-selected 164 that have the best correlation of their monthly accumulated rainfall are -*(BN007) - Observatori Fabra* and *(BN032) - Barcelona - SMN (Travessera de Dalt)* with $r = 0.95$.

ACMANT accomplishes an automatic creation of a reference series for each candidate by combining available series after checking Pearson's correlation coefficient (using a minimum of 0.4) but the geographic distance is not taken into account. Given the fact that the 3^{rd} version of ACMANT allows a maximum of 99 series per execution of the program, the selected series have been grouped before the application of the homogenisation procedure.

In particular, two different approaches to the reduction in the number of series have been followed. In the first approximation, the executable ACMANP3month has been applied to groups of 50 series in which a selection of reference series has been performed for a specific candidate and the results of the homogenisation are read only for that candidate at each execution; hence, in this approach there have been 164 executions of ACMANT after careful selection of the series that were fed to the automatic procedure. On a second approach, ACMANT has been applied to four groups which divide the territory in three zones plus a central zone which covers the borders of the other three (see Fig[3.3]). Therefore, there have been a total of 168 executions of ACMANT homogenisation procedure and the same number of results have been analysed; Fig[3.2] shows a screenshot of one of the performed executions of ACMANP3month.

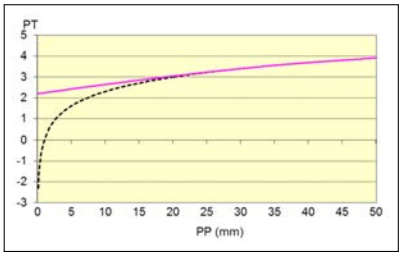


Figure 3.4: transformation of precipitation data from PP (original value of precipitation) to PT (transformed data) following PT = log(PP) if PP $\geqslant$ 30 mm and PT $= \log(0.4\text{PP} + 0.01\text{PP}^2 + 9)$ if PP < 30 mm (solid line) compared to the pure logarithmic transformation. Source: Domonkos (2015).

In Table[3.2] the resulting number of breaks is summarised. The magnitude of the

break is indicated in ACMANT as a the multiplicative factor that adjusts the monthly accumulated precipitation in the pre-break period to the last period of the series. The highest magnitude of a break found in the 164 series is 1.64 and it is the only break that indicates a correction over 60% (either because the original series underestimates or overestimates rainfall).

	Total number of breaks	Between 0.50 and 0.75	Between 0.75 and 1.00	Between 1.00 and 1.25	Between 1.25 and 1.50	Over 1.50
By groups of selected series	249	13	125	124	29	3
Southern zone	49 (in 56 series)	0	23	21	5	0
Northern zone	95 (in 43 series)	9	42	26	14	4
Eastern zone	84 (in 65 series)	5	39	30	10	0
Overlapping Central zone	137 (in 97 series)	3	59	64	10	1

Table 3.2: number of detected breaks by ACMANT in the 164 pre-selected series for the period 1881-2016.

The cases with a proposed correction over 1.5 (i.e., ACMANT proposes to increase rainfall in 50%) are five series in the 4 zones, considering any of the zones, namely: BG011 (1.58), BG028 (1.53), VA012 (1.53), AD006 (1.57) and AG001 (1.6) and three series in the approach with selected reference series, namely: VA012 (1.51), AD006 (1.58) and AG001 (1.64). All these series are considered to need an excessive correction, therefore, they are disregarded from the selection because of their inhomogeneity.

In a further step to select the best series for the subsequent project, the Mann-Kendall rank test (Mann 1945, Kendall 1975) was applied to series of annual maximum precipitation in 24h obtained from selected years of 380 measuring sites (a broader set selected from available data than the series which underwent the ACMANT homogenisation procedure, consisting in series with at least 20 years of data in the period 1942-2016). The results show that 29 series had a statistically significant trend, hence they are not appropiate for the subsequent study as they could influence the outcome dependending on which years were considered; from these 29 series, only 11 series were in the 164 series pre-selection.

3.3 Selection of homogeneous daily rainfall series

Considering the results of the quality control and the homogenisation checks a selection of series has been performed. This selection has been adjusted to achieve a spatially uniform selection (incorporating some series outside from Catalonia's borders and rejecting series too close to other series, selecting the best or longest ones). The finally selected series are shown in Fig[3.5] (in blue and green dots); the same map displays the rest of the series that fulfil the criteria of the pre-selection performed before homogenisation but were disregarded in order to obtain a more uniform density in the territory (black dots) and also the series that were rejected by the homogenisation procedure or the Mann-Kendall rank test (yellow and red crosses respectively).

Series were pre-selected when they had a minimum of 25 years in the period 1942-2016 that fulfil all of the following requirements:

- None of the months of the year were marked as suspect in the metadata.
- The year scores at least 80 in the global Quality Index.
- A maximum of 1% of days in the year have missing data.
- The individual quality indices ${Q_{zero}}^m$ and ${Q_{zero}}^w$ score at least 50 (the score of the other individual quality indices was not considered, in the case of Q_{gaps} because it does not make sense when the restriction of 99% completeness is already fulfilled and in the case of $Q_{outliers}$ because it was used to classify the available series according to their quality and proceed to the automatic relative quality control, but now, when the results of that quality control are taken into account, outliers are already filtered).
- A maximum of 1% of days in the year are marked as "N" (not valid data) in the automatic quality control that labels data according to a relative comparison with auxiliary stations.

Finally, the selection of series to be used in subsequent chapters consists in 163 series that fulfil the pre-selection criterion, summarised above, in the period 1942-2016 and have also passed a homogenisation test or a tendency test in the following way:

- 150 of them are considered homogeneous because the inhomogeneities detected by ACMANT have a magnitude under the chosen threshold and they also pass the Mann-Kendall rank test, therefore, they do not have any detectable tendency in the series of annual maximum precipitation in 1-day (green dots in Fig[3.5]).
- 13 of them pass the Mann-Kendall rank test, although the homogeneisation with ACMANT has not been performed (blue dots in Fig[3.5]). All of them are outside the borders of the study area (Catalonia) but they have interest for the spatial interpolation.

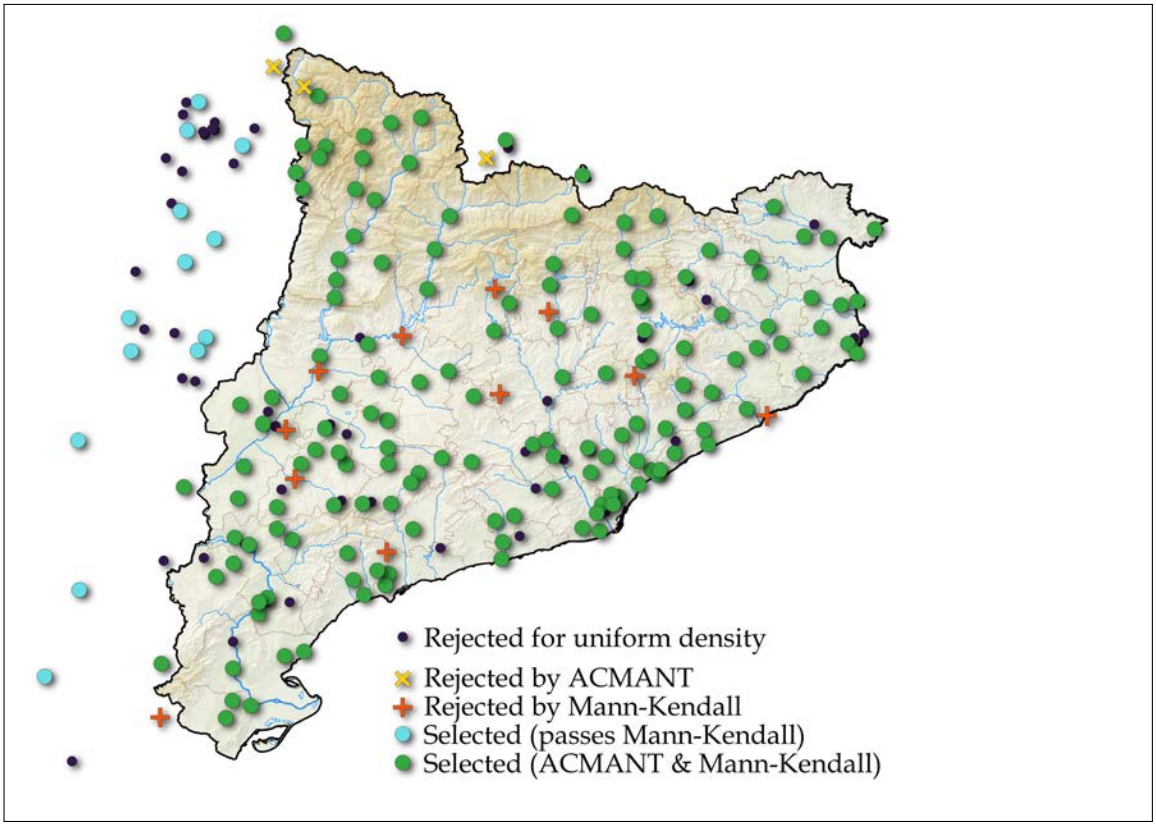


Figure 3.5: final selection of 163 series (green and blue dots) and rejected pre-selected series by homogeneity checks (yellow crosses) or Mann-Kendall rank test (red crosses) as well as disregarded series that comply with the pre-selection criteria but are too close to other series that are considered to be better (after checking their quality and completeness in the period 1942-2016).

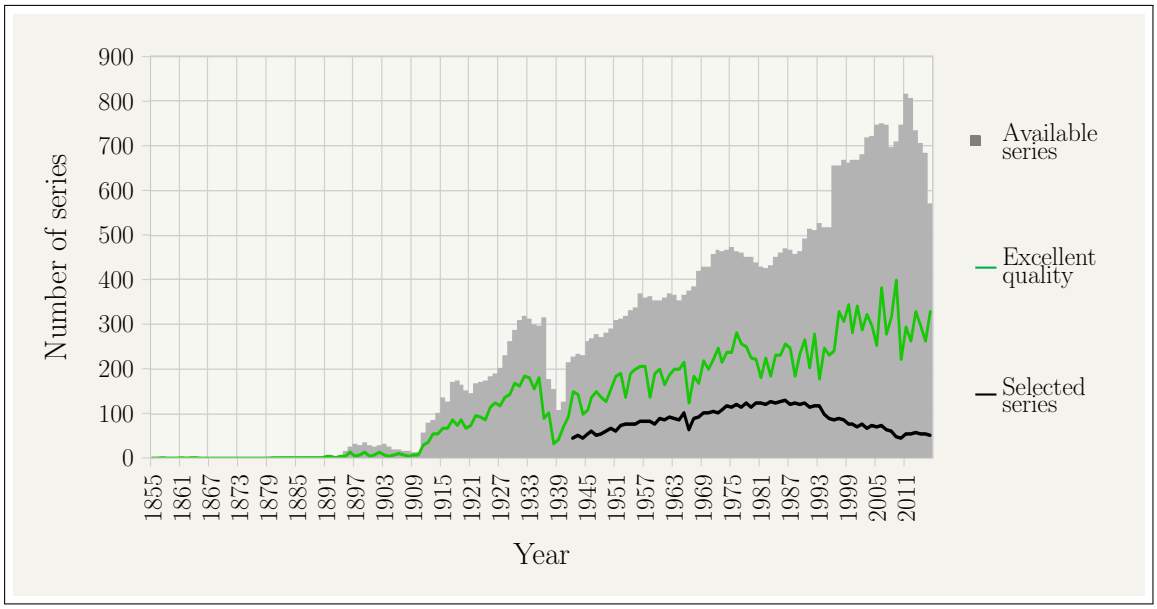


Figure 3.6: number of rainfall series available over the 1855-2016 period (grey bars), number of excellent quality series as defined in the previous chapter (green line) and number of selected series in the period 1942-2016 (black line).

Only the years of a selected series that fulfil each of the aforementioned requirements for the pre-selection of series will be considered in subsequent parts of the project, the rest of the years, even if they are part of a selected series, are discarded because they might contain suspect data. The 1942-2016 period has been chosen because it allows a pre-selection of series with constant density and completion of data.

Fig[3.6] displays the availability of selected series at each considered year in comparison to the total number of series considered in the project and the number a filter of excellent quality series. In recent years, even if there is an increase in the number of available series as well as in their quality, the number of selected series declines because the good quality series are shorter than the minimum 25 years requirement.

Table[3.3] shows the characteristics of the 163 selected series, the code used for identification, the name county and name of the station, the altitude and coordinates, and characteristics of length, completeness and selected years.

ID	County	Name	Altitude (m)	X (UTM) ETRS89 / UTM zone 31N	Y (UTM) ETRS89 / UTM zone 31N	Initial data	Last data	Series length (number of years)	Completness (%)	Number of selected years	First selected year	Last selected year
AC010	Alt Camp	Valls - El Fornàs	232	354383	4572742	01-10-1990	31-12-2016	26.3	99.8	25	1992	2016
AE010	Alt Empordà	Castelló d'Empúries - 2	17	505932	4678988	01-11-1951	31-12-1994	43.2	98.2	38	1952	1994
AE024	Alt Empordà	Darnius - 2	193	486183	4690380	01-04-1969	31-12-2016	47.8	98.7	25	1970	2008
AE041	Alt Empordà	Figueres - Escola La Salle	40	497272	4679572	01-06-1943	31-03-1984	40.8	100.0	40	1944	1983
AE052	Alt Empordà	Cadaqués	30	523076	4682161	01-01-1911	30-11-2016	105.9	93.2	55	1942	2015
AP011	Alt Penedès	Vilafranca del Penedès - Estació Enològica	227	391295	4577663	01-04-1914	31-12-2013	99.8	99.5	70	1942	2013
AP023	Alt Penedès	Gelida	120	405213	4587360	01-01-1917	31-12-1995	79.0	84.7	40	1950	1995
AP024	Alt Penedès	Pantà de Foix	104	387150	4568030	01-01-1911	31-12-2016	106.0	93.4	51	1942	2016
AP027	Alt Penedès	Castellví de la Marca	190	384319	4575694	01-01-1949	30-11-2016	67.9	93.8	44	1950	2014
AU013	Alt Urgell	Adrall	627	367796	4686950	01-07-1927	31-01-2002	74.6	87.3	50	1942	1995
AU022	Alt Urgell	Organyà - 3	549	362302	4674780	01-03-1950	30-11-1999	49.8	100.0	29	1970	1998
AU024	Alt Urgell	Oliana - Casa de la CHE	468	359629	4660265	01-05-1950	31-10-1999	49.5	99.0	44	1951	1997
AR001	Alta Ribagorça	Central de Senet	1093	313991	4712747	01-10-1915	31-12-1997	82.3	73.2	35	1952	1996
AR002	Alta Ribagorça	Central de Vilaller	925	311508	4702924	01-03-1928	30-09-1999	71.6	77.2	36	1942	1998
AR004	Alta Ribagorça	Caldes de Boí - Central	1298	322386	4712379	01-01-1954	31-12-1998	45.0	97.5	35	1954	1996
AR008	Alta Ribagorça	Boí	1096	320134	4708355	01-10-1923	31-12-1998	75.3	61.7	31	1949	1998
AR014	Alta Ribagorça	el Pont de Suert - Central Elèctrica	830	313967	4696969	01-06-1945	31-12-1998	53.6	96.2	44	1946	1998
AD004	ANDORRA	Ransol	1645	388277	4714714	01-10-1934	31-12-2015	81.3	100.0	74	1942	2015
AN004	Anoia	El Bruc	273	398259	4603715	01-01-1948	31-12-2016	69.0	99.1	61	1948	2016
AN007	Anoia	Calaf - 2	715	376541	4621109	01-08-1960	31-08-1993	33.1	89.2	25	1961	1992
AN027	Anoia	Santa Maria de Miralles	640	375737	4597277	01-09-1930	31-05-1975	44.7	85.3	27	1942	1974
ES021	ARAGON	ID: 9559 - FID: 4466 - SANTOLEA (EMBALSE)	644	219821	4518836	01-02-1923	31-07-1996	73.5	89.8	45	1943	1995
ES029	ARAGON	ID: 9572 - FID: 4475 - ALCÃ¤roÃ®IZ LA ESTANCA	344	232502	4550486	01-09-1932	31-08-2015	83.0	65.0	37	1943	2012
ES033	ARAGON	ID: 9576 - FID: 4479 - LA ALMOLDA	474	232026	4605141	01-03-1965	31-07-2015	50.4	97.9	40	1966	2014
ES036	ARAGON	ID: 9579 - FID: 4482 - MEQUINENZA	241	270662	4588122	01-08-1949	31-07-2015	66.0	97.5	35	1952	2014
ES048	ARAGON	ID: 9783 - FID: 4592 - URDICETO CENTRAL	1897	276081	4728627	01-11-1932	31-07-1990	57.7	64.8	32	1950	1985
ES057	ARAGON	ID: 9794 - FID: 4601 - SALINAS DE BIELSA	799	271913	4718263	01-02-1928	30-11-1989	61.8	74.4	34	1952	1986
ES069	ARAGON	ID: 9829 - FID: 4628 - MEDIANO (EMBALSE)	516	269342	4688764	01-01-1922	30-09-2008	86.7	81.7	44	1943	2007
ES073	ARAGON	ID: 9833 - FID: 4632 - EL GRADO (EMBALSE)	415	271164	4670115	01-01-1928	30-04-1999	71.3	61.2	31	1961	1998
ES077	ARAGON	ID: 9842 - FID: 4636 - VILLANOVA (PRESA)	930	292099	4712690	01-01-1946	31-07-1985	39.6	96.6	32	1946	1983
ES080	ARAGON	ID: 9849 - FID: 4639 - GRAUS (VENTAS DE SANTA LUCIA)	498	281858	4678545	01-01-1928	30-09-1990	62.7	88.6	33	1942	1974
ES102	ARAGON	ID: 9874 - FID: 4662 - BERBEGAL	460	250656	4649623	01-08-1951	31-08-2015	64.1	99.1	42	1952	2013
ES103	ARAGON	ID: 9875 - FID: 4663 - EL TORMILLO-TERREU	417	251387	4637552	01-03-1949	31-07-2015	66.4	98.0	38	1953	2013
ES113	ARAGON	ID: 9913 - FID: 4696 - SAN ESTEBAN DE LITERA	438	278715	4642549	01-04-1931	31-07-2015	84.3	76.0	38	1952	2010
ES115	ARAGON	ID: 9915 - FID: 4698 - BINEFAR GRANJA	289	275750	4637763	01-08-1935	31-12-1989	54.4	83.8	32	1942	1989
ES121	ARAGON	ID: 9927 - FID: 4708 - BECEITE	567	262460	4523549	01-07-1933	31-07-2015	82.1	72.9	44	1946	2013
BA007	Bages	Sallent - Cabrianes	258	409005	4628041	01-03-1916	31-12-2004	88.8	82.7	42	1942	2004
BA053	Bages	Abadia de Montserrat	740	403253	4605397	01-10-1901	31-12-2016	115.3	69.4	47	1942	2016
BC011	Baix Camp	Montbrió del Camp - C. de Barcelona	133	332587	4554125	01-07-1950	31-07-2009	59.1	98.2	41	1952	2006
BC013	Baix Camp	Cambrils	18	336501	4548704	01-01-1916	31-05-1993	77.4	81.8	34	1943	1992
BC015	Baix Camp	Alforja	365	330331	4563975	01-01-1982	31-12-2016	35.0	98.9	28	1982	2012
BC023	Baix Camp	Aeroport de Reus	73	345491	4556454	01-02-1945	31-12-2016	71.9	99.5	68	1946	2016
BC026	Baix Camp	Reus - Centre de Lectura	133	341333	4557610	01-03-1934	31-12-2016	82.8	84.5	57	1951	2016
BB003	Baix Ebre	el Perelló - 3	142	307471	4526340	18-02-1960	31-12-2016	56.9	99.2	50	1961	2016
BB006	Baix Ebre	l'Ametlla de Mar - 3	22	314541	4528011	01-10-1948	31-08-2007	58.9	98.2	40	1949	2004
BB017	Baix Ebre	Observatori de l'Ebre	49	288608	4521819	01-01-1905	31-12-2016	112.0	99.4	74	1942	2016
BE004	Baix Empordà	Monestir de Santa Maria de Solius	75	496877	4629205	01-02-1973	31-12-2016	43.9	100.0	42	1974	2016
BE016	Baix Empordà	Palafrugell - 4	81	513318	4640267	01-10-1956	31-01-2012	55.3	98.4	39	1957	2010
BE019	Baix Empordà	Calella de Palafrugell	15	516500	4636719	01-02-1947	31-03-1989	42.2	98.6	32	1949	1988
BE031	Baix Empordà	La Bisbal d'Empordà - 5	50	503504	4646201	15-04-1988	31-12-2016	28.7	100.0	28	1989	2016
BE039	Baix Empordà	Jafre - 1	90	499907	4657055	01-12-1928	30-11-1998	70.0	90.2	45	1942	1997
BE047	Baix Empordà	Torroella de Montgrí - 3	23	510730	4654390	01-09-1977	31-12-2016	39.3	96.6	33	1978	2016
BE050	Baix Empordà	l'Estartit - 2	2	516672	4655775	01-01-1976	31-12-2016	41.0	100.0	41	1976	2016
BL006	Baix Llobregat	Gavà - Miranda	25	416259	4573096	01-03-1977	31-12-2016	39.8	88.1	31	1978	2016
BL009	Baix Llobregat	Aeroport de Barcelona	6	422668	4572008	01-03-1924	31-12-2016	92.8	84.4	69	1944	2016
BL021	Baix Llobregat	Esparreguera - Aigües d'Esparreguera	204	405663	4599382	01-02-1979	31-12-2016	37.9	98.5	32	1980	2016
BL045	Baix Llobregat	Esplugues de Llobregat - Finestrelles	145	423446	4581839	01-03-1979	31-12-2016	37.8	98.4	26	1980	2016
BL047	Baix Llobregat	Cornellà de Llobregat	13	421553	4578589	01-01-1917	31-12-2002	86.0	70.9	34	1942	1998
BN007	Barcelonès	Observatori Fabra	410	426770	4585408	01-07-1913	31-12-2016	103.5	100.0	75	1942	2016
BN022	Barcelonès	Barcelona - Can Bruixa	61	427507	4581733	01-06-1987	31-12-2016	29.6	100.0	29	1988	2016
BN032	Barcelonès	Barcelona - SMN (Travessera de Dalt)	94	429434	4584581	01-08-1936	31-07-1970	34.0	100.0	28	1942	1969
BN060	Barcelonès	Badalona - Museu Municipal	38	436912	4589080	01-06-1966	09-01-2003	36.6	96.1	32	1967	2002
BG011	Berguedà	Central Tèrmica de Cercs	672	405648	4669470	01-04-1936	31-05-2002	66.2	78.9	29	1943	1998
BG017	Berguedà	Berga - Cal Corneta	760	404424	4661711	01-01-1956	30-11-1988	32.9	100.0	32	1956	1987
BG028	Berguedà	Puig-reig - 2	434	407172	4645817	01-06-1972	31-12-2016	44.6	96.2	29	1973	2012
CE002	Cerdanya	Llívia - 2	1220	416297	4701912	01-12-1948	31-12-1977	29.1	99.3	27	1950	1977
CE010	Cerdanya	la Molina	1704	412370	4687275	01-04-1927	19-09-2002	75.5	66.4	38	1957	1997
CB003	Conca de Barberà	Vimbodí - Riudabella	595	335962	4582033	01-01-1912	30-06-2000	88.5	68.0	28	1957	1999
CB011	Conca de Barberà	Montblanc - 4	353	346353	4582058	01-10-1949	31-12-2016	67.3	96.9	55	1950	2016
CB015	Conca de Barberà	Rocafort de Queralt - 2	560	356457	4593220	01-09-1941	31-12-2016	75.3	96.0	61	1942	2014
CB017	Conca de Barberà	Sarral - 2	473	353598	4589634	01-01-1951	31-12-2016	66.0	89.6	46	1952	2016
CB022	Conca de Barberà	Santa Coloma de Queralt - 4	687	364881	4598753	01-05-1982	28-11-2010	28.6	99.6	25	1983	2007
AG003	FRANCE	Cierp - Gaud	500	307089	4753715	01-03-1952	31-03-2013	61.1	100.0	60	1953	2012
GF001	Garraf	Cubelles - les Salines	6	386684	4561776	01-07-1953	31-05-2015	61.9	96.0	48	1954	2014
GA002	Garrigues	els Omellons - Cal Teresa	377	329756	4596456	20-02-1944	31-07-1991	47.4	99.9	44	1945	1990
GA004	Garrigues	Arbeca - 1	309	327211	4600767	01-11-1968	31-12-2016	48.2	98.8	40	1969	2016
GA011	Garrigues	Juneda - 2	262	318679	4601733	01-11-1959	31-12-2016	57.2	95.8	52	1960	2016
GA013	Garrigues	Castelldans	352	313495	4596560	01-03-1984	31-12-2016	32.8	99.9	32	1985	2016
GA017	Garrigues	la Pobla de Cérvoles - 2	672	325504	4581506	01-11-1988	31-12-2016	28.2	100.0	27	1989	2016
GA027	Garrigues	La Granadella	526	304676	4580803	01-04-1954	31-08-1990	36.4	99.1	31	1955	1989
GX009	Garrotxa	la Vall d'en Bas - Mas Can Gronxa	478	453974	4664702	01-07-1979	31-12-2016	37.5	99.1	26	1981	2016
GX021	Garrotxa	Castellfollit de la Roca	286	462603	4674424	01-07-1916	31-12-2016	100.5	59.4	40	1955	2016
GX030	Garrotxa	Maià de Montcal - Can Garriga	159	478148	4671814	01-07-1985	31-12-2016	31.5	99.7	29	1986	2016
GI012	Gironès	Llambilles	143	488849	4640410	01-10-1945	30-06-1980	34.7	93.0	28	1946	1979
GI017	Gironès	Girona - Col·legi Bell-lloc	90	484094	4646587	01-01-1970	31-10-2005	35.8	97.7	25	1974	2004
MA005	Maresme	Alella - Camí de Teià	109	441479	4594377	01-02-1917	31-12-1994	77.9	99.3	49	1943	1994
MA009	Maresme	el Masnou - Sindicat de Pagesos	10	443857	4592637	01-03-1952	31-12-1994	42.8	98.1	33	1955	1994
MA015	Maresme	Teià - Can Barrera	140	444632	4594353	01-01-1911	31-12-2004	94.0	93.8	38	1942	2004
MA026	Maresme	Argentona - 2	85	449938	4600482	01-06-1945	31-01-1994	48.7	98.5	32	1946	1993
MA042	Maresme	Arenys de Munt - Collsacreu	390	460686	4608803	01-07-1977	31-07-2006	29.1	99.9	28	1978	2005
MA048	Maresme	Arenys de Mar - el Xifré	37	462197	4603598	01-12-1985	31-12-2016	31.1	99.9	30	1986	2016
MA059	Maresme	Tordera - 3	30	476330	4616384	01-11-1953	31-10-2004	51.0	99.5	46	1954	2002
MI006	Moianès	Moià - EEPP	740	424859	4629439	01-01-1911	08-02-2016	105.1	87.8	57	1942	2015
MO014	Montsià	Godall - 5	170	285989	4503709	01-10-1969	31-12-2016	47.3	97.9	27	1971	2016
MO016	Montsià	Santa Bàrbara	79	288424	4510011	01-03-1945	31-01-1996	50.9	99.0	42	1947	1994
MO018	Montsià	Amposta - 2	8	295124	4508163	01-01-1944	31-05-1983	39.4	97.7	32	1944	1982
NO002	Noguera	Artesa de Segre	317	337848	4640027	01-11-1968	31-12-2016	48.2	98.5	39	1969	2016
NO011	Noguera	Presa de Sant Llorenç de Montgai	240	320300	4635739	01-01-1927	31-03-1996	69.2	77.2	36	1942	1993
NO024	Noguera	Penelles - Granja de Sant Vicenç Ferrer	242	327637	4621997	01-05-1930	31-12-1974	44.7	88.8	32	1942	1974
OS005	Osona	Prats de Lluçanès - C. de l'Hospital	701	419397	4651026	31-05-1933	28-02-2009	75.8	92.9	58	1942	2007
OS009	Osona	Malla - Torrellebreta	573	438364	4633823	01-04-1929	31-12-2016	87.8	98.1	66	1942	2016
OS016	Osona	la Farga de Bebié	608	434406	4664455	01-01-1917	31-03-1986	69.2	89.8	40	1942	1985
OS017	Osona	Santa Maria de Besora	861	438630	4664055	01-09-1964	31-12-2016	52.3	97.9	27	1969	2016

ID	County	Name	Altitude (m)	X (UTM) ETRS89 / UTM zone 31N	Y (UTM) ETRS89 / UTM zone 31N	Initial data	Last data	Series length (number of years)	Completness (%)	Number of selected years	First selected year	Last selected year
OS023	Osona	Borgonyà - Colònia Fabra i Coats	524	437113	4657362	01-02-1912	31-12-2016	104.9	92.6	68	1943	2016
OS026	Osona	Torelló	350	438341	4655073	01-01-1947	31-12-2016	70.0	98.5	35	1949	2015
OS041	Osona	Taradell - Avda. de Catalunya	632	440861	4635605	01-01-1976	31-12-2007	32.0	99.7	31	1976	2007
OS055	Osona	Gurb - la Cabra	466	439029	4645148	01-06-1961	31-12-2016	55.6	97.0	44	1964	2015
OS060	Osona	Vilanova de Sau - el Tortadès	832	453358	4638720	01-09-1956	31-12-2016	60.3	96.4	31	1957	2013
PJ003	Pallars Jussà	Estany Gento	2142	335876	4708096	01-08-1925	31-12-1985	60.4	98.8	42	1942	1985
PJ006	Pallars Jussà	Molinos	981	333294	4696777	01-02-1928	31-12-1995	67.9	96.3	46	1942	1995
PJ015	Pallars Jussà	la Pobla de Segur - Central Hidráulica	525	332759	4679579	01-01-1927	28-02-1994	67.2	77.4	37	1942	1993
PJ017	Pallars Jussà	Talarn - Central de Nerets	430	327167	4671155	01-09-1915	29-02-1996	80.5	97.2	49	1942	1994
PJ024	Pallars Jussà	Abella de la Conca	956	342954	4669908	01-11-1928	31-05-1971	42.6	88.6	25	1942	1969
PJ026	Pallars Jussà	Gavet de la Conca - Central de Reculada	382	326312	4663610	01-07-1937	31-03-1994	56.8	96.5	50	1942	1993
PJ029	Pallars Jussà	Central Hidroelèctrica de Terradets	360	325758	4657087	01-09-1936	31-05-1996	59.7	76.0	37	1942	1993
PS006	Pallars Sobirà	Esterri d'Àneu - Central Hidroelèc. i Bombers	952	346296	4721003	01-05-1955	30-06-2005	50.2	85.9	42	1956	2004
PS009	Pallars Sobirà	Estany de Sant Maurici	1906	336496	4716178	01-07-1953	31-12-2016	63.5	65.1	26	1954	2011
PS014	Pallars Sobirà	Tavascan - Central Elèctrica	1126	357323	4722967	01-12-1920	31-12-1994	74.1	79.9	32	1942	1993
PS018	Pallars Sobirà	Llavorsí	802	353107	4706430	01-12-1915	31-05-1999	83.5	64.0	38	1960	1998
PS025	Pallars Sobirà	Escós	790	340378	4692866	01-12-1920	29-02-1988	67.2	98.0	46	1942	1987
PE008	Pla de l'Estany	Fontcoberta - Melianta	224	481048	4666347	01-01-1984	31-12-2016	33.0	91.2	27	1984	2016
PU007	Pla d'Urgell	Mollerussa - IES Agrària l'Urgell	253	322273	4609458	01-04-1961	31-12-2016	55.8	98.4	45	1962	2016
PR011	Priorat	Cabacés	362	310237	4568904	01-08-1951	31-12-2016	65.4	99.5	60	1952	2016
RE001	Ribera d'Ebre	Riba-roja d'Ebre - 1	62	289245	4569650	01-12-1931	31-07-1980	48.7	90.5	31	1942	1979
RE005	Ribera d'Ebre	Flix - la Central	42	294495	4567016	01-06-1956	30-04-1999	42.9	90.5	35	1960	1998
RE006	Ribera d'Ebre	la Palma d'Ebre	340	304623	4572854	01-04-1952	29-02-1988	35.9	99.2	31	1954	1986
RE023	Ribera d'Ebre	Benissanet	30	301215	4547698	01-09-1911	30-11-2013	102.3	90.3	48	1942	2008
RE029	Ribera d'Ebre	Rasquera - 2	179	297937	4541725	01-04-1959	31-12-2016	57.8	91.8	42	1960	2016
RE031	Ribera d'Ebre	Miravet - 1	40	298143	4546014	01-07-1949	31-12-2016	67.5	98.6	47	1950	2016
RI006	Ripollès	Vilallonga de Ter	1056	443608	4686892	01-12-1928	31-01-1979	50.2	84.2	28	1942	1977
RI023	Ripollès	Ribes de Freser	940	431570	4684625	01-03-1916	31-05-1988	72.3	77.5	26	1942	1987
RI025	Ripollès	Campdevànol - 2	734	431279	4674834	01-06-1945	31-12-2016	71.6	99.9	70	1946	2016
SG001	Segarra	Torà	434	367305	4630258	01-02-1930	28-02-2003	73.1	91.2	55	1942	1999
SG006	Segarra	Torrefeta - Granges Nial	478	356775	4626280	01-06-1981	31-12-2016	35.6	90.5	26	1982	2016
SE007	Segrià	Vilanova de Segrià	222	302854	4620647	01-06-1983	31-12-2016	33.6	97.2	29	1984	2016
SE020	Segrià	Lleida - Observatori 2	192	299600	4611106	01-03-1983	31-12-2016	33.8	100.0	33	1984	2016
SE031	Segrià	Maials - Mas de Marquet	326	290419	4583961	01-07-1979	31-12-2016	37.5	98.7	33	1981	2015
SE039	Segrià	Pantà d'Utxesa	170	292487	4595611	01-01-1927	31-03-1988	61.2	81.8	33	1942	1983
SE044	Segrià	Raimat - Caseta del Canal	320	291296	4618046	01-02-1938	31-12-2016	78.9	85.1	40	1942	2015
SV002	Selva	Breda - C. Joan XXIII	176	463530	4622461	01-01-1977	31-07-2009	32.6	99.9	31	1977	2008
SV018	Selva	Santa Coloma de Farners - Can Font de Glòria	149	472161	4634641	01-02-1978	31-12-2016	38.9	93.1	28	1979	2016
SV035	Selva	Amer	189	467155	4651035	01-01-1962	31-12-2016	55.0	86.1	40	1962	2016
SV048	Selva	Aeroport de Girona - Costa Brava	127	480022	4638732	01-01-1973	31-12-2016	44.0	99.0	42	1973	2016
SO005	Solsonès	Riner - Xixons	611	384044	4644945	01-07-1961	31-12-2016	55.5	98.9	50	1962	2016
SO011	Solsonès	Navès - Tentellatge	845	389554	4655091	01-04-1936	31-05-2001	65.2	92.0	50	1942	1996
TG001	Tarragonès	Vila-seca	36	344301	4552210	01-01-1929	31-12-2016	88.0	98.1	47	1942	2016
TA011	Terra Alta	Vilalba dels Arcs	456	282449	4555323	01-12-1949	31-12-2016	67.1	99.1	62	1950	2016
TA013	Terra Alta	la Fatarella - 2	486	288766	4560189	01-01-1965	31-12-2003	39.0	97.0	32	1965	2002
UG004	Urgell	Agramunt - Plaça de l'Amball	342	341843	4627918	14-10-1966	31-12-2015	49.2	99.8	48	1967	2015
UG013	Urgell	Tàrrega - Estació Meteorològica Municipal	375	344942	4612344	01-07-1936	31-12-1981	45.5	91.0	35	1942	1981
UG017	Urgell	Anglesola	313	338980	4614881	01-05-1971	31-01-2001	29.8	95.8	26	1972	1999
UG019	Urgell	Rocallaura	637	345288	4596651	01-12-1947	31-12-2016	69.1	97.7	62	1948	2016
UG020	Urgell	Ciutadilla	514	345106	4602513	01-06-1983	31-12-2016	33.6	98.7	29	1984	2016
VA010	Val d'Aran	Vielha - Central Hidroelèctrica	951	319631	4730870	01-03-1945	31-08-1993	48.5	98.6	44	1946	1992
VC010	Vallès Occidental	Terrassa - Placeta de la Creu	311	418244	4602501	01-01-1965	31-12-1995	31.0	100.0	30	1965	1995
VC017	Vallès Occidental	Rubí - Escola Montserrat	135	419282	4593462	01-04-1963	31-12-2005	42.8	97.0	35	1965	2005
VC022	Vallès Occidental	Sabadell - Can Marcet	180	424979	4599185	01-01-1937	30-11-1979	42.9	98.7	30	1944	1978
VR030	Vallès Oriental	Cardedeu - Creu de la Serreta	193	446649	4609404	01-11-1950	31-12-2016	66.2	99.2	61	1951	2016
VR040	Vallès Oriental	Santa Eulàlia de Ronçana	210	436062	4611078	01-06-1943	30-11-1981	38.5	97.4	29	1944	1978
VR051	Vallès Oriental	Martorelles - Avda. de Piera	98	436337	4597634	01-01-1963	31-12-2000	38.0	99.3	31	1965	2000
VR057	Vallès Oriental	Caldes de Montbui - Torre Marimon	175	430713	4607107	01-11-1932	22-11-2011	79.1	92.7	44	1942	2008
VR063	Vallès Oriental	Turó de l'Home	1699	453055	4625105	20-11-1932	31-08-2004	71.8	92.5	50	1942	2000
VR078	Vallès Oriental	Santa Maria de Palautordera - C. de l'Hospital	210	453777	4615951	01-03-1977	31-12-2016	39.8	99.8	37	1978	2016

Table 3.3: main characteristics of the 163 selected series.

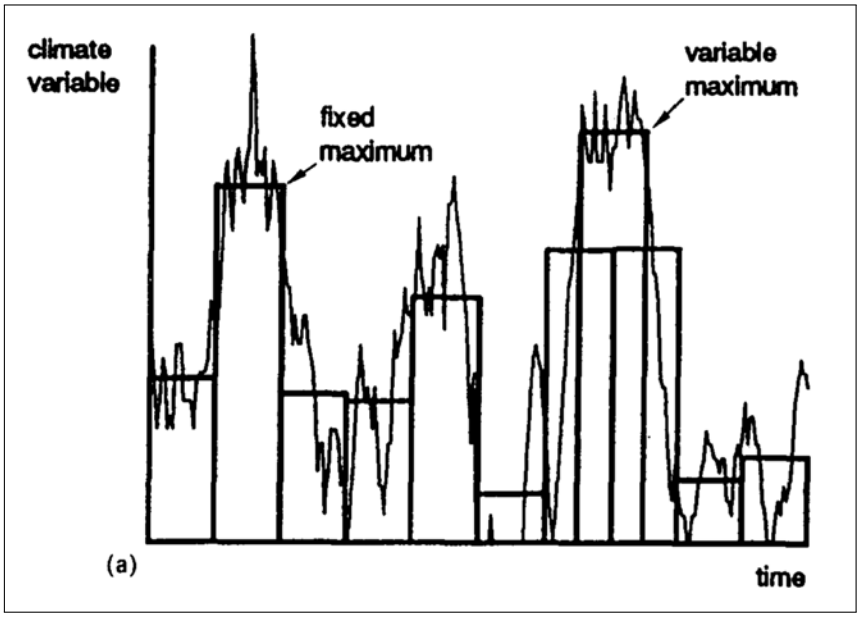
Chapter 4

SERIES OF MAXIMUM RAINFALL

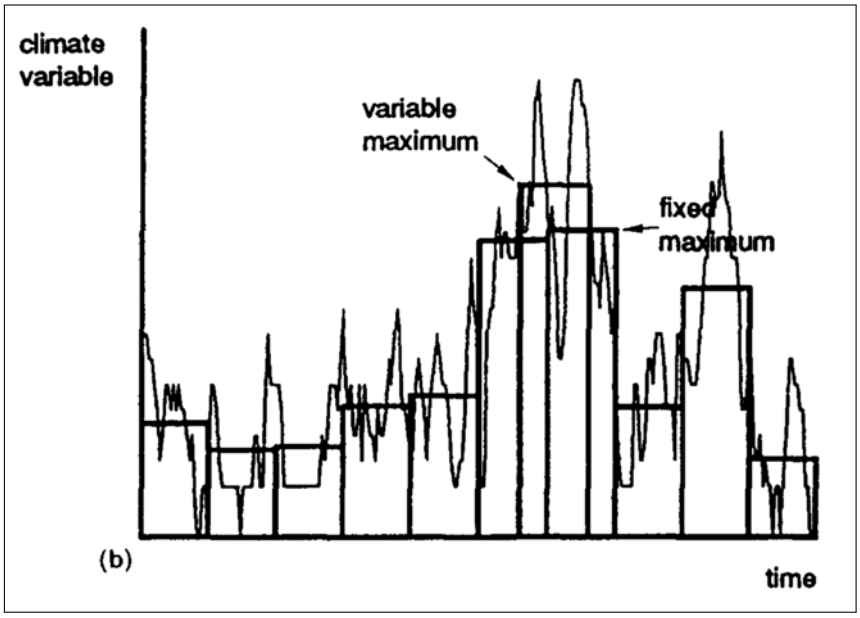
In maximum rainfall studies, as in the obtainment of Intensity-Duration-Frequency relationships, it is common to use systematically recorded data of accumulated precipitation. In the present project, series of annual maximum rainfall in 24 hours are used to obtain expected subdaily precipitation. However, most historical data have been recorded by collecting daily rainfall with a fixed starting time (e.g., local 8:00 a. m.) and it is known that this fact conditions the measure of the true maxima values in unrestricted intervals of 24 hours (Dwyer and Reed 1995, see in Fig[4.1] how easily can a true maximum in an arbitrary interval differ from the maximum in an interval of the same duration that has a restricted starting time). In order to correct the maximum precipitation obtained from measures taken at fixed times as if they had been obtained from a sliding window of the considered duration, it is recommended to carry out a correction which is usually applied through a multiplicative factor.

In the present chapter an empirical correction factor has been obtained to be used as a multiplicative correction of annual maxima of one day. The empirical correction factor is calculated as a ratio of mobile interval over fixed interval maxima for durations of one or more days. Hereafter, maximum rainfall in an interval of one day with arbitrarily fixed starting time might be referred to as *maximum in one day*, whereas true maximum rainfall, maximum rainfall in a sliding interval or corrected maximum rainfall (hence, estimated true maximum) in an interval of the same duration might be referred to as *maximum in 24 hours*. Rainfalls in sliding intervals have been obtained from available hourly records in Catalonia. Moreover, a comparison with existing correction factors and a study of the possible dependence of the ratio between sliding maxima and fixed interval maxima on the characteristic rainfall pattern of the study area have been performed. In particular, several factors, explored in subsequent sections, which might influence the value of this ratio are the following:

- Starting time of the fixed interval. Some maxima can be caused by events centred on a time of the day that is conditioned by the effect of the diurnal cycle, in case rainfall arises from land heating. In the region of study, the diurnal cycle affects especially in the case of summer storms that appear mainly in the afternoon. If



(a) true and fixed maxima as part of the same event.



(b) true and fixed maxima arising from different events.

Figure 4.1: source: Dwyer and Reed (1995)

rainfall episodes occur with this kind of influence, the correction needed would depend on the time at which daily measures were taken.

- Season. The climatological definition of season has been followed, with winter comprising the months of December of the previous year, January and February; spring comprising the months from March to May; summer from June to August and Autumn from September to November.
- Geographical location. The spatial distribution of empirical correction factors has been analysed.
- Actual duration of rainfall events.

4.1 Correction from daily to 24 hours maxima

One of the first studies facing the matter of estimating true maximum rainfall from restricted time measures was performed by Hershfield (1961) who, using data of the United States, proposed an empirical multiplicative factor of 1.13 to correct fixed maxima obtained from daily measures as if a sliding 1440-min window had been used, as well as for correcting fixed hourly measures as if a sliding 60-min window had been used. The fact that Hershfield was one of the first authors to propose a multiplicative correction factor is the reason that this factor is widely known as Hershfield factor regardless of the exact methodology used to obtain it. Indeed, the Hershfield factor was also derived by Weiss (1964) following a theoretical formulation; Weiss (1964) proposed a model based on simple assumptions which yielded different correction values depending on the sampling discretisation. Later studies were, among others, those performed by Huff and Angel (1992), who applied an empirical factor to develop a rainfall frequency atlas for the Midwest of the United States; Dwyer and Reed (1995), who studied rainfall records from the UK and Australia at a sampling resolution of 1 hour to determine empirical correction factors proposing a range from 1.15 to 1.17; van Montfort (1997), who used monthly maxima of daily rainfall from a Chinese dataset and found a dependence with location, autocorrelation and the fraction of wet days; Asquith (1998), who determined an empirical value of 1.13 for the state of Texas; and Young and McEnroe (2003), who used high-temporal resolution automatic rain gauges to derive the empirical correction factor for Kansas city. Morbidelli et al. (2017) studied the error involved in the estimation of true maxima in several stations of central Italy and Barcelona and proposed a correction factor derived from empirical ratios. An extensive overview of the matter was recently presented by Papalexiou et al. (2016) in their large analysis of the behaviour of the correction factor at different time scales using hourly records from 7127 automatic gauges across the United States, shedding light to the probabilistic characteristics of fixed and sliding maxima.

In the present work, empirical correction factors have been derived using available hourly data. The methodology, explained in the next section, leads to a slightly different

factor than the one described by the classical definition where the correction factor is defined as the ratio of the estimator of the sliding interval maximum and estimator of the fixed interval maximum (which is different form the averaged ratios of sliding and fixed interval maxima). Moreover, in the present study, sliding maxima has been calculated only for the same episode as fixed time maxima. These discrepancies with the classical definition of the Hershfield factor are due to issues that followed in the implementation of the calculation. Specifically, due to the use of daily data restricted to one month interval (with the aim of obtaining the empirical factor at different months and seasons) and the fact that the sliding maximum cannot be calculated by hourly data forbidden to the fixed maximum (i. e., data from the previous or following month). The empirical factor presented in this chapter has been obtained by the implementation of the methodology described in the next section, with the mentioned deficiencies. However, some tests of different implementations performed with less data indicated that the results would be extremely similar, always inside of the uncertainty margins shown in the presented results. The ultimately presented results are chosen because of the consistency of the used methodology.

Table[4.1] shows the obtained empirical results for one day correction of maximum rainfall starting the fixed interval at a selection of different times; Table[4.2] shows the empirical correction factor for different durations, with 8 a. m. the starting time of the fixed interval. Annual and seasonal values are the average of the corresponding monthly results. The presented correction factors are obtained as the average of the individual ratios calculated at a selection of considered locations in Catalonia and the uncertainty is obtained from the dispersion of the results at the different locations calculated as the sample standard deviation over the square root of the number of locations.

1 day	**Annual**	**Spring**	**Summer**	**Autumn**	**Winter**
0 a. m.	1.116 ± 0.002	1.120 ± 0.002	1.075 ± 0.003	1.119 ± 0.003	1.151 ± 0.003
8 a. m.	1.129 ± 0.002	1.161 ± 0.003	1.093 ± 0.003	1.125 ± 0.003	1.139 ± 0.003
12 a. m.	1.127 ± 0.002	1.151 ± 0.003	1.091 ± 0.003	1.135 ± 0.003	1.129 ± 0.003
4 p. m.	1.137 ± 0.002	1.152 ± 0.003	1.104 ± 0.004	1.145 ± 0.003	1.148 ± 0.003

Table 4.1: empirical correction factors obtained in the present study obtained to correct one day maximum rainfall for different starting times of the fixed interval and for different seasons.

8 a. m.	Annual	Spring	Summer	Autumn	Winter
1 day	1.129 ± 0.002	1.161 ± 0.003	1.093 ± 0.003	1.125 ± 0.003	1.139 ± 0.003
2 days	1.038 ± 0.001	1.046 ± 0.002	1.024 ± 0.002	1.048 ± 0.002	1.034 ± 0.002
3 days	1.030 ± 0.001	1.031 ± 0.001	1.025 ± 0.002	1.035 ± 0.001	1.028 ± 0.002
4 days	1.020 ± 0.001	1.026 ± 0.001	1.020 ± 0.002	1.019 ± 0.001	1.016 ± 0.001
5 days	1.018 ± 0.001	1.024 ± 0.001	1.017 ± 0.001	1.017 ± 0.001	1.014 ± 0.001
6 days	1.016 ± 0.001	1.019 ± 0.001	1.014 ± 0.001	1.019 ± 0.001	1.012 ± 0.001
7 days	1.014 ± 0.001	1.012 ± 0.001	1.013 ± 0.001	1.016 ± 0.001	1.015 ± 0.001
8 days	1.012 ± 0.001	1.013 ± 0.001	1.012 ± 0.001	1.012 ± 0.001	1.010 ± 0.001
9 days	1.012 ± 0.001	1.015 ± 0.001	1.011 ± 0.001	1.012 ± 0.001	1.009 ± 0.001
10 days	1.011 ± 0.001	1.011 ± 0.001	1.011 ± 0.001	1.011 ± 0.001	1.009 ± 0.001
11 days	1.012 ± 0.001	1.016 ± 0.001	1.011 ± 0.001	1.013 ± 0.001	1.007 ± 0.001
12 days	1.011 ± 0.001	1.012 ± 0.001	1.008 ± 0.001	1.013 ± 0.001	1.010 ± 0.001
13 days	1.008 ± 0.001	1.012 ± 0.001	1.008 ± 0.001	1.007 ± 0.001	1.006 ± 0.001
14 days	1.009 ± 0.001	1.009 ± 0.001	1.008 ± 0.001	1.008 ± 0.001	1.011 ± 0.001
15 days	1.007 ± 0.001	1.010 ± 0.001	1.006 ± 0.001	1.007 ± 0.001	1.007 ± 0.001

Table 4.2: empirical correction factors for fixed intervals starting at 8 a. m.; the annual and seasonal values have been averaged from monthly results. The empirical factor has been obtained for durations of several days using hourly data in all cases.

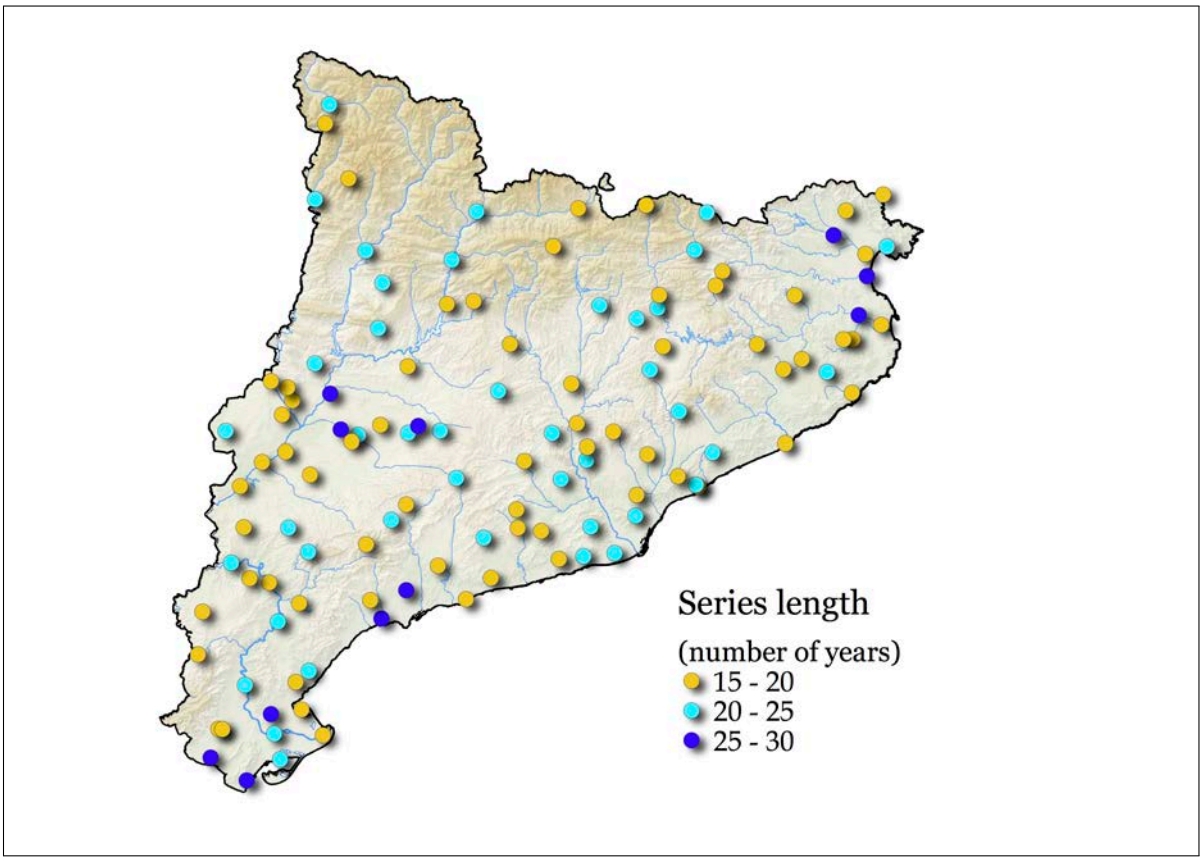


Figure 4.2: selected 120 automatic weather stations. Hourly data at these locations in the period 1988-2016 has been used to obtain the empirical correction factor.

4.2 Study of the empirical correction factor

A selection of 120 series has been drawn from the available dataset. The selected series are from the XEMA network, hence, they are automatic weather stations that measure data at subdaily resolution (data are used at hourly resolution). The set of 120 has been chosen because all the stations have been operational for at least 15 years. The 120 selected AWS have a global quality index over 80 (actually, over 84.5), although some individual years have a lower quality index (but they are less than a 5% of the total years of data that have been used). The temporal period is 1988-2016. Fig[4.2] shows the location of the 120 AWS used in the study of the correction factor.

Empirical correction factors have been obtained from the ratio of rainfall amounts measured at unrestricted against fixed intervals in the following way:

- The monthly maximum rainfall in one day has been calculated using a 24-hour fixed time interval. The monthly maximum in one day has been found at each AWS for every month with rainfall data.

- The maximum rainfall for unrestricted intervals of 24 hours has been found using a sliding window of 1 hour on every episode detected in the previous step, with the condition of having at least one hour inside the considered fixed time interval.
- The maximum value obtained using the sliding window has been divided by the value for the fixed time interval, thus, yielding a ratio for each AWS at each month in which there is available data.
- The average of the ratio found for every month with data at each AWS yields the annual global value for the AWS. The seasonal correction factor is found as the average of the months that correspond to the season considering all the available years of the AWS.

This process has been repeated for different fixed starting times as well as for durations other than 24 hours. Specifically, four different starting times along the day have been considered in order to explore the influence of the diurnal cycle on the empirical ratio: 0 a.m., 8 a.m., 12 a.m. and 4 p.m. where time is indicated in Coordinated Universal Time (as registered by the AWS); official local time is 1 hour ahead of UT in standard time (roughly November to March) and 2 hours ahead in daylight saving time (roughly April to October).

It is known (Weiss, 1964) that an increase of the number of fixed time intervals in which each duration has been discretised leads to a decrease in the correction factor. To investigate this dependence, the ratio between amounts measured using fixed and sliding intervals has been calculated also for duration higher than 1 day, from 2 days up to 15 days.

The global averaged value obtained at each AWS has been interpolated using the simple kriging methodology in order to display a spatial distribution of the correction factor over Catalonia and compare it with the mean annual and seasonal rainfall maps.

The actual duration of the episodes with maximum daily rainfall registered at every AWS has been measured with the aim of exploring dependence of the ratio on meteorological time scales, and thus, on physics behind rainfall generation.

4.2.1 Comparison with previous factors

Empirical factors obtained by the previously described methodology agree with factors previously obtained by several authors. The value of the global averaged correction factor for a duration of 24 hours for the whole studied territory, calculated as the mean correction factor yielded at the each one of the 120 considered AWS, resulted 1.129 in this study. The obtained averaged correction factor is in quite agreement with the Hershfield factor (Hershfield, 1961) and other studies listed in Table[4.3] (Huff and Angel 1992, Young and McEnroe 2003)

Duration (in days)	Hershfield (1961)	Weiss (1964)	Huff Angel (1992)	Young McEnroe (2003)	Present project (2018)
1	1.13	1.143	1.13	1.13	1.129
2		1.067	1.05	1.05	1.038
3		1.044	1.02	1.03	1.030
4		1.032			1.020
5		1.026	1.01		1.018
6		1.022		1.01	1.016
10		1.013	1.01		1.011
12		1.011		1.00	1.011

Table 4.3: comparison between factors obtained by several authors.

Weiss's factor Weiss (1964) for 1-day duration had a slightly higher factor (1.143). The difference between the empirical factors and the one proposed by Weiss might be explained by the fact that Weiss derived the factor theoretically, assuming a uniform rainfall intensity which is an unrealistic rainfall behaviour. Comments on the significant overestimation of the correction factor obtained from the calculation carried out by Weiss can be found at Dwyer and Reed (1995) and Young and McEnroe (2003), and it is also pointed out by Asquith (1998) that Weiss's assumption that the probability of a rainfall event occurring is independent of the time of the day cannot be applied to any location.

Table[4.3] also shows the correction factor calculated by several authors for durations over 1 day. As commented before, durations of 2 days and more were considered to explore how the correction factor decreases along with the increase of the number of fixed time intervals N in which each duration has been discretised. This calculation, in the present study, has been done by considering accumulated rainfall over several N days measured using a 1-day fixed time interval and then comparing measures using a sliding 1-day interval. Weiss (1964) formulated the dependence of the correction factor F with the number of intervals N as Eq[4.1], whereas Young and McEnroe (2003) fitted their experimental results to Eq[4.2]. Papalexiou et al. (2016), among other statistics, compared the mean and standard deviation of the annual maxima sample measured using fixed intervals and sample using sliding intervals, obtaining an empirical function for the mean which formulated in terms of the ratio F would be Eq[4.3].

$$F = N/(N - 0.125) \tag{4.1}$$

$$F = 1 + 0.13N^{-1.5} \tag{4.2}$$

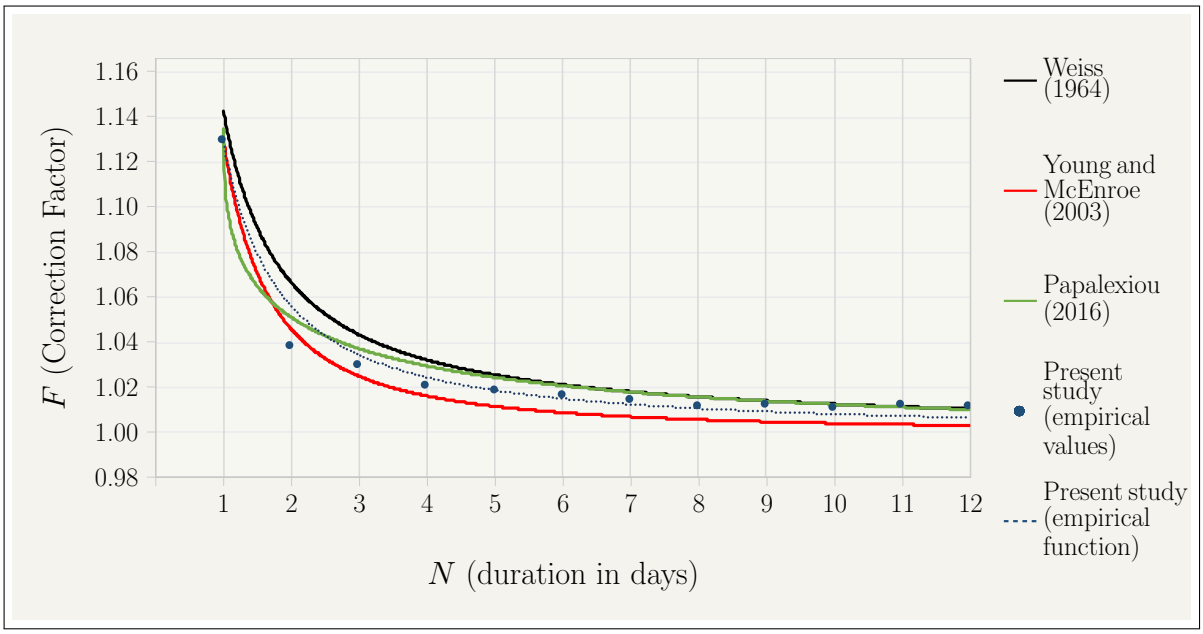


Figure 4.3: graphical comparison between the empirical correction factor obtained in the present study and the correction factor obtained by several other authors, namely, Weiss (1964), Young and McEnroe (2003) and Papalexiou et al. (2016).

$$F = 1 + 0.135e^{-(\frac{N-1}{1.078})^{0.408}} \tag{4.3}$$

Regarding the results obtained for those durations, the empirical correction factor obtained is lower than the theoretical factor proposed by Weiss (1964) for durations shorter than 10 days, but the agreement improves for accumulations longer than 10 days. This result is caused by the fact that the more fixed intervals are accumulated the less important are the problems caused by Weiss's unrealistic assumptions.

The type of empirical equation proposed by Young and McEnroe (2003) has been used to fit the values of the correction factor obtained in the present study, for durations between 1 and 15 days, forcing the function to 1.129 for the results of one day period and resulting in an exponent of -1.2, see Eq[4.4].

$$F = 1 + 0.129N^{-1.2} \tag{4.4}$$

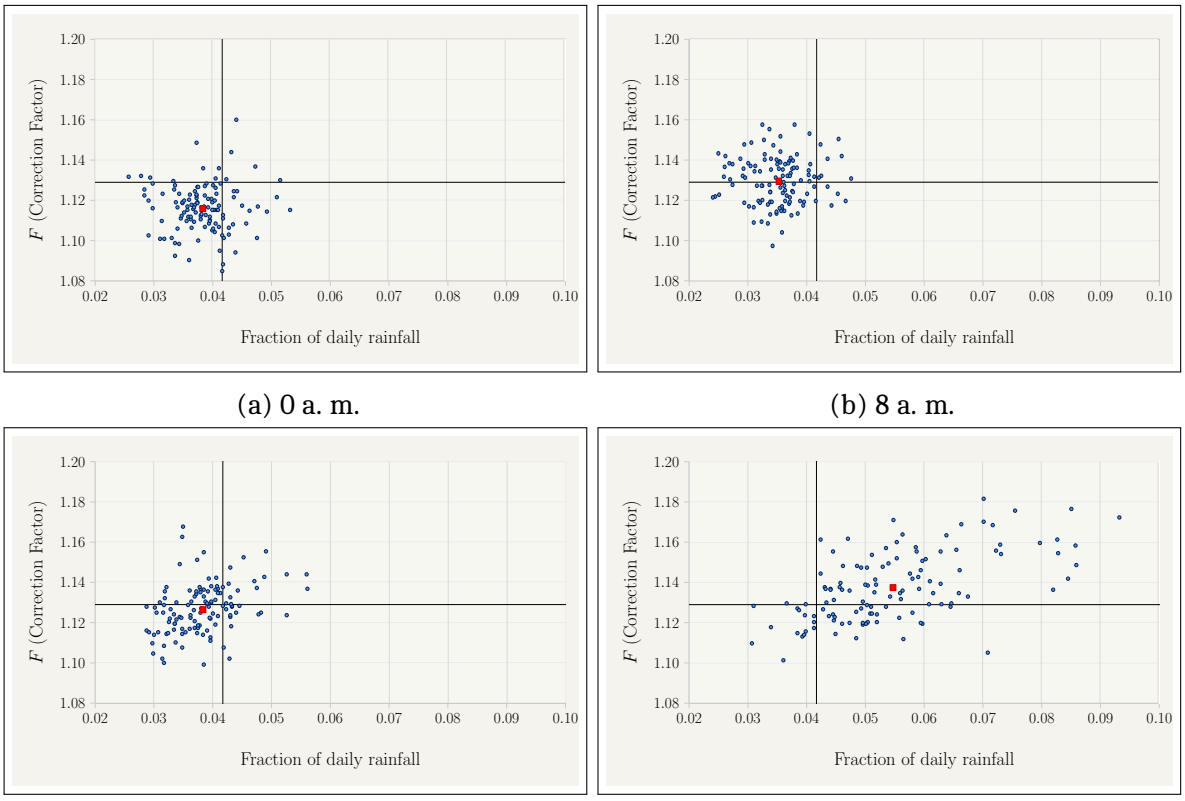
Fig[4.3] shows the comparison between the empirical factors obtained in the present study and curves given by Eq[4.1], Eq[4.2], Eq[4.3] and Eq[4.4].

4.2.2 Dependence on the fixed starting time

Traditionally, rainfall is most commonly measured daily at a fixed time in the morning. Nonetheless, as a part of the present study, the difference in the empirical correction factor depending on the sampling time has been analysed. Specifically, the fixed intervals that start in the following time of the day have been considered: 0 a.m., 8 a.m., 12 a.m. and 4 p.m., all of them at Universal Time (which is between one and two hours before local time, depending on daylight saving time).

In order to graphically compare the correction factors resulting from this calculation at each one of the 120 AWS considered, and moreover, to show which fixed starting time causes a worst measure compared to the sliding maxima, the empirical correction factors have been plotted against the fraction of daily rainfall collected in the hour centred at the time at which the fixed starting interval is starting. The fraction of daily rainfall has been calculated for every AWS as the total amount collected in 1 hour at the starting time of the fixed interval at each day with available data and averaged over the total amount of daily rainfall collected by the AWS. The vertical line at 0.0417 represents the fraction of daily rainfall that would be collected during 1 hour if the rain were uniform over time (i.e., $1/24$). The horizontal line at 1.129 indicates the global averaged empirical correction factor for measures at 8 a.m. obtained in this work. Low values of the fraction of daily rainfall indicate that it is rare at that particular AWS to be raining at the time of the fixed interval measure or that collected amounts are low. On the contrary, high values of the fraction of daily rainfall indicate that at that particular time it is frequently raining with a chance of collecting large amounts, hence, it may be not the most appropiate starting time for accumulated rainfall measurements, because rain episodes would be frequently divided into two parts.

In Fig[4.4] the empirical factors obtained for each one of the 120 AWS have been represented against the fraction of daily rainfall registered in each station at the considered starting time. Comparing the four representations, it can be seen that almost at every location considered it rains less in the morning than what would be collected by a constant rain, whereas in the afternoon, at 4 p.m., most locations present a higher fraction of daily rainfall around the measuring time. In the same way, empirical correction factors indicate a need for a higher correction in the case of measures taken in the afternoon in favour of those taken in the morning. Even more, Fig[4.4] shows that daily measures would, in general, result slightly better (lower correction factors and lower fraction of daily rainfall) if they were taken at 0 a.m. instead of 8 a.m., which is the most common measuring time of daily rainfall in the study area. Nonetheless, 8 a.m., seems to be a good starting time for the fixed interval. Correction factors for daily measures taken at fixed time intervals starting at 8 a.m. yielded a mean value of 1.129, lower than the mean value of 1.137 corresponding to factors obtained for measures taken at fixed time intervals starting at 4 p.m., proving that rainfall events would be more often split if measures were taken in the afternoon. Most locations in mountainous areas of Catalonia present a very high fraction of daily rainfall in the afternoon. As an example, Fig[4.5] compares this fraction for every hour of the day for two different locations: Núria, which



(a) 0 a. m.

(b) 8 a. m.

(c) 12 a. m.

(d) 4 p. m.

Figure 4.4: empirical correction factors for fixed intervals of 24 hours, starting at different times, represented against the fraction of daily rainfall at the given starting time of the interval. Red squares are the centre of gravity of each sample.

is located in the Pyrenees at 1971 m a.s.l. and present very high values of the fraction of daily rainfall after midday, linked to solar heating of mountain slopes (Palau and Rovira, 2015), and low for the rest, and the city of Barcelona in the coast, showing a fraction almost uniform in comparison, around the value $1/24$.

4.2.3 Seasonality and spatial distribution

Due to several geographical and aerological drivers, rainfall in Catalonia is highly contrasted from one area to another, with some mountainous areas where the mean annual precipitation exceeds the value of 1200 mm in contrast to other areas, mainly in the west of the territory, where the mean annual value is lower than 400 mm. Regarding seasonality, and in brad terms, pluviometric maxima are observed in Catalonia especially in autumn and in spring, depending on the area. In the present study, the 48% of the annual maxima considered occurred in autumn, 22% in spring, 16% in summer and 14% in winter.

The mean empirical correction factors at each AWS represented in Fig[4.4] were ob-

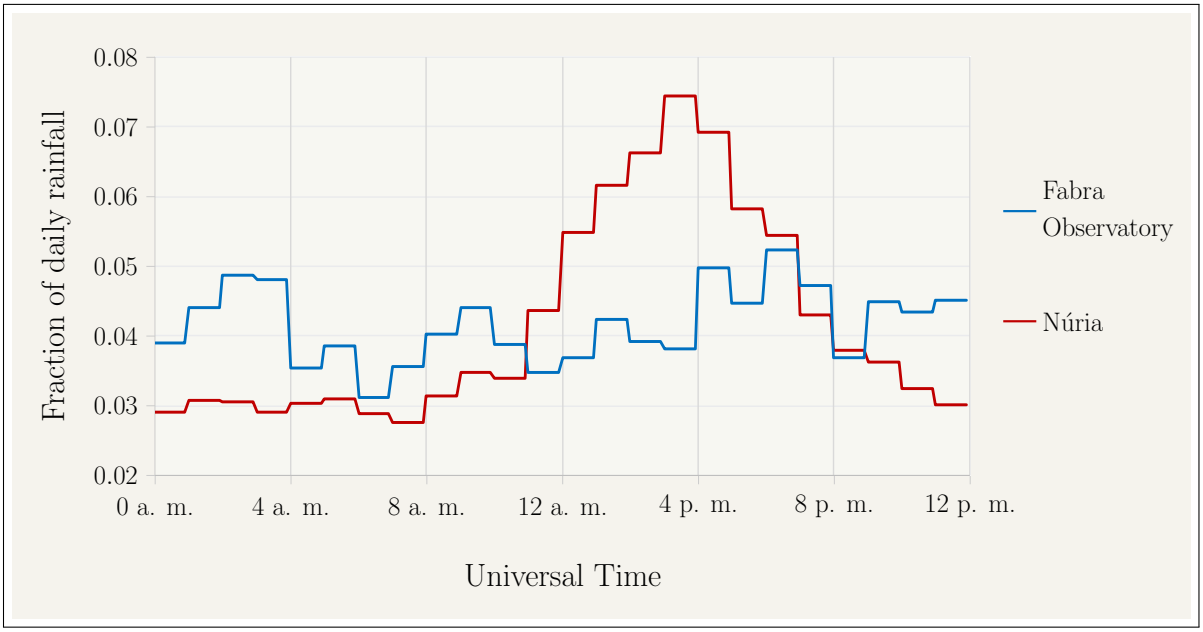
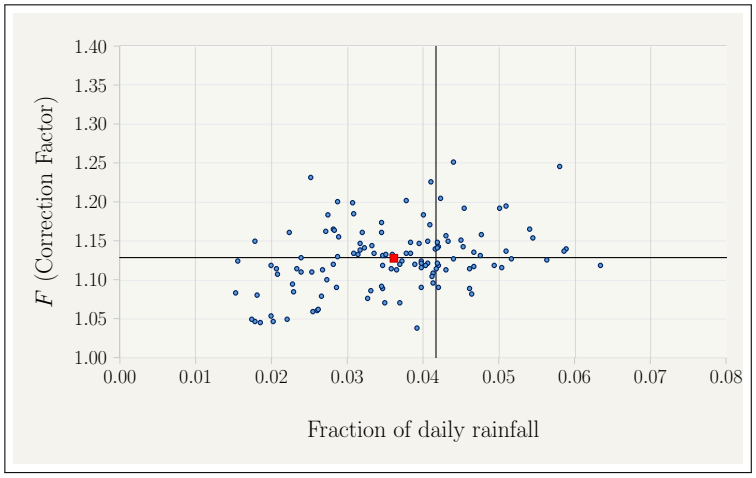


Figure 4.5: fraction of daily rainfall registered every hour of the day at Núria (a station in the Pyrenees, at 1971 m a.s.l.) and at the coastal city of Barcelona (Fabra Observatory).

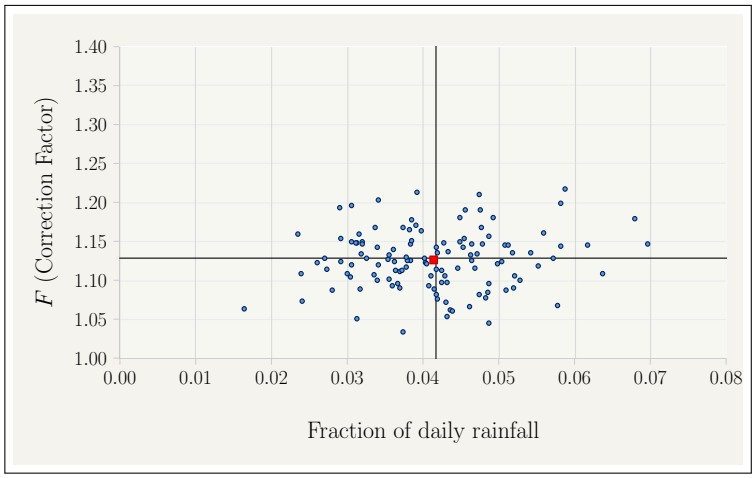
tained after averaging the ratios calculated for every monthly maximum amount recorded. With the aim of showing the monthly variations of the frequency factors against the fraction of daily rainfall, as well as seasonality, in Fig[4.6] are represented, for every month, the averaged ratios calculated from monthly maxima at every station, considering all measures of 8 a.m. starting fixed interval. In Fig[4.7] these points have been averaged by season, with spring comprising the months form March to May, summer June to August, autumn September to November and winter December to February.

Fig[4.6] and Fig[4.7] show that summer is the season for which lower correction factors are needed because a very low fraction of the daily rainfall is registered at 8 a.m. This is an expected result since rainfall maxima recorded in summer in Catalonia are often produced by local storms taking place after midday, due to a clear diurnal surface warming effect in their convective development, as well as mesoscale formations also triggerd in the afternoon (Casas et al. 2004, Pérez-Zanón et al. 2016). In fact, the prevalence in summer of mesoscale circulations with marked diurnal cycles has been reported for the Western Mediterranean area (Palau et al., 2017). The same result is shown in Fig[4.8], where empirical correction factors have been averaged by month. Fig[4.8] shows also that factors for the climatological spring (March, April and May) are high, all of them over the value of 1.14.

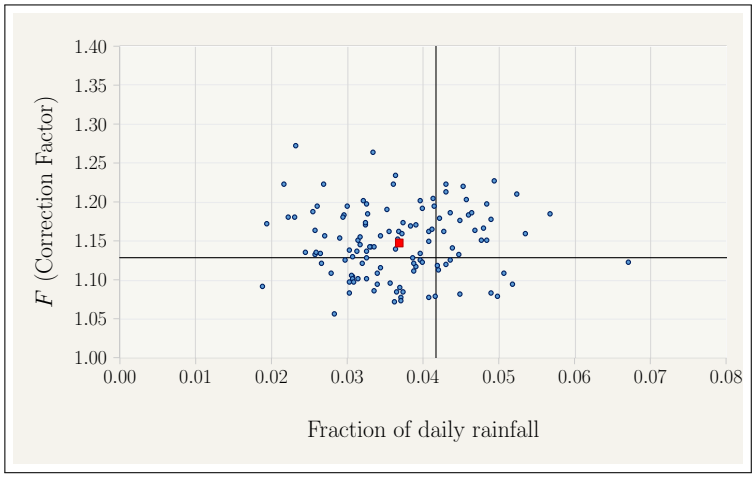
Regarding the influence of seasons, empirical factors are found to have a mean seasonal global value in Catalonia of 1.161 in spring, 1.093 in summer, 1.124 in autumn and 1.139 in winter. The fact that correction factors are found to be highest in spring means that it is the season when rain episodes are more often split when measures are taken



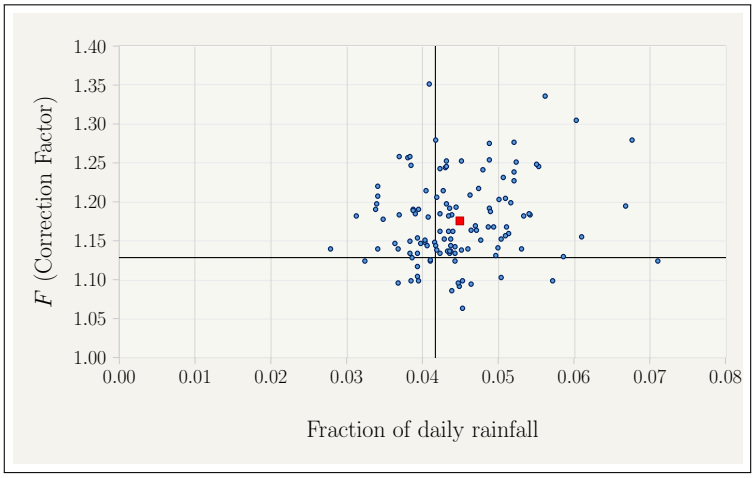
(a) January



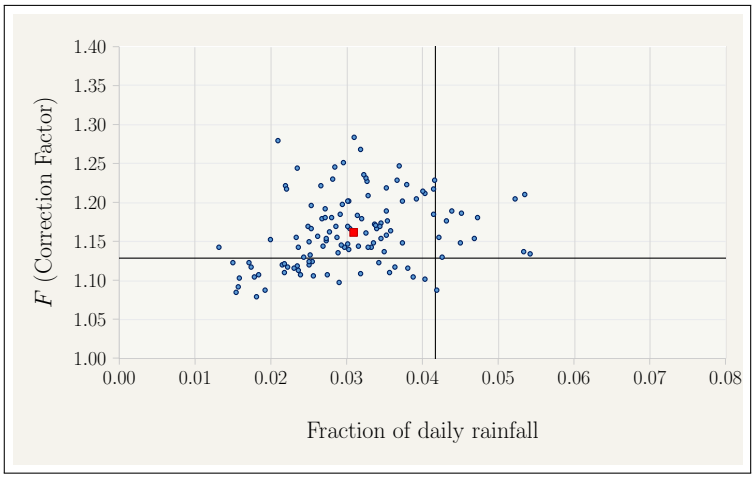
(b) February



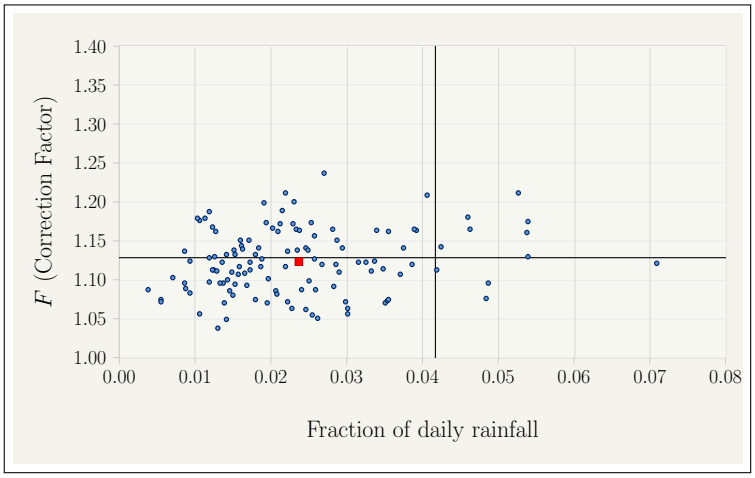
(c) March



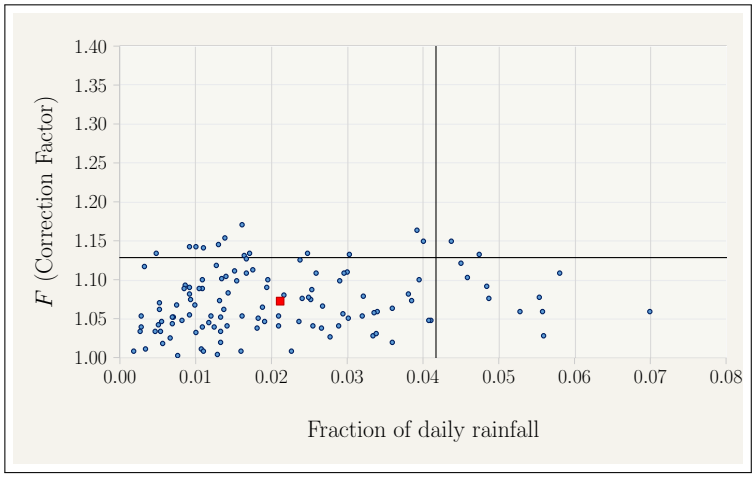
(d) April



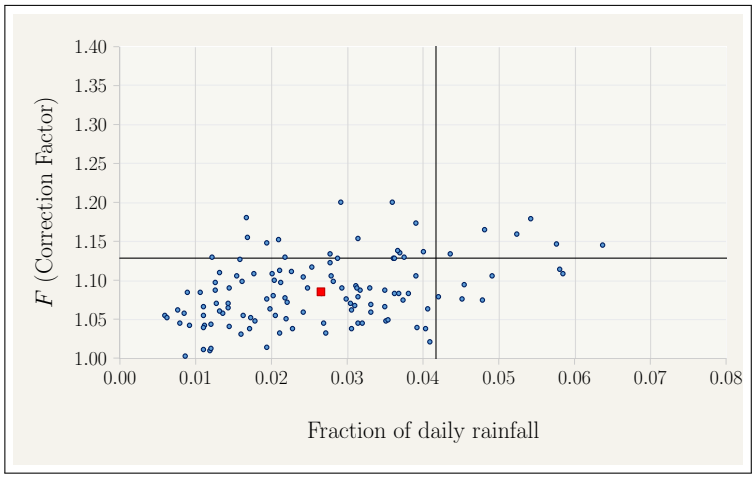
(e) May



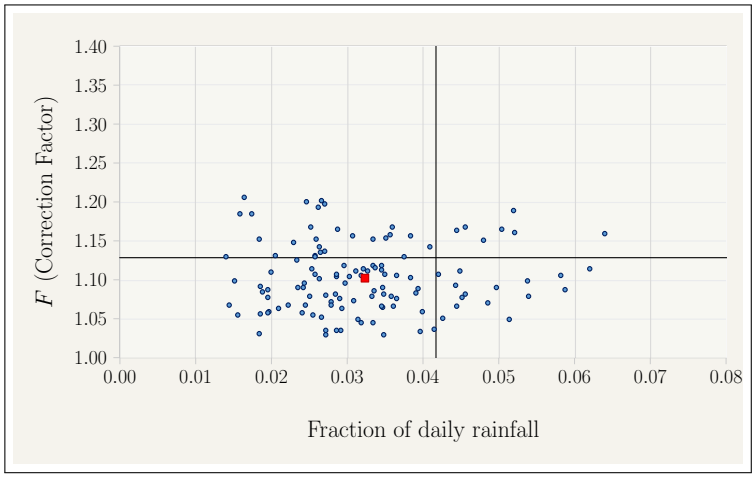
(f) June



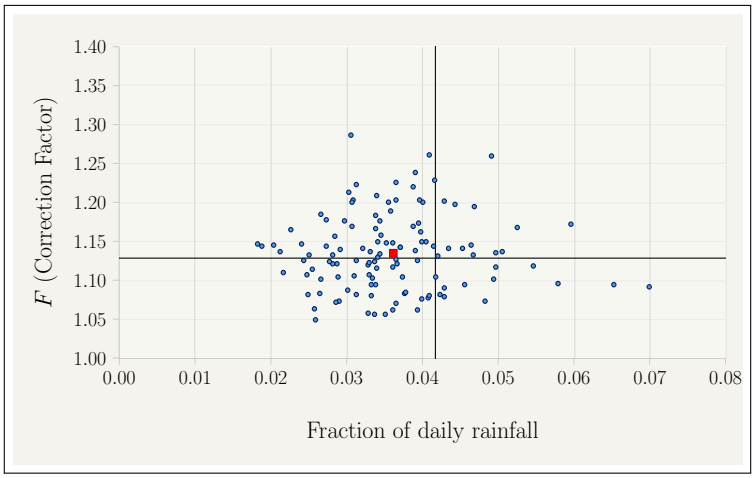
(g) July



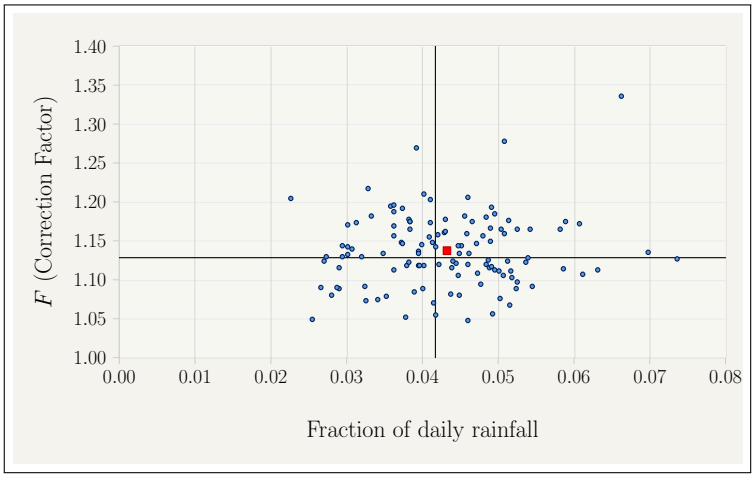
(h) August



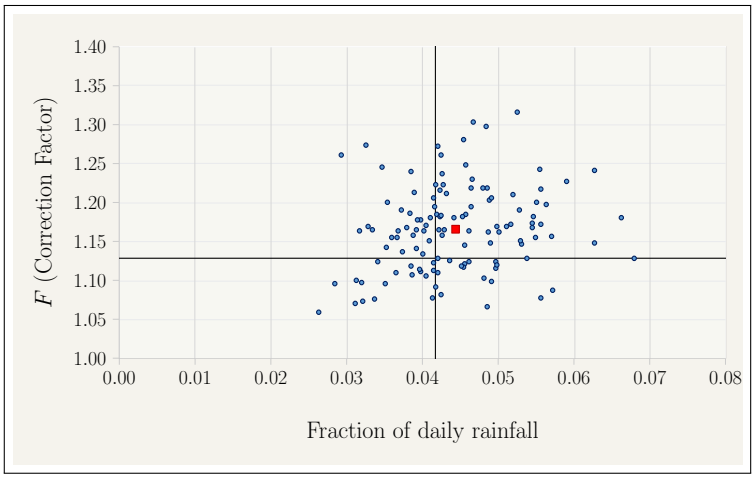
(i) September



(j) October

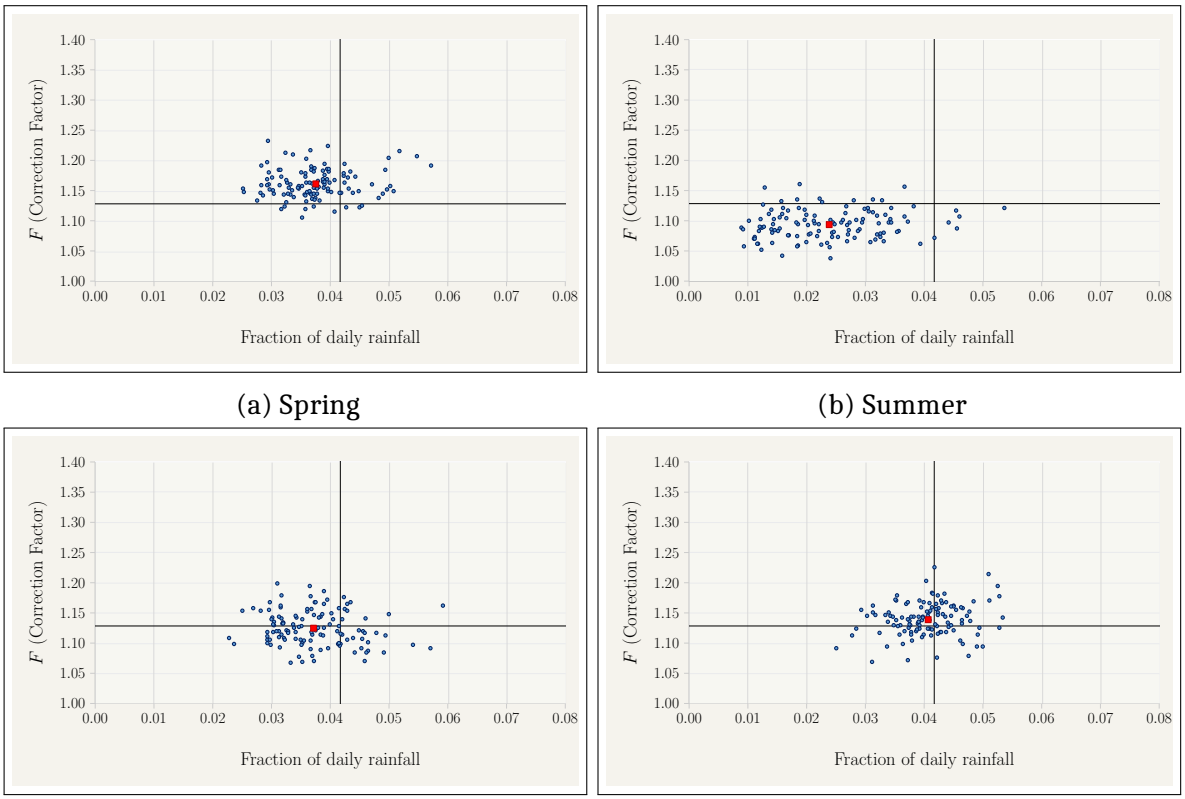


(k) November



(l) December

Figure 4.6: empirical correction factors for fixed intervals of 24 hours starting at 8 a. m., obtained from data of different months, represented against the fraction of daily rainfall at the given starting time of the interval. Red squares are the centre of gravity of each sample.



(a) Spring

(b) Summer

(c) Autumn

(d) Winter

Figure 4.7: Empirical correction factors for fixed intervals of 24 hours starting at 8 a. m., obtained from averaging the correction factor of the three months of each season, represented against the fraction of daily rainfall at the given starting time of the interval. Red squares are the centre of gravity of each sample.

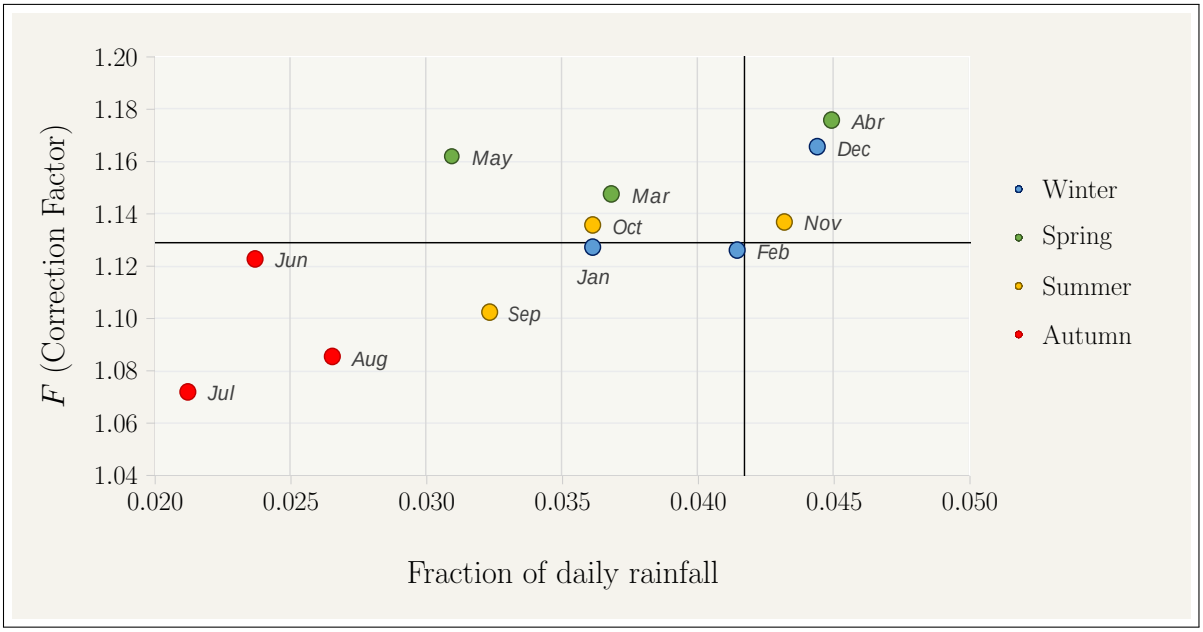


Figure 4.8: empirical correction factors averaged by month and represented against the averaged fraction of daily rainfall of the 120 AWS corresponding to each month. The climatological seasons are indicated by different colours.

at 8 a.m., hence, a higher ratio is obtained between rainfall at unrestricted intervals and rainfall at fixed time intervals. On the contrary, as commented, episodes in summer are the ones that need the lowest corrections; indeed, summer episodes in the study area are often raised by the influence of the diurnal cycle and the 24 hour maxima are not usually split by a measure taken in the morning.

A spatial distribution of the correction factor in Catalonia has been obtained using a simple kriging methodology for which the global averaged value at each AWS has been interpolated. Fig[4.9] shows this spatial distribution for the four climatological seasons (see Fig[1.4]). The general pattern of the spatial distribution resembles the mean annual and mean seasonal rainfall patterns in Catalonia, with a need for lower correction in drier regions. However, it must be noticed that this spatial patter is highly variable depending on the season, up to the point of having a region (in central Catalonia and slightly to the west) which needs clearly high correction factors in winter while the same region needs low correction factors in autumn.

Both spring and summer distributions present low spatial variability and the mean ratio yields the highest and lowest seasonal correction factors respectively. Autumn's distribution presents a high spatial variability of empirical ratios, with the lowest correction coinciding with the driest regions. Winter presents a high spatial variability as well, but the relation to the mean rainfall distribution is not so clear.

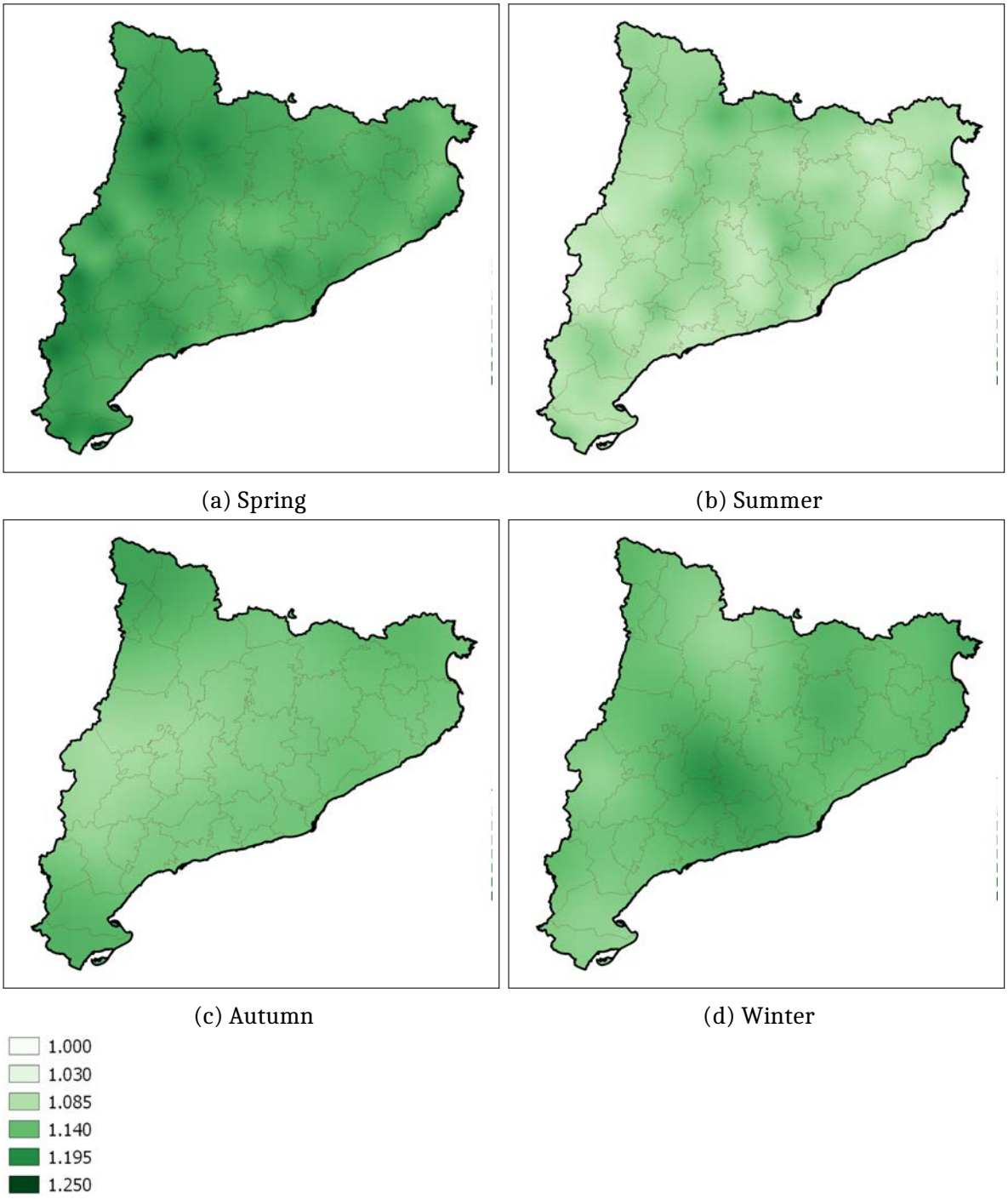
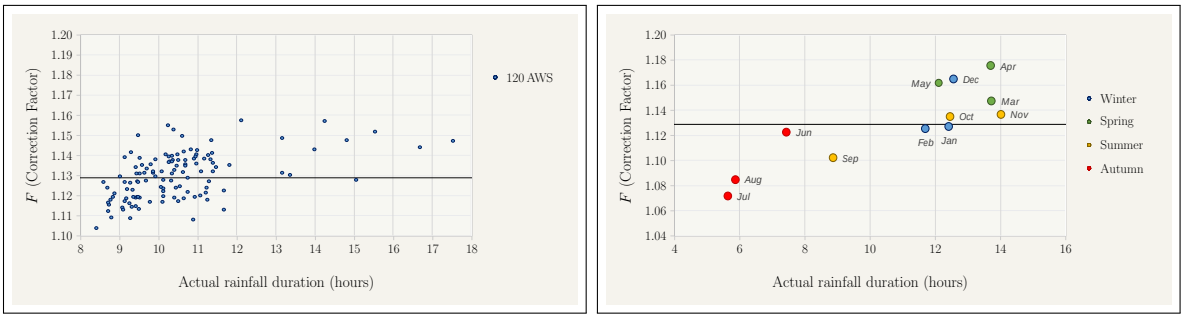


Figure 4.9: spatial interpolation of empirical correction factor obtained in 120 AWS using a simple kriging.

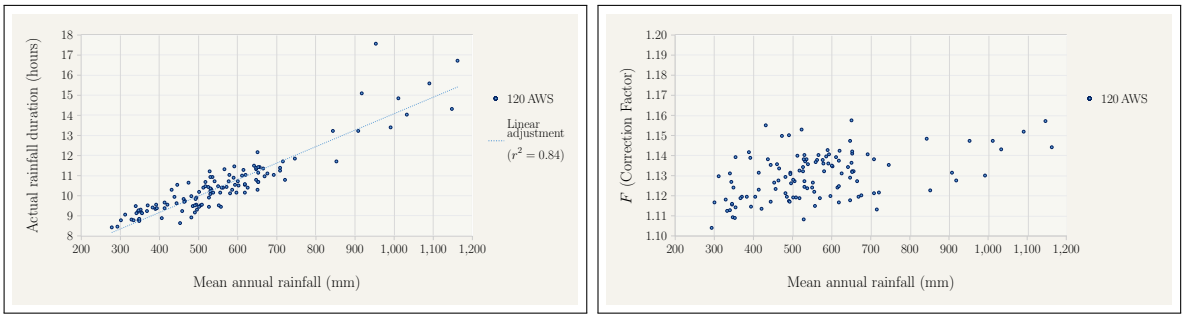


(a) (b)

Figure 4.10: empirical correction factors represented against the actual rainfall duration of the corresponding location (calculated with data of the same AWS); dispersion with all data (a) and averaged by months using data from all stations (b).

4.2.4 Actual rainfall duration

In many areas of Catalonia, the meteorological situations producing abundant rainfall amounts contributing highly to monthly and annual totals are not always the same as those involved in daily maxima (Casas et al., 2008). Thus, while synoptic-scale situations have a greater influence on the annual mean rainfall, daily maxima are usually related to smaller organisations for which local and mesoscale drivers as orography, distance to the sea or temperature and humidity advections at low levels and between sea and land, are decisive. With the aim of investigating the possible influence of meteorological scales on the ratio between fixed and sliding intervals, and thus, on the correction factor, the actual duration of rainfall episodes which produced monthly maxima at every AWS has been taken into account. Fig[4.10] shows that the mean value of the actual duration of the monthly maxima events averaged by year is mostly between 8 and 12 hours. Nonetheless, different durations are contributing to these mean values depending on the season. In Fig[4.10] the actual duration averaged by month has been displayed also, showing that the monthly maxima in summer months, as well as in the first month of autumn (September), were the shortest events: between 6 and 9 hours according to their mainly relationship to highly convective mesoscale organisations. For the rest of autumn, winter and especially spring, with high correction factors and durations above 12 hours, synoptic scale situations, sometimes with smaller mesoscale systems developed inside, seem to have more influence in the rainfall pattern. In both graphics of Fig[4.10] a certain increasing dependence between the correction factor and the actual duration can be glimpsed. Due to the expected relationship between the mean annual rainfall and the mean actual duration of the monthly maxima (Fig[4.11]), an increasing dependence between the correction factor and the mean annual rainfall has also been found.



(a) (b)

Figure 4.11: relation of mean annual rainfall with actual rainfall duration (a) and empirical correction factor (b) in the same location.

4.3 Obtainment of series of maximum rainfall in sliding 24 hours from daily measures at fixed times

The series of annual maximum rainfall in one day (with fixed starting time at 8 a.m.) were obtained from the 163 selected series in the selected years (all of them in the period 1942-2016 and at least 25 years per series). These series of maximum rainfall in one day need to be corrected to obtain annual series of rainfall in 24 hours (as if they had been measured with a sliding window). To correct the annual series, the empirical correction factor has been applied as multiplicative factor.

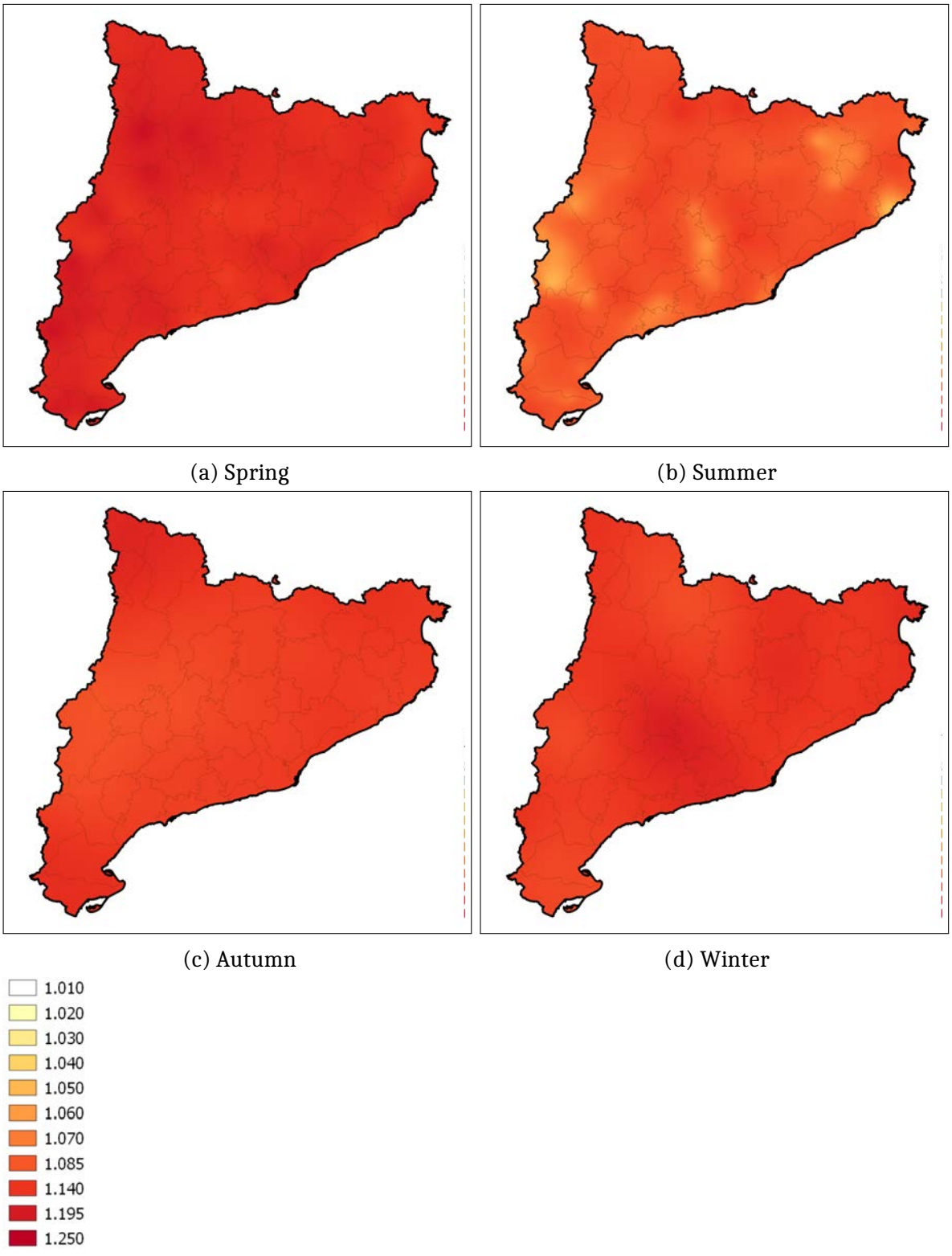
After the study of the correction factor presented in this chapter, it has been seen that the empirical factor is similar to other proposed factors in the literature, hence, it has been decided to use the empirical factor obtained in the present study as it is representative of Catalonia.

It has been shown that the empirical factor depends on the location and on the season. Therefore, it has been chosen to use a specific factor depending on each location and depending on the season in which the annual maximum of rainfall was collected. The seasonal empirical factor had been obtained from monthly maximum daily values in 120 AWS, however, the location of the AWS does not coincide exactly with the location of any of the 163 selected manual series.

The followed implementation to solve this issue has been to extract the interpolated value of a kriging performed with the empirical factors of the 120 AWS. The kriging interpolation has been performed for each season (see Fig[4.12] that is the same interpolation shown in Fig[4.9] except that the legend is different, using now the exact same legend as in Fig[4.13], Fig[4.14] and Fig[4.15]). In this way, the applied correction factor does not depend on the empirical calculation in one particular station but it is smoothed by all the values in the surrounding locations. Nevertheless, the factor in a particular location depends highly on the season and the corresponding factor has been applied depending on the day that led to maximum annual rainfall.

Nevertheless, not only the series of annual maximum rainfall in 24 hours are need in subsequent chapters. The series of annual maximum rainfall in 48 hours, 76 hours, etc. (up to aggregations of 15 days) are needed in order to calculate the relation between intensity and frequency at different durations. Therefore, the same procedure has been applied using the spatial interpolation of the empirical correction factor at different durations (for each season). Fig[4.13] shows the interpolation for 48 hours, Fig[4.14] for 72 hours and Fig[4.15] for 240 hours.

Finally, series of annual maximum rainfall in 24, 48, 72, ..., and 360 hours were obtained at each of the 163 selected locations for the available years in the period 1942-2016.



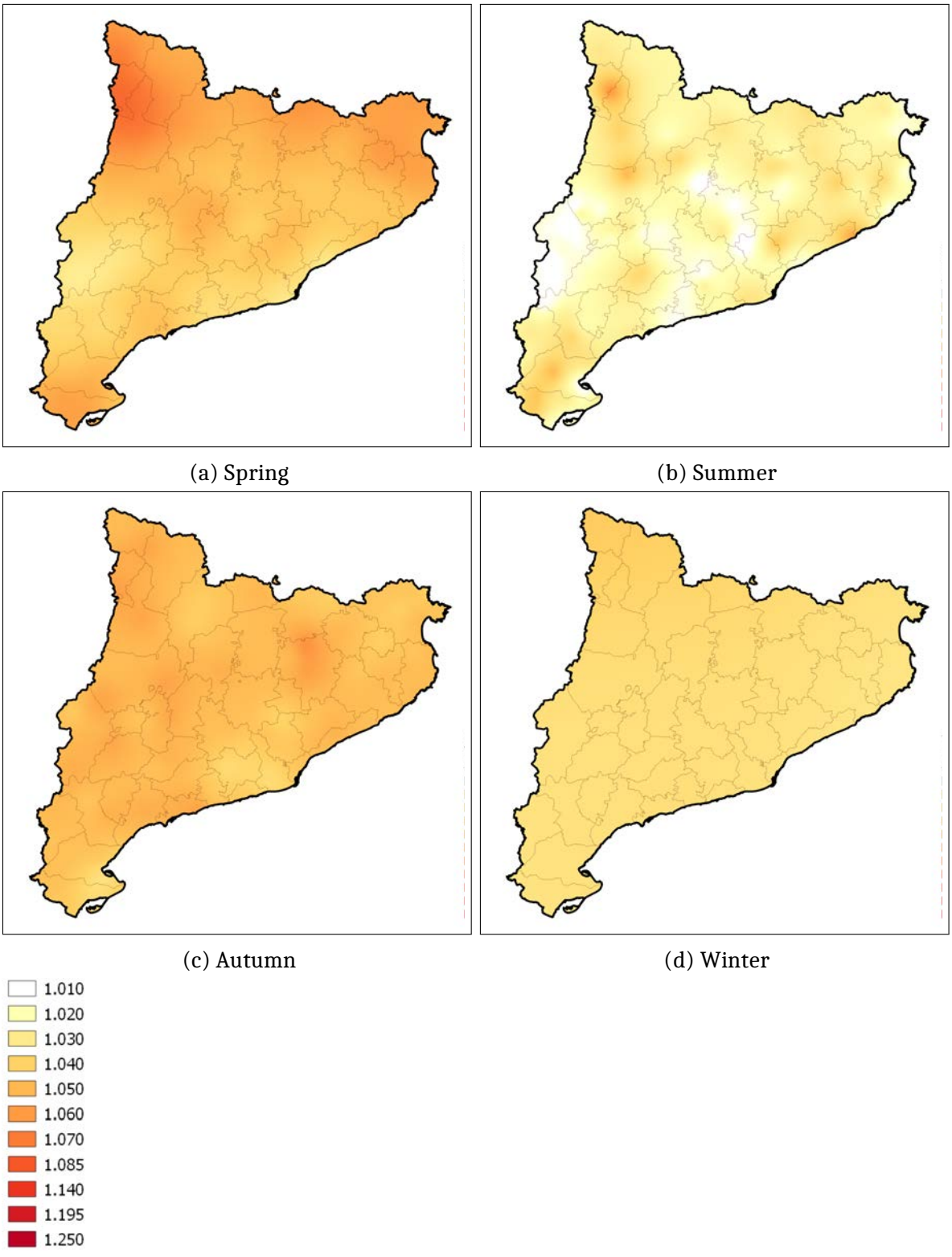
(a) Spring

(b) Summer

(c) Autumn

(d) Winter

Figure 4.12: spatial interpolation (using a simple kriging) of empirical correction factor obtained to correct 1 day to 24 hours maxima (originally obtained using 120 AWS).



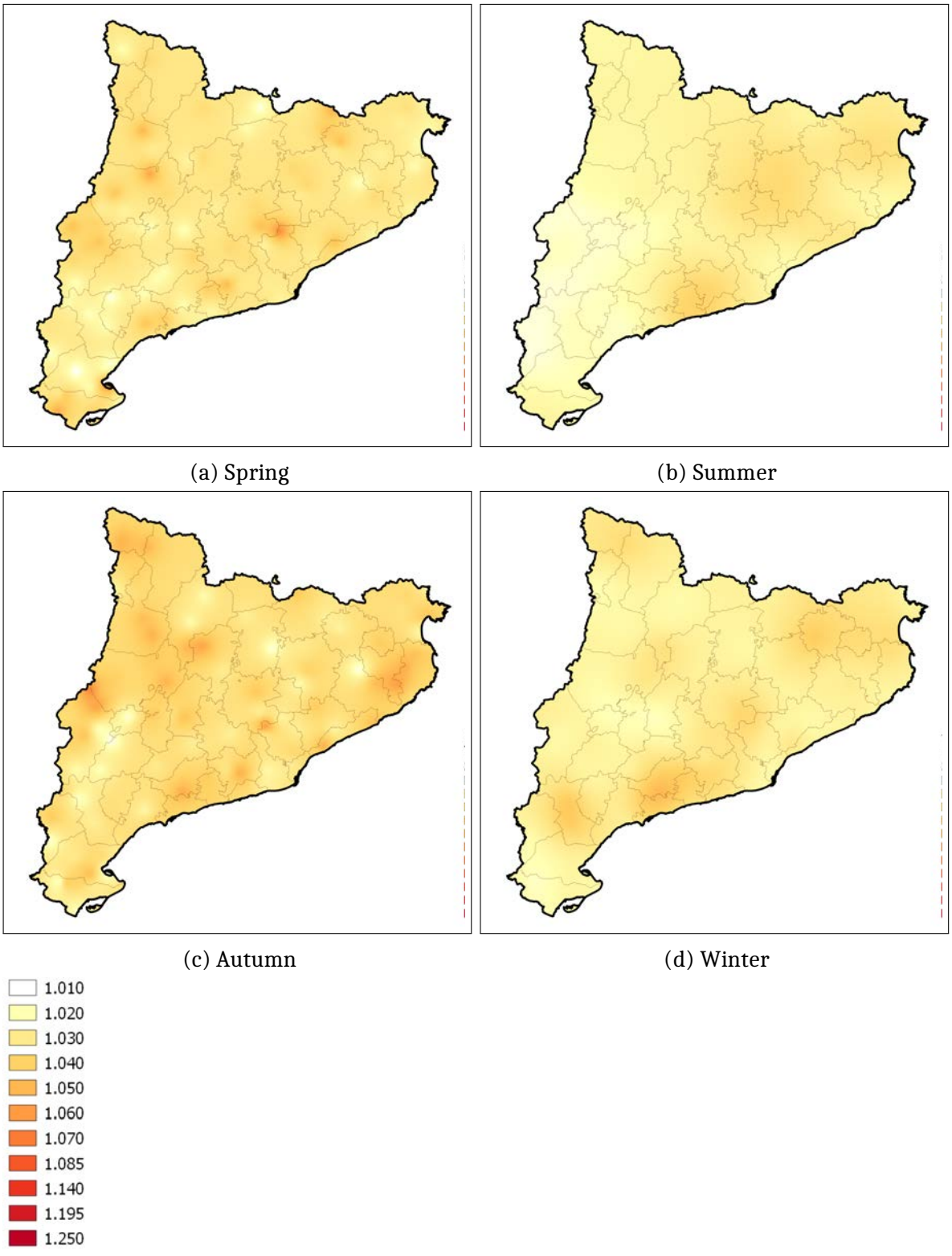
(a) Spring

(b) Summer

(c) Autumn

(d) Winter

Figure 4.13: spatial interpolation (using a simple kriging) of empirical correction factor obtained to correct 2 days to 48 hours maxima (originally obtained using 120 AWS).



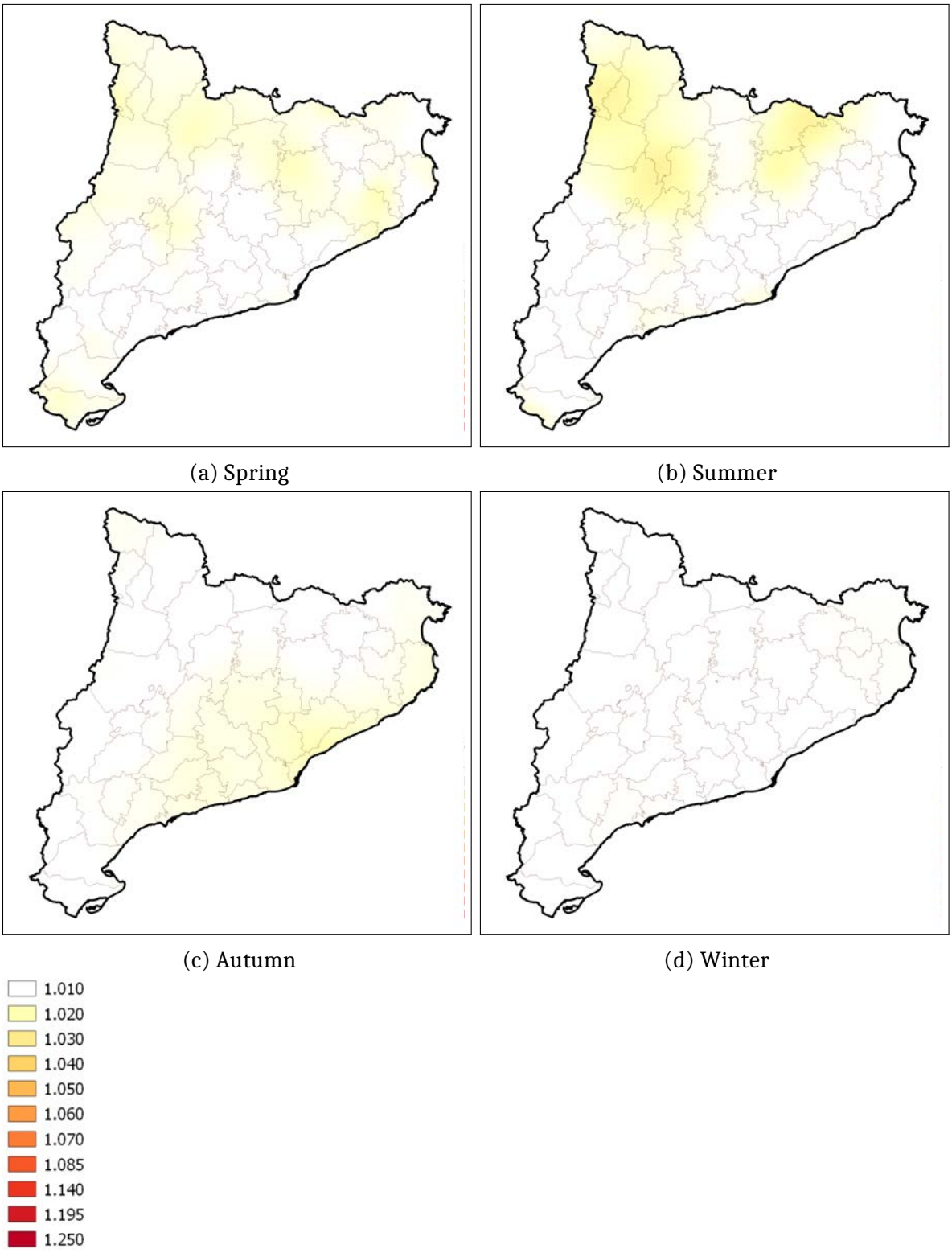
(a) Spring

(b) Summer

(c) Autumn

(d) Winter

Figure 4.14: spatial interpolation (using a simple kriging) of empirical correction factor obtained to correct 3 days to 72 hours maxima (originally obtained using 120 AWS).



(a) Spring

(b) Summer

(c) Autumn

(d) Winter

Figure 4.15: spatial interpolation (using a simple kriging) of empirical correction factor obtained to correct 10 days to 240 hours maxima (originally obtained using 120 AWS).

Chapter 5

TEMPORAL DOWNSCALING

The present chapter deals with the underling fractal theory in the methodology used in this project to obtain Intensity - Duration - Frequency relationships. The monofractal methodology is not the only one ready for the obtainment of IDF relationships; classic methodologies, that calculate the return period of a rainfall depth of certain duration from the probability of exceedance calculated by ranking the observations or adjusting them to an extreme value probability distribution, are used to infer IDF relationships at a lower temporal resolution than the initial data. Therefore, IDF relationships can be obtained, by aggregation methods, from daily data at a resolution of days, weeks, months or years; however, in order to obtain IDF relationships at a subdaily duration, hourly or minute data are needed in classic methodologies. The monofractal downscaling methodology is suitable in the present project, where the initial data is at daily resolution and the ambition is to obtain IDF relationships at a higher temporal resolution (i.e., subdaily).

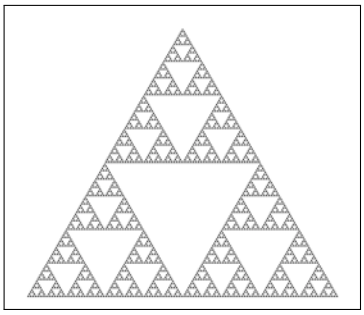
The intensity-duration-frequency curves (IDF curves), which have been a matter of considerable interest to engineers and hydrologists for over a century, remain nowadays as an important tool to analyse the risk of natural hazards for hydrological purposes. The mathematical relationships more often used to describe the IDF curves are empirical, sometimes in the form of a generalised equation for the rainfall intensity $I(t,T)$, valid for all durations, t, and return periods, T, considered. This equation usually has the generalised form $\frac{a(T)}{b(t)}$, where $a(T)$ and $b(t)$ are functions independent of each other. The function $a(T)$ can be found empirically (Casas et al., 2004), although there are authors (Koutsoyiannis et al., 1998) who proposed to use a function of statistical probability of the maximum rainfall intensity to determine it. Other authors (Burlando and Rosso 1996, Menabde et al. 1999) considered the fractal property of scale invariance of the rainfall series to find an analytical relationship for the IDF curves taking into accound the scaling behaviour. Burlando and Rosso (1996) were pioneers in applying scaling relationships to the statistical moments of annual maximum rainfall series. There is also a methodology based on the property of scale invariance (Menabde et al. 1999, Yu et al. 2004, Desramaut 2008, Bara et al. 2010) to obtain IDF curves in those places where daily rainfall data are the only available. For instance, Aronica and Freni (2005) anal-

ysed extreme rainfall data from a rain gauge network within the metropolitan area of Palermo (Italy) with the aim of combining and taking advantage of high-resolution rain gauges with a short working period along with low-resolution rain gauges with longer data records to obtain plausible depth-duration-frequency (DDF) curves. Applying this scaling approach, Aronica and Freni (2005) found better results than those coming from the classical subhourly rainfall regression formulas (Bell 1969, Ferreri and Ferro 1990). Likewise, studying the scaling properties of selected rainfall quantiles and applying this methodology, Bara et al. (2009) derived the IDF curves for durations shorter than a day, calculated from a historical dataset covering the whole territory of Slovakia. In a recent paper, Rodríguez-Solà et al. (2017) used this methodology also to reproduce the well-known empirical IDF curves at three Spanish locations: Barcelona (Casas et al. 2004, Rodríguez-Solà et al. 2014), the Ebre Observatory (Pérez-Zanón et al., 2016), and Madrid (Casas-Castillo et al., 2018b), taking into consideration the scaling behaviour of rainfall.

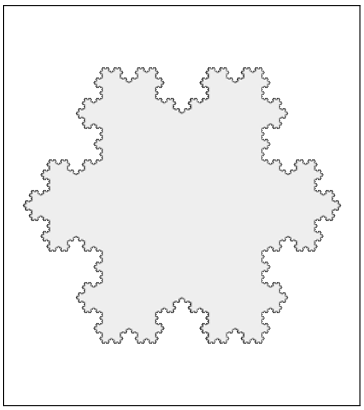
5.1 Fractals and rainfall

The term *fractal* was first coined by Benoît Mandelbrot (1924-2010), the mathematician known as the father of fractal geometry. The term comes from the Latin *fractus* that means irregularly broken. Fractal objects already existed in geometry at the time, like the Sierpiński triangle or the Koch curve (see, Fig[5.1a] and Fig[5.1b]). Mandelbrot, in his two most known books (Fractals: Form, Chance and Dimension, Mandelbrot 1977 and The Fractal Geometry of Nature, Mandelbrot 1982), showed the existence of the fractal geometry in nature (see Fig[5.1c]) and laid the foundations for the development of fractal theory. Nowadays, fractal theory has applications in many fields, from the modelling of stock market prices to computer simulation of artificial landscapes. Indeed, natural landscapes have fractal characteristics and by knowing their fractal dimension, a computer can represent artificial mountains and clouds that look like real ones. As Mandelbrot stated: "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line".

Schertzer and Lovejoy (1991) studied the spatial scaling properties of rain and cloud fields and proved those fields to have a multifractal behaviour and supported the idea that the atmosphere has a multifractal structure across meteorological scales. The authors concluded that their results could be implemented in a method for correcting radar measurements of rain (Schertzer and Lovejoy, 1991)). Indeed, rainfall presents a spatial self-similarity where synoptic areas could present low precipitation intensities but they would have embedded mesoscale areas with higher intensity and, in turn, those areas contain smaller cells of different intensities in a hierarchic structure that can be modelled in a multiplicative cascade model (a fractal distribution of points produced by a multiplicative and iterative random process).



(a) Sierpiński triangle is an equilateral triangle that has been subdivided in four triangles and the central one has been removed. The process repeated infinitely with the remaining triangles leads to an object that is in a plane (2 dimensional) but does not fill it completely. Its fractal dimension is 1.585.



(b) Koch curve (also known as Koch snowflake). It is a line of infinite length that circles a finite area. It is more than a line (1 dimensional) but cannot cover a plane (2 dimensional). Its fractal dimension is 1.262.

(c) romanesco broccoli has a structure that repeats itself at different scales (source: photographed by the author). Vegetables usually have fractal structures on the outside while animals have fractals on the inside (e.g., the ramifications of the lungs or the circulatory system).

Figure 5.1: examples of classical fractal objects in mathematical geometry and in nature.

This processes are considered to have structures of large characteristic scale that are fragmented in a random number of subestructures of a smaller characteristic scale. In a similar way that turbulent processes are considered to transfer energy from large to small scales, in a multiplicative cascade model of rainfall it is assumed that water is transferred across scales in the atmosphere (see Fig[5.2]).

Rainfall also presents temporal self-similarity characteristics across scales that can be analysed by considering the fractal theory. Temporal fractality of rainfall is less well studied and defined than spatial fractality, but a box-counting of occurrence or exceeding a threshold can be used to determine the fractal temporal dimension (see Fig[5.3]).

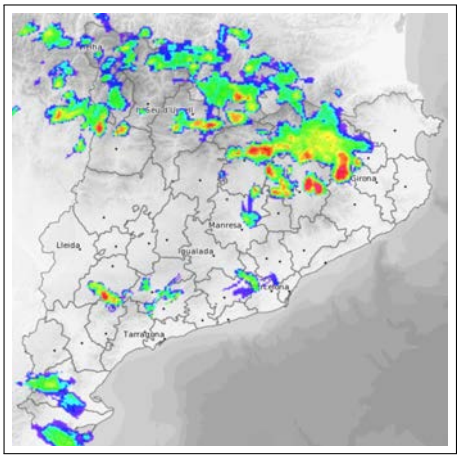


Figure 5.2: radar image over Catalonia on the 19^{th} of September 2019 (source: Meteorological Service of Catalonia). Rainfall intensity is spatially distributed following a fractal pattern, i.e., large scales present embedded cells of high intensity that alternate with areas without rain showing a self-similarity across scales.

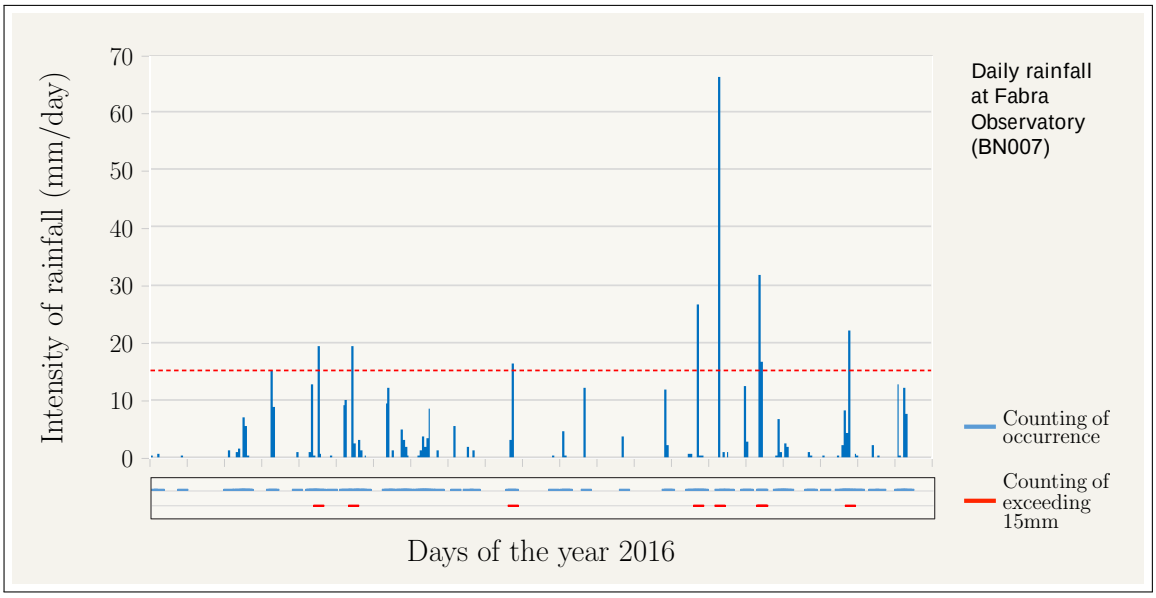


Figure 5.3: daily rainfall collected at Fabra Observatory (Barcelona) during the year 2016. Below the plot, a box-counting of occurrence (blue) and exceedance (red) is displayed. The box-counting presents a fractal pattern, but it depends on the chosen threshold.

5.2 Fractal downscaling methodology

Many atmospheric processes, rainfall generation among them, act in a wide temporal range giving rise to phenomena which accomplish self-similarity, i.e., that behave in the same way regardless of the temporal scale at which they are observed. This kind of processes can be considered of fractal type, with properties manifesting power laws of the scale λ, which is the ratio t/t_0 between any two durations t and t_0 within a scaling regime. In general, the fractal self-similarity of natural processes has a statistical nature; thus, the scaling properties of rainfall can be expressed by statistical relationships (Schertzer and Lovejoy 1987, Grupta and Waymire 1990, Schertzer and Lovejoy 2011). For instance, it has been widely observed (Koutsoyiannis and E.Foufoula-Georgiou 1993, Burlando and Rosso 1996, Menabde et al. 1999) that the probability distribution of the annual maximum rainfall intensity satisfies scale relationships, meaning that the probability distribution of the annual maximum intensity for a duration t, I_t, and the distribution at an other time scale $t_0 = \lambda t$, $I_{\lambda t}$, can be related by a factor that is a power function of the scale parameter λ. This property, usually referred as *simple scaling in the strict sense* (Grupta and Waymire 1990, Yu et al. 2004), can be expressed by Eq[5.1], where the symbol $\overset{\text{dist}}{=}$ indicates equality of probability distributions, and β is a scaling parameter, known hereafter as the *scaling exponent*. Eq[5.1] implies that the statistical moments of these two distributions fulfil the equality, as well as their quantiles and the rest of statistical features. In terms of the statistical moments of order q of the rainfall intensity for

a duration t, $\langle I_t^q \rangle$ defined in Eq[5.2], the scaling relationship can be expressed as Eq[5.3].

$$I_t \overset{\text{dist}}{=} \lambda^\beta I_{\lambda t} \tag{5.1}$$

$$\langle I_t^q \rangle = \frac{\sum_{i=1}^{n} I_{t_i}^q}{n} \tag{5.2}$$

$$\langle I_t^q \rangle = \lambda^{\beta q} \langle I_{\lambda t}^q \rangle \tag{5.3}$$

The exponent βq can be considered as the linear case of a general scaling function $K(q)$, a function resulting nonlinear in the multifractal case. The simplest procedure to determine the scaling exponent β from daily data is to calculate the statistical moments using Eq[5.2] of maximum annual series calculated by aggregation from daily series (with rainfall amounts for 2, 3, 4... days) for different values of the order q, and perform a linear regression between the logarithmic values of these moments and the logarithm of the duration t for every value of q. The straight lines obtained (in the case of simple scaling), each one of them with a slope of value βq, evidence scale invariance.

The equality of the quantiles of the probability distribution of Eq[5.1] implies that these quantiles are also be related by the same scaling relationship. In particular, the scaling relationship which corresponds to an extreme rainfall intensity $I(t, T)$, with a return period T, and a duration t, i.e. IDF curves, can be expressed by Eq[5.4], where daily duration appears as the reference duration $t_0 = 24h$.

$$I_{(t,T)} = \frac{t}{24}^{\beta} I_{(24,T)} \tag{5.4}$$

Once known the scaling exponent β, Eq[5.4] can be used to downscale daily values $I(24, T)$ to IDF values for sub-daily durations t; always under the assumption that the simple scaling relationship is fulfilled by sub-daily durations, which can only be considered as an approximation.

5.3 Results from other authors

In Catalonia and Spain, there has has been some recent research on the implementation of the monofractal theory for the downscaling of intensity-frequency relationships at a duration shorter than one day.

Rodríguez-Solà et al. (2017) obtained the IDF curves for a hundred Spanish locations for which only daily rainfall data were available, and found a spatial distribution of the observed scaling behaviour over Spain in concordance with the characteristic rainfall pattern in diverse areas (see Fig[5.4]). Rodríguez-Solà et al. (2017) also reproduced the well-known IDF curves of three Spanish locations (Fabra and Ebre observatories in Catalonia and Retiro observatory in Madrid) using the monofractal technique. The curves

reproduced satisfactorily the classical results when downscaling daily records for durations above 1 h, with mean relative differences lower than 7%. Discrepancies between the downscaled values and the known values $I(t, T)$ for durations shorter than 1 h resulted slightly higher (around 20% in the worst cases) and seemed to depend on the kind of measuring instrument.

Meseguer-Ruiz et al. (2019) studied the temporal fractality of rainfall in peninsular Spain and the Balearic Islands by calculating the fractal dimension of rainfall at 48 observatories. Meseguer-Ruiz et al. (2014) calculated the fractal dimension, a numerical expression of the temporal distribution of rainfall that can be understood as an indicator of self-similarity characteristics, in 20 observatories in Spain. The authors linked the fractal dimension to the concentration index of precipitation.

García-Marín et al. (2013) studied the multifractal properties of subdaily rainfall data and obtained the IDF curves for Málaga (a location in southern Spain); they found that the intensity of rainfall in Málaga followed a scaling behaviour that was valid between 1 hour and 21 days.

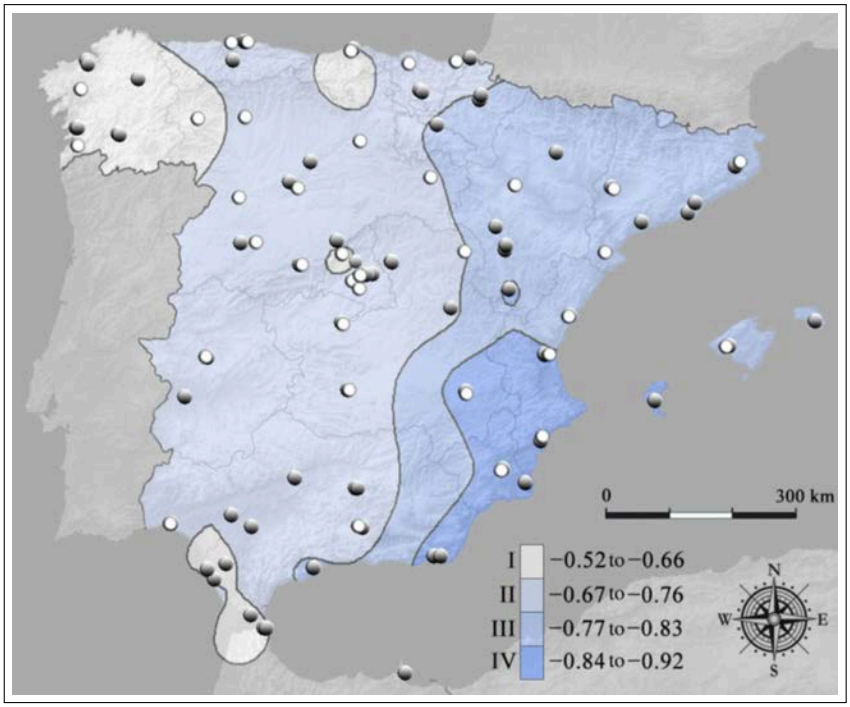


Figure 5.4: spatial distribution of the simple scaling exponents found for 100 Spanish stations. In white, test stations used by the authors to check the reliability of the applied methodology. Source: Rodríguez-Solà et al. (2017).

5.4 Applied procedure for the calculation of the scaling exponent

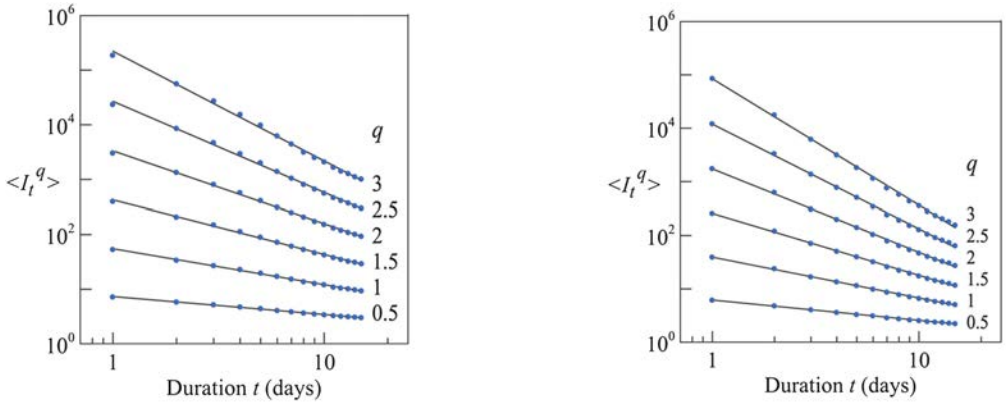
The phenomenon arising from the rainfall generation process presents a fractal self-similarity of statistical type, as it has been also observed in other natural processes (Schertzer and Lovejoy, 2011). Therefore, rainfall series generally show space and time scaling properties, being the self-similarity at different time scales the cornerstone of the present chapter. The simple scaling methodology is based on the statistical self-similarity of rainfall time series and it is quite broadly used to find or reproduce the Intensity-Duration-Frequency curves of a location.

In the present chapter, the obtained empirical scaling exponent (β), which will be used in subsequent chapters to obtain Intensity-Duration-Frequency relationships based on the monofractal methodology, is analysed and its relation to the characteristic rainfall patterns of the study area is investigated. The empirical scaling exponent has been obtained for the 163 series selected in Chapter 3 using daily data on the selected years in the period 1942-2016. For a particular part of the study of the scaling exponent, concerning the dependence of β on the chosen period and trends over the past decades, the years of used data have been extended.

Because of its definition (see, Eq[5.4]), the values of the simple scaling exponent are expected to be higher than -1. The limit value of $\beta = -1$ would correspond to rainfall samples with isolated annual maximum values; in the case of daily rainfall, a maximum value for a specific day P_1 surrounded by dry days. Thus, the process of aggregation leads to series where the precipitation for n days, P_n is the same regardless of the number of days. Then, in terms of intensity, $I_n = I_1/n$, which corresponds to a scaling exponent of $\beta = -1$ compared to the scaling relationship for $q = 1$ (mean), $I_1/n = n^\beta I_1$. The opposite (and hypothetical) case would be a totally regular sample where all days (within the n-aggregation) present the same rainfall amount, which implies the same intensity for all durations, and consequently, $\beta = 0$. In real rainfall cases, the scaling exponent ranges between $\beta = -1$ and a value close to $\beta = -0.5$. It seems reasonable then to expect some relationship between the scaling exponent values and rainfall series regularity, with the lowest values close to -1 corresponding to areas where rainfall is usually very irregular and convective, with sudden isolated maximum values, and higher values for rainy areas with a more regular rainfall pattern and uniform rain. For instance, Menabde et al. (1999) compared two sets of rainfall data representing two examples of quite different climate types, and found a scaling exponent of -0.65 for Melbourne (Australia), a city with a mid-latitude temperate climate and rainfall throughout the year, and a value of -0.76 for Warmbaths (South Africa), having a semiarid climate with summer convective rainfall, and concluded that the scaling exponent appears to be dependent on the rainfall and climate characteristics. In addition, Bara et al. (2009) found scaling exponents around -0.75 for three locations representative of the western (Kuchyňa-Nový Dvor), central (Liptovský Hrádok), and eastern (Humenné) areas of Slovakia, and Yu et al. (2004) found three types of rainfall scaling behaviour over northern

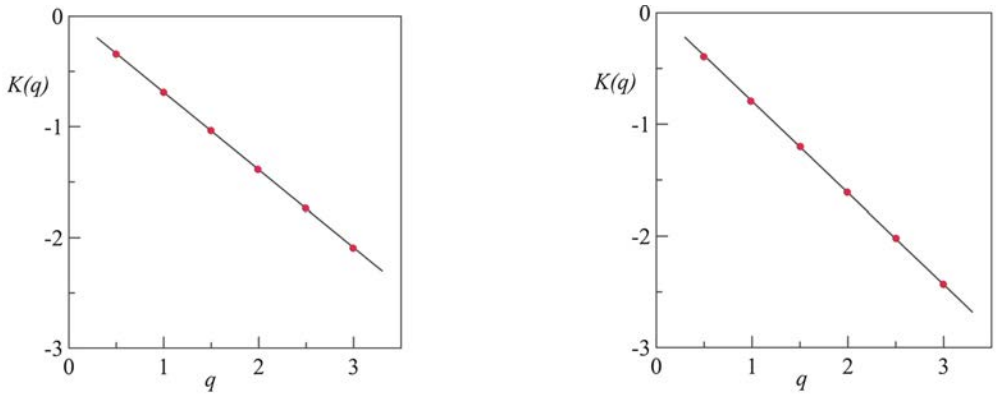
Taiwan, related to the change in topography and the influence of the northeast monsoon. Rodríguez-Solà et al. (2017) found a general concordance between the spatial distribution of β over the Iberian Peninsula and the mean annual precipitation distribution, with high values between -0.55 and -0.66 in rainy areas and low between -0.84 and -0.92 for the dry ones, with some discrepancies related to the kind of precipitation contributing to high rainfall amounts and the proportion of convective rainfall in total. In particular for Catalonia, Rodríguez-Solà et al. (2017) assigned a range of the scaling exponent between -0.77 and -0.83, based on the analysis of series from few stations located in this area. To investigate, in more detail for this area, the influence of geographical location and the different mechanisms of rainfall generation in the scaling behaviour, a new calculation of the simple scaling exponent has been performed in the present project from the selected 163 daily rainfall series of Catalonia and surroundings. Finally, a spatial distribution of its values is presented and analysed.

The scaling analysis presented hereafter has been performed for the 163 series selected on Chapter 3. Firstly, series of annual maximum of accumulated rainfall on 1 to 15 days were obtained from aggregation of daily data. Afterwards, these series were corrected by the multiplicative correction factor that corresponded to the location of the series and the season in which the maximum was produced at each year. The corrected series of annual maximum rainfall in 1 to 15 days were analysed and the q order statistical moments of rainfall intensities were calculated for values of q 0.5, 1, 1.5, 2 2.5 and 3, following Eq[5.2]. A linear regression between their logarithmic values and the logarithm of the duration (t) was performed for every value of q with the aim to evidence scale invariance and to find the value of the scaling exponent β for each case. As an example Fig[5.5] shows the log-log plots of the statistical moments against duration corresponding to two stations: Vielha and Lleida. These two stations have been selected, because they correspond to two different climates in Catalonia and they present quite different values of empirical β (Vielha -0.71 and Lleida -0.83). Vielha is located in the Val d'Aran County, a part of Catalonia where Atlantic influence dominates over Mediterranean and its climate is characterised by a regular precipitation through the year with high accumulated total amounts. On the other hand, Lleida's climate, in Segrià County, is classified as Dry Continental Mediterranean and rainfall characteristics imply low annual amounts in an irregular pattern with seasonal maxima in spring and autumn. In Fig[5.5], straight lines fitted by linear regression, each of them with a slope of value $K(q)$, indicate scaling invariance. The values of these slopes haven been displayed in Fig[5.6], where the linear behaviour of the scaling function $K(q)$ ($K(q) = -0.71q$ for Vielha and $K(q) = -0.83q$ for Lleida) shows the monofractal or simple scaling behaviour of these two specific series, which is a general result for all the analysed locations.



(a) VA010 Vielha (b) SE020 Lleida

Figure 5.5: logarithmic plot of statistical moments for different orders q of the annual maximum intensity.



(a) VA010 Vielha (b) SE020 Lleida

Figure 5.6: slope of statistical moments of the annual maximum intensity against the order q.

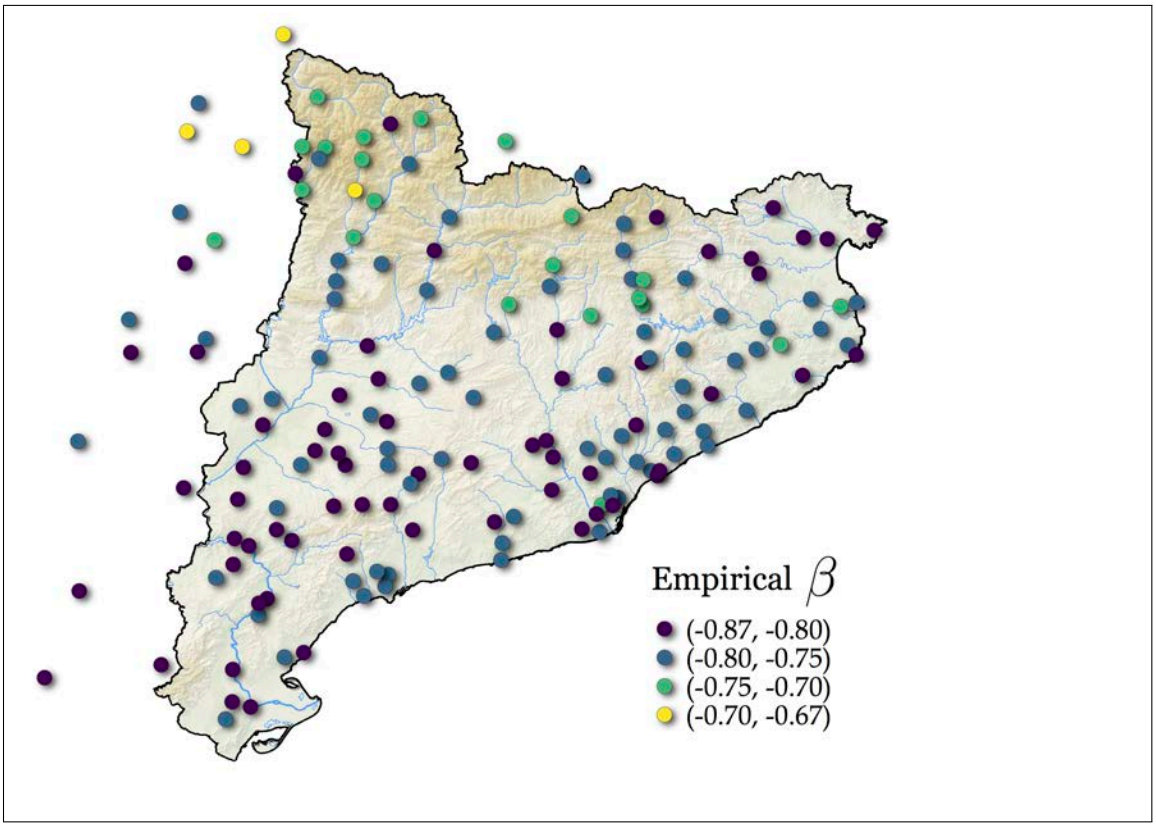
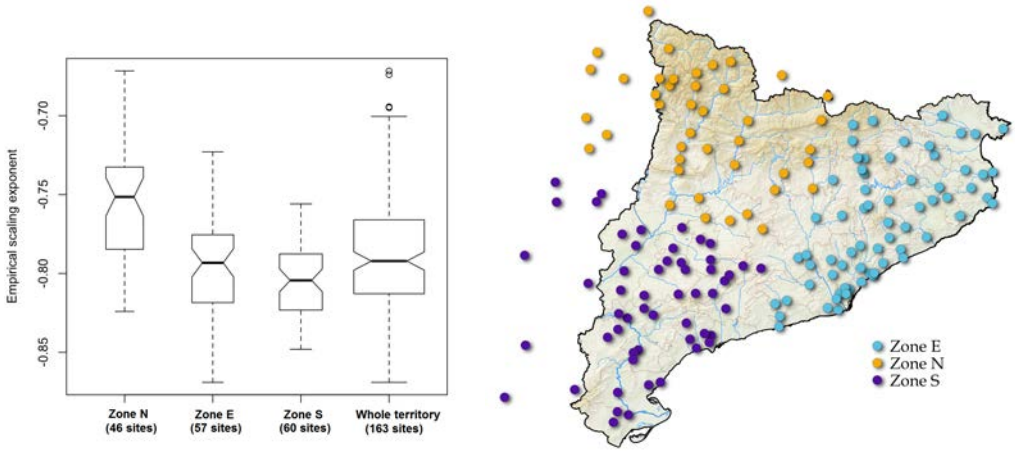


Figure 5.7: results of the empirical scaling exponent obtained in 163 locations with series of annual maxima in 24 hours in the period 1942-2016 (selected years).

5.5 Study of the empirical scaling exponent

The empirical results found for β show a wide variability within a short distance with a standard deviation value of 0.04 for the whole territory. The observed variability might be caused by the fact that different series have a different number of years available, being the shortest series (with 25 to 30 years) more influenced by particular episodes. However, the performed spatial analysis yields a certain pattern, namely higher values mainly concentrated in the northwest (see Fig[5.7]). While 90% of the empirical values range between -0.84 and -0.72 and have a mean of -0.79, there are two distinct zones which outstand: a northern area with a mean value of -0.75, matching a mountainous area with some Atlantic influence at its most northwestern corner, and a southwestern area with a mean value of β of -0.81, in great concordance with the driest areas in Catalonia. Indeed, a statistical analysis of the values obtained in these zones, i.e., Kolmogorov-Smirnov test for two samples, shows that the northern zone (Zone N in Fig[5.8]) and the southwestern zone (Zone S) come from different distributions, and hence, the samples are distinct. The different results of β values in these two distinct zones can be seen in Fig[5.8] where empirical values are shown grouped in boxplots according to separated zones.



(a) boxplot of obtained β in 163 locations grouped by zones.

(b) division of the territory in three zones that roughly correspond to regions with different climates and, especially, to zones with different empirical scaling exponents.

Figure 5.8: boxplot of empirical values of the scaling exponent (a) grouped in three zones according to their location (b).

5.5.1 Temporal trends of β

In the present subsection, a study of the dependence of the scaling exponent on the chosen time period is performed. Only for this subsection data will be used outside from the selected years in the period 1942-2016 and the specific period will be indicated in each case.

The objective of analysing the temporal evolution of the scaling exponent is to determine if it presents trends over time or if some specific years present distinctive values from other years. In the case that the β parameter were to depend on the specific years in which it is calculated, it would not be useful to infer long lasting characteristics of the study area because it would mean that the processes underlying rainfall generation are changing.

Fig[5.9] shows the temporal evolution of the scaling exponent calculated with 30 consecutive years of data starting at each year in the period 1911-1987 (where the last period of 30 years is 1987-2016 and it used the most recent available data in the project). The scaling exponent has been calculated in these cases using all available data in the selected 163 locations, regardless of the quality of the data. In each of the 30-year periods there were at least 10 locations where β could be calculated and all the available data are used to calculate the percentiles shown in Fig[5.9]. The percentile 0.5 of all the evaluated data presents a very low temporal trend that indicates, with a confidence level of 95%, a reduction in the value of β of 0.002 each decade. This is such a small trend that should not compromise the subsequent results of the project, but the observation of this behaviour seems worth studying.

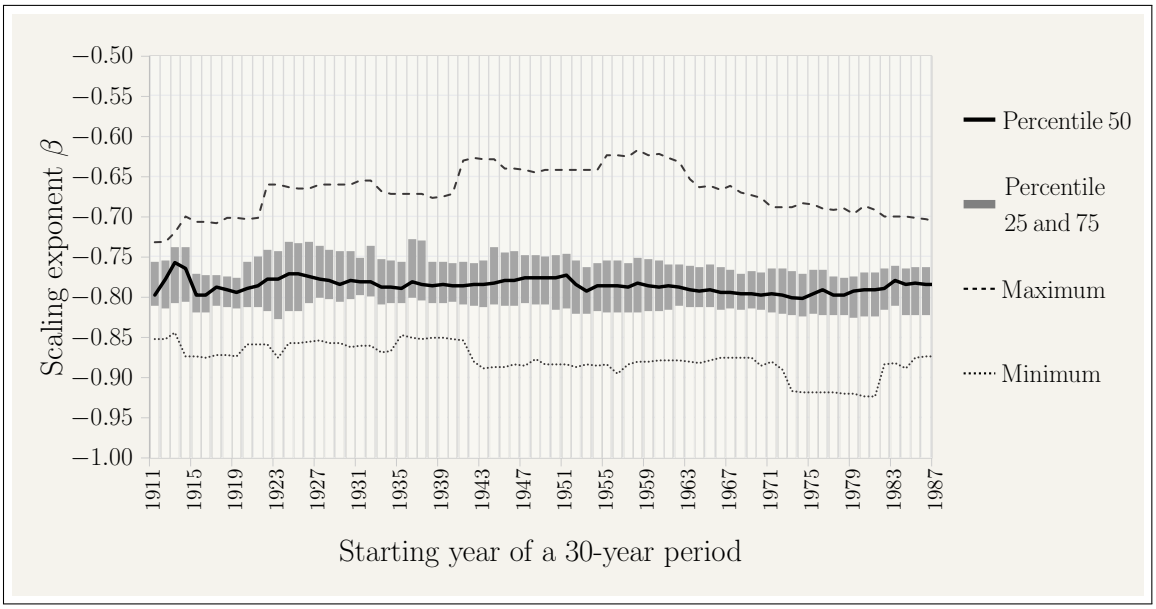


Figure 5.9: temporal evolution of the maximum, minimum and several percentiles of the scaling exponent calculated at 163 locations using data sliding periods of 30 years.

Fig[5.10] shows the temporal evolution of the scaling exponent obtained at mobile periods of 30 years considering the differentiated zones introduced in Fig[5.8b]. In this figure the maximum and minimum values obtained in the whole territory as well as the percentiles 0.25 and 0.75 of the whole territory plotted in Fig[5.9] are maintained for reference. The distinction of three zones in the territory reveals different behaviours of the scaling exponent: the average value of the northern zone presents a decreasing statistically significant trend of -0.006 per decade in the value of β and in the southern zone an increasing, also statistically significant, trend of 0.004 per decade. The eastern zone does not present a trend. As mentioned previously, these trends are two small to interfere in the results in a period of the order of 50 years, but it should be noticed that they exist. Recent studies in Catalonia hardly show significant trends of precipitation or its climatic indices (Servei Meteorològic de Catalunya, 2019), but a decreasing trend of β in the northern part of the territory could be related to an increase of convective rainfall in that region and the opposite in the southern part that presents an increase of the scaling exponent.

Finally, the temporal evolution of the scaling exponent has been analysed in the two longest and most complete series of the territory, i.e., Fabra and Ebre observatories.

In Fig[5.11] both observatories present a non neglectable variability depending on the 30-year period used in the calculation. The value of the scaling exponent in time has similar variations in both observatories, although the values are greatly differentiated especially in the most recent period. The variability of the scaling exponent in Fabra and Ebre supports the idea of using a spatial interpolation of β values to obtain IDF relationships at high spatial resolution (creating a grid) in a manner that the used

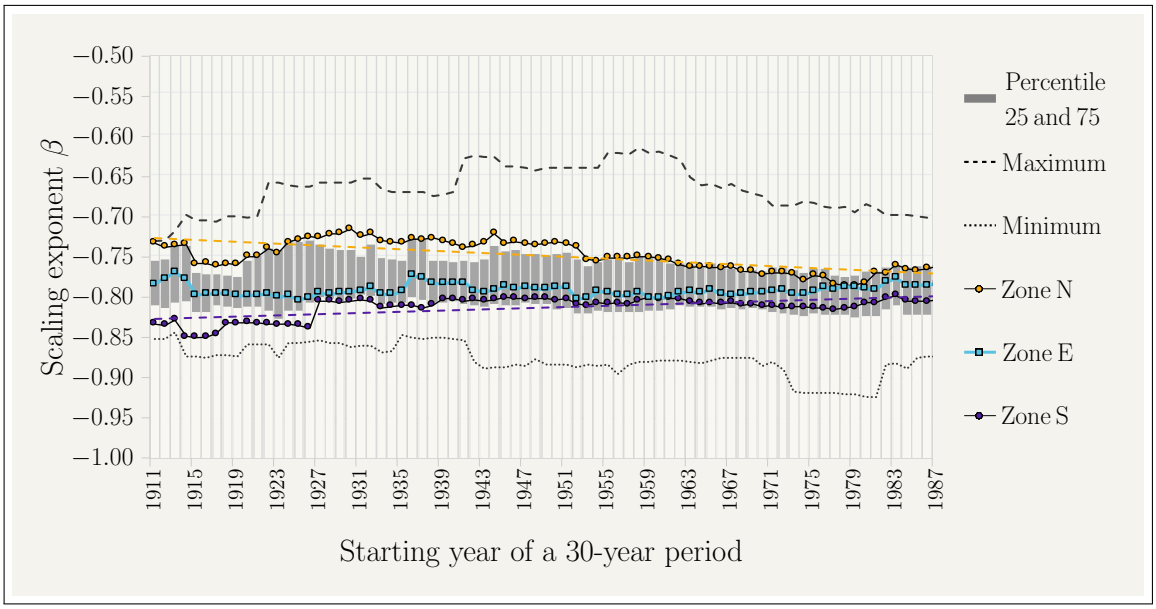


Figure 5.10: temporal evolution of the maximum, minimum and percentile 0.25 and 0.75 of the scaling exponent calculated at 163 locations using data sliding periods of 30 years. Also average value of the scaling exponent obtained in the same periods selecting three different groups of locations.

scaling exponent at each grid point is representative of the region, capturing its scaling characteristics and reducing the influence of using a short and specific time period.

5.5.2 Spatial distribution of β

As it has been discussed before and as it can be seen in Fig[5.7], the northern part of the study area has higher β values, especially at its most northwestern corner where some features of Atlantic climate are observed. In the northwestern area of Catalonia the climate is less influenced by the Mediterranean because of its distance to the coastline and the blockage of the high mountains of the Pyrenees. Moreover, the northwestern corner if often influenced by Atlantic fronts and its climate is characterised by high amounts of rain collected regularly through the year.

On the other hand, the southwestern zone, and especially the west of the study area, has a climate characterised by scarce rainfall, recorded mainly during Autumn and Spring. This western dry area presents some of the lowest β values obtained over Catalonia supporting its climatic observations of irregular rainfall patterns. Not so differently from the north, the western zone is not extremely influenced by the Mediterranean, being far from the coastline, it has a more continental climate than other areas in Catalonia; the main difference between Zone S and Zone N (see Fig[5.8b]) is the dryness regarding total annual amounts and the irregularity of the rainfall pattern with two seasons when most of the rain is collected.

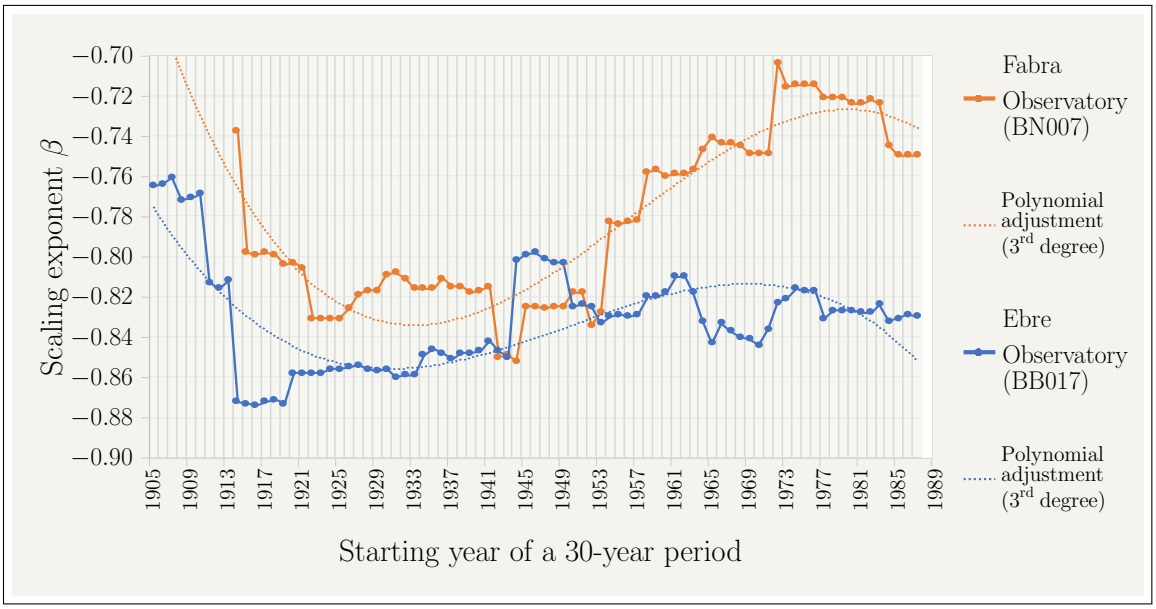


Figure 5.11: temporal evolution of the scaling exponent obtained at two different locations for sliding periods of 30 years.

Nevertheless, the pattern of β values empirically obtained shows other areas characterised by high or low values, such as low values in the eastern coastline where the rainfall pattern is known to be irregular, but these areas are more difficult to spot because of the high variability of the empirical results. To better analyse the spatial pattern of the scaling exponent and its relationship to climatic characteristics, the empirical values have been interpolated. The interpolation has been performed using a simple kriging technique through a hole effect model to fit the variogram (see Fig[5.12] for fitting details).

The spatial distribution of β after interpolation is shown in Fig[5.13]. The distinct zones previously discussed, namely, Zone N of high β values (less negative) and Zone S of low β values (more negative) are clearly captured by the interpolation, being the distinction between the northwestern area and the southwestern zone the most clear feature. Moreover, a third zone in the east, near the north part of Catalonia's Mediterranean coast, presents low values of interpolated β. The eastern coastline is clearly influenced by the Mediterranean and heavy

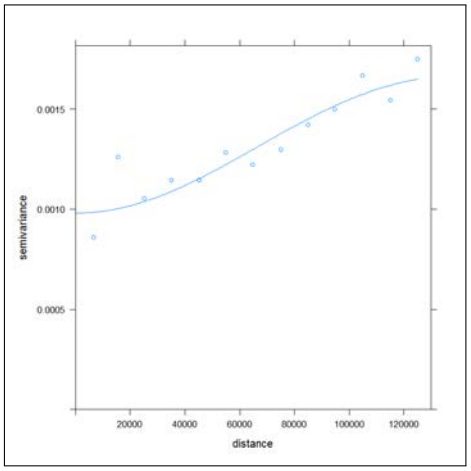


Figure 5.12: variogram of the empirical scaling exponent and fitting to a model with Hole effect that denotes anisotropy (different change of the parameter in different directions). The simple kriging interpolation, using this model and the mean value of empirical β in the whole territory results in a mean value of -0.7833, first quartile -0.7964, median -0.7838 and third quartile -0.7749.

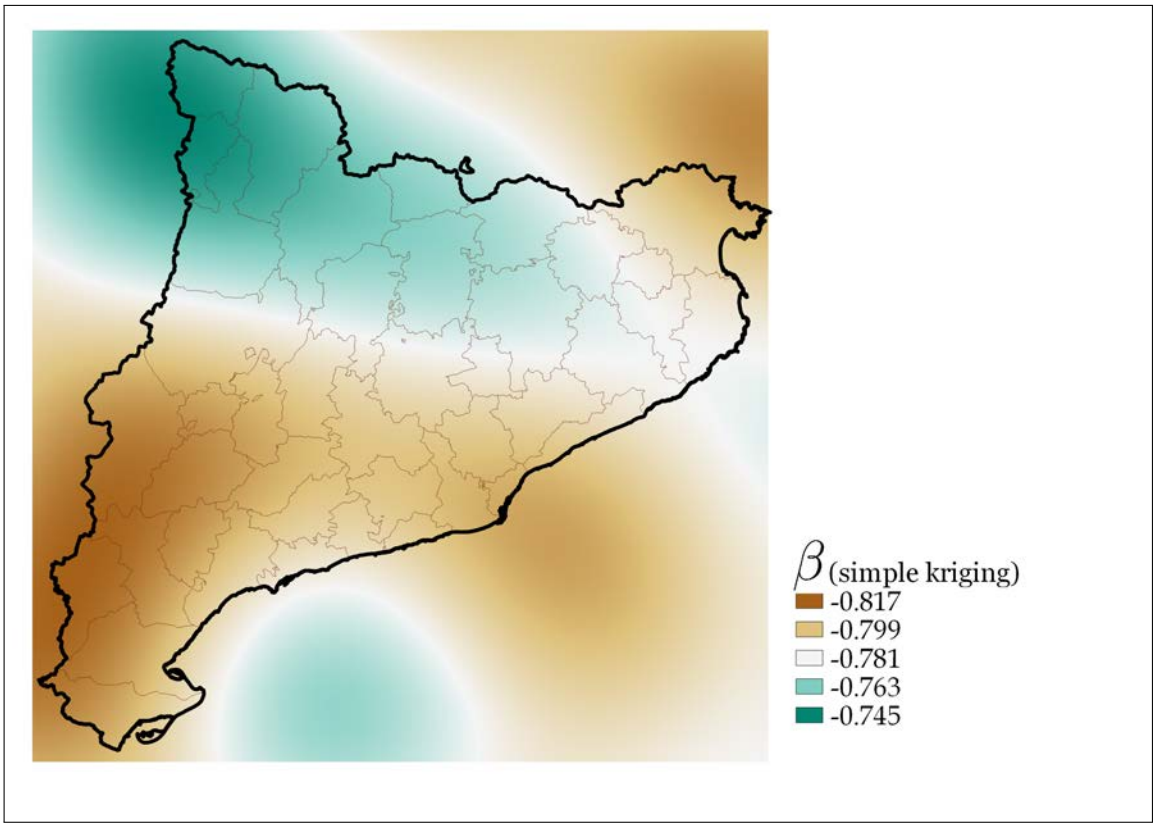


Figure 5.13: interpolated results of the empirical scaling exponent using a simple kriging technique from the calculation of β at 163 locations (the specific locations are shown, for instance, in Fig[5.8b]).

rain episodes occur often associated with Mediterranean low-pressure systems that bring wet eastern winds to this area. The rainfall pattern of this eastern area is characterised by low total annual amounts collected mainly in Autumn. Therefore, the climatic characteristics of this area suport the rainfall irregularity suggested by the empirical β values.

Overall, β values are observed to become more negative towards the coastline where the rainfall pattern becomes more influenced by the Mediterranean Sea, with the exceptions of the southern coastline. This exception is actually caused by an area of lower β values in the west associated with an irregularity of the rainfall pattern not caused, in this case, by the Mediterranean influence but by the droughtiness of the region. The regularity of the rainfall pattern associated with high β values is clearly supported by the climatic characteristics of the northwestern area.

As it can be seen in Fig[5.14a], there is some correlation between the scaling exponent and the mean annual rainfall, β increasing with mean annual rainfall. This correlation has been especially noticed in the two distinct zones highlighted before (high β values and annual rainfall amounts in the NW in contrast to western low β values

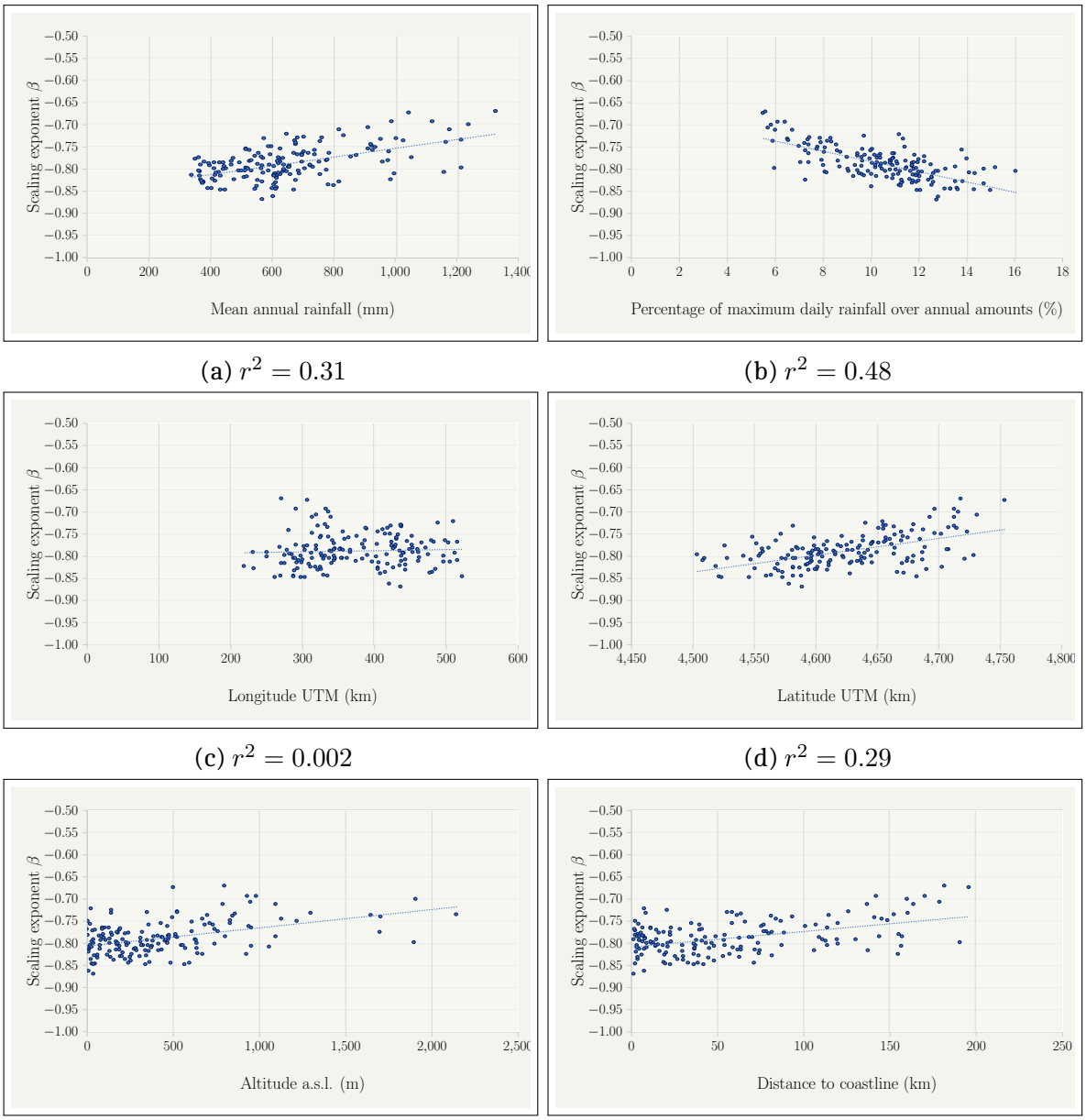
and dry areas). Fig[5.14b] shows that β slightly decreases with the percentage of collected annual precipitation during the maximum one day rainfall of the year. This contribution of the maximum rainfall day to the total amount is related to rainfall irregularity as a higher contribution of one day to annual amounts corresponds to heavy rain episodes sparse in time (the same case that corresponds to most negative values of β, i.e., $\beta \to -1$ following the scaling exponent's definition). Fig[5.14c] show that no dependence has been observed between β and geographical longitude, whereas there is some correlation, although small, with latitude (Fig[5.14d]), β increasing towards north, altitude (Fig[5.14e]), β being higher to greater height, and with the distance to coastline (Fig[5.14f]), β showing less irregularity of precipitation away from the Mediterranean influence.

5.6 Spatial grid of the empirical scaling exponent

The scaling exponent obtained and analysed in the present chapter is needed to transform intensity-frequency relationships of rainfall in 24 hours to intensity-duration-frequency relationships that are valid for several durations. This transformation is done using the scaling properties of statistical type following Eq[5.1]. IDF relationships can be obtained at the exact 163 previously selected locations using the same series of annual maximum rainfall in 24 hours (corrected by the multiplicative factor presented in Chapter 4). However, in order to obtain IDF relationships at high spatial resolution, the results will need to be spatially interpolated; as the performed interpolation of the scaling exponent has been analysed giving satisfactory results that are coherent with the climatic characteristics of the region, it has been decided to use the interpolation presented in this chapter and apply it to separately interpolated results of intensity at selected return periods (i.e., intensity-frequency relationships). In this way, the value that will be used of the scaling exponent will depend on the location but it will be smoothed by all the nearby results and, hence, not extremely dependent on the available data and time period of a specific location (as it has been seen in subsection 7.2.1 that it have been a potential issue).

Therefore, the spatially interpolated results presented in Fig[5.13] have been used to extract the scaling exponent at grid points at 1 km x 1 km resolution. In this manner, once the results of intensity-frequency are obtained, those can be interpolated and extracted at the same grid points for a selection of return periods and, then, the intensity-duration-frequency relationship can be calculated at high spatial resolution for any desired duration without needing any new interpolation.

Fig[5.15] shows the extracted values of the scaling exponent in a grid of 1 km x 1 km. In fact, having no power to discern the details, Fig[5.15] looks the same as Fig[5.13] (with a different legend), but it is actually a collection of spatial points with the assignation of a specific scaling exponent. Fig[5.15] may look as if the β values had been discretised, but it is not the case, they present a continuum of values just as Fig[5.13] did, only the spatial locations and the colour scale (to aid visualisation) have been discretised in Fig[5.15].



(a) $r^2 = 0.31$ (b) $r^2 = 0.48$

(c) $r^2 = 0.002$ (d) $r^2 = 0.29$

(e) $r^2 = 0.20$ (f) $r^2 = 0.22$

Figure 5.14: dependence of the empirical scaling exponent β (at 163 locations) on different rainfall and geographic characteristics of the location.

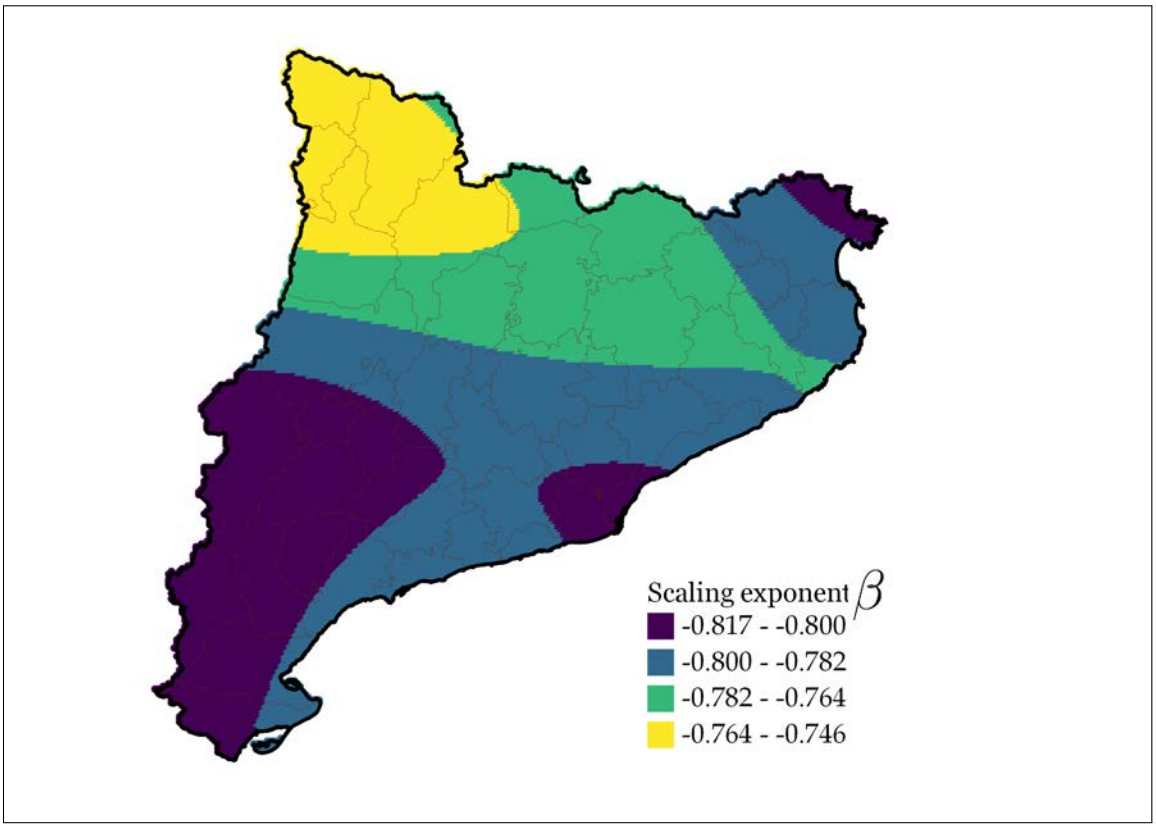


Figure 5.15: grid of 1 km x 1 km of scaling exponent values. The colour scale is discretised but the values used in the project have a continuum spectrum just as the results of the spatial interpolation.

Chapter 6

INTENSITY - FREQUENCY RELATIONSHIPS FOR 24 HOURS

The first step to obtain the IDF relationships, before applying the monofractal downscaling using Eq[5.1] ($I_t \overset{\text{dist}}{=} \lambda^{\beta} I_{\lambda t}$), is to obtain the distribution of the expected intensity at a reference duration, t_0. In this case, the reference duration is $t_0 = 24h$ as base rainfall data used in the project was measured at daily intervals. In the present chapter, the frequency distribution of maximum intensity is obtained for 24 hours duration and expected maximum precipitation in 24 hours at different return periods are presented.

The intensity-frequency relationship at each of the 163 locations of the study area selected in the project is obtained by adjusting the series of annual maximum intensity to a probability distribution function. The series of annual maxima are used given that the interest is in the maximum expected rainfall; the used series are the corrected series of annual maximum rainfall in 24 hours presented in Chapter 4 (after applying the correction factor). The extreme value theorem states that the Generalised Extreme Value Distribution (GEV) is the probability distribution to which a sequence of independent and identically distributed random variables converge.

A probability distribution function gives the probability of occurrence of a random variable. Whereas the cumulative distribution function (CDF) represents the probability that a random variable (in this case, rainfall intensity) will take values less than or equal to a given value. The GEV distribution is the general case of three probability distribution functions: Gumbel (or type I extreme value distribution function), Fréchet (or type II extreme value distribution function) and Weibull (or type III extreme value distribution function). The cumulative distribution function of the GEV distribution is defined by Eq[6.1], where s is the standardised variable ($s = (x - \mu)/\sigma$, where μ is the location parameter and σ the scale parameter) and ξ is the shape parameter.

$$F(s;\xi) = \begin{cases} e^{-(1+\xi s)^{-1/\xi}} \text{ if } \xi \neq 0 \\ e^{e^{-s}} \text{ if } \xi = 0 \end{cases} \tag{6.1}$$

The return period of an event of certain magnitude is defined as the mean time between occurrences of events that equal or exceed that magnitude (Chow et al., 1988). The return period (t) is related to the probability of exceedance by Eq[6.2] when the number of considered extreme events equals the number of years of the series, which is the case as the extremest event of each year is taken.

$$Prob(X \geqslant x) = \frac{1}{T} \tag{6.2}$$

6.1 Selection of probability distribution functions

As a general rule, the Generalised Extreme Value distribution was used to fit the series of annual maximum rainfall in 24 hours and obtain expected rainfall at different return periods for that fixed duration. However, the GEV distribution was not always suitable because for some series of maximum rainfall the adjustment to the distribution yielded a positive shape parameter that implies a very small increase of expected intensity at long enough return periods; this case is not a physical reflect of the behaviour of rain but probably a lack of rare heavy episodes in the sample. Fig[6.1] shows that the spatial distribution of the shape parameter of fitting a GEV distribution to the samples of maximum rainfall is not related to geographic characteristics or the climatology of rainfall in the territory in this case, although, Ragulina and Reitan (2017) concluded in their study that the shape parameter could be related to the dominant precipitation systems, but there was need of more research on the subject, and they found that it decreased with altitude (Ragulina and Reitan, 2017).

The fitting of the probability distributions was performed using the software R and the packages *lmom* (Hosking, 2019a) and *lmomRFA* (Hosking, 2019b); the last one was used to evaluate the goodness of fit. L-moments first developed by Hosking (1990) are robust in the presence of outliers, hence, suitable in the use of series of maxima. The analysis of L-moments was performed on the annual series of maximum rainfall to determine their statistical characteristics, mainly L-skewness and L-kurtosis parameters. The parameters of skewness and kurtosis of a sample can be used to chose a probability distribution function to which the sample can be fitted. L-skewness and L-kurtosis for the 163 series of maximum rainfall are plotted in Fig[6.2] that also shows the parameters that a sample should have to have an underlying probability distribution of the most common kinds.

As shown in Fig[6.3] the GEV distribution is not the distribution that has been used to finally fit all the samples. Fig[6.2] already showed that not all the samples can be fitted to a GEV distribution, but there are other distributions more appropiate instead; the number of annual maximum rainfall series fitted to each distribution is shown in Table[6.1].

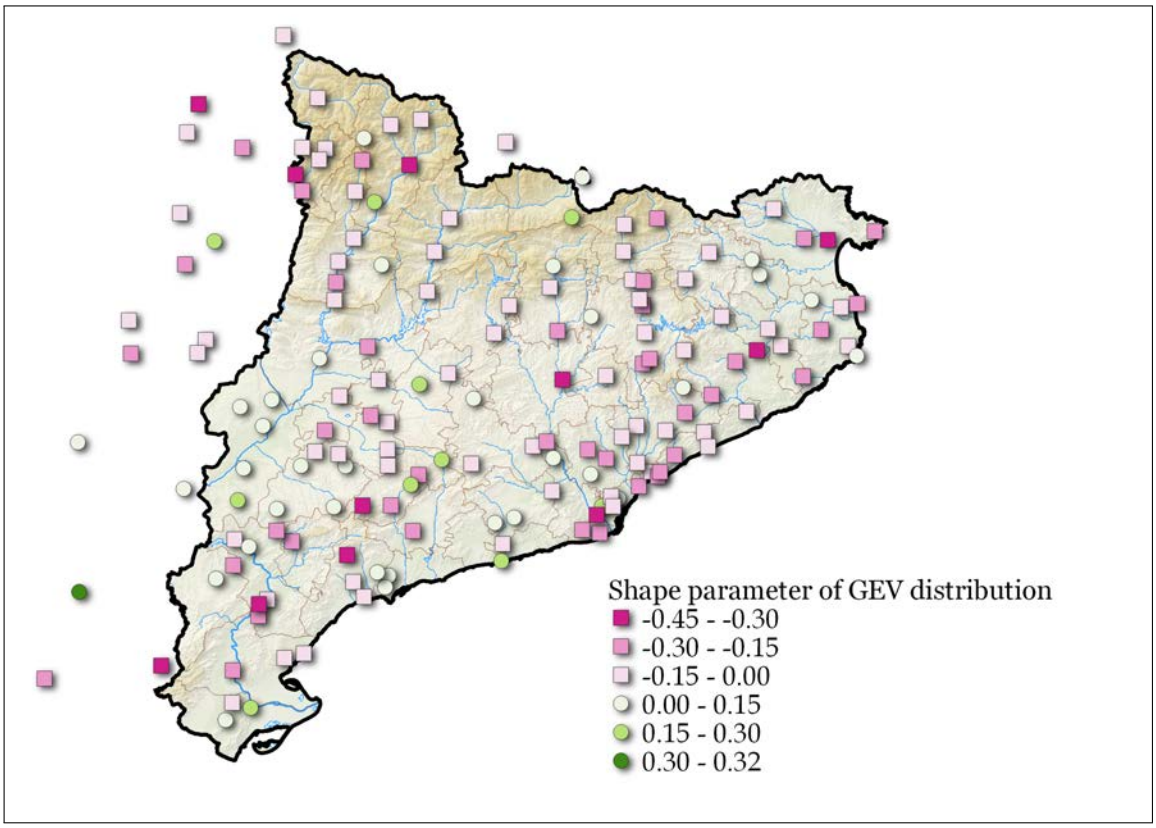


Figure 6.1: spatial distribution of the value of the shape parameter that would be needed to fit the selected series of annual maximum rainfall in 24 hours to a GEV distribution.

Probability Distribution name	**Number of series fitted to the distribution**
GEV	100
GUMBEL	44
GNO	7
GLO	5
GPA	4
PE3	3

Table 6.1: list of probability distributions used and number of series that have been fitted to each distribution. The GEV distribution was chosen unless data had a much better fit to another distribution or the results yielded unrealistic intensities at high return periods.

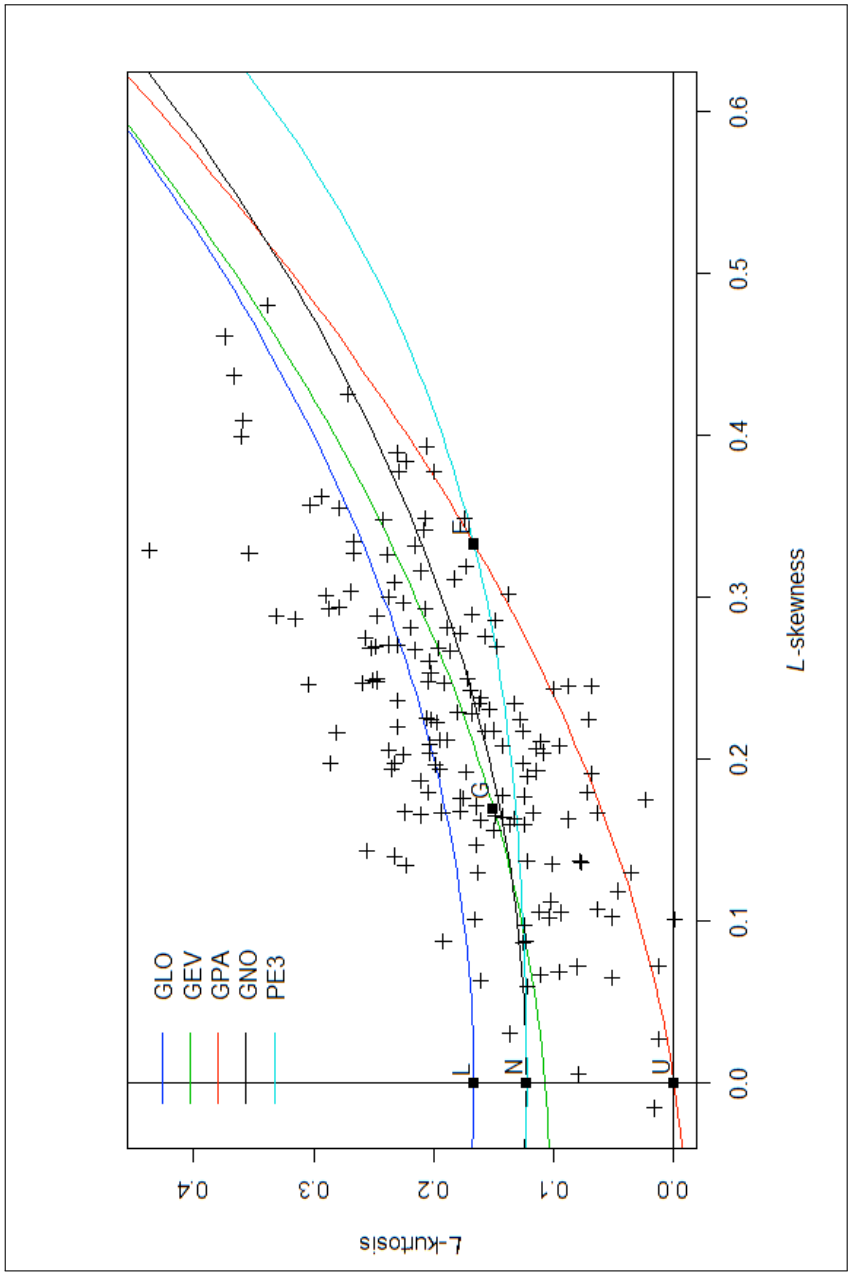


Figure 6.2: output of the software *lmom* (Hosking, 2019a) displaying the L-skewness and L-kurtosis of each of the samples and the combination of those parameteres that could have a sample that followed a GLO distribtuion (blue line), a GEV distribution (green line), a GPA distribution (red line), a GNO distribution (black line), a PE3 distribution (light blue line) or particular cases of the general distributions like the Gumbel distribution (central square).

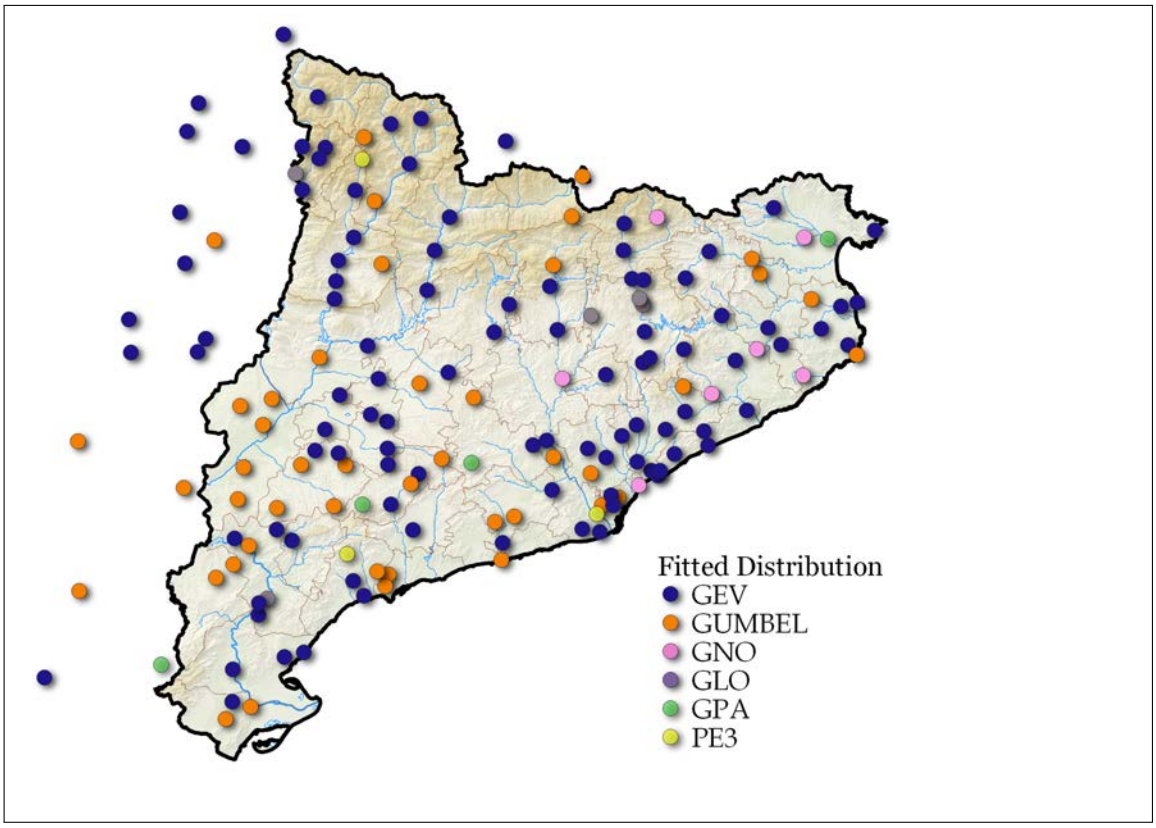
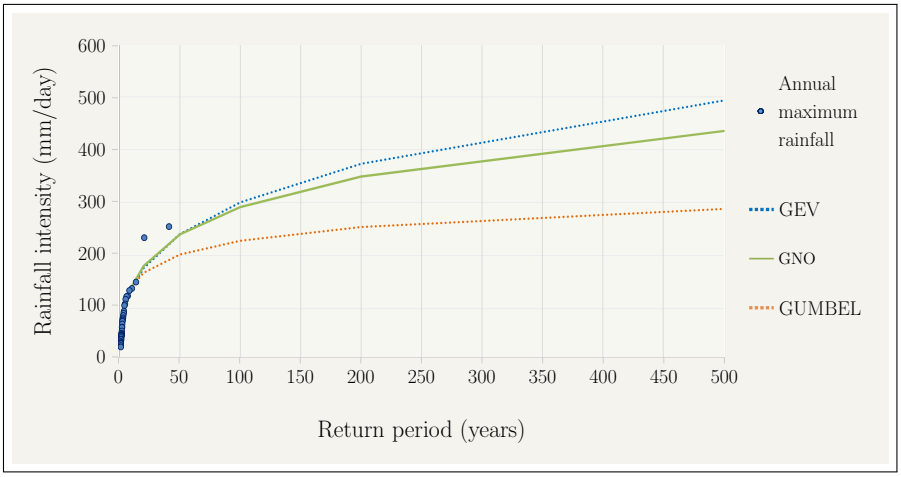
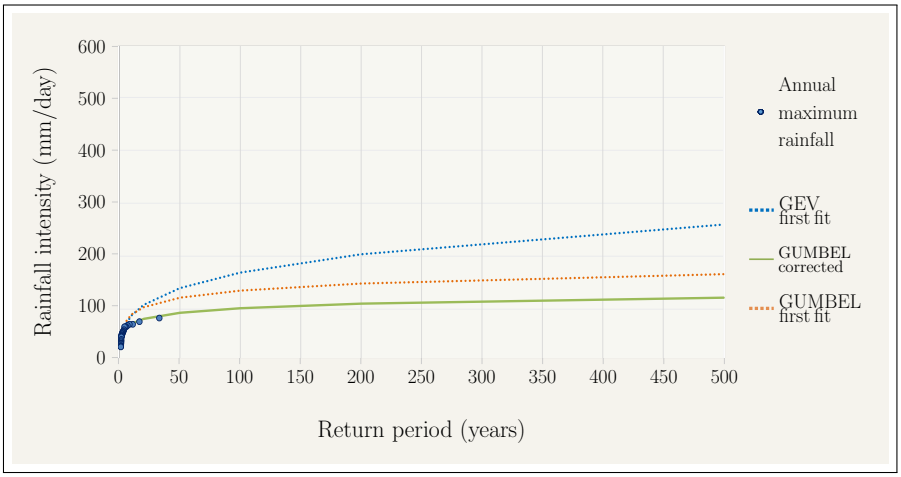


Figure 6.3: map of the study area showing the probability distribution fitted to the series of annual maximum rainfall in 24 hours at each of the selected locations.

The finally chosen distribution to fit each of the available series was selected after a careful analysis of the spatial distribution of expected intensity given by the prediction of different distributions. The visual inspection of the spatial distribution of the results allowed the detection of two errors in the original series that had previously been undetected despite the rigorous quality control: the series of *TA013 - La Fatarella - 2* had an absolute maximum of 199.0 mm and the series of *ES048 - Urdiceto Central* had an absolute maximum of 307.2 mm which, after checking daily data, were found to be erroneous values; original data was corrected and the series of maxima were recalculated (the scaling exponent was not recalculated as it was considered that by using the smoothed value with a spatial interpolation, the difference would be marginal). Fig[6.4] illustrates the difference in expected maximum rainfall that would be obtained depending on the chosen probability distribution; intensity at high return periods yields the most outstanding difference but the difference in expected intensity at low return periods can also be significant and, hence, it was also considered.



(a) annual maximum rainfall in *AE041 - Figueres - Escola La Salle* (dots) and expected maximum intensity according to the prediction of several distribution to which the sample can be fitted. In this case, the GNO distribution was used instead of GEV because otherwise the expected intensity at low return periods gave low values that were outstanding in a spatial interpolation. The GUMBEL distribution was not used, despite its goodness of fit for low return periods, because the absolute maximum and the second maximum of the series showed higher intensities better predicted by GEV or GNO and these two maxima were checked and considered correct values.



(b) annual maximum rainfall in *TA013 - La Fatarella - 2* (dots) and expected maximum intensity predicted by several distributions: dotted lines show the prediction before correcting the absolute maximum of the series that was found to be an error, despite being an area with low expected intensities, predicted intensity at high return periods had an outstanding value relative to nearby locations and this fact led to spot an erroneous value in daily data (a maximum of 119.0 mm); the solid green line shows the prediction of the fitted GUMBEL distribution after correcting the series of maxima.

Figure 6.4: examples of expected rainfall intensity depending on the return period, (i.e., intensity-frequency relationship) predicted by different probability distributions at two locations.

6.2 Expected rainfall in 24 hours

Once the series of annual maximum rainfall in 24 hours were fitted to a suitable probability distribution at each location, as discussed in the previous section, the prediction of rainfall intensity of the chosen probability distributions was obtained for several return periods. The return periods chosen to obtain expected maximum intensity were 2, 5, 10, 20, 50, 100, 200 and 500 years. The results of expected rainfall intensity at the mentioned return periods are shown in Table[6.2]. From these data a spatial interpolation was performed for each selected return period, thus obtaining the intensity - frequency relationship of precipitation at high spatial resolution. The spatial interpolation was performed, in this case, using the software ArcGIS to implement a geographically weighted regression that uses altitude as an independent variable and kriging to correct the residuals; this interpolation methodology is the one that is more often used in the SMC to spatially interpolate precipitation data.

The results of the spatial interpolation of the maximum expected intensity for rainfall events of 24 hours at three of the selected return periods are shown in Fig[6.5], Fig[6.6] and Fig[6.7]. There is an overall similarity of the spatial patterns of maximum rainfall with maps of annual rainfall depth; the most outstanding feature is the lowest maxima in the western zone in concordance with the driest areas of Catalonia. However, discrepancies in some areas indicate the effect of convective rainfall. Val d'Aran, in the northwestern corner of Catalonia, is a rainy region (in total annual depth) with low daily and subdaily maxima instead; continuous rain with low convection produce high total amounts without short time maxima. Maximum expected rainfall is mainly found with in the southern part of the territory (Montsià county) and in the northeastern part (Alt Empordà and Ripollès) with similar values (see Fig[6.14] to localise the mentioned counties).

Maximum values found in the southern part shift their location depending on the return period: for lower return periods, maximum values are found in the Ebre's delta, while for high return periods maximum values are displaced 30 to 40 km to the west (where El Port mountainous area reaches altitudes of 1,400 m a.s.l.). A relative maximum found in the Montseny mountains (east of central territory) is important for low return periods and in Fig[6.5] it can be seen the high expected intensities are not isolated in that part but are linked to high intensities in the eastern Pyrenees by the values found in Garrotxa and in between Osona and La Selva counties. This maximum expected intensity in Montseny area for low return periods is mainly caused by the intensities registered at the weather station of *VR063 - Turó de l'Home* (1,669 m a.s.l.). Series of maximum rainfall in this particular location have been adjusted to a GUMBEL distribution that predicted the absolute maximum expected rainfall intensities in Catalonia for low return periods, although the predicted intensity does not increase as much as in other locations with the increase of the return period and it presents the 15^{th} maximum for 100 years return period and the 25^{th} for 500 years.

ID	Fitted Distribution	Intensity (mm/dia) T = 2 years	Intensitat (mm/dia) T = 5 years	Intensitat (mm/dia) T = 10 years	Intensitat (mm/dia) T = 20 years	Intensitat (mm/dia) T = 50 years	Intensitat (mm/dia) T = 100 years	Intensitat (mm/dia) T = 200 years	Intensitat (mm/dia) T = 500 years
AC010	GEV	58.1	76.6	92.0	109.8	138.2	164.2	195.1	245.0
AE010	GPA	72.8	124.9	168.0	214.5	281.7	337.3	397.3	484.0
AE024	GEV	102.4	149.2	180.4	210.6	249.9	279.6	309.3	348.9
AE041	GNO	62.7	105.1	140.6	180.3	240.2	291.8	349.4	435.4
AE052	GEV	83.0	125.9	159.2	195.3	249.3	295.7	347.7	426.5
AP011	GUMBEL	64.1	84.2	97.4	110.2	126.6	139.0	151.3	167.5
AP023	GEV	82.4	112.8	135.7	159.8	194.8	224.0	255.9	302.8
AP024	GEV	64.0	85.6	99.9	113.7	131.7	145.2	158.8	176.7
AP027	GUMBEL	80.8	104.5	120.2	135.2	154.7	169.2	183.8	202.9
AU013	GEV	54.1	73.1	86.4	99.9	118.3	132.9	148.2	169.4
AU022	GEV	64.8	89.3	106.5	123.6	146.9	165.2	184.1	210.2
AU024	GEV	59.0	76.3	88.5	100.9	117.9	131.4	145.6	165.5
AR001	GEV	74.4	95.1	109.6	124.0	143.5	158.8	174.6	196.4
AR002	GLO	68.9	92.4	114.7	143.7	197.4	254.8	332.7	480.1
AR004	GEV	68.6	85.5	98.1	111.2	130.1	145.7	162.7	187.4
AR008	GEV	73.7	95.4	111.9	129.6	155.5	177.3	201.5	237.6
AR014	GEV	66.9	87.5	103.5	120.7	146.2	168.0	192.4	229.1
AD004	GEV	62.2	82.9	98.6	115.2	139.5	159.9	182.3	215.5
AN004	GEV	70.7	95.1	111.8	128.2	150.1	166.9	184.1	207.5
AN007	GUMBEL	53.6	69.0	79.2	89.0	101.6	111.1	120.6	133.0
AN027	GPA	69.7	98.9	116.8	131.8	147.9	157.8	166.0	174.9
BA007	GNO	53.3	74.9	94.6	117.9	154.9	188.3	226.9	286.8
BA053	GEV	69.6	92.4	111.6	134.0	169.9	203.2	242.9	307.6
BC011	GEV	69.2	94.5	113.3	132.9	161.1	184.4	209.6	246.3
BC013	GEV	71.1	100.0	121.0	142.6	172.9	197.5	223.7	261.2
BC015	PE3	75.3	118.2	158.3	201.7	262.4	309.9	358.4	423.7
BC023	GUMBEL	66.9	89.6	104.6	119.0	137.6	151.5	165.5	183.8
BC026	GUMBEL	71.7	96.8	113.4	129.3	149.9	165.3	180.7	201.0
BB003	GEV	82.5	117.8	143.0	168.7	204.2	232.7	262.7	305.0
BB006	GEV	98.1	146.3	178.8	210.3	251.7	283.1	314.8	357.2
BB017	GEV	77.1	110.1	137.0	167.4	214.5	256.7	305.7	383.0
BE004	GNO	85.4	127.7	161.1	197.0	249.1	292.5	339.7	408.2
BE016	GEV	77.9	113.9	140.4	168.1	207.4	239.6	274.5	324.8
BE019	GUMBEL	79.3	111.5	132.8	153.2	179.7	199.5	219.2	245.3
BE031	GEV	74.4	97.5	115.5	135.1	164.5	189.9	218.4	261.8
BE039	GUMBEL	75.5	103.1	121.3	138.8	161.4	178.4	195.3	217.6
BE047	GEV	75.8	105.1	125.9	147.0	175.8	198.7	222.7	256.2
BE050	GEV	70.6	103.9	130.7	160.7	206.9	247.9	295.2	369.1
BL006	GEV	75.1	105.6	131.4	161.4	209.7	254.3	307.7	394.8
BL009	GEV	67.3	101.1	128.1	158.0	203.5	243.6	289.4	360.3
BL021	GUMBEL	74.6	100.5	117.7	134.2	155.6	171.5	187.5	208.5
BL045	GUMBEL	68.4	91.9	107.4	122.3	141.6	156.1	170.5	189.5
BL047	PE3	71.1	117.1	154.6	193.3	245.6	285.7	326.2	380.1
BN007	GEV	70.2	97.9	118.3	139.5	169.5	194.0	220.5	258.6
BN022	GEV	67.6	90.3	106.0	121.7	142.8	159.4	176.4	199.9
BN032	GUMBEL	70.0	101.1	121.7	141.5	167.1	186.3	205.4	230.6
BN060	GNO	68.9	107.6	140.3	177.2	233.0	281.4	335.5	416.7
BG011	GUMBEL	81.1	107.1	124.3	140.8	162.2	178.3	194.2	215.3
BG017	GEV	77.7	101.5	118.4	135.4	158.7	177.2	196.6	223.6
BG028	GEV	59.7	79.0	94.0	110.3	134.5	155.2	178.4	213.4
CE002	GUMBEL	48.5	64.4	74.9	85.0	98.1	107.9	117.7	130.6
CE010	GUMBEL	97.1	127.9	148.4	168.0	193.3	212.3	231.3	256.3
CB003	GPA	73.5	117.9	157.0	201.5	270.0	330.3	399.1	504.9
CB011	GEV	62.0	84.8	102.6	122.0	151.0	176.1	204.4	247.3
CB015	GEV	53.6	72.9	87.7	103.7	127.5	147.8	170.5	204.6
CB017	GUMBEL	53.7	68.0	77.6	86.7	98.5	107.4	116.2	127.8
CB022	GUMBEL	56.7	68.7	76.6	84.2	94.1	101.4	108.8	118.5
ES021	GEV	47.7	63.6	76.2	90.2	111.4	130.0	151.2	184.0
ES029	GUMBEL	52.3	71.9	84.9	97.3	113.4	125.5	137.5	153.4
ES033	GUMBEL	44.6	60.5	71.0	81.1	94.2	104.0	113.7	126.6
ES036	GUMBEL	56.6	74.6	86.5	98.0	112.8	123.9	135.0	149.5
ES048	GEV	77.7	106.6	128.5	151.8	185.5	213.9	245.1	291.2
ES057	GEV	78.3	102.0	118.0	133.7	154.4	170.3	186.3	208.0
ES069	GEV	63.9	85.3	101.1	117.6	141.1	160.5	181.4	211.7
ES073	GEV	56.3	76.5	93.4	112.7	143.4	171.5	204.6	258.0
ES077	GEV	73.5	97.8	116.3	136.1	165.2	189.9	217.1	257.8
ES080	GUMBEL	51.9	68.2	79.0	89.3	102.7	112.8	122.8	136.0
ES102	GEV	48.8	65.9	78.0	90.2	106.9	120.2	134.1	153.5
ES103	GEV	47.0	65.4	80.7	98.3	126.2	151.7	181.8	230.2
ES113	GEV	48.0	65.5	78.2	91.2	109.3	124.0	139.6	161.8
ES115	GEV	44.0	60.2	71.7	83.4	99.7	112.7	126.4	145.8
ES121	GPA	76.2	124.5	164.2	206.7	267.8	317.9	371.7	448.9
AG003	GEV	61.8	79.8	92.3	104.7	121.5	134.6	148.1	166.6
GF001	GUMBEL	64.1	84.6	98.2	111.3	128.1	140.8	153.4	170.0
GA002	GUMBEL	46.2	61.1	71.0	80.4	92.7	101.8	111.0	123.1
GA004	GEV	43.9	57.8	67.6	77.4	90.8	101.3	112.3	127.5
GA011	GEV	42.4	58.6	70.4	82.4	99.2	112.7	127.1	147.5
GA013	GUMBEL	43.9	55.9	63.9	71.5	81.4	88.8	96.2	105.9
GA017	GUMBEL	61.8	85.3	100.9	115.8	135.2	149.7	164.1	183.2
GA027	GUMBEL	50.5	65.6	75.5	85.1	97.4	106.7	115.9	128.1
GX009	GEV	89.0	121.5	143.1	163.7	190.5	210.5	230.5	256.9
GX021	GEV	101.2	143.1	174.3	207.2	254.5	293.8	336.6	399.1
GX030	GUMBEL	96.0	133.7	158.7	182.6	213.6	236.9	260.0	290.6
GI012	GEV	82.7	117.1	141.3	165.7	198.9	225.2	252.6	290.6
GI017	GEV	75.4	103.4	123.5	144.1	172.7	195.7	220.0	254.4
MA005	GEV	67.4	91.1	108.8	127.5	154.3	176.6	200.9	236.4
MA009	GEV	63.1	88.7	108.5	130.0	162.0	189.4	220.2	266.7
MA015	GEV	69.2	102.6	130.9	163.6	216.3	265.0	323.2	418.1
MA026	GEV	75.7	105.1	128.1	153.2	190.8	223.2	259.8	315.5
MA042	GEV	75.1	101.3	119.1	136.5	159.4	176.9	194.6	218.5
MA048	GEV	65.9	85.6	99.9	114.7	135.4	152.3	170.4	196.2
MA059	GEV	77.8	106.1	124.9	143.1	166.8	184.6	202.5	226.3
MI006	GEV	70.9	95.1	111.6	127.9	149.6	166.2	183.2	206.3
MO014	GUMBEL	105.3	154.0	186.2	217.1	257.0	287.0	316.8	356.2
MO016	GEV	83.9	120.8	147.2	174.2	211.5	241.5	273.1	317.8
MO018	GUMBEL	85.7	118.0	139.5	160.0	186.6	206.6	226.5	252.7
NO002	GEV	48.5	66.1	80.5	96.7	122.0	144.7	171.1	212.9
NO011	GUMBEL	49.4	65.3	75.9	86.0	99.2	109.0	118.8	131.8
NO024	GEV	44.4	59.8	70.3	80.7	94.5	105.1	115.9	130.6
OS005	GLO	58.8	74.5	85.1	95.9	111.4	124.2	138.1	158.7
OS009	GEV	66.7	93.0	113.7	136.5	170.9	200.9	235.0	287.2
OS016	GEV	76.6	103.2	122.4	142.0	169.2	191.0	214.2	246.9
OS017	GEV	71.2	91.1	107.4	125.9	154.7	180.7	210.9	258.9

ID	Fitted Distribution	Intensity (mm/dia) T = 2 years	Intensitat (mm/dia) T = 5 years	Intensitat (mm/dia) T = 10 years	Intensitat (mm/dia) T = 20 years	Intensitat (mm/dia) T = 50 years	Intensitat (mm/dia) T = 100 years	Intensitat (mm/dia) T = 200 years	Intensitat (mm/dia) T = 500 years
OS023	GLO	63.0	79.5	91.2	103.5	121.7	137.2	154.6	181.1
OS026	GLO	58.5	73.5	85.3	98.7	120.1	140.0	163.9	203.2
OS041	GEV	71.5	91.4	106.5	122.8	146.7	167.0	189.5	223.0
OS055	GEV	65.0	85.4	100.3	115.6	137.2	154.7	173.5	200.3
OS060	GEV	93.0	125.4	148.0	170.5	200.9	224.7	249.2	283.0
PJ003	PE3	79.8	117.0	143.9	170.3	204.7	230.5	256.1	289.8
PJ006	GEV	64.6	81.6	93.4	105.1	120.7	132.9	145.3	162.4
PJ015	GEV	56.0	70.5	80.4	90.1	103.0	112.9	123.0	136.6
PJ017	GEV	50.6	67.9	80.2	92.8	110.1	124.0	138.6	159.2
PJ024	GUMBEL	60.8	81.5	95.2	108.3	125.4	138.1	150.8	167.6
PJ026	GEV	53.2	71.4	85.7	101.3	124.9	145.4	168.6	204.0
PJ029	GEV	65.6	87.1	102.2	117.5	138.4	155.0	172.3	196.4
PS006	GEV	54.6	73.3	87.4	102.4	124.0	142.0	161.8	190.9
PS009	GUMBEL	77.2	104.7	122.9	140.4	163.0	180.0	196.8	219.1
PS014	GEV	52.7	72.0	84.8	97.1	113.0	124.9	136.7	152.4
PS018	GEV	52.0	68.0	82.8	101.4	133.9	166.6	208.5	282.8
PS025	GUMBEL	58.3	73.0	82.7	92.0	104.1	113.1	122.1	134.0
PE008	GUMBEL	93.0	126.1	148.1	169.1	196.4	216.8	237.2	264.0
PU007	GEV	41.8	57.3	69.3	82.2	101.3	117.6	135.7	162.9
PR011	GEV	65.1	92.5	113.6	136.4	170.1	198.9	230.9	279.1
RE001	GEV	56.7	74.2	86.4	98.6	115.3	128.5	142.1	161.0
RE005	GUMBEL	59.7	79.8	93.1	105.9	122.4	134.7	147.1	163.3
RE006	GEV	50.5	72.3	90.4	111.1	143.9	173.8	209.0	265.4
RE023	GLO	61.9	91.3	114.0	139.5	179.7	216.4	259.9	330.3
RE029	GEV	75.9	114.3	144.7	178.0	228.5	272.4	322.3	398.9
RE031	GEV	64.3	94.0	120.2	151.9	205.1	256.5	320.1	428.2
RI006	GNO	92.1	144.7	183.4	221.6	271.6	309.3	346.7	396.2
RI023	GEV	89.9	115.7	132.8	149.3	170.7	186.8	202.9	224.2
RI025	GEV	69.1	92.0	109.4	127.9	154.9	177.7	202.8	240.0
SG001	GEV	50.7	67.2	78.8	90.5	106.5	119.2	132.4	150.8
SG006	GUMBEL	47.2	61.8	71.5	80.8	92.8	101.8	110.8	122.7
SE007	GUMBEL	42.9	55.6	64.0	72.0	82.4	90.2	98.0	108.2
SE020	GUMBEL	41.2	54.4	63.2	71.6	82.5	90.6	98.7	109.4
SE031	GUMBEL	47.3	61.4	70.6	79.6	91.1	99.8	108.4	119.8
SE039	GUMBEL	49.7	67.2	78.8	89.9	104.2	115.0	125.7	139.9
SE044	GUMBEL	40.2	54.5	64.0	73.1	84.8	93.6	102.4	113.9
SV002	GNO	72.4	101.4	126.0	154.8	201.2	244.4	296.3	381.2
SV018	GEV	75.3	105.4	128.7	154.0	191.8	224.1	260.4	315.3
SV035	GEV	86.0	119.5	141.6	162.9	190.5	211.2	231.8	259.1
SV048	GNO	65.7	95.4	121.1	150.3	195.1	234.3	278.5	345.4
SO005	GEV	55.4	73.9	86.7	99.3	116.2	129.3	142.7	161.0
SO011	GEV	65.4	85.9	99.6	112.8	130.1	143.1	156.2	173.5
TG001	GUMBEL	69.3	91.9	106.8	121.1	139.6	153.5	167.3	185.6
TA011	GUMBEL	63.0	83.7	97.4	110.6	127.6	140.4	153.1	169.9
TA013	GUMBEL	49.4	63.4	72.7	81.6	93.2	101.8	110.4	121.8
UG004	GEV	45.9	61.8	73.0	84.1	99.2	111.1	123.4	140.3
UG013	GEV	41.5	56.1	65.9	75.3	87.7	97.2	106.6	119.3
UG017	GEV	41.4	53.2	62.2	72.1	86.6	99.1	113.0	134.0
UG019	GEV	52.3	69.5	81.8	94.3	111.7	125.6	140.3	161.1
UG020	GEV	46.2	59.7	69.5	79.6	93.7	105.0	117.1	134.4
VA010	GEV	56.3	70.3	80.6	91.3	106.4	118.8	132.1	151.2
VC010	GEV	75.9	105.4	128.6	154.0	192.4	225.8	263.6	321.5
VC017	GUMBEL	69.4	97.1	115.4	132.9	155.7	172.7	189.7	212.0
VC022	GEV	68.4	93.1	112.8	134.7	168.0	197.2	230.7	282.4
VR030	GEV	76.0	100.5	117.2	133.6	155.3	172.0	188.9	211.8
VR040	GEV	67.6	96.2	117.1	138.8	169.5	194.5	221.3	259.8
VR051	GEV	72.1	99.0	117.2	135.0	158.4	176.2	194.2	218.4
VR057	GEV	65.5	84.9	99.2	114.3	135.8	153.5	172.7	200.6
VR063	GUMBEL	114.0	157.2	185.9	213.4	249.0	275.6	302.2	337.3
VR078	GEV	68.3	92.5	111.7	132.9	165.1	193.3	225.4	274.9

Table 6.2: maximum expected intensity at 24 hours for several return periods at the 163 selected locations (more information on the location and used series, like spatial coordinates or number of years, can be found in Table[3.3]).

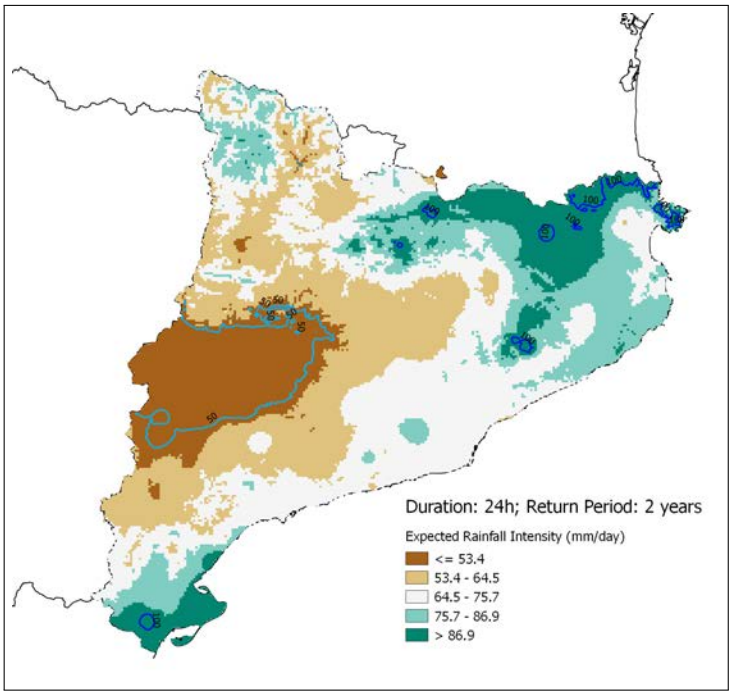


Figure 6.5: expected maximum rainfall in 24 hours with a return period of 2 years.

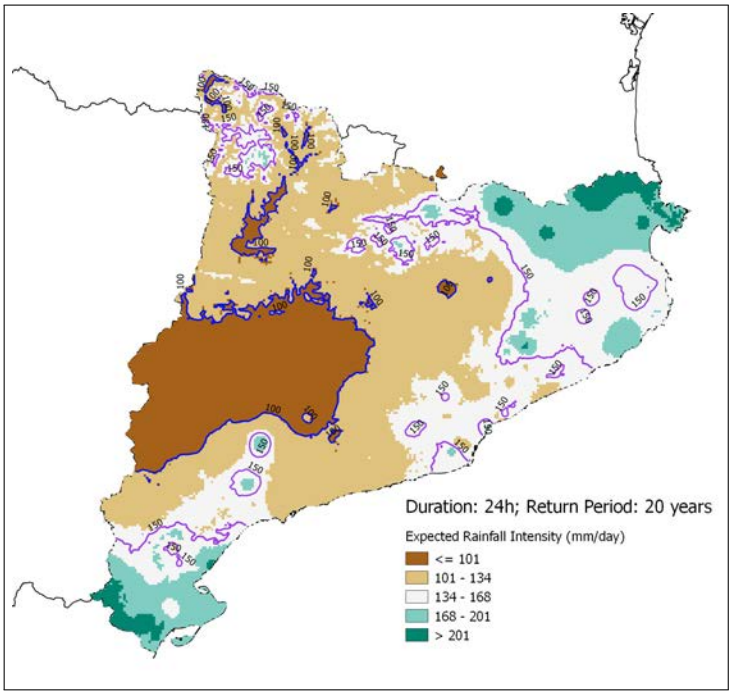


Figure 6.6: expected maximum rainfall in 24 hours with a return period of 20 years.

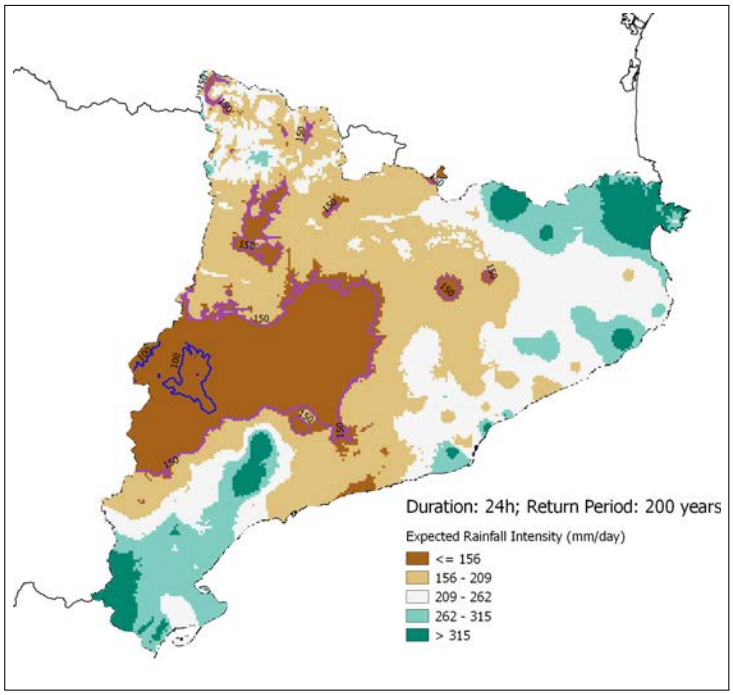


Figure 6.7: expected maximum rainfall in 24 hours with a return period of 200 years.

6.3 Comparison with other author's results

In this section, expected maximum rainfall in 24 hours obtained in the present project is compared to the results of previous works that had been accomplished in Catalonia, namely, the works done by Casas et al. (2007) and also presented in Casas et al. (2005), which are the intensity - frequency relationships for 24 hours currently used by the SMC.

Fig[6.8], Fig[6.10] and Fig[6.12] show the results published by Casas et al. (2005) that display the maximum expected intensity for 24 hours rainfall at several return periods. Fig[6.9], Fig[6.11] and Fig[6.13] show the spatial differences in maximum expected rainfall calculated as the results provided by Casas et al. (2005) and subtracting the results obtained in the present project.

Results obtained in both cases are quite similar, although they have been obtained with different databases and by different methodologies. The weighted interpolation including altitude in the results of the present project gives a better spatial resolution with distinct values between the peaks and the valleys in the western Pyrenees; other important mountain ranges like Montsec (between Pallars Jussà and Noguera counties) can also be spotted to have slightly higher expected intensities than their surroundings at lower altitudes.

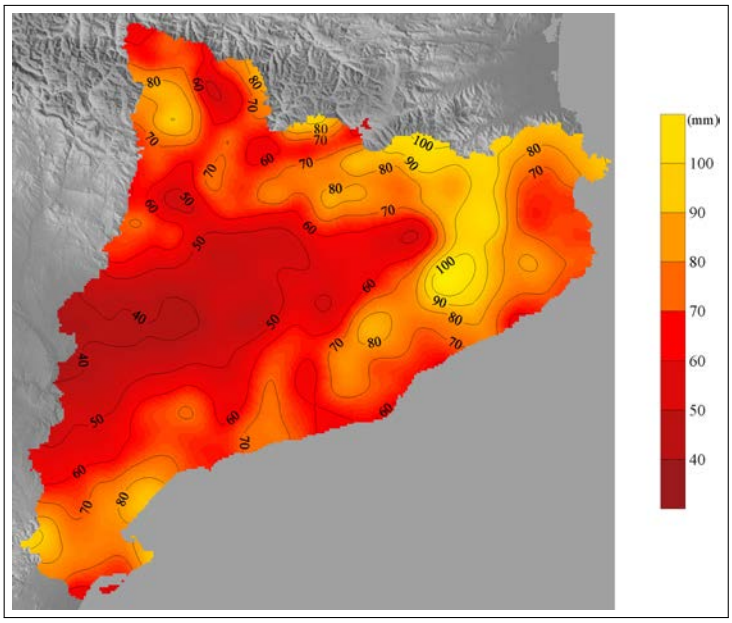


Figure 6.8: expected maximum rainfall in 24 hours with a return period of 2 years. Source: Casas et al. (2005).

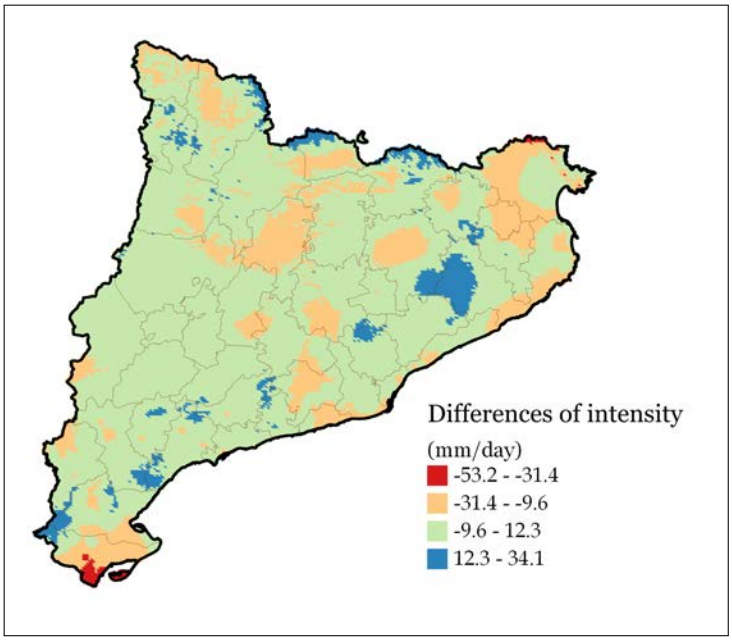


Figure 6.9: differences in maximum expected rainfall in 1 day and return period of 2 years: results of the present project subtracted from results provided by Casas et al. (2005).

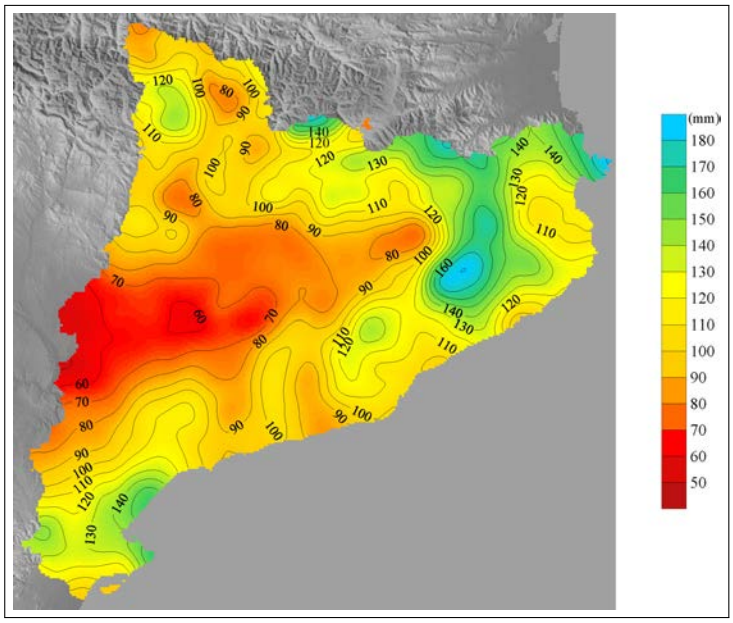


Figure 6.10: expected maximum rainfall in 24 hours with a return period of 10 years. Source: Casas et al. (2005).

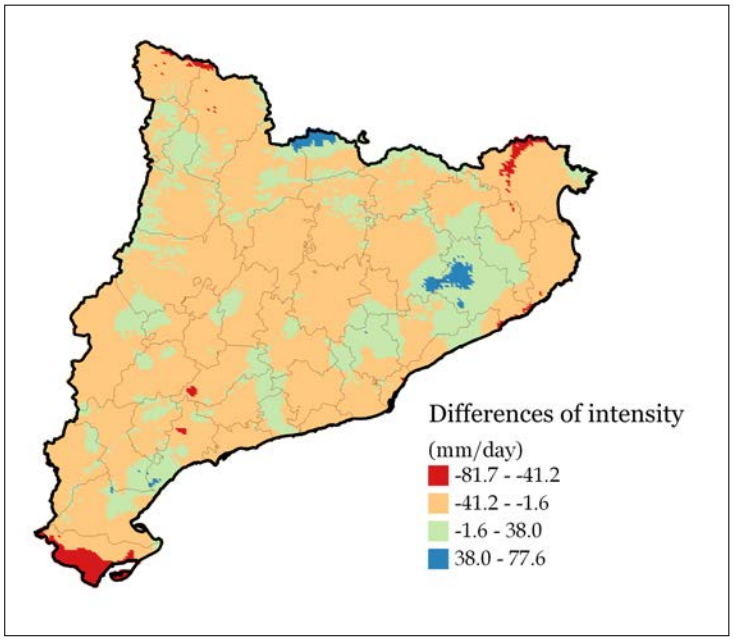


Figure 6.11: differences in maximum expected rainfall in 1 day and return period of 10 years: results of the present project subtracted from results provided by Casas et al. (2005).

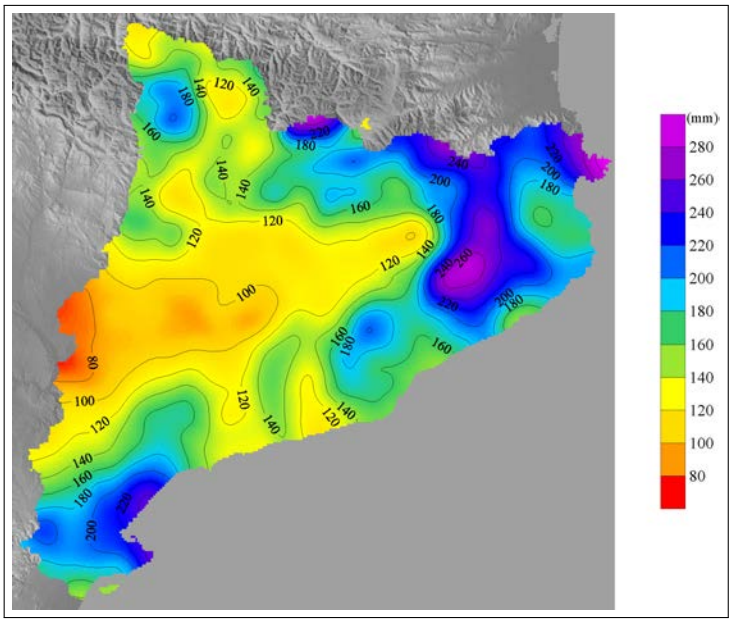


Figure 6.12: expected maximum rainfall in 24 hours with a return period of 100 years. Source: Casas et al. (2005).

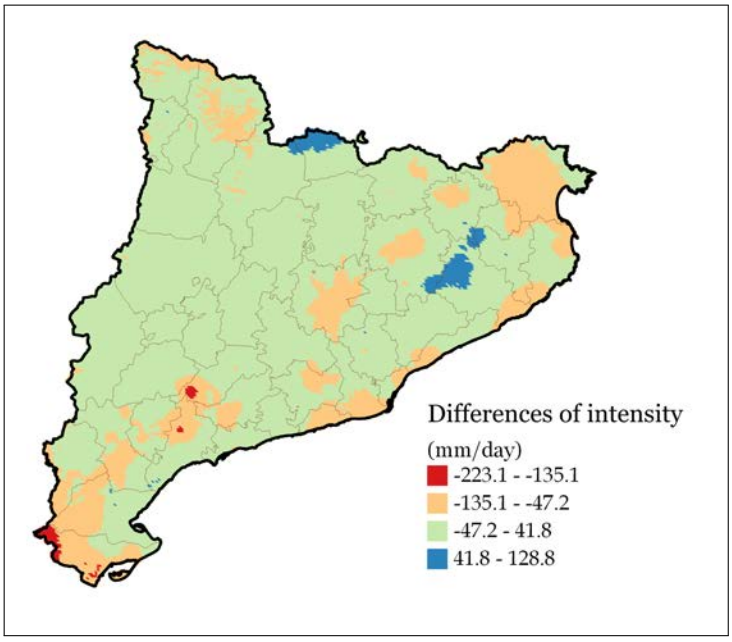


Figure 6.13: differences in maximum expected rainfall in 1 day and return period of 100 years: results of the present project subtracted from results provided by Casas et al. (2005).

Difference in expected intensity (mm/day)	Return period: 2 years	Return period: 10 years	Return period: 100 years
[-240, -220)			0.01%
[-220, -200)			0.02%
[-200, -180)			0.03%
[-180, -160)			0.10%
[-160, -140)			0.19%
[-140, -120)			0.42%
[-120, -100)			1.28%
[-100, -80)			2.79%
[-80, -60)		0.5%	6.81%
[-60, -40)	0.17%	1.56%	14.62%
[-40, -20)	2.13%	15.90%	32.08%
[-20, 0)	67.30%	61.28%	28.53%
[0, 20)	29.22%	16.60%	9.10%
[20, 40)	1.18%	3.33%	2.43%
[40, 60)		0.65%	1.16%
[60, 80)		0.08%	0.22%
[80, 100)			0.14%
[100, 120)			0.07%
[120, 140)			0.02%

Table 6.2: percentage of the territory affected by differences of a certain magnitude (in mm/day) in expected maximum intensity predicted by Casas et al. (2007) and subtracting the prediction of the present project for a duration of 24 hours at several return periods.

As it can be seen both in the maps of Fig[6.9], Fig[6.11] and Fig[6.13], and in Table[6.2], spatial differences in maximum expected rainfall are usually low, with over 90% of the territory having differences under 20 mm/day for return periods of 2 years, the same maximum differences at 78% of the territory for return periods of 10 years and differences up to 40 mm/day over 70% of the territory for return periods of 100 years. The results of the present project produce overall slightly higher expected intensities and the spatial regions with most significant differences are the spots of extremest values, namely, the north of Alt Empordà and the south and west of Montsià that present local higher expected intensities according to the results of the present project; the maxima in northeastern Montseny, roughly west of La Selva county, is a region where the values have been underestimated in the present project, although it produced the absolute expected maxima for low return periods.

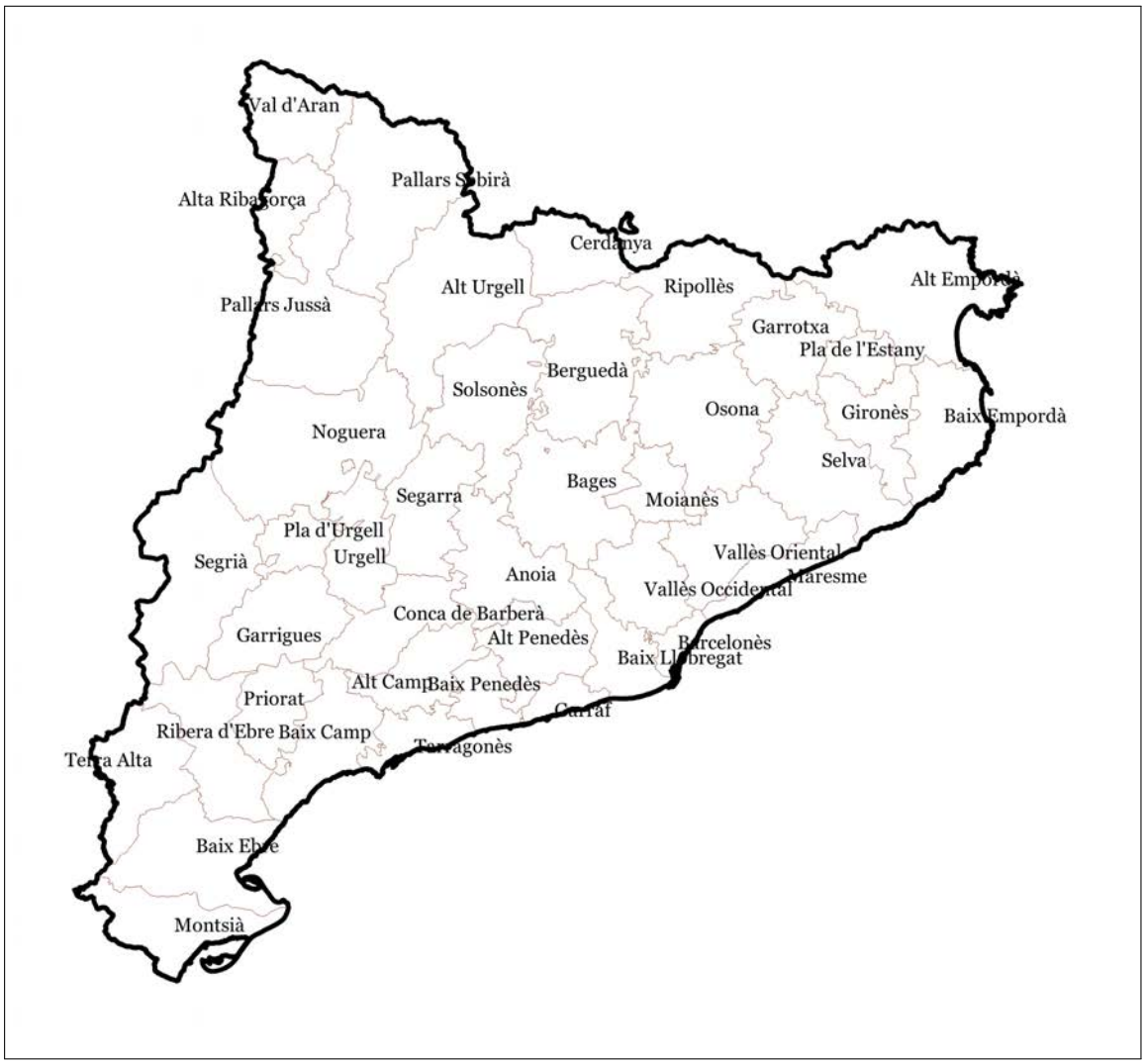


Figure 6.14: names of the 42 counties in Catalonia for reference of the regions mentioned in the discussion.

Chapter 7

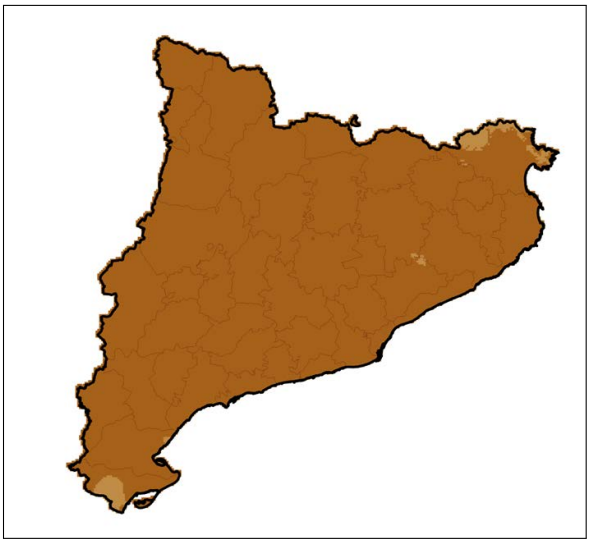
IDF RELATIONSHIPS

The obtainment of IDF relationships comes directly from applying the downscaling distribution equality (Eq[5.1]) to the high spatial resolution grids of maximum expected intensity in 24 hours, using the scaling exponent β obtained and discussed in Chapter 5 (see section 5.6). In the present chapter, a collection of maps that display the results of the project are presented in the first section. The second section of this chapter consists in a validation of the results.

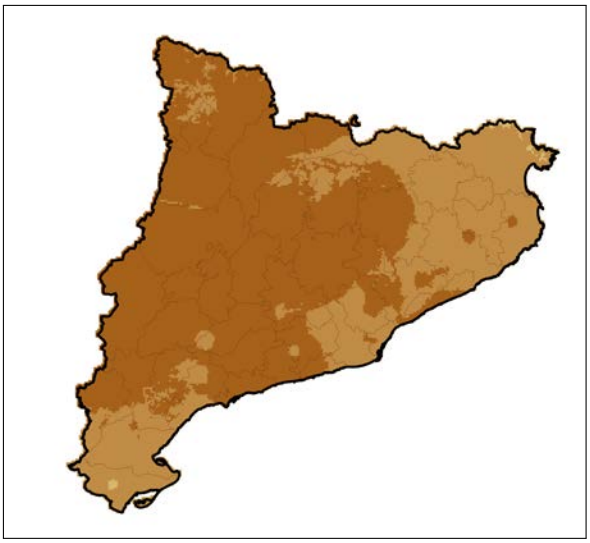
7.1 Spatial distribution of expected rainfall

Fig[7.1] shows the spatial distribution of maximum expected intensity in 1 hour for eight different return periods. As it can also be seen for a different duration (Fig[7.2], Fig[7.3], Fig[7.4] and Fig[7.5]), maximum expected intensity presents low spatial variability at low return periods; at high return periods there are regions where expected intensity rises considerably while in other regions (mainly the western area) expected maximum intensity is low.

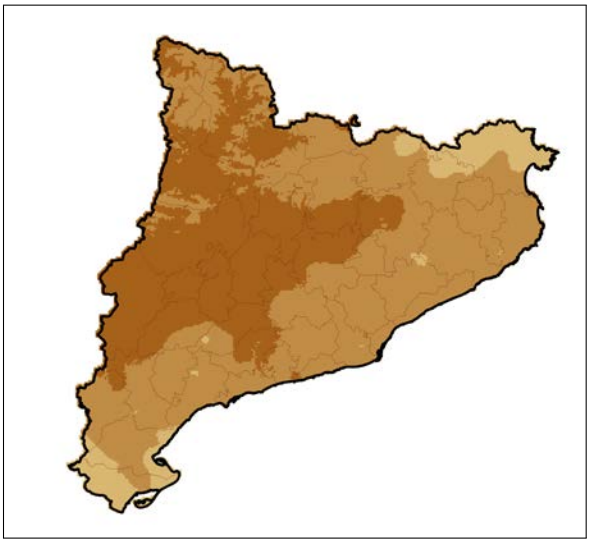
The SMC has defined weather warnings for risk of precipitation episodes both for intensities and accumulation. The threshold to issue a warning of precipitation intensity is at 20 mm/30 min (low level) and 40 mm/30 min (high level). The intensity warning defined in this way cannot be compared to the presented results of this project as the temporal downscaling applied is not necessarily valid for a subhourly duration; however, a threshold of 40 mm/1 h has been analysed. The threshold to issue a warning of precipitation accumulation is 100 mm/24 h (low level) and 200 mm/24 h (high level). Table[7.1] summaries the percentage of the territory where the mentioned thresholds are expected to be exceeded depending on the return period.



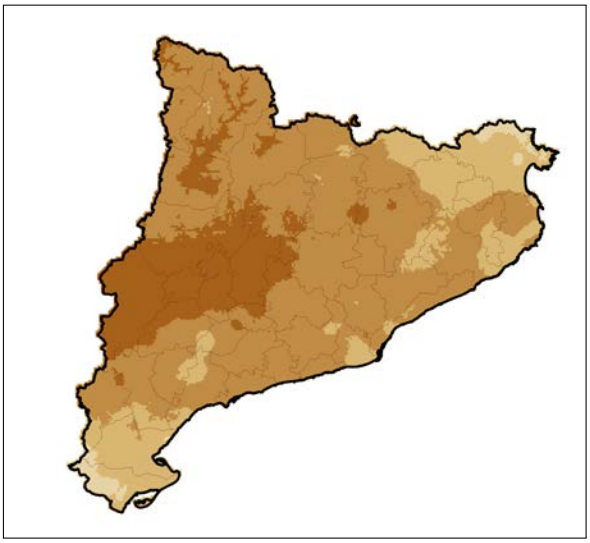
(a) return period: 2 years.



(b) return period: 5 years.

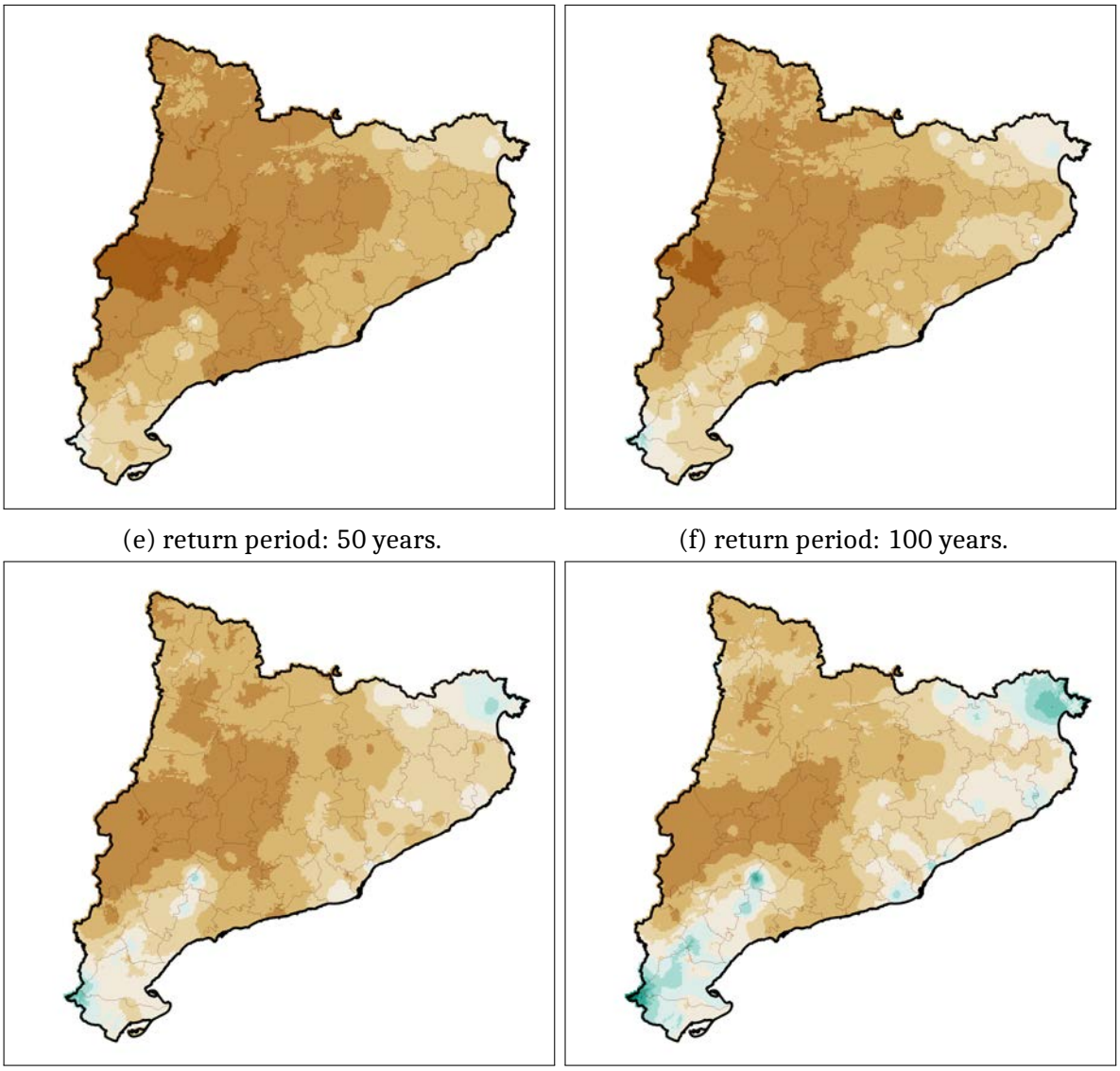


(c) return period: 10 years.



(d) return period: 20 years.

(mm per 1 h)
- 22 - 51
- 51 - 80
- 80 - 109
- 109 - 139
- 139 - 168
- 168 - 197
- 197 - 226
- 226 - 256
- 256 - 285
- 285 - 314



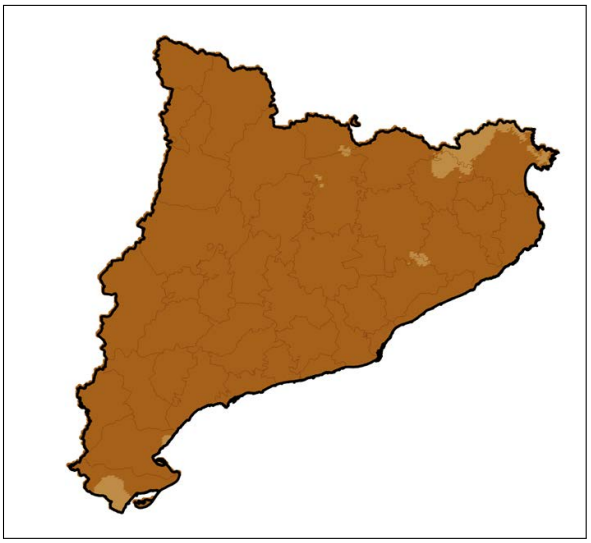
(e) return period: 50 years.

(f) return period: 100 years.

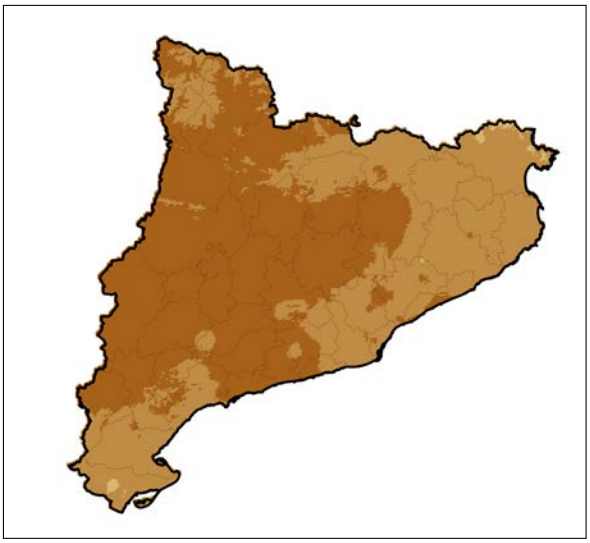
(g) return period: 200 years.

(h) return period: 500 years.

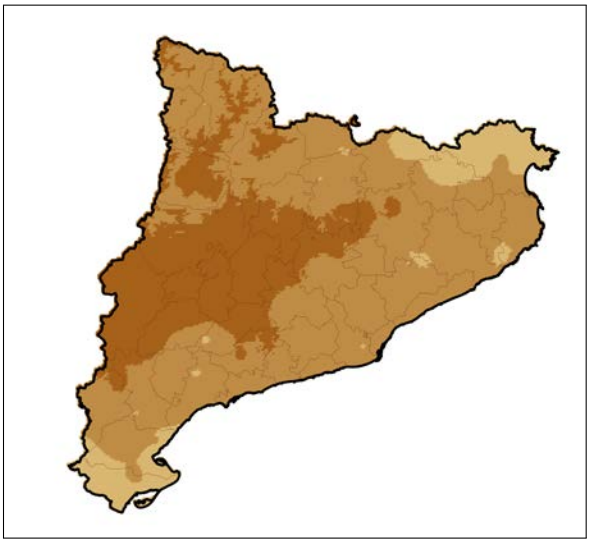
Figure 7.1: maximum expected rainfall for 1h. Results of the project after temporal monofractal scaling at high spatial resolution (grid of 1 km x 1 km) from spatial interpolation at 163 daily precipitation measuring sites.



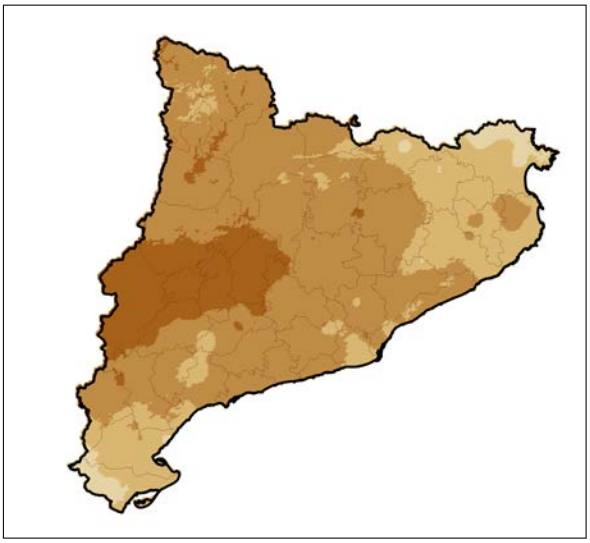
(a) return period: 2 years.



(b) return period: 5 years.

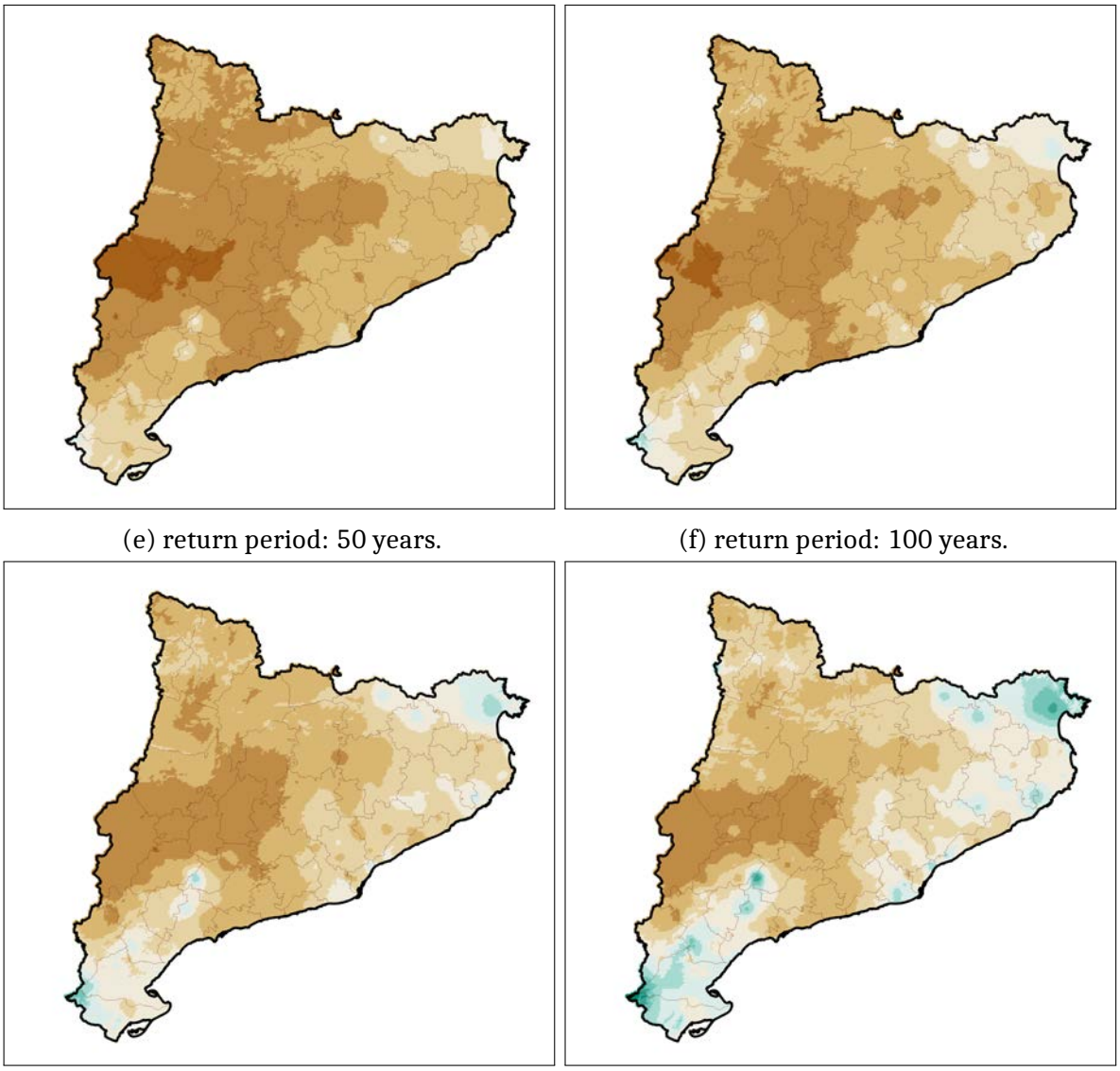


(c) return period: 10 years.



(d) return period: 20 years.

(mm per 6 h)
- 31 - 71
- 71 - 112
- 112 - 153
- 153 - 194
- 194 - 234
- 234 - 275
- 275 - 316
- 316 - 357
- 357 - 398
- 398 - 438



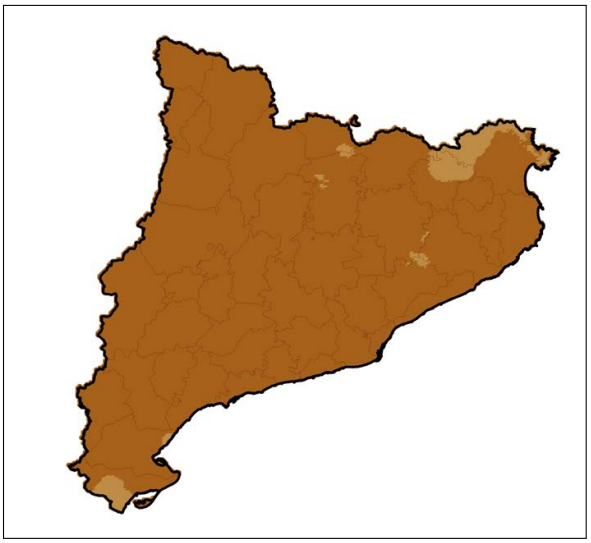
(e) return period: 50 years.

(f) return period: 100 years.

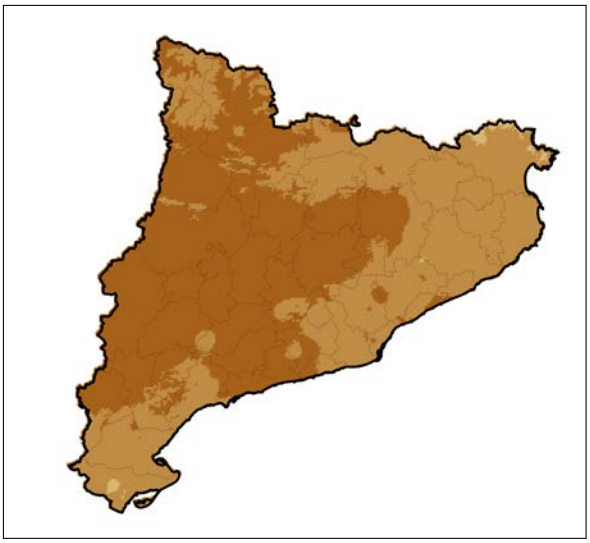
(g) return period: 200 years.

(h) return period: 500 years.

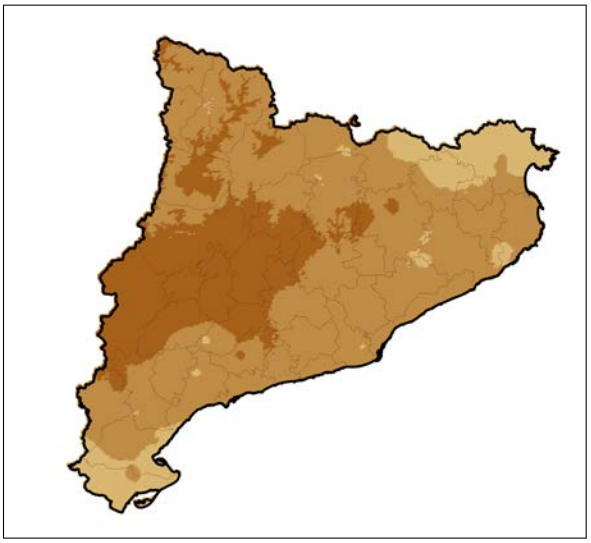
Figure 7.2: maximum expected rainfall for 6h. Results of the project after temporal monofractal scaling at high spatial resolution (grid of 1 km x 1 km) from spatial interpolation at 163 daily precipitation measuring sites.



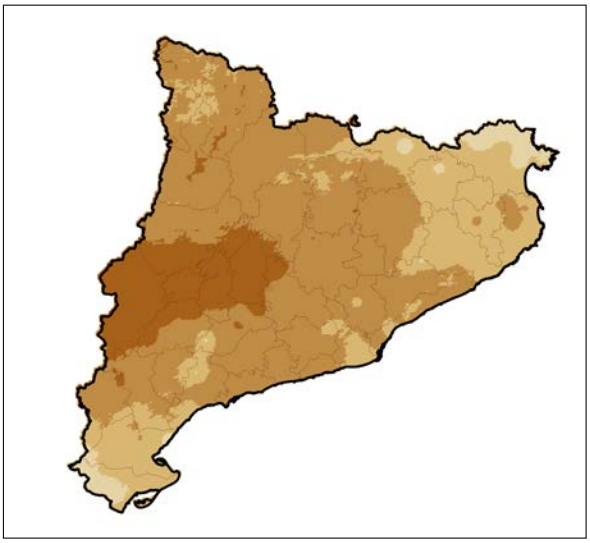
(a) return period: 2 years.



(b) return period: 5 years.

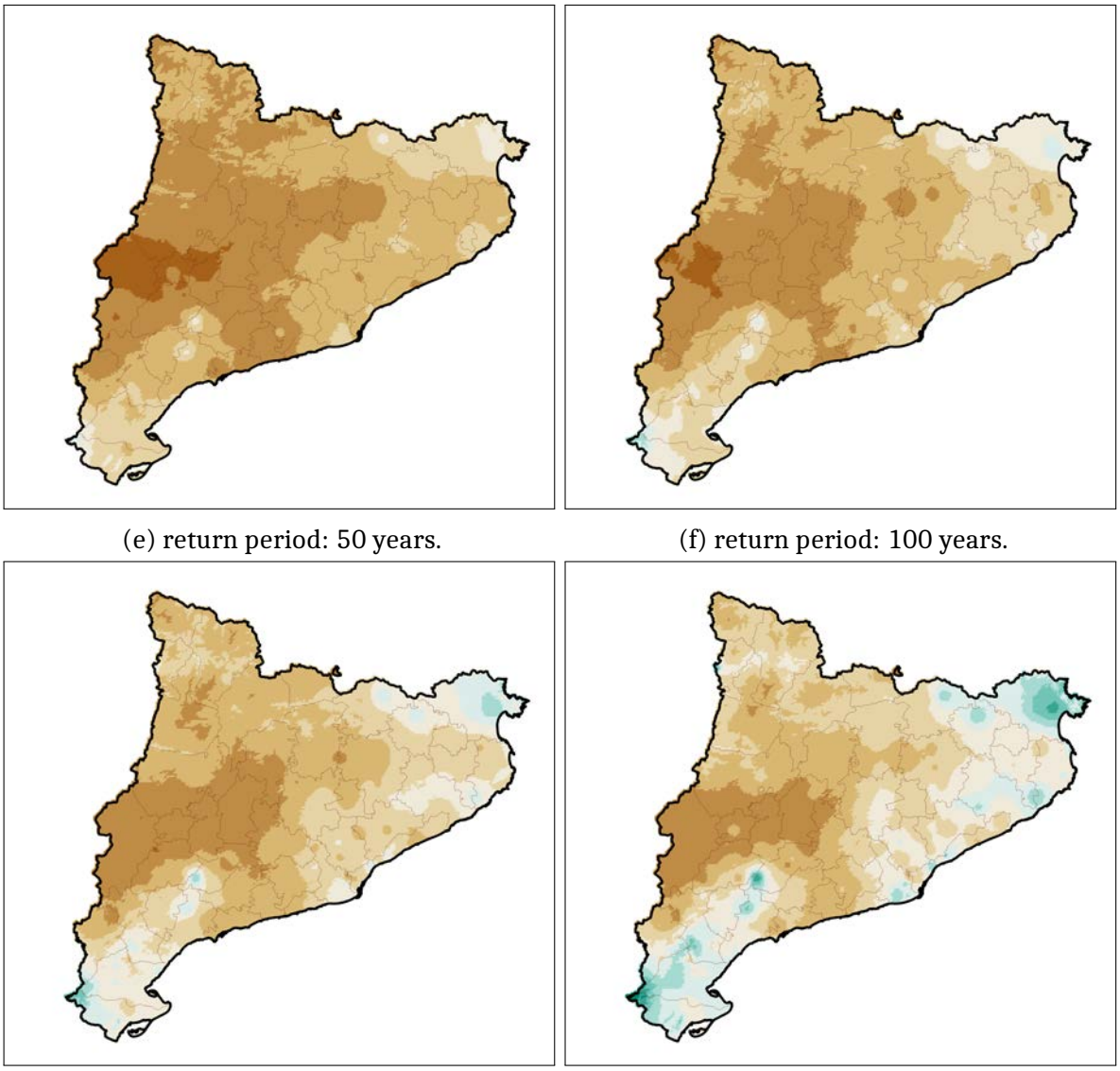


(c) return period: 10 years.



(d) return period: 20 years.

(mm per 12 h)
- 35 - 81
- 81 - 128
- 128 - 174
- 174 - 220
- 220 - 267
- 267 - 313
- 313 - 360
- 360 - 406
- 406 - 452
- 452 - 499



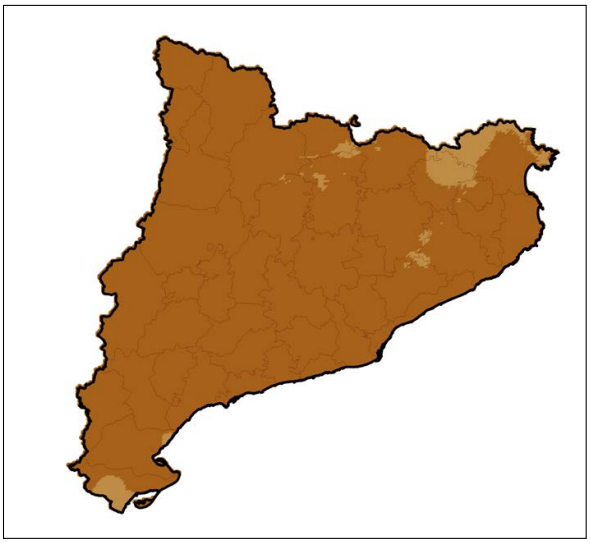
(e) return period: 50 years.

(f) return period: 100 years.

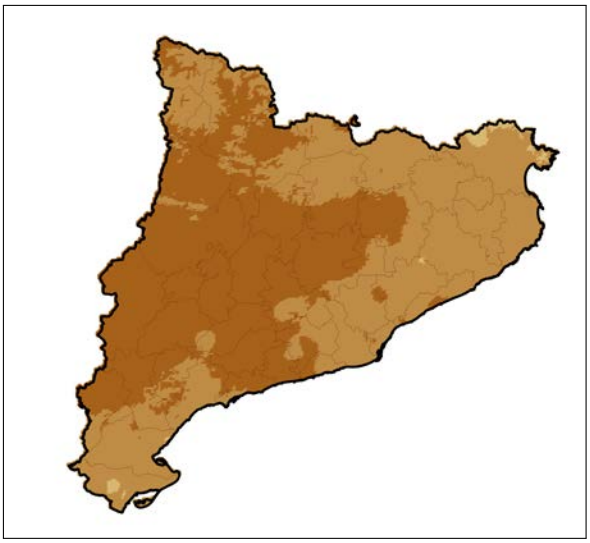
(g) return period: 200 years.

(h) return period: 500 years.

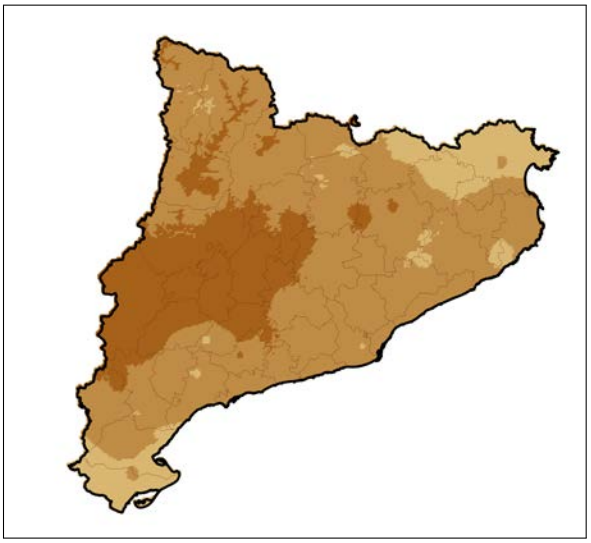
Figure 7.3: maximum expected rainfall for 12h. Results of the project after temporal monofractal downscaling at high spatial resolution (grid of 1 km x 1 km) from spatial interpolation at 163 daily precipitation measuring sites.



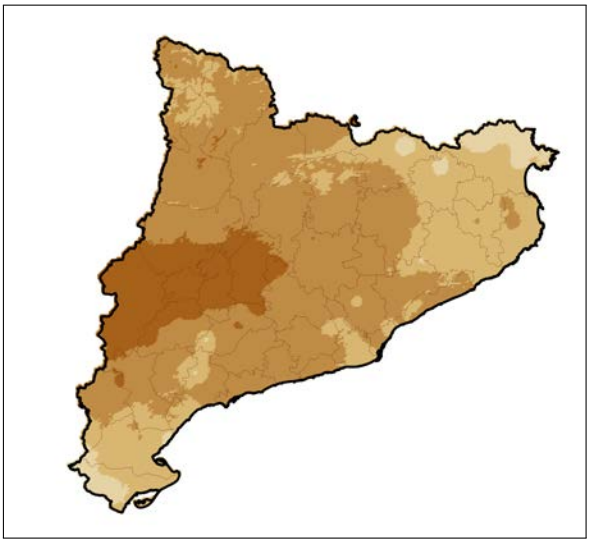
(a) return period: 2 years.



(b) return period: 5 years.

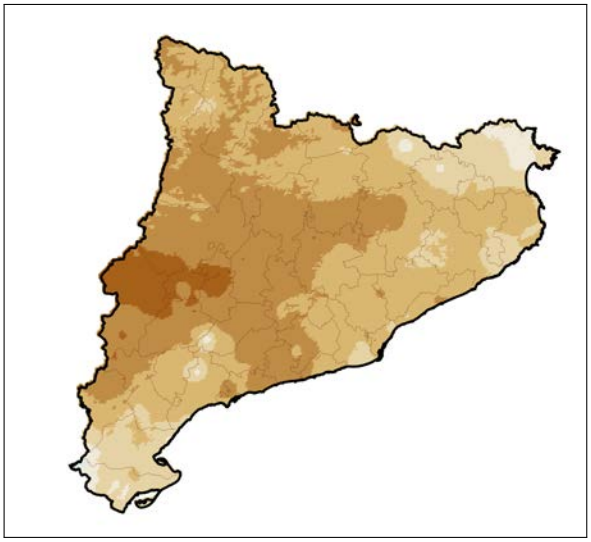


(c) return period: 10 years.

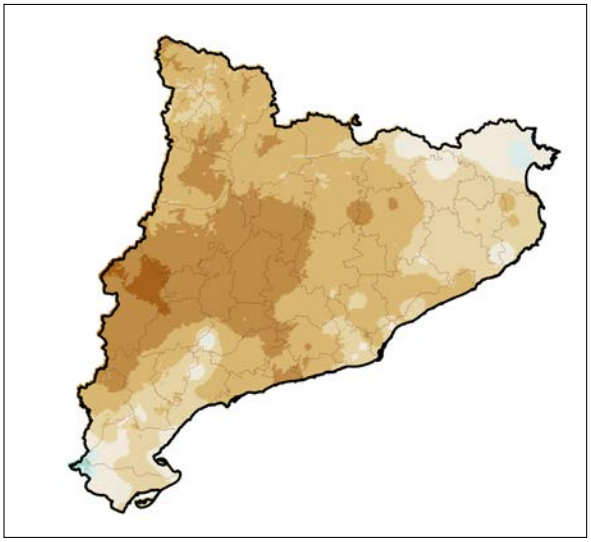


(d) return period: 20 years.

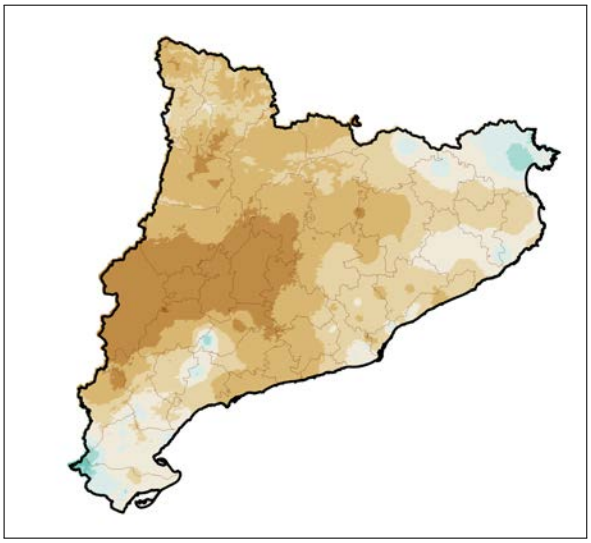
(mm per 24 h)
- 40 - 93
- 93 - 145
- 145 - 198
- 198 - 251
- 251 - 304
- 304 - 356
- 356 - 409
- 409 - 462
- 462 - 514
- 514 - 567



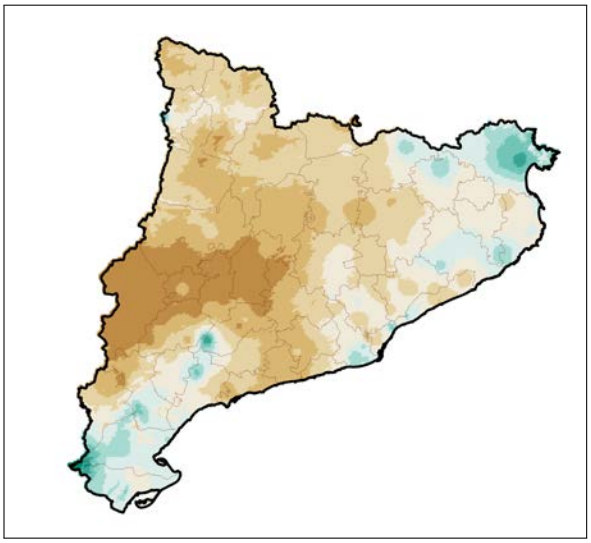
(e) return period: 50 years.



(f) return period: 100 years.

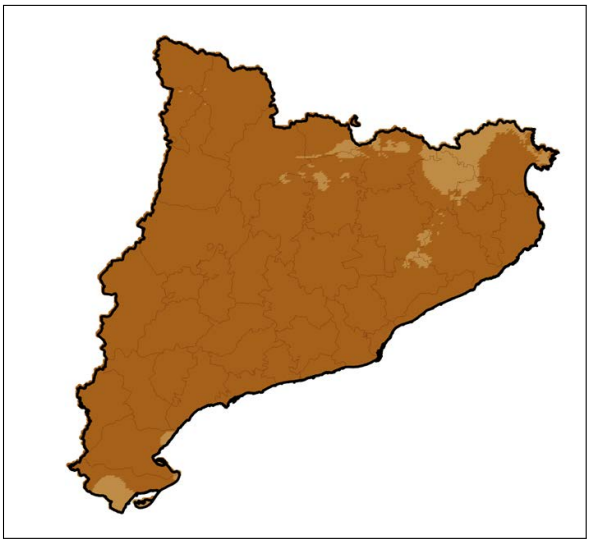


(g) return period: 200 years.

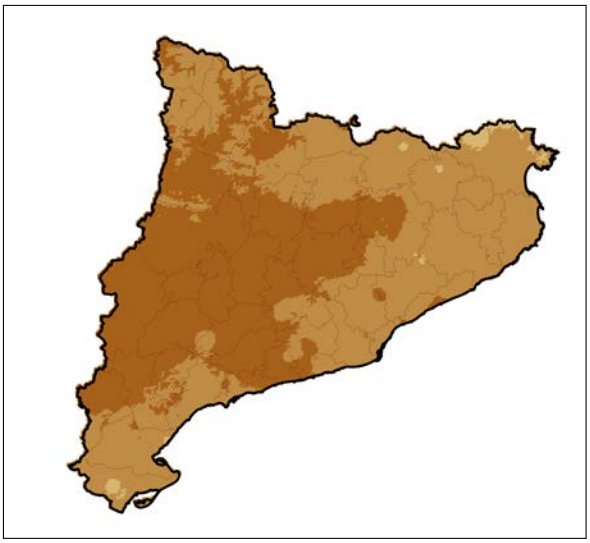


(h) return period: 500 years.

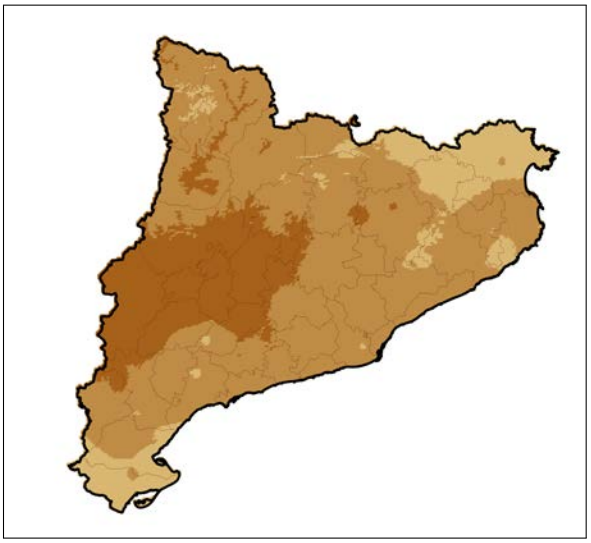
Figure 7.4: maximum expected rainfall for 24h. Results of the project at high spatial resolution (grid of 1 km x 1 km) from spatial interpolation at 163 daily precipitation measuring sites.



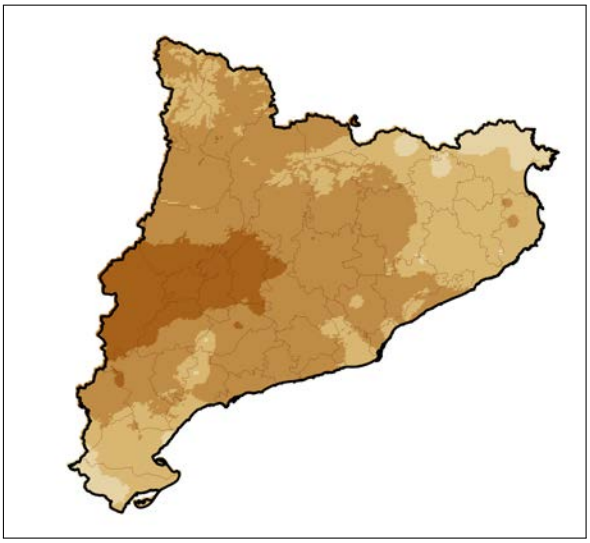
(a) return period: 2 years.



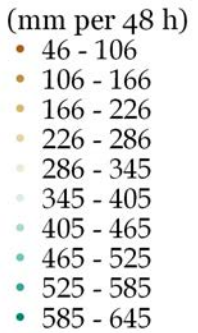
(b) return period: 5 years.

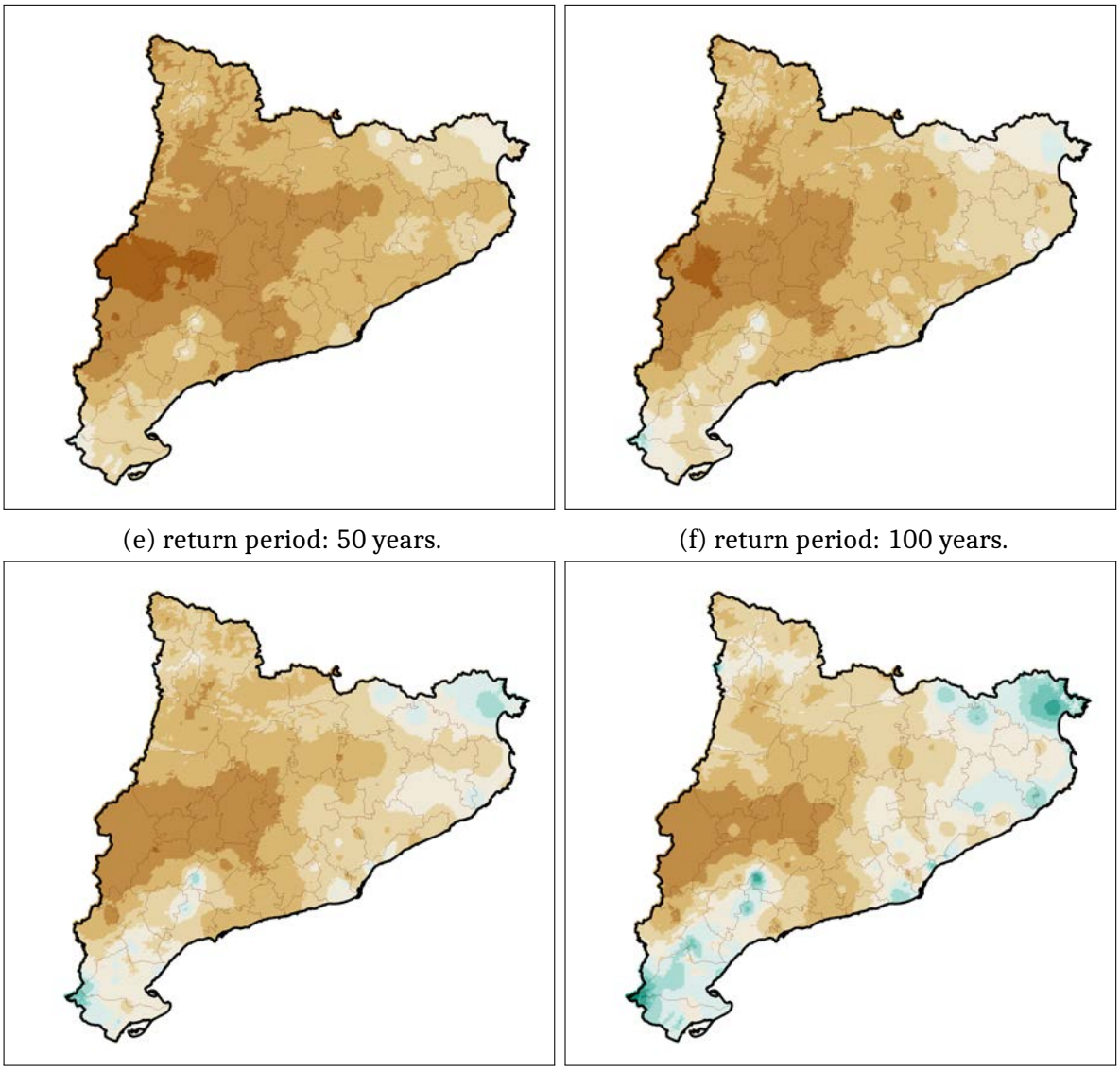


(c) return period: 10 years.



(d) return period: 20 years.





(e) return period: 50 years.

(f) return period: 100 years.

(g) return period: 200 years.

(h) return period: 500 years.

Figure 7.5: maximum expected rainfall for 48h. Results of the project after temporal monofractal scaling at high spatial resolution (grid of 1 km x 1 km) from spatial interpolation at 163 daily precipitation measuring sites.

Return period	Intensity threshold: 40 mm/h	Accumulation threshold: 100 mm/24h	Accumulation threshold: 200 mm/24h
2 years	19.1%	1.2%	0.0%
5 years	72.0%	36.0%	0.0%
10 years	90.5%	65.4%	0.0%
20 years	97.5%	81.8%	3.6%
50 years	100%	90.9%	15.4%
100 years	100%	96.2%	30.6%
200 years	100%	98.7%	47.3%
500 years	100%	100%	65.8%

Table 7.1: percentage of the territory where rainfall in 1 hour or in 24 hours is expected to exceed a threshold related to a weather warning.

In Table[7.1] the percentage of the territory in which rainfall is expected to exceed certain thresholds is calculated for each of the considered return periods.

The threshold of intensity (mm/h) is the threshold used by most home insurance companies and it is exceeded in most of the territory for medium return periods and it is exceeded in the whole territory for return periods of 50 years and higher.

The two thresholds of accumulation in 24 hours (low level risk, as defined by the SMC, of 100 mm and high level risk at 200 mm) are not as generally exceeded as the intensity threshold. Accumulations of 100 mm in 24 hours are expected to be exceeded in over 80% of the territory only for return periods of at least 20 years. Accumulations of 200 mm in 24 hours are only expected to be exceeded in around half the territory for long return periods (over 200 years).

7.2 Validation of subdaily downscaling

In order to validate the results obtained from the temporal downscaling, a comparison with results of a different methodology and database has been performed. In particular, maximum expected rainfall in 1 hour presented in the previous section (see Fig[7.1]) is compared to maximum expected rainfall in the same duration obtained from hourly data at a selection of 105 AWS. The selected 105 locations used to check the performance of the downscaling are series with a minimum of 15 years of complete hourly data, with a spatial density similar to the one of the 163 daily series and covering the whole study area (see Fig[7.6]). The results obtained from the 105 AWS series of hourly data, will be, form now on, referred to the *control results*.

Expected maximum intensity of rainfall in 1 hour at several return periods have been obtained following a similar procedure to the one followed in Chapter 6 for daily data. However, in the case of the control results, the obtainment of intensity - frequency relationships has not been so careful given that the objective is to gain an overall idea of the magnitude of the differences between the two methodologies (i.e., the control results will not be presented as actual results, nor used afterwards with any other intention than to validate the primary results of the project).

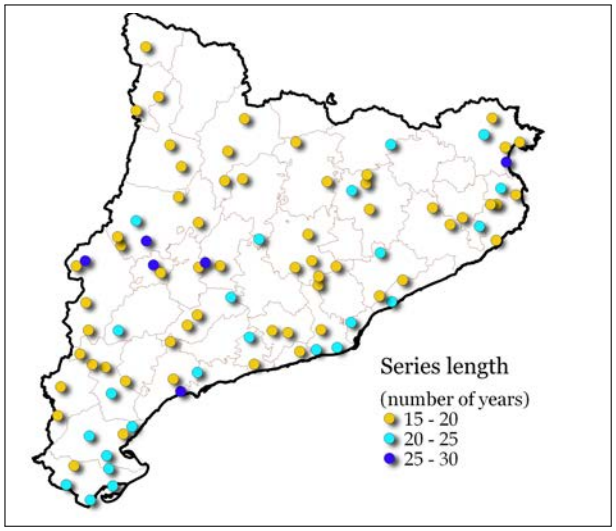


Figure 7.6: selection of 105 AWS that have complete hourly data for at least 15 years in the period 1990-2017.

The procedure followed to obtain the control results from 105 selected AWS (see Fig[7.6]) is to obtain the series of annual maximum rainfall in 1 hour, correct these series from fixed to sliding interval, adjust these series to a probability distribution function and obtain the expected maximum intensity predicted by the distribution. The steps of adjustment to a distribution has not been applied so carefully as in the primary results of the project, as series of maxima have been adjusted to either a GUMBEL or GEV distribution without considering other distributions, nor spatially checking the results before deciding the final adjustment. The multiplicative correction from fixed to sliding intervals has also been implemented in a general manner: all values have been corrected by a multiplicative factor of 1.129 (the global correction factor obtained for daily data in the whole study area) because the study of the correction factors has not been accomplished in the present project for subdaily data (hence, a better precision is not possible), but correcting factors for 1 hour and 1 day should be the same according to Weiss (1964).

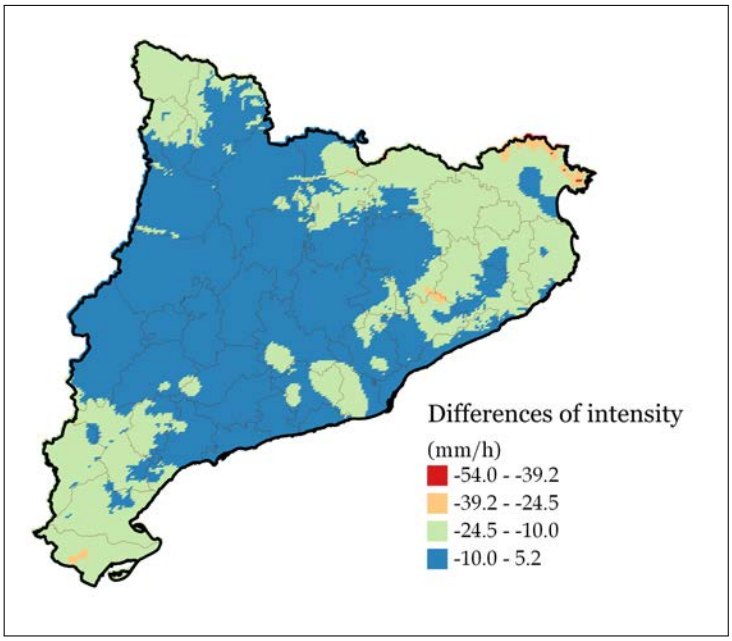


Figure 7.7: difference of expected maximum rainfall (in 1 hours and return period of 2 years) between the results obtained with AWS or downscaling daily data following the monofractal methodology.

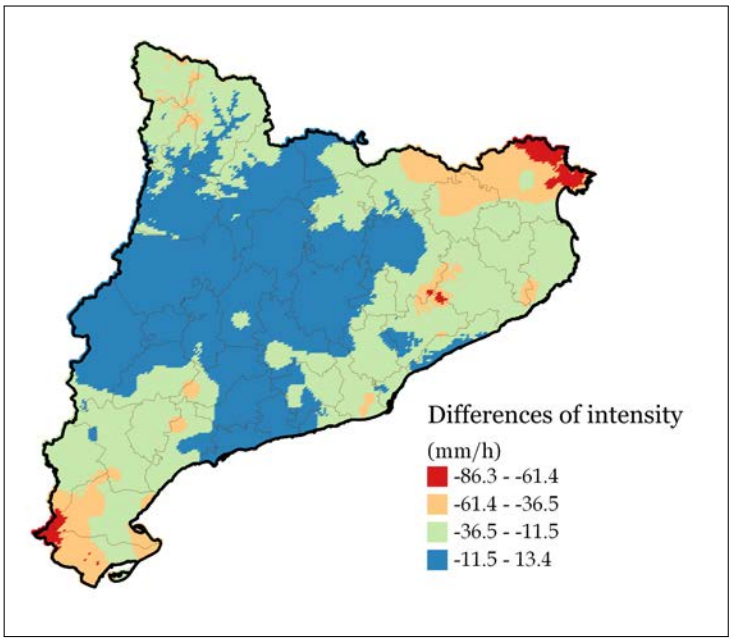


Figure 7.8: difference of expected maximum rainfall (in 1 hours and return period of 20 years) between the results obtained with AWS or downscaling daily data following the monofractal methodology.

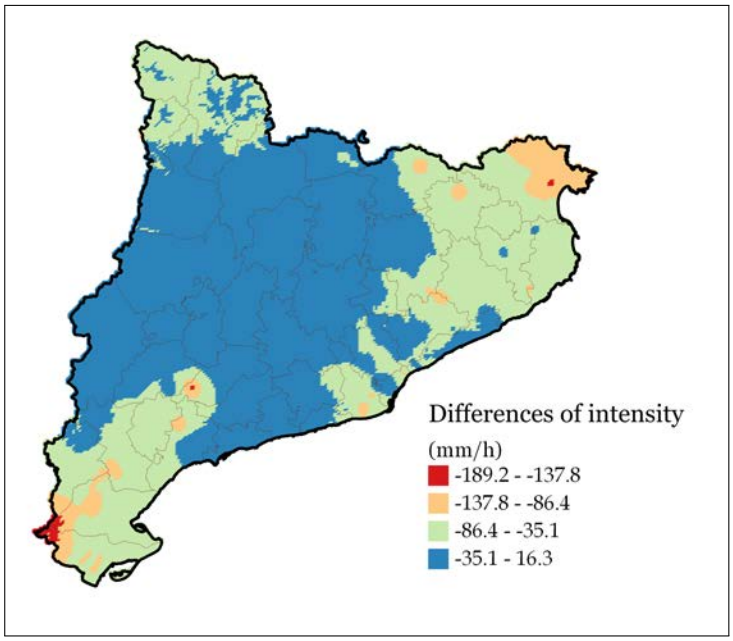


Figure 7.9: difference of expected maximum rainfall (in 1 hours and return period of 200 years) between the results obtained with AWS or downscaling daily data following the monofractal methodology.

Fig[7.7], Fig[7.8] and Fig[7.9] prove a good working of the monofractal temporal downscaling implemented in the present project. Indeed, differences are low in most of the territory (under 12 mm/h). The greatest differences in expected maximum intensity are found for long return periods (e.g., differences over 86 mm/h can be found at return periods of 200 years, see Fig[7.9]).

It is at both the northeastern and southwestern corners, where the highest intensities are expected, that the maximum differences are found, although in very localised areas. It should be noticed that, especially regarding the differences at he southwestern corner, the results of the present project are more reliable as weather stations outside the borders of Catalonia were also considered. Maximum expected intensity in Montseny is higher in the primary results of the project and the observed differences are considered to be caused by an underestimation of the prediction in that region given the short length of the control samples; especially considering that in the comparison with the results of Casas et al. (2007) for 24 hours, the primary results of the project underestimated intensity instead of overestimating it.

Table[7.2] shows that most differences between the control results and the primary results are negative, meaning that the primary results predict higher hourly maxima. Also, the vast majority of the territory, over 95% of the spatial area, presents maximum differences of 20 mm/h for return periods of 2 years (when the control results are most reliable due to short available series). At higher return periods, differences with the control results grow extremely big (up to 180 mm/h) but only in localised areas, while

over 90% of the territory presents differences under 80 mm/h.

Difference in expected intensity (mm/h)	Return period: 2 years	Return period: 20 years	Return period: 200 years
[-180, -160)			0.02%
[-180, -160)			0.11%
[-160, -140)			0.15%
[-140, -120)			0.93%
[-120, -100)			1.42%
[-100, -80)		0.07%	4.50%
[-80, -60)		0.78%	9.52%
[-60, -40)	0.05%	6.67%	20.36%
[-40, -20)	3.45%	31.47%	20.01%
[-20, 0)	94.25%	51.62%	36.95%
[0, 20)	2.25%	8.39%	6.03%

Table 7.2: percentage of the territory affected by differences of a certain magnitude (in mm/h) in expected maximum intensity predicted by both methodologies for a duration of 1 hour at several return periods. Differences are obtained by subtracting temporal downscaled results from control results, i.e., negative differences are found when control values are lower than primary values of the project.

Automatic weather stations will be more reliable to construct IDF relationships in the future, when these stations will be able to provide longer time series. In the meantime, the monofractal temporal downscaling is a good methodology to estimate subdaily maxima from daily rainfall data.

Conclusions

The first part of the project was dedicated to acquire a knowledge of available historic daily rainfall data in Catalonia and select the series that are used in the rest of the project. Daily rainfall data was quality controlled and the homogeneity of the series was checked.

Instead of correcting daily data, which is controversial in the study of extremes and, in any case, it could compromise the final results if the correction were not adequate, the policy has been to select the best data to be used in the project. The quality control was implemented to approach the selection of quality series, starting by the definition of a quality index that is calculated for each year of each series using only daily rainfall data in that year. Afterwards, a relative quality control was designed where each daily value of each rainfall series ($14,144,395$ values in total) were compared to spatially nearby measures of the same day and labelled according to their reliability. The implementation of this designed quality control is satisfactory, although a 25.7% of the days could not be checked for lack of an adequate reference; a 74% of rainfall daily values in the database were considered reliable ("valid" label).

The quality control carried out in the scope of the present project is useful for the Meteorological Service of Catalonia, as it has now a tool to clear historical data from major errors, as well as to select the best quality rainfall data for specific purposes. The work done in the design and implementation of the quality control methodology of the present project led to a published paper: Llabrés-Brustenga et al. (2019).

The selection of data to be used in the rest of the project is also carried out in this first part of the project, in Chapter 3. The selection took into account the results of the quality checks: the quality indices and the results of the relative spatial comparison of rainfall measures. But also a homogeneity check was performed. As it had been the policy during the quality control, after the homogeneity checks the values were not corrected but selected or disregarded. The homogeneity correction is extremely important in the study of climate, but it could be problematic in the study of extreme values; hence, only fairly homogeneous series have been used and inhomogeneous series were disregarded.

The finally selected dataset consists in 163 series uniformly spread over the territory and with daily data in the temporal period 1942-2016. Some years of data in selected series have been disregarded in case they did not comply with the proposed filters, however, all selected series have a minimum of 25 years with complete and high quality daily precipitation data.

The second part of the project (Chapters 4, 5 and 6) consisted in preparing the obtainment of intensity-duration-frequency relationships. Firstly, in Chapter 4, the series of annual maximum rainfall in 24 hours for 163 locations were obtained after a correction of the maximum values in one day (fixed time interval) with a multiplicative factor. The choice of the multiplicative factor used in the project is supported by the analysis of the empirical correction factors in Catalonia obtained as the averaged ratios of rainfall amounts measured using fixed time intervals and rainfall amounts measured using unrestricted (sliding) intervals of the same duration using a selection of 120 AWS. The mean value for the whole study area results in a multiplicative factor of 1.129 ± 0.002 for intervals of one day. However, the empirical value presents spatial differences that were related to climatic features in this study. The empirical correction factor was analysed taking into account seasonality and geographical location. In general, the correction factor was found to be lower in the driest areas of Catalonia, but a high spatial variability was observed, which depends clearly on the season. Rainfall maxima recorded in summer in Catalonia are often produced by local and mesoscale formations with remarkable diurnal cycles for which the diurnal heating surface effect is decisive in their convective development. In concordance with that, summer is the season when lower correction factors are needed due to very low fraction of the daily rainfall registered at fixed interval starting time (8 a.m.). Series of annual maximum rainfall in one day were corrected at each location using the empirical correction factor spatially interpolated at that particular location (from the 120 empirical results) and depending on the season of the particular maximum at each year. The same procedure was implemented to correct series of annual maximum rainfall at longer duration, up to 15 days, which were needed to calculate the scaling exponent of Chapter 5.

In Chapter 5, the implemented fractal theory was presented. The monofractal downscaling methodology can be applied as long as in the calculation of the scaling exponent, the logarithm of statistical moments of annual maximum rainfall series and the logarithm of order q of the moments have a linear relationship, hence, there is a single scaling exponent. The monofractal scaling theory was implemented and the value of the scaling exponent was obtained and analysed; this part of the project resulted in a published paper: Casas-Castillo et al. (2018a).

In most of Catalonia, empirical values of the scaling exponent β range between -0.82 (percentile 0.20) and -0.76 (percentile 0.80), with a mean of -0.79. Despite of the high variability of this parameter, which seems to depend on the longitude of the sample and the presence of particular episodes in it (the specific chosen time period), a spatial analysis reveals a certain configuration which can be related to some specific climatic rainfall characteristics. Analysing the spatial distribution of this parameter, two distinct zones were detected: a northern area with a mean value of -0.75, matching a mountainous area with some Atlantic influence at its most northwestern corner, and a southwestern area with a mean value of -0.81, in great concordance with the driest cares of Catalonia. Apart from these distinct zones, at the eastern zone, low β values are distributed over areas of clear Mediterranean influence where convective rainfall occurs often. This results are in agreement with those obtained by Rodríguez-Solà et al. (2017), that found a gen-

eral concordance between the spatial distribution of β over the Iberian Peninsula and the mean annual precipitation distribution, with high values in rainy areas and low for the dry ones, with some discrepancies related to the kind of precipitation contributing to high rainfall events and the proportion of convective rainfall in total.

The scaling exponent analysed in Chapter 4 was spatially interpolated and its corresponding value was obtained in a grid of 1 km x 1 km that can be used in the temporal downscaling of expected intensity obtained for 24 hours in the nodes of the grid. In this way high spatial and temporal resolution is achieved. The grid of spatially interpolated values maintains the characteristics of the results obtained at the 163 individual locations: the mean value of β in the grid nodes is -0.79 and most of the territory presents values between -0.80 (percentile 0.20) and -0.77 (percentile 0.80).

In Chapter 6, intensity - frequency relationships were found for a duration of 24 hours and several return periods. The results were spatially interpolated using a geographically weighted regression with altitude as an independent variable and a correction of residuals by kriging. Maximum expected rainfall in 24 hours was compared to the results of Casas et al. (2007). This comparison (Fig[6.9], Fig[6.11] and Fig[6.13], and Table[6.2]), shows that spatial differences in maximum expected rainfall in the two projects are usually low, with over 90% of the territory having differences under 20 mm for return periods of 2 years, the same maximum differences at 78% of the territory for return periods of 10 years and differences up to 40 mm over 70% of the territory for return periods of 100 years. The results of the present project produce overall slightly higher expected intensities and the spatial regions with most significant differences are the spots of extremest values, namely, the north of Alt Empordà and the south and west of Montsià that present local higher expected intensities according to the results of the present project; the maxima in northeastern Montseny, roughly west of La Selva county, is a region where the values have been underestimated in the present project, although it produced the absolute expected maxima for low return periods.

Obtained results of maximum expected rainfall in 24 hours were also analysed in relation to current thresholds for weather warnings. Accumulations of 100 mm in 24 hours (low level risk according to SMC warnings) are expected to be exceeded in over 80% of the territory only for return periods of at least 20 years. Accumulations of 200 mm in 24 hours (high level risk according to SMC warnings) are only expected to be exceeded in around half the territory for long return periods (over 200 years).

Finally, maximum expected rainfall at high spatial resolution was temporally downscaled to a selection of subdaily durations (1h, 6h and 12h) and a temporal scaling to 48h was also performed using the same methodology. IDF results were presented in a collection of maps of expected intensity at the considered return periods. The final results of the project for a duration of 1 hour were validated with a comparison to maximum expected rainfall obtained from automatic weather stations (without need of any temporal downscaling methodology from daily data). The comparison revealed some differences that could be attributed to the short length of automatic series (hence, low probability of registering extreme events) and less rigorousness in the quality controls that AWS data

has undergone (except for data registered in the most recent period, 2008-2016, that is already of high quality).

Differences found in the section dedicated to validate the obtained IDF relationships were already expected and are completely acceptable; these differences prove that the temporal downscaling methodology was needed because AWS are not yet ready to produce the same results. Automatic weather stations will be more reliable to construct IDF relationships in the future, when these stations are able to provide longer time series. In the meantime, the monofractal temporal downscaling is a good working methodology to estimate subdaily maxima from daily rainfall data.

List of Acronyms

ACA Agència Catalana de l'Aigua

AEMet Agencia Española de Meteorología

AWS Automatic Meteorological Station

CDF Cumulative Distribution Function

DDF Depth-Duration-Frequency

ETRS89 European Terrestrial Reference System 1989

GEV Generalised Extreme Value

GLO Generalised Logistic

GNO Generalised Normal

GPA Generalised Pareto

ICGC Institut Cartogràfic i Geològic de Catalunya

IDF Intensity-Duration-Frequency

PE3 Pearson type III

SAIH Sistema Automático de Información Hidrológica

SMC Servei Meteorològic de Catalunya

SNHT Standard Normal Homogeneity Test

UTM Universal Transverse Mercator

XEMA Xarxa d'Estacions Meteorològiques Automàtiques
(network of AWS managed by the SMC)

XOM Xarxa d'Observadors Meteorològics
(network of manned stations managed by the SMC)

List of Publications

Published papers:

- Llabrés-Brustenga A, A. Rius, R. Rodríguez-Solà, M. C. Casas-Castillo and A. Redaño. 2019. Quality control process of the daily rainfall series available in Catalonia from 1855 to the present, *Theoretical and Applied Climatology*, 137:2715-2729, doi 10.1007/s00704-019-02772-5 (https://rdcu.be/bhhSp).

- Casas-Castillo M. C., A. Llabrés-Brustenga, A. Rius, R. Rodríguez-Solà and X. Navarro. 2018. A single scaling parameter as a first approximation to describe the rainfall pattern of a place: application on Catalonia. *Acta Geophysica*, 66(3):415-424, doi 10.1007/s11600-018-0122-5 (http://rdcu.be/HdU3).

Summitted papers:

- Llabrés-Brustenga, A, A. Rius, R. Rodríguez-Solà, M. C. Casas-Castillo and A. Redaño. Influence of regional and seasonal rainfall patterns on the ratio between fixed and unrestricted measured intervals of rainfall amounts. Summitted to *Theoretical and Applied Climatology*.

Communications in workshops:

- Llabrés-Brustenga, A., M. C. Casas-Castillo, R. Rodríguez-Solà, A. Rius, A. Redaño and M. Barnolas. 2018. A study of the scaling properties of rainfall and its application to obtain Intensity-Duration-Frequency relationships in Catalunya. Poster presented in *European Geosciences Union General Assembly (EGU, Vienna, April 2018)*.

- Llabrés-Brustenga, A., M. C. Casas-Castillo, R. Rodríguez-Solà and A. Rius. 2017. A study of the dependence of the ratio between rainfall amounts measured using fixed and unrestricted intervals on the characteristic rainfall pattern of a place. Poster presented in *European Meteorological Society annual meeting (EMS, Dublin, September 2017)*.

- Llabrés-Brustenga, A., M. C. Casas-Castillo, R. Rodríguez-Solà and A. Rius. 2017. Ajustando el factor de corrección para medidas de precipitación realizadas con intervalos fijos según el régimen pluviométrico de la localidad. Poster presented in

XXXVI Reunión Bienal de la Real Sociedad Española de Física (Santiago de Compostel·la, July 2017).

- Llabrés-Brustenga, A., M. C. Casas-Castillo, R. Rodríguez-Solà, A. Rius, M. Barnolas and A. Redaño. 2017. A study of the scaling properties of rainfall and their relationship to climate characteristics in Catalonia. Poster presented in *10th HyMeX (hydrological cycle in the Mediterranean) workshop (Barcelona, June 2017)*.
- Llabrés-Brustenga, A., M. Herrero, M. Perpinyà and A. Rius. 2015. Quality control of rainfall data at a daily scale for 1855-2014 in Catalonia. Poster presented in *10th EUMETNET Data Management Workshop (St. Gallen, October 2015)*.

Bibliography

Aguilar, E., I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa
2003. Guidelines on climate metadada and homogenization. *World Meteorological Organization, WMO/TD No. 1186*.

Alexandersson, H.
1986. A homogeneity test applied to precipitation data. *Journal of Climatology*, 6:661–675.

Anduaga, A.
2012. Meteorología, ideología y sociedad en la España contemporánea. *Consejo Superior de Investigaciones Científicas, Madrid*, ISBN 978-84-00-09421-8.

Aronica, G. T. and G. Freni
2005. Estimation of sub-hourly DDF curves using scaling properties of hourly and sub-hourly data at partially gauged site. *Atmospheric Research*, 77(1-4):114–123.

Asquith, W. H.
1998. Depth-duration frequency of precipitation for Texas. *Water Resources Investigation (U.S. Geological Survey) Report 98-4044*.

Bara, M., L. Gaál, S. Kohnová, J. Szolgay, and K. Hlavčová
2010. On the use of the simple scaling of heavy rainfall in a regional estimation of IDF curves in Slovakia. *Journal of Hydrology and Hydromechanics*, 58(1):49–63.

Bara, M., S. Kohnová, L. Gaál, J. Szolgay, and K. Hlavčová
2009. Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia. *Contributions to Geophysics and Geodesy*, 39(3):187–206.

Beaulieu, C., O. Seidou, T. B. M. J. Ouarda, X. Zhang, G. Boulet, and A. Yagouti
2008. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research*, 44:W02425.

Bell, F. C.
1969. Generalised rainfall-duration-frequency relationships. *Journal of Hydraulic Division ASCE*, 95(1):311–327.

Branisavljević, N., D. Prodanović, M. Arsić, Z. Simić, and J. Borota
2009. Hydro-meteorological data quality assurance and improvement. *Journal of the Serbian Society for Computational Mechanics*, 3(1):228–249.

Burlando, P. and R. Rosso
1996. Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187(1-2):45–64.

Casas, M. C., B. Codina, A. Redaño, and J. Lorente
2004. A methodology to classify extreme rainfall events in the western Mediterranean area. *Theoretical and Applied Climatology*, 77:139–150.

Casas, M. C., J. Cunillera, X. Amo, M. Herrero, M. Ninyerola, X. Pons, A. Redaño, A. Rius, and R. Rodríguez
2005. Mapes de precipitació màxima diària esperada a catalunya per a diferents períodes de retorn. *Servei Meteorològic de Catalunya, Departament de Medi Ambient i Habitatge, Generalitat de Catalunya*.

Casas, M. C., M. Herrero, M. Ninyerola, X. Pons, R. Rodríguez, A. Rius, and A. Redaño
2007. Analysis and objective mapping of extreme daily rainfall in catalonia. *International Journal of Climatology*, 27(3):399–409.

Casas, M. C., R. Rodríguez, R. Nieto, and A. Redaño
2008. The estimation of probable maximum precipitation: The case of Catalonia. *Trends and Directions in Climate Research (Annals of the New York Academy of Sciences)*, 1146:291–302.

Casas-Castillo, M. C., A. Llabrés-Brustenga, A. Rius, R. Rodríguez-Solà, and X. Navarro
2018a. A single scaling parameter as a first approximation to describe the rainfall pattern of a place: application on Catalonia. *Acta Geophysica*, 66(3):415–424.

Casas-Castillo, M. C., R. Rodríguez-Solà, X. Navarro, B. Russo, A. Lastra, P. González, and A. Redaño
2018b. On the consideration of scaling properties of extreme rainfall in Madrid (Spain) for developing a generalized intensity-duration-frequency equation and assessing probable maximum precipitation estimates. *Theoretical and Applied Climatology*, 131(1):573–580.

Caussinus, H. and O. Mestre
2004. Detection and correction of artificial shifts in climate series. *Royal Statistical Society*, Pp. 405–425.

Chen, M., P. Xie, and CPC Quality Control Working Group
2008. Quality control of daily precipitation reports at NOAA/CPC. *AMS 12th conferences on IOAS-AOLS, New Orleans, LA*, (3.3).

Chow, V., D. R. Maidment, and L. W. Mays
1988. Applied hydrology. *McGraw-Hill*.

COST Action ES0601
2011. Progress report from 03/05/2007 to 01/06/2011.

Desramaut, N.
2008. Estimation of intensity duration frequency curves for current and future climates (thesis of master). *Department of Civil Engineering and Applied Mechanics, McGil University, Montreal, Quebec (Canada)*.

Domonkos, P.
2011. Adapted Caussinus-Mestre algorithm for networks of temperature series (ACMANT). *International Journal of Geosciences*, 2:293–309.

Domonkos, P.
2015. Homogenization of precipitation time series with ACMANT. *Theoretical and Applied Climatology*, 122:303–314.

Domonkos, P. and J. Coll
2016. Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests. *International Journal of Climatology*, 37(4):1910–1921.

Dwyer, I. J. and D. W. Reed
1995. Allowance for discretization in hydrological and environmental risk estimation. *Institute of Hydrology, Wallingford, report 123*.

Einfalt, T., N. Gerlach, C. Podlasly, and N. Demuth
2008. Rainfall and climate data quality control. *11th International Conference on Urban Drainage, Edinburgh, Scotland, UK*.

Esteban-Parra, M. J., F. S. Rodrigo, and Y. Castro-Diez
1998. Spatial and temporal patterns of precipitation in Spain for the period 1880-1992. *International Journal of Climatology*, 18:1557–1574.

Feng, S., Q. Hu, and W. Qian
2004. Quality control of daily meteorological data in China 1951-2000: a new dataset. *International Journal of Climatology*, 24:853–870.

Ferreri, G. B. and V. Ferro
1990. Short-duration rainfalls in Sicily. *Journal of Hydraulic Engineering*, 116(3).

García-Marín, A. P., J. L. Ayuso-Muñoz, F. J. Jiménez-Hornero, and J. Estévez
2013. Selecting the best IDF model by using the multibractal approach. *Hydrological Processes*, 27:433–443.

Grupta, V. K. and E. Waymire
1990. Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research*, 95(D3):1999–2009.

Guijarro, J. A.
2011. Influence of network density on homogenization performance. *World Meteorological Organization, WCDMP-No.78 (Climate Data and Monitoring). Seventh Seminar for Homogenization and Quality Control in Climatological Databases jointly organized with the Meeting of COST ES0601 (HOME) Action MC Meeting*, Pp. 11–18.

Guijarro, J. A.
2018. Homogenization of climatic series with climatol. *Version 3.1.1 https://CRAN.R-project.org/package=climatol*.

Herrera, S., J. M. Gutiérrez, R. Ancell, M. R. Pons, M. D. Frías, and J. Fernández
2012. Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *International Journal of Climatology*, 32:74–85.

Hershfield, D. M.
1961. Rainfall frequency atlas of the United States for durations from 30 minutes to 24 hours and return periods from 1 to 100 years. *U.S. Weather Bureau Technical Paper (Washington, D.C.)*, 40:115 pp.

Hosking, J. R. M.
1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society B*, 52(1):105–124.

Hosking, J. R. M.
2019a. L-Moments. *CRAN repository (last updated version: 2019-03-11)*.

Hosking, J. R. M.
2019b. Regional frequency analysis using L-moments. *CRAN repository (last updated version: 2019-03-10)*.

Huff, F. and J. R. Angel
1992. Rainfall frequency atlas of the midwest. *Illinois State Water Survey, Champaign Bulletin 71*.

Kendall, M. G.
1975. Rank correlation methods. *Charles Griffin, London*.

Koutsoyiannis, D. and E.Foufoula-Georgiou
1993. A scaling model of storn hyetograph. *Water Resources Research*, 29(7):2345–2361.

Koutsoyiannis, D., D. Kozonis, and A. Manetas
1998. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1-2):118–135.

Lana, X. and A. Burgueño
2000. Some statistical characteristics of monthly and annual pluviometric irregularity for the Spanish Mediterranean coast. *Theoretical and Applied Climatology*, 65:79–97.

Lana, X., M. D. Martínez, C. Serra, and A. Burgueño
2004. Spatial and temporal variability of the daily rainfall regime in Catalonia (north-eastern Spain), 1950-2000. *International Journal of Climatology*, 24:613–641.

Llabrés-Brustenga, A., A. Rius, R. Rodríguez-Solà, M. C. Casas-Castillo, and A. Redaño
2019. Quality control process of the daily rainfall series available in Catalonia from 1855 to the present. *Theoretical and Applied Climatology*, 137:2715–2729.

Llasat, M. C. and M. Puigcerver
1994. Meteorological factors associated with floods in the north-eastern part of the Iberian Peninsula. *Natural Hazards*, 9:81–93.

Mandelbrot, B. B.
1977. Fractals: form, chance and dimension. *W. H. Freeman and Co.*

Mandelbrot, B. B.
1982. The fractal geometry of nature. *W. H. Freeman and Co.*

Mann, H. B.
1945. Nonparametric tests against trend. *Econometrika*, 13:245–259.

Martín Vide, J. and J. M. Raso Nadal
2008. Atles climàtic de Catalunya: període 1961-1990. *Institut Cartogràfic de Catalunya & Servei Meteorològic de Catalunya, Barcelona*.

Menabde, M., A. Seed, and G. Pegram
1999. A simple scaling model for extreme rainfall. *Water Resources Research*, 35(1):335–339.

Meseguer-Ruiz, O., J. Martín-Vide, J. Olcina Cantos, and P. Sarricolea Espinoza
2014. La distribución espacial de la fractalidad temporal de la precipitación en la España peninsular y su relación con el índice de concentración. *Investigaciones Geográficas (Chile)*, 48:73–84.

Meseguer-Ruiz, O., T. J. Osborn, P. Sarricolea, P. D. Jones, J. O. Cantos, R. Serrano-Notivoli, and J. Martín-Vide
2019. Definition of a temporal distribution index for high temporal resolution precipitation data over Peninsular Spain and the Balearic Islands: the fractal dimension; and its synoptic implications. *Climate Dynamics*, 52(1-2):439–456.

Mestre, O., C. Gruber, C. Prieur, H. Caussinus, and S. Jourdain
2011. Splidhom: a method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, 50:2343–2359.

Morbidelli, R., C. Saltalipi, A. Flammini, M. Cifrodelli, T. Picciafuoco, C. Corradini, M. C. Casas-Castillo, J. H. Fowler, and S. M. Wilkinson
2017. Effect of temporal aggregation on the estimate of annual maximum rainfall depths for the design of hydraulic infrastructure systems. *Journal of Hydrology*, 554:710–720.

Nie, S. P., Y. Luo, W. P. Li, T. W. Wu, X. L. Shi, and Z. Z. Wang
2012. Quality control and analysis of global gauge-based daily precipitation dataset from 1980 to 2009. *Advances in Climate Change Research*, 3(1):45–53.

Palau, J. L. and F. Rovira
2015. Meso-alpha scale tropospheric interactions within the Western Mediterranean Basin: Statistical results using 15-year NCEP/NCAR reanalysis dataset. *Advances in Meteorology*, 2015:11 pp.

Palau, J. L., F. Rovira, and M. J. Sales
2017. Satellite observations of the seasonal evolution of total precipitable water vapor over the Mediterranean sea. *Advances in Meteorology*, 2017:9 pp.

Papalexiou, S. M., Y. G. Dialynas, and S. Grimaldi
2016. Hershfield factor revisited: Correcting annual maximum precipitation. *Journal of Hydrology*, 542:884–895.

Peterson, T. C., D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker
1998. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology*, 18:1493–1517.

Prohom Duran, M.
2006. La contribución de la sociedad astronómica de Barcelona en la difusión de las observaciones meteorológicas en Cataluña (1910-1923). *Investigaciones Geográficas*, (40):141–155.

Pérez-Zanón, N., M. C. Casas-Castillo, R. Rodríguez-Solà, J. C. Peña, A. Rius, J. G. Solé, and A. Redaño
2016. Analysis of extreme rainfall in the Ebre Observatory (Spain). *Theoretical and Applied Climatology*, 124:935–944.

Pérez-Zanón, N., J. Sigró, P. Domonkos, and L. Ashcroft
2015. Comparison of HOMER and ACMANT homogenization methods using a central Pyrenees temperature dataset.

Ragulina, G. and T. Reitan
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.

Ribeiro, S., J. Caineta, and A. C. Costa
2015. Review and discussion of homogenisation methods for climate data. *Physics and Chemistry of the Earth*, 94:167–179.

Rodríguez-Solà, R., M. C. Casas-Castillo, X. Navarro, and A. Redaño
2017. A study of the scaling properties of rainfall in Spain and its appropriateness to generate intensity-duration-frequency curves from daily records. *International Journal of Climatology*, 37(2):770–780.

Rodríguez-Solà, R., X. Navarro, M. C. Casas, J. Ribalaygua, B. Russo, L. Pouget, and A. Redaño
2014. Influence of climate change on IDF curves for the metropolitan area of Barcelona (Spain). *International Journal of Climatology*, 34:643–654.

Romero, R., J. A. Guijarro, C. Ramis, and S. Alonso
1998. A 30-year (1964-1993) daily rainfall data base for the Spanish Mediterranean regions: first exploratory study. *International Journal of Climatology*, 18:541–560.

Saladié, O., M. Brunet, E. Aguilar, J. Sigró, and D. López
2007. Observar la lluvia en Cataluña: 150 años de registros. *Publicacions URV, Tarragona*, ISBN 978-84-84-24093-8.

Schertzer, D. and S. Lovejoy
1987. Physical modelling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research*, 92(D8):9693–9714.

Schertzer, D. and S. Lovejoy
1991. Multifractal analysis techniques and the rain and cloud fields from 10^{-3} to 10^{6} m. *Non-Linear Variability in Geophyisics*, Pp. 111–144.

Schertzer, D. and S. Lovejoy
2011. Multifractals, generalized scale invariance and complexity in geophysics. *International Journal of Bifurcation and Chaos*, 21(12):3417–3456.

Sciuto, G., B. Bonaccorso, A. Cancelliere, and G. Rossi
2009. Quality control of daily rainfall data with neural networks. *Journal of Hydrology*, 364:13–22.

Serrano-Notivoli, R., M.
de Luis, M. A. Saz, and S. Beguería
2017a. Spatially based reconstruction of daily precipitation instrumental data series. *Climate Research*, 73(3):167–186.

Serrano-Notivoli, R., S.
Beguería, M. A. Saz, L. A. Longares, and M. de Luis
2017b. SPREAD: a high-resolution daily gridded precipitation dataset for Spain - an

extreme events frequency and intensity overview. *Earth System Science Data*, 9:721–738.

Servei Meteorològic de Catalunya
2019. Butlletí anual d'indicadors climàtics. *Departament de Territori i Sostenibilitat*.

Shearman, R. J.
1975. Computer quality control of daily and monthly rainfall data. *Meteorological Magazine*, 104:102–108.

Stepanek, P.
2003. Anclim software documentation. *http://www.climahom.eu/AnClim.html*.

Szentimrey, T.
2006. Development of mash homogenization procedure for daily data. *Fifth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary*, Pp. 123–130.

van Montfort, M.
1997. Concomitants of the hershfield factor. *Journal of Hydrology*, 194:357–365.

Venema, V. K. C., O. Mestre, E. Aguilar, I. Auer, J. A. Guijarro, P. Domonkos, G. Vertacnik, T. Szentimrey, P. Stepanek, P. Zahradnicek, J. Viarre, G. Müller-Westermeier, M. Lakatos, C. N. Williams, M. J. Menne, R. Lindau, D. Rasol, E. Rustemeier, K. Kolokythas, T. Marinova, L. Andersen, F. Acquaotta, S. Fratianni, S. Cheval, M. Klancar, M. Brunetti, C. Gruber, M. Prohom Duran, T. Likso, P. Esteban, and T. Brandsma
2012. Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8:89–115.

Vicente-Serrano, S. M., S. Beguería, J. I. López-Moreno, M. A. García-Vera, and P. Stepanek
2010. A complete daily precipitation database for northeast Spain: reconstruction, quality control and homogeneity. *International Journal of Climatology*, 30(8):1146–1163.

Vigneau, M. J. P.
1987. 1986 dans les Pyrénées orientales: deux perturbations méditerranéennes aux effets remarquables. *Revue géographique des Pyrénées de du Sud-Ouest*, 58(1):23–54.

Viney, N. R. and B. C. Bates
2004. it never rains on Sunday: the prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. *International Journal of Climatology*, 24:1171–1192.

Wang, X. L. and Y. Feng
2009. RHtestV3 user manual. *http://cccma.seos.uvic.ca/ETCCDMI/software.shtml*.

Weiss, L. L.
1964. Ratio of true to fixed-interval maximum rainfall. *Journal of Hydraulics Division*, 90:77–82.

World Meteorological Organization
2008. Guide to meteorological instruments and methods of observation. *Weather · Climate · Water*, ISBN 978-92-63-10008-5.

You, J., K. G. Hubbard, S. Nadarajah, and K. E. Kunkel
2007. Performance of quality assurance procedures on daily precipitation. *Journal of Atmospheric and Oceanic Technology*, 24:821–834.

Young, C. B. and B. M. McEnroe
2003. Sampling adjustment factors for rainfall recorded at fixed time intervals. *Journal of Hydrologic Engineering*, 8(5):294–296.

Yu, P. S., T. C. Yang, and C. S. Lin
2004. Regional rainfall intensity formulas based on scaling property of rainfall. *Journal of Hydrology*, 295(1-4):108–123.