

UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Scienze Statistiche
SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE STATISTICHE
CICLO XXVI

Quantile Regression and Bass Models in Hydrology

Direttore della Scuola: Ch.ma Prof.ssa Monica Chiogna

Supervisore: Ch.mo Prof. Renato Guseo

Co-supervisore: Dott.ssa Claudia Furlan

Dottorando: Luca Sartore

31 Gennaio 2014

To those relatives who are no longer among us.

Acknowledgements

Foremost, I would like to thank my supervisor Prof. Renato Guseo, for his support and guidance during these years of research, for his perspective of statistics in science and for the long meetings and stimulating discussions. I would like expressing my gratitude also to Dr. Claudia Furlan for her patience and questions, which contributed in inspecting furthermore the solutions of some problematic topics. I am also grateful to Prof. Cinzia Mortarino, Dr. Mariangela Guidolin and Dr. Alessandra Dalla Valle, who contributed in different ways to inspire new ideas or improve others. I owe a particular thanks to Prof. Paul Eilers, whom I met during the third joint statistical meeting D.A.G.Stat. in Freiburg: the new method, that he presented, inspired one technique proposed in this thesis.

I thank my fellow classmates Roberta Pappadà, Erlis Ruli, Ivan Luciano Danesi, Lorenzo Maragoni, Md Abud Darda, Akram Yazdani for the hard work during the sleepless nights of the first year to be able to respect all the deadlines. My deepest appreciation goes to Shireen Assaf for her patience in reading part of the thesis and trying to improve my oral and written English.

I would express my gratitude to the technical and administrative staff, in particular to Elena Zamborlin, Delfina Di Monte, Giuliana Fanin, Patrizia Piacentini, Annamaria Pieretti and Tiziano Schiavo. I am also grateful for the technical assistance given by Gianluca Moro, Silvia Sartorelli, Mirko Moro, and especially by Angelo Scarano, whose friendship, patience and kindness were essential in these years.

I have greatly benefited also from the time spent with my friends Luca Pierbon, Marco Simioni, and the Cecilian friends in Vicenza, in particular Marcello Meneguzzo and Francesco Dal Lago, who always ask about the ongoing of my research. In addition, I would like to thank Oksana Žurakovska and the Medoacus choir for those hours spent in creating music together, and Fr. Ettore Simioni, whose insightful advices have been guiding me since I started my Ph.D.

Last but not the least, I would like to express my gratitude to my family: my parents Paolo Sartore and Giulia Campagnaro, and my brother Luigi Sartore, who are supporting me as no one can. I also have had the great encouragement of my aunt Lina Sartore.

LUCA SARTORE

Padova

20th January 2014

Abstract

Spatiotemporal phenomena related to the rainfall measurements can be characterised by statistical models grounded on physical concepts instead of being identified by spatiotemporal patterns based on standard correlations and related analytical tools. This perspective is useful in understanding if the relationships among neighbouring zones and consecutive years are attributable to latent physical mechanisms. Satellite data are used to examine this theory and provide evidence on empirical basis. A recent hydrological theory, which is based on the concept of self-organisation, consists of simplified physical mechanisms that are essential for the explanation of local data relationships. The regression models inspired by the diffusion of innovation can approximate the evolution of the rainfall process within a year through a more straightforward perspective. However, the multitude of collected data requires innovative techniques of data management and advanced analytical solutions, in order to achieve optimal results in reasonable time. Indeed, the nonlinear least squares and nonlinear quantile regression are considered to make inference on the response variable given some covariates. A new quantile regression technique is developed in order to provide simultaneous estimates that do not violate the monotonicity property of quantiles. The nonlinear least squares highlight strong connections among rainfall and the salient features of the measurements areas. Furthermore, the quantile regression analyses quantify the intrinsic variability of the data.

Sommario

I fenomeni spazio-temporali relativi alle misurazioni di piovosità possono essere caratterizzati da modelli statistici fondati su concetti fisici invece di essere identificati da modelli standard basati su correlazioni spazio-temporali e i relativi strumenti analitici. Questa prospettiva è utile per capire se i rapporti tra zone confinanti e anni consecutivi sono attribuibili a meccanismi fisici latenti. Dati satellitari vengono utilizzati per esaminare questa teoria e fornire prove su base empirica. Una recente teoria idrologica, basata sul concetto di auto-organizzazione, è caratterizzata da meccanismi fisici semplificati che sono essenziali per la spiegazione delle relazioni locali presenti nei dati osservati. I modelli di regressione, che si ispirano alla teoria della diffusione di innovazioni, sono in grado di approssimare l'evoluzione del processo di precipitazione di un singolo anno attraverso una più semplice prospettiva. Tuttavia, la moltitudine di informazioni raccolte richiede tecniche innovative di gestione dei dati e soluzioni analitiche avanzate con lo scopo di ottenere risultati ottimali in tempi ragionevoli. Infatti, i minimi quadrati e la regressione quantilica per modelli non-lineari vengono utilizzati per fare inferenza sulla variabile risposta condizionatamente ad alcune covariate. Una nuova tecnica di regressione quantilica è stata sviluppata *ad hoc* al fine di fornire stime simultanee che non violino la proprietà di monotonicità dei quantili. I minimi quadrati non lineari evidenziano un forte legame tra le precipitazioni e alcune caratteristiche salienti delle zone di misurazione. Inoltre, le analisi ottenute tramite la regressione quantilica quantificano la variabilità intrinseca nei dati.

Contents

1	Introduction	1
1.1	Overview	1
1.2	Summary and main contributions of the thesis	3
2	A simplified regression model for rainfall data	5
2.1	The hydrological cycle	6
2.2	The aerosol-cloud-precipitation system	8
2.3	The simplification based on the diffusion of the innovation	10
2.4	Conclusion	13
3	Constrained quantile regression	15
3.1	Literature review of estimation methods	17
3.1.1	Separate quantile regressions	18
3.1.2	Restricted quantile regression	19
3.1.3	Rearrangement quantile estimator	20
3.1.4	Multi-objective quantile estimators	22
3.2	New estimation procedures	25
3.2.1	Simultaneous linear quantile regression	30
3.2.2	Simultaneous nonlinear quantile regression	32
3.2.3	Semi-parametric quantile sheets	36
3.3	Testing multi-objective quantile estimators	38
3.3.1	Goodness of fit tests	39
3.3.2	Analyses of residuals	42
3.3.3	Confidence regions and intervals	44
3.4	Conclusion	48
4	Preliminary Data Analysis	49
4.1	Precipitation data	49
4.1.1	General Information	50
4.1.2	Data adjustments	51
4.2	Methodology	51

4.2.1	Data transformations	53
4.2.2	Kohonen networks (self-organising maps)	55
4.2.3	Cluster validation	58
4.3	Application	61
4.4	Conclusion	66
5	Least squares regression analyses	69
5.1	Data reduction	69
5.2	The Bass model	70
5.3	The generalised Bass model	77
5.4	Conclusion	80
6	Quantile regression analyses	87
6.1	The Bass model	87
6.2	The generalised Bass model	89
6.3	Conclusion	94
7	Conclusions	101
A	Notes on quantile regression	105
A.1	Proof of the homogeneity property	105
A.2	Proof of quantile level identification	105
A.3	Proofs of critical values consistency	106
A.4	Simulation studies	108
A.4.1	A linear regression model	110
A.4.2	A nonlinear regression model	110
A.5	Notes on the simulation accuracy	113
A.6	Simulation results	115
A.6.1	Tables related to the linear regression	115
A.6.2	Tables related to the non linear regression	131
A.6.3	Tables related to the tests	147
B	Cluster Analysis Results	152
B.1	Clustering with random initialisation	152
B.1.1	Resulting clusters	152
B.1.2	Clustering validation graphs	153
B.1.3	Resulting centroids	155
B.1.4	Centroid hierarchical clustering	157
B.1.5	Merging Costs	159
B.2	Clustering with linear initialisation	161
B.2.1	Resulting clusters (PCA)	161

B.2.2	Clustering validation graphs (PCA)	162
B.2.3	Resulting centroids (PCA)	164
B.2.4	Centroid hierarchical clustering (PCA)	166
B.2.5	Merging Costs (PCA)	168
B.3	Clustering with informative indexes	170
B.3.1	Clustering validation graphs (indexes)	170
B.3.2	Resulting centroids (indexes)	171
B.3.3	Centroid hierarchical clustering (indexes)	171
B.3.4	Merging Costs (indexes)	172
B.3.5	Resulting clusters (indexes)	172
Bibliography		173
Curriculum Vitae		183

List of Tables

4.1	Cluster validation statistics	65
4.2	Cluster validation statistics for the model-based approach . . .	65
A.1	Average of the estimates for the Normal distribution	115
A.2	Estimator variance for the Normal distribution	116
A.3	Squared bias for the Normal distribution	117
A.4	Mean squared error for the Normal distribution	118
A.5	Average of the estimates for the Laplace distribution	119
A.6	Estimator variance for the Laplace distribution	120
A.7	Squared bias for the Laplace distribution	121
A.8	Mean squared error for the Laplace distribution	122
A.9	Average of the estimates for the Cauchy distribution	123
A.10	Estimator variance for the Cauchy distribution	124
A.11	Squared bias for the Cauchy distribution	125
A.12	Mean squared error for the Cauchy distribution	126
A.13	Average of the estimates for the Beta-Normal distribution . . .	127
A.14	Estimator variance for the Beta-Normal distribution	128
A.15	Squared bias for the Beta-Normal distribution	129
A.16	Mean squared error for the Beta-Normal distribution	130
A.17	Average of the estimates for the Gamma distribution	131
A.18	Estimator variance for the Gamma distribution	132
A.19	Squared bias for the Gamma distribution	133
A.20	Mean squared error for the Gamma distribution	134
A.21	Average of the estimates for the Log-Normal distribution . . .	135
A.22	Estimator variance for the Log-Normal distribution	136
A.23	Squared bias for the Log-Normal distribution	137
A.24	Mean squared error for the Log-Normal distribution	138
A.25	Average of the estimates for the Uniform distribution	139
A.26	Estimator variance for the Uniform distribution	140
A.27	Squared bias for the Uniform distribution	141
A.28	Mean squared error for the Uniform distribution	142

A.29 Average of the estimates for the Kumaraswamy-Gamma distribution	143
A.30 Estimator variance for the Kumaraswamy-Gamma distribution	144
A.31 Squared bias for the Kumaraswamy-Gamma distribution . . .	145
A.32 Mean squared error for the Kumaraswamy-Gamma distribution	146
A.33 Quantities related to the distribution of \bar{R}^1 test statistic . . .	147
A.34 Coverage probabilities for testing data from a Gamma distribution	148
A.35 Critical values for testing data from a Gamma distribution . .	148
A.36 Coverage probabilities for testing data from a Log-Normal distribution	149
A.37 Critical values for testing data from a Log-Normal distribution	149
A.38 Coverage probabilities for testing data from a Uniform distribution	150
A.39 Critical values for testing data drawn a Uniform distribution .	150
A.40 Coverage probabilities for testing data from a Kumaraswamy-Gamma distribution	151
A.41 Critical values for testing data from a Kumaraswamy-Gamma distribution	151

List of Figures

4.1	Map of the informative areas for the analyses.	61
4.2	Three dimensional grid to define relationships among centroids.	62
4.3	Clustered areas with random initialisation.	64
4.4	Clustered areas with linear initialisation.	66
4.5	Clustered areas with informative indexes.	67
5.1	Selected region to analyse.	71
5.2	Smoothed maps of the parameter μ of the Bass model.	74
5.3	Smoothed maps of the parameter ζ of the Bass model.	75
5.4	Smoothed maps of the parameter ξ of the Bass model.	76
5.5	Approximations of the intervention function and its integral.	79
5.6	Smoothed maps of the parameter μ of the generalised Bass model.	81
5.7	Smoothed maps of the parameter ζ of the generalised Bass model.	82
5.8	Smoothed maps of the parameter ξ of the generalised Bass model.	83
5.9	Smoothed maps of the parameter α of the generalised Bass model.	84
5.10	Smoothed maps of the parameter β of the generalised Bass model.	85
6.1	Smoothed quantile maps of the parameter μ of the Bass model.	90
6.2	Smoothed quantile maps of the parameter ζ of the Bass model.	91
6.3	Smoothed quantile estimates of the parameter ξ of the Bass model.	92
6.4	Smoothed quantile maps of the parameter μ of the generalised Bass model.	95
6.5	Smoothed quantile maps of the parameter ζ of the generalised Bass model.	96

6.6	Smoothed quantile maps of the parameter ξ of the generalised Bass model.	97
6.7	Smoothed quantile maps of the parameter α of the generalised Bass model.	98
6.8	Smoothed quantile maps of the parameter β of the generalised Bass model.	99
B.1	Different precipitation areas after cluster validation.	152
B.2	Cluster validation maps (original and standardised data).	153
B.3	Cluster validation maps (log and range standardised data).	154
B.4	Centroids patterns (original and standardised data).	155
B.5	Centroids patterns (log and range standardised data).	156
B.6	Dendrograms (original and standardised data).	157
B.7	Dendrograms (log and range standardised data).	158
B.8	Merging costs (original and standardised data).	159
B.9	Merging costs (log and range standardised data).	160
B.10	Precipitation areas after cluster validation (PCA).	161
B.11	Cluster validation maps via PCA (original and standardised data).	162
B.12	Cluster validation maps via PCA (log and range standardised data).	163
B.13	Centroids patterns via PCA (original and standardised data).	164
B.14	Centroids patterns via PCA (log and range standardised data).	165
B.15	Dendrograms (original and standardised data).	166
B.16	Dendrograms (log and range standardised data).	167
B.17	Merging costs (original and standardised data).	168
B.18	Merging costs (log and range standardised data).	169
B.19	Cluster validation maps via indexes	170
B.20	Cluster validation maps via indexes	171
B.21	Dendrogram via indexes	171
B.22	Merging Costs via indexes.	172
B.23	Precipitation areas after cluster validation (indexes).	172

Chapter 1

Introduction

The quantitative clouds analysis and explanations of the processes which involve the formation of cloud and precipitations started relatively late. In fact, the developments obtained in this scientific field between the 17th century and the 1940's were mainly based on philosophical concepts that could not find an empirical evidence of proves.

In the recent years, the rapid advances in technological developments have led to a continuous improvement of the tools to collect data. From a statistical perspective, the analysis of images and measurements produced by satellites is the most challenging issue due to the large amount of data and the complexity of the models adopted to describe the observed phenomena.

A part of the statistical disciplines has centred on producing estimation techniques and forecasting methods, which have a marginal role in other scientific field, such as physics; as a matter of fact, modelling an error or data distributions introduces assumptions which might generate inaccurate results. To obviate this problem, inference based on empirical quantiles can be used for assumption free distribution analyses.

A brief review of the outstanding literature is given in Section 1.1, while Section 1.2 introduces the structure of the thesis and its main contributions.

1.1 Overview

Analysis of rainfall data has strong relevance not only in meteorology. It helps people to organise their lives and activities. This type of data are related to the water-cycle and they are collected and studied mainly by hydrologists. There are many different methods to study rainfall data, some focus on the analysis of extreme values and others on the identification and interpretation of proper dynamical systems. The physical principles of the

process that generates these data and their main characteristics are well documented in hydrology, meteorology and physics of the atmosphere. Mostly, self-organisation is the main feature of this complex system, wherein aerosol particles and water molecules play crucial roles. The so-called aerosol-cloud-precipitation system (Koren and Feingold, 2011) is based on simultaneous differential equations describing the interaction between clouds and aerosol particles. This dynamical perspective can be considered as a mean field approximation of the Nagel and Raschke's proposal (1992), which was the first attempt to model clouds dynamics and local interactions among particles by the use of cellular automata.

The dynamics among agents of a cellular automaton are well described by the innovation diffusion models introduced by Bass (1969) as a mean field approximation (Tang and Bak, 1988) of the underlying process. Few years ago, Guseo and Guidolin (2010) developed a new extension for the generalised Bass model (Bass *et al.*, 1994) which deals with cellular automata and, if it is correctly simplified, can be applied to rainfall data on large-scales. However, these models are not able to catch seasonal effects due to the radiation intensity and, hence, variation of temperatures and precipitations. Recently, several solutions for this problem were studied. The proposal of Guidolin and Guseo (2013) is particularly useful for this purpose, because it allows to specify harmonic functions which are introduced in the generalised Bass model as seasonal shocks. Of course, the physical interpretation of this model is meaningful according to the theories developed in these recent years.

From an inferential point of view, the parameters of these models are usually estimated on cumulative quantities by nonlinear least-squares regression techniques (Seber and Wild, 2003). The most popular algorithm to minimise the deviance of the model was developed by Marquardt (1963), who improved the method previously introduced by Levenberg (1944). Nonlinear least squares are generally free distribution methods that allows to estimate an expected trajectory. However, quantile regression techniques are more suitable and robust procedures to study the characteristics of the distribution of a response variable with respect some covariates. Without making assumption on the underlying distribution, Koenker and Bassett Jr. (1978) introduced the linear quantile regression methodology, wherein the deviance to be minimised is based on a summation of linear asymmetric loss functions. In order to obtain reliable results and improve the convergence to the optimal point for nonlinear quantile regression problems, Koenker and Park (1996) developed a further technique based on both the Busovaca's procedure (1985) and the Powell's algorithm (1969).

In most cases, this initial quantile regression techniques suffer from the so-called crossing quantile problem. This happens especially when each quantile

trajectory is separately fitted to the data; therefore the information of other curves position is not considered, so that it is possible to obtain parameters estimates which lead to curves that violate the monotonicity property of quantiles. The first solution of this problem was presented in the literature by He (1997), who proposed some forcing conditions based on the location-scale paradigm. Chernozhukov *et al.* (2010) proposal is based on a monotonic reorganisation of the quantile levels for each point in the domain of the estimated curves. Furthermore, the method of Bondell *et al.* (2010) consists of an algorithm to estimate simultaneously all parameters for each quantile curve according to some constraints and, more recently, the technique proposed by Schnabel and Eilers (2013) allows to fit simultaneous quantile curves by avoiding a constrained optimisation. However, none of these techniques can be adopted to fit parametric nonlinear models, because they are all based on linear model or linear combination of basis functions (such as splines).

1.2 Summary and main contributions of the thesis

The complexity of the phenomena and process that generates rainfall data is described in Chapter 2. However, the large amount of data that are collected by meteorologists and hydrologists requires simple models that are able to catch most of the physical properties of the process. The rationale of the innovation diffusion theory is here discussed and interpreted in the light of a new scientific context.

In order to describe the spatiotemporal variability of the data by avoiding distributive assumptions, several quantile regression techniques are considered in Chapter 3, and a particular emphasis is given to alternative solutions of the quantile crossing curves for nonlinear models. In this chapter, the main contribution focuses on simultaneous quantile regression procedures and to the development of reliable parametric estimators. Both the computational efficiency and accuracy of these techniques are taken into account; in fact, the Marquardt's algorithm (1963) is considered from a new perspective.

From a practical point of view, a preliminary analysis of rainfall data is performed and documented in Chapter 4. It mostly deals with a brief review of cluster analysis theory in order to choose an appropriate algorithm and validation methods. The selected procedures are subsequently applied to understand which macro-area can be used for the successive analyses.

Chapter 5 deals with the application of least square regression in order to estimate the parameters and perform a model selection based on the data

coming from a reduced area. Several innovation diffusion models are fitted; in particular, the Bass' model (1969), its generalisations (Bass *et al.*, 1994; Guidolin and Guseo, 2013) and different mixtures of shifted Bass models (Guseo, 2004) are considered. The results highlight the role of the innovation diffusion models in interpreting the basic mechanisms underlying the formation of precipitation, i.e. the initialisation of the condensation through the aerosol particles and subsequent coalescence which leads to heavier droplets. Nonetheless, other connections between the parameters and the orography of the considered area emerge by adopting this model-based perspective.

All the analyses and the results exposed in Chapter 6 are obtained through the application of the theory developed in Chapter 3. Particular emphasis is given to the simultaneous estimation techniques, which are applied to the Bass' model and its extension (Bass *et al.*, 1994). A more detailed description of the data variability is obtained via quantile regression, which shows the effects of parameters in correspondence of probability variations. Most of the parameters are quite constant or show little significant variations, however only the parameter controlling the scale of the trajectory presents the largest discrepancies between probability levels.

Final conclusions and remarks are presented in Chapter 7.

Chapter 2

A simplified regression model for rainfall data

Meteorology and weather forecast play a major role in peoples' lives and can be very important for planning purposes. For instance, in agriculture it can be useful to establish irrigation periods and quantity of water distribution, or in population safety it is possible to predict the path and the area covered by hurricanes in order to alert population. Weather forecast is mostly used to determine the time and the duration of precipitation events. However, special meteorological measures such as rainfall can be much more useful to determine how much freshwater is available for human consumption, agriculture and industry.

Since precipitation is observed in a specific spatio-temporal coordinate system, the techniques used to analyse these kinds of data are based on both spatial statistics and time series methodologies. For forecasting purposes, even if most of the statistical analyses are quite advanced, the physical description of the atmospheric phenomena related to the precipitation process is mostly considered on a macroscopic scale rather than a microscopic. The recorded rainfall, in fact, depends on several aspects: the type of clouds, the humidity, the land-use, the orography, the wind speed and direction, the atmospheric pressure and many others. At the same time, the microphysical processes consist mainly in the interactions among water molecules and different kind of aerosol particles (e.g. biomass burning and air pollutant). Both microphysical and macrophysical processes coexist in nature, so that the interactions among the particles of the clouds are affected by many macroscopic factors. The aerosols, under some conditions, initialise the formation of cloud droplets or ice crystals. The precipitation is consequently the result of the formation of raindrops by the collision and coalescence of many small droplets (Levin and Cotton, 2008).

A brief review and explanation of the most used hydrological terms and concept is given in Section 2.1. The physics of the aerosol-cloud-precipitation system is explained in Section 2.2, while the approximation based on the diffusion of the innovation is presented in Section 2.3. Brief conclusions of this Chapter are exposed in Section 2.4.

2.1 The hydrological cycle

The *water cycle*, or *hydrological cycle*, explains how the water is continuously moving across the earth and the atmosphere. It consists of several physical processes, which describe both the state transition of water and its various movements. The *state transition*, or *phase transition*, processes have in common the transformation of a thermodynamical system. E.g. water becomes ice or vapour according to temperature variations, which are due to several factors (e.g. the solar radiation and altitude). So that we can speak about

evaporation, when water passes from liquid to gas states;

condensation, which is the opposite transition of the evaporation;

sublimation, when the phase transition occurs from solid to gas states;

deposition, which is the opposite transition of the sublimation;

solidification, when water changes from liquid to solid states; and

fusion, which denotes the opposite transition of the solidification.

Other physical processes are behind the relocation of water on the earth and in the atmosphere. We can refer to

precipitation for liquid (rainfall or cloud water droplets) or solid water (snow, hail, etc.) that fall on the surface of the planet;

canopy interception for the precipitation that does not fall on the ground because it is intercepted by plant foliage;

runoff for the movements across the land (e.g. rivers) also including artificial canals;

infiltration for the water-flows from the surface into the ground;

subsurface flow for those flows that happen in the underground (e.g. the aquifers);

advection for the atmospheric movements of water in any state;

transpiration for the release of water vapour from vegetation and soil into the atmosphere;

percolation for horizontal flows through the soil under the influence of pressure forces.

The solar radiation provides heat and energy, which evaporates water from the oceans, lakes, rivers, moisturised soil and so on. By transpiration, plants release water into the air, while the ice in snowfields and glaciers may also sublimate. By advection, the water vapour is taken high in the atmosphere, where cooler temperatures make it to condense. Small droplets are so formed inside the clouds, which move from a place to another by advection. When the droplets become too heavy, the water returns to the ground as a form of precipitation. Part of this water is frozen in ice caps or glaciers. Part of the precipitation is intercepted by the vegetation and some flows to the seas and oceans as a runoff. Another part infiltrates and slowly flows underground among impermeable rock layers, eventually it returns to the seas as salty water (Han, 2011).

The hydrological cycle is known to be a complex and self-organising system. In particular, *self-organisation* is defined as the generation of a coherent global pattern, which spontaneously arises from local interactions among the components (or the *agents*) of the system. Self-organisation is an evident characteristic of many natural systems, and it is a useful concept adopted in different scientific fields, such as physics, environmental sciences, biology and economics (Heylighen, 2001). The complex nature of the water cycle consists of different and simultaneous self-organising processes, whereas the water molecules act as agents in constant change, whose interactions emerge from the necessity to adapt autonomously with the surrounding environment.

Modelling the water cycle denotes the capacity of being able to describe each single process and its relationships with the others. The limitation of tools to develop a general model led to simplified solutions, which focus on some processes. These solutions are often the results of approaches based on either a large- or micro-scale point of view. The *large-scale modelling* is oriented to explain the macroscopic behaviour of a phenomenon and studying its effects by ignoring the behaviour of the agents involved in the process. On the other hand, the *micro-scale modelling* concentrates the attention on the role of the particles, and the propagation of their effects.

2.2 The aerosol-cloud-precipitation system

As part of the water cycle, the *precipitation process* can be considered as a *self-organising subsystem*. It is characterised by specific tasks, which are related to the formation and the release of rain droplets, or other kind of solid precipitation (such as snow and hail). In order to simplify the complexity of the problem, such a process can be studied by ignoring most of the other subsystems, which form the water cycle (such as infiltration and percolation). For instance, the cloud formation mechanism of Nagel and Raschke (1992) is described in terms of self-organisation by the use of few clouds physical processes; in fact, they did not consider solar radiation and other micro-scale effects.

Condensation and solidification are the most known processes that lead to the formation of precipitation in the troposphere (i.e. the closer part of the atmosphere to the planet surface). These phase transitions arrange atmospheric molecules as a response to temperature variations. In the recent years, the nature, the physical properties and the chemical composition of such molecules has become clearer. Usually, we refer to the *aerosols* for a joint definition of solid or liquid particles in the atmosphere. The aerosols contribute to the formation of cloud droplets and ice particles and their role and variations have become crucial for a sensible explanation of clouds and precipitation changes (Stevens and Feingold, 2009).

The aerosol sources and the micro-physical effects are described in an extensive review by Andreae and Rosenfeld (2008). Almost all the varieties of aerosols are constituted by some deliquescent component, which has the potential to start the formation of *cloud condensation nuclei* (i.e. liquid cloud droplets) or *ice nuclei* (i.e. ice crystals). This potential differs according to the origin of aerosols particles and their size, and even if such particles are internally mixed, their composition should be taken into account. Due to the controversial role of human beings in climate change causes, the aerosols are mainly divided in anthropogenic and natural sources type. A more informative classification is based on the chemical composition and the original location of aerosol emissions. In particular, we can speak about soil dust, sea salt and sea spray aerosols, biological particles (e.g. from plants pollen, fungi, microbes or other organic matter), biomass burning, fossil fuel combustion, and other particles (such as nitrates and sulfates).

The interaction among aerosols, clouds and precipitations have been studied by several scientists. They often refer to the so-called *aerosol-cloud-precipitation system*, which consists in a variety of complex mechanisms. Stevens and Feingold (2009) proposed that these interactions should be considered as a buffered system, wherein a chain of subsequent events is given as

a system response to compensate for external changes. Feingold *et al.* (2010) considered a *cellular cloud field* characterised by stable oscillations (related to the Rayleigh–Bénard convective theory; Rayleigh, 1916), wherein precipitation affects and transforms cloud patterns. Koren and Feingold (2011) employed some principles of population dynamics to reduce the interactions complexities. In fact, by the use of *Lotka–Volterra equations* (Lotka, 1910; Volterra, 1927), they were able to model and mimic through simulations the main features of the aerosol-cloud-precipitation system.

The Lotka–Volterra equations can describe the instantaneous variations of two reference populations within a specific environment. According to the interpretation of Volterra (1927), one population acts as a resource for the other, and the transfer of this resource is the unique connection between these two populations. Indeed, we can refer to these equations as the *predator-prey model*, which is generally defined as two simultaneous differential equations, i.e. (Guseo, 2004)

$$\begin{aligned}\frac{\partial x(t)}{\partial t} &= \Phi\{x(t)\} - y(t)\pi\{x(t)\}; \\ \frac{\partial y(t)}{\partial t} &= -\nu y(t) + \eta y(t)\pi\{x(t)\},\end{aligned}$$

where x and y respectively represent the populations of preys and predators. The system-dependent constants ν and η respectively denote the mortality rate of predators and the efficiency of the resource usage. The system-dynamic $\Phi(\cdot)$ defines the growth of the preys and $\pi(\cdot)$ describes their consumption rate. According to the interpretation of Feingold *et al.* (2010), the quantity $-y(t)\pi\{x(t)\}$ denote the loss of cloud depth due to rainfall, and hence the reduction of water vapour within the clouds. The dynamic $\Phi\{x(t)\}$ defines the cloud depth increase, i.e. it describes how the water goes from the planet into the atmosphere (e.g. the processes of evaporation, sublimation, advection and many others). On the other equation, $-\nu y(t)$ is associated to the loss of cloud drop concentration, while $\eta y(t)\pi\{x(t)\}$ can be interpreted as the concentration of the aerosol particles and their renewal. More specifically, the quantities involved in these equations are

$x(t)$, which defines the total amount of water vapour in the clouds at the time t ;

$\Phi\{x(t)\}$, which describes the dynamics of the evaporation and sublimation of the water in the planet;

$\pi\{x(t)\}$, which denotes the varying rate of the cloud depth reduction, and it might be considered as a measure of the strength of the *coalescence*

effect, i.e. the aggregation strength of two or more particles to form a bigger one until the rain drop fall;

$y(t)$, which represents the total amount of the precipitation at the time t ;

ν , which refers to the gravitational consumption rate of the rain drops in the clouds;

η , which is the renewal rate of aerosol particles.

This formulation permits to study cyclical patterns characterised by evident oscillations, whose features are estimated from the data of the two populations. In many cases, the data of both populations are not always available and either the amount of precipitation or the cloud depth is usually missing. A simplified model, which takes into account the physical properties of the process, is required because rainfall data are more easily accessible.

2.3 The simplification based on the diffusion of the innovation

The fundamental idea of this viewpoint is based on the study of a sequence of annual oscillations obtained by splitting the cyclical patterns experimented by Koren and Feingold (2011). Even if this perspective is less accurate from a physical point of view, it can be considered a possible solution to describe the evolution of the amount of precipitation within a year. The *Riccati differential equations* (1758) were used by Bass (1969) to explain the diffusion of the innovations. Such models are here adopted in order to simplify the Lotka–Volterra system within a year window. In fact, the Lotka–Volterra equations can be used to describe the dynamics of precipitation and the cloud depth continuously for several years. On the other hand, the Riccati equations expresses a behaviour which leads to the saturation after a fixed period. For this reason, each year must be studied separately and, eventually, the time trends of the parameters estimates can show useful variation to forecast the precipitation for the successive years. From a physical viewpoint, the annual fluctuations are mostly due to the seasonal changes of the solar radiation, which synergically affects temperatures and clouds systems.

The simplest formulation of the Riccati equation can be expressed as

$$\frac{\partial y(t)}{\partial t} = \mu \left\{ \zeta - \xi \frac{y(t)}{\mu} \right\} \left\{ 1 - \frac{y(t)}{\mu} \right\}$$

whose solution is known as the *Bass model*, i.e.

$$y(t) = \mu \zeta \frac{1 - \exp\{-t(\zeta + \xi)\}}{\zeta + \xi \exp\{-t(\zeta + \xi)\}}, \quad (2.1)$$

which is a nonlinear function over time wherein μ refers to the total amount of adoptions as $t \rightarrow \infty$, ζ represents the innovation effect and ξ denotes the imitation effect. The physical explanation of these parameters can be given from two different perspectives. In both of them, μ approximates the total rainfall measured during the year. However, from a macroscopic point of view, the parameter ζ represents the proportion of observed rainfall generated in the analysed spatial unit, while ξ denotes the proportion of observed rainfall generated in other neighbouring areas. From a microscopic viewpoint, ζ indicates the effect of the aerosol in initialising the formation of the precipitation; in fact, the estimated value of this parameter may vary according to the chemical composition of such particles (e.g. see Petters and Kreidenweis, 2007) which is strictly linked to the land use of the area, human activities and transpiration of vegetation (Andreae and Rosenfeld, 2008). The microscopic interpretation of ξ is based on the behaviour of water droplets, which become bigger and heavier by collision and coalescence; in particular, when a critical mass of these rain-drops is reached, the gravity makes them fall. Since all these fallen droplets behave in this way, then the collisions, the coalescence and the gravitational effect may be considered as part of a unique imitating behaviour: the inertia.

The model in (2.1) can further be extended by the adoption of the so-called *generalised Bass model* (Bass *et al.*, 1994), which is formulated in terms of a differential equation as

$$\frac{\partial y(t)}{\partial t} = \mu \left\{ \zeta - \xi \frac{y(t)}{\mu} \right\} \left\{ 1 - \frac{y(t)}{\mu} \right\} a(t),$$

whose solution is given by

$$y(t) = \mu \zeta \frac{1 - \exp\left\{ -(\zeta + \xi) \int_0^t a(z) dz \right\}}{\zeta + \xi \exp\left\{ -(\zeta + \xi) \int_0^t a(z) dz \right\}}, \quad (2.2)$$

where the *intervention function* $a(t)$ is a non-negative nonlinear function over time and its role contribute to accelerate or decelerate the adoption process. In practise, the function $a(t)$ allows to modify the shape of the first derivative of $y(t)$ and, simultaneously, it keeps constant the total amount of adoptions

μ as $t \rightarrow \infty$. This function can be useful in modelling possible seasonal patterns within the year, and it is also possible by the use of harmonic functions (Guidolin and Guseo, 2013) to estimate the seasonal frequency and the intensity of the precipitations, e.g. if

$$a(t) = 1 + 2\pi\alpha\beta \cos(2\pi\beta t), \quad (2.3)$$

then the integral is

$$\int_0^t a(z) dz = t + \alpha \sin(2\pi\beta t),$$

where α corresponds to the amplitude of the shock and β indicates how frequent is the seasonal effect in the year. However, the quantity $2\pi|\alpha\beta|$ strongly affects the variations of the instantaneous precipitation rate and better identifies the magnitude of the seasonality. The inverse of the parameter β , i.e. $1/\beta$, denotes the time-length of the shock. When either $\alpha = 0$ or $\beta = 0$, there is no seasonal effect within the year and, therefore, such a model coincides with the standard Bass model in (2.1).

An alternative model can be considered instead of the generalised Bass model, because it seems reasonable to decompose the process in two parts. In this way, the rainfall observed during the autumn can be modelled differently than the rainfall observed during spring. Hence, the seasonal phenomena can be described as a *finite mixture of Bass models* (Guseo, 2004), i.e.

$$\begin{aligned} y(t) = & \mu_1 \zeta_1 \frac{\exp\{t(\zeta_1 + \xi_1)\} - 1}{\zeta_1 \exp\{t(\zeta_1 + \xi_1)\} + \xi_1} + \\ & + \mu_2 \zeta_2 \frac{\exp\{(t - \gamma)(\zeta_2 + \xi_2)\} - 1}{\zeta_2 \exp\{(t - \gamma)(\zeta_2 + \xi_2)\} + \xi_2} \mathbb{1}_{[\gamma, +\infty)}(t), \end{aligned} \quad (2.4)$$

where the parameters μ_1 , ζ_1 and ξ_1 refers to the autumnal part of the process, while μ_2 , ζ_2 and ξ_2 are related to the springtime rainfall, and γ denotes the time origin of the *shifted Bass model* (Guseo, 2004), namely the evolution explained by the second addend in (2.4). However, the reasons to justify the use of this mixture are not consistent with the nature of the rainfall process. As matter of fact, the meaning of the parameters of such a mixture are hardly interpretable, because they are time related (e.g. μ_1 refers to the former part of the period to analyse, while μ_2 to the latter). In addition, the lack of information about the most adequate values to initialise the estimation procedure will generate very unstable estimates, especially when the rainfall time-series are defined for a high number of micro-areas (or location cells).

The inference from observed rainfall data must take into account a proposed theoretical model, e.g. the simplified trajectory in (2.1), and few weak

assumptions on the residual factors that determine the dynamics of the observed process. A possible way is based on the following regression model:

$$r(t) = y(t) + \varepsilon(t),$$

where $r(t)$ is the observed measurement of precipitation in an area at the time t , $y(t)$ denotes a deterministic model describing the evolution of the process, and $\varepsilon(t)$ indicates the residual mixture of stochastic and deterministic effects acting at the time t on the response $r(t)$.

Many non-parametric approaches to inference may be a reasonable way to address the problem. However, it seems more suitable to adopt nonlinear least squares and quantile regression tools, since the nature of $y(t)$ is here based on nonlinear parametric models.

2.4 Conclusion

Standard statistical methods used to model and forecast rainfall data mainly concern with distributive assumptions and inferential paradigms. This perspective mostly focuses on the dependence of each observation on the covariance structure of the process which is usually dependent on spatiotemporal distances among observations. The trend in scientific areas is reasonably based on improving models by introducing more flexible tools and including more explicative variables. However, this approach increases the complexity of the models, which leads to unfeasible estimation and forecast techniques.

In this chapter, the proposed simplification of the dynamical system related to the Lotka–Volterra equations is characterised by a physical interpretation of the innovation diffusion models in a hydrological context. In so doing, other inferential consideration can be achieved and can be linked with other physical processes and measurements. In particular, this new perspective reduces the flexibility of the model and the number of covariates, and hence it leads to a computational improvement by decreasing the model complexity.

The adoption of a non-parametric approach in the subsequent regression analyses requires suitable estimation methods. While the nonlinear least squares are well studied and optimal procedure are well developed, the current quantile regression tools for parametric nonlinear models are not suitably developed and they suffer of a methodological lack in the literature. In the next chapter, nonlinear simultaneous quantile regression techniques are studied and new methods are developed in order to achieve proper results in reasonable time.

Chapter 3

Constrained quantile regression

Historically, the first attempt to regression was done in order to calculate the earth ellipticity (Boscovich, 1779). The physicist, astronomer, mathematician and Jesuit priest Roger J. Boscovich (Ruđer J. Bošković, 1711–1787), who was director of the Brera observatory in Milan and served as chair of mathematics in Pavia, referred for the first time to the *least absolute deviations* procedure, also known as L^1 -norm minimisation. Successively, the Boscovich approach was studied deeper by Laplace (1846) and after by Edgeworth (1888), who argue how the L^1 -norm minimisation may perform better than the *least squares*. Since *quantile regression* has been introduced by Koenker and Bassett Jr. (1978), it becomes a very useful statistical tool which provides a more informative description of the phenomenon under study without making any assumption on its underlying distribution. Usually, the standard regression techniques are based on least square methods, which were first introduced by Legendre (1806, with the name of *méthode des moindres carrés*) and after by Gauss (1809). If the least square methods focus on the conditional expectation of a *random variable* (r.v.) of interest, then the methods based on the minimum sum of absolute deviations are more robust and they focus on the conditional median. In order not to restrict the attention to the median, quantile regression provides a conditional quantile response for each probability level. In particular, this method is based on the minimisation of asymmetric loss functions, where the symmetric L^1 -norm is a particular case. This optimisation problem corresponds to the *asymmetric absolute deviations* and it can be viewed as a *constrained programming* problem. As described in Chen and Wei (2005), the *simplex algorithm* can be used to compute the estimates for data sets with few observations. The *interior point methods* (Mehrotra, 1992) is a faster alternative to the simplex algorithm which was developed to deal with large data sets. It was also proved that interior point methods give better results than the simplex algo-

rithm. In addition, the collinearity in the covariates will not cause a failure during the estimation (Wagner, 1959).

When each quantile level is separately estimated, it is possible to obtain fitted curves which violate the monotonicity property of quantiles. Many works are presented in the literature to provide simultaneous estimation methods in order to solve this problem, which is usually called *crossing quantile curves*. In fact, when each curve is independently estimated, the information related to the position of the other curves is completely ignored. Many estimation techniques were recently proposed to perform quantile regression for each curve by including this further information. In so doing, both the locations of the curves and the inference on the parameters improve. The first attempt to solve such a problem was made by He (1997). His *restricted version of quantile regression*, described in Section 3.1.2, forces some ordering conditions on the quantile curves. The basic idea is founded on a location model shifted by the scale. The proposal of Chernozhukov *et al.* (2010) focuses on a *rearranged estimator*, explained in Section 3.1.3. This method sorts (according to an increasing or decreasing order) the estimates obtained by separate regressions in order to get non crossing quantile curves. Among the methods explained in Section 3.1.4, Wu and Liu (2009) proposed a *step-wise multiple quantile regression* based on an ordered estimation sequence which forces the non crossing constraints on previous estimates. Similarly, *constrained simultaneous quantile regression* of Bondell *et al.* (2010) consists of a procedure to estimate simultaneously all parameters for each quantile curve. This method, however, is limited to the cases wherein the regression models have bounded domains; e.g. it can not be applied for linear models defined on an unbounded domain, because they have constant slope and location shift determined by the quantiles of the error distribution.

The new estimation techniques exposed in Section 3.2, are a particular case of simultaneous quantile regression. In some cases, when there is the necessity to estimate unknown parameters by fixing them for each quantile level, equality constraints must be imposed to obtain more reliable results. This can be due to the properties of the model, e.g. for linear models with unbounded domain.

It is explained in Section 3.3 how to get the test statistics to evaluate the goodness of fit for all the simultaneously estimated levels. Such statistics are directly derived from those presented in the literature for separate quantile levels. New test based on the analysis of the residuals are also obtained by the use of non parametric techniques, such as the *generalised additive models* (Hastie and Tibshirani, 1990). Further methods for the construction of the confidence regions are studied as well.

In order to compare the regression methods and the tools provided for

the tests, two simulation studies are described in Appendix A.4.

3.1 Literature review of estimation methods

Let $Y|X = x$ be a conditional stochastic response variable with *cumulative distribution function* (c.d.f.) $F_{Y|X=x}: \mathbb{R} \rightarrow [0, 1]$, where $x \in \mathcal{X} \subseteq \mathbb{R}^K$ is a deterministic or stochastic covariate K -dimensional vector. A simple model to describe such a response may be based on a position index $\nu(x; \tau)$ of order $\tau \in (0, 1)$, which defines the τ -*quantile function* with reference to $F_{Y|X=x}$. If $Y|X = x$ is continuous the following relationships are satisfied:

$$\tau = \Pr\{Y \leq \nu(x; \tau)|X = x\} = F_{Y|X=x}\{\nu(x; \tau)\},$$

and vice versa the *conditional quantile function* $Q_{Y|X=x}: [0, 1] \rightarrow \mathbb{R}$ is defined as

$$Q_{Y|X=x}(\tau) = F_{Y|X=x}^{-1}(\tau) = \nu(x; \tau).$$

In the discrete case, infinite multiple solutions are available and different methods described by Koenker (2005, Section 8.2) are not considered here.

A τ -quantile arises as the solution of the following decision theory problem:

$$\min_{\nu(x; \tau)} \mathbf{E} [\rho_{\tau}\{(Y|X) - \nu(X; \tau)\}|X = x], \quad (3.1)$$

where the *check function* (Koenker, 2005) $\rho_{\tau}(\cdot)$ is defined as

$$\rho_{\tau}(z) = \begin{cases} z\tau, & \text{if } z \geq 0, \\ z(\tau - 1), & \text{if } z < 0, \end{cases}$$

and the previous expectation is computed as

$$\mathbf{E} [\rho_{\tau}\{(Y|X) - \nu(X; \tau)\}|X = x] = \int_{-\infty}^{+\infty} \rho_{\tau}\{y - \nu(x; \tau)\} dF_{Y|X=x}.$$

Suppose, without loss of generality, that the τ -quantile function does not depend on x , $\nu(x; \tau) = \nu_{\tau}$, so that the minimisation problem in (3.1) can be reformulated as

$$\min_{\nu_{\tau}} (\tau - 1) \int_{-\infty}^{\nu_{\tau}} (y - \nu_{\tau}) dF_Y + \tau \int_{\nu_{\tau}}^{+\infty} (y - \nu_{\tau}) dF_Y.$$

By differentiating with respect to ν_{τ} , the solution of the equation

$$(1 - \tau) \int_{-\infty}^{\nu_{\tau}} dF_Y - \tau \int_{\nu_{\tau}}^{+\infty} dF_Y = 0$$

is identified by the τ -quantile, when $F_Y^{-1}(\tau)$ is unique.

A τ -quantile regression model $g(x; \theta_\tau)$ defines a family of functions, where $\theta_\tau \in \mathbb{R}^r$ is a vector of parameters. Such a regression model may approximate the τ -quantile function $\nu(x; \tau)$ and could not include it as a special case. In other words, the regression model $g(x; \theta_\tau)$ is used to better identify $\nu(x; \tau)$; e.g. each observation in a sample of size n can be decomposed as the sum of two quantities:

$$y_i = g(x_i; \theta_\tau) + \varepsilon_i(\tau),$$

for any $i = 1, \dots, n$, where the error term $\varepsilon(\tau)$ is around the regression model $g(x; \theta_\tau)$, which might be either a linear or a general nonlinear function of θ_τ . In order to achieve optimal results, the error term must satisfy the following properties:

- the τ -quantile of $\varepsilon(\tau)$ must be zero, and
- the expectation $\sum_{i=1}^n \mathbb{E}[\rho_\tau\{\varepsilon_i(\tau)\}]$ is minimum when $g(x_i; \theta_\tau)$ is replaced with $\nu(x_i; \tau)$.

3.1.1 Separate quantile regressions

The procedure of Koenker and Bassett Jr. (1978) refers to the quantile regression for linear models. After the introduction of the interior point optimisation methods (Mehrotra, 1992), a general approach for nonlinear quantile regression was studied by Koenker and Park (1996). The estimator of the conditional quantile function $\hat{Q}_{Y|X=x}(\tau) = g(x; \hat{\theta}_\tau)$ can be expressed as follows:

$$\hat{\theta}_\tau = \arg \min_{\theta_\tau} \sum_{i=1}^n \rho_\tau\{y_i - g(x_i; \theta_\tau)\}. \quad (3.2)$$

A special case might be considered when $\tau = 0.5$, because such a conditional quantile estimator corresponds to the conditional median or to the estimator based on the L^1 -norm.

When a finite set of possible values for τ is available, each regression model is separately fitted through the use of the estimator in (3.2). Even if this procedure is theoretically strong to detect the conditional quantile $\nu(x; \tau)$, the fitted regression models may violate the monotonicity property of quantiles. This problem is known as *quantile crossing curves* and it arises when there exist at least a pair of indexes $i, j = 1, \dots, m$, such that the inequality $g(x; \theta_{\tau_i}) > g(x; \theta_{\tau_j})$ is satisfied in a non empty subset of \mathcal{X} when $\tau_i < \tau_j$. Since the regression models are separately estimated without any further information related to the position of the other models, this problem

is quite common in linear regression and even more frequent in nonlinear settings (He, 1997).

3.1.2 Restricted quantile regression

The idea of *restricted quantile regression* (He, 1997) can be adopted in order to estimate quantile curves by avoiding the occurrence of crossing quantiles. This method is used for estimating the parameters for a specific class of models

$$y = g(x) + s(x)\varepsilon, \quad (3.3)$$

where ε denotes the error term, $g(\cdot)$ and $s(\cdot)$ are two generic functions which represent the conditional location and scale respectively. This method introduces some other assumptions on the error term distribution and on the function controlling the scale. In particular,

- the median of ε must be zero, which is required for the identifiability of the location function $g(\cdot)$;
- the median of $|\varepsilon|$ must be one, so that the scale function $s(\cdot)$ can correctly be identified;
- $s(\cdot)$ must be non negative, since $|\varepsilon| \geq 0$ is always satisfied.

In practice, once $\hat{g}(\cdot)$, which is correspondent to the median curve, and $\hat{s}(\cdot)$, which denotes the median dispersion around $\hat{g}(\cdot)$, are estimated, then a scale-shifting parameter $\gamma_\tau \in \mathbb{R}$ is estimated for each $\tau \in [0, 1]$, in order to place the quantile curve $\hat{g}(\cdot) + \hat{\gamma}_\tau \hat{s}(\cdot)$ in its optimal position. Since $s(\cdot)$ is non negative and γ_τ increases monotonically with respect to τ , the resulting curves do not intersect. More precisely, the algorithm is based on three simple steps (He, 1997):

1. the estimation of the conditional median function is performed to find $\hat{Q}_{Y|X=x}(0.5) = \hat{g}_{0.5}(x)$ and, in so doing, the residuals can be computed as $\hat{r}_i = y_i - \hat{g}_{0.5}(x_i)$, for any $i = 1, \dots, n$;
2. since $\hat{r}_i \approx s(x_i)\varepsilon_i$, the median of $|\hat{r}_i|$ is approximately $s(x_i)$; therefore, the least absolute deviations method is applied to estimate the function $s(\cdot)$, so that the quantity $\sum_{i=1}^n \rho_{0.5}\{|\hat{r}_i| - s(x_i)\}$ is minimised;
3. the computation of the quantity γ_τ is followed in order to minimise the sum of asymmetric loss functions computed for each probability level, i.e. $\hat{\gamma}_\tau = \arg \min_{\gamma_\tau} \sum_{i=1}^n \rho_\tau\{\hat{r}_i - \gamma_\tau \hat{s}_{0.5}(x_i)\}$, where $\hat{s}_{0.5}(x_i)$ is the median scale function estimated at the previous step. After this, the conditional quantile is computed as $\hat{Q}_{Y|X=x}(\tau) = \hat{g}_{0.5}(x) + \hat{\gamma}_\tau \hat{s}_{0.5}(x)$.

He (1997) discussed the main properties of results obtained through this procedure. In particular, he stated that the conditional quantile estimator is consistent. However, this method can not be used in those cases wherein the data do not satisfy the location and scale paradigm as in (3.3).

3.1.3 Rearrangement quantile estimator

The estimator proposed by Chernozhukov *et al.* (2010) is based on the application of the *rearrangement operator* introduced by Hardy *et al.* (1934). Such a method takes a particular permutation of the original estimates of a non monotone curve, namely it sorts them so that the resulting curve is monotone. To be more precise, the rearrangement operator is defined on a function $f: \mathcal{T} \rightarrow \mathbb{R}$, where the space \mathcal{T} is bounded and it can be replaced by the interval $[0, 1]$ without any loss of generality, i.e.

$$f^*(\tau) = \inf \left\{ y \in \mathbb{R} : \int_0^1 \mathbb{1}_{(-\infty, f(u)]}(y) du \geq \tau \right\}, \quad (3.4)$$

where the *indicator function* is defined as

$$\mathbb{1}_{\mathcal{A}}(z) = \begin{cases} 1, & \text{if } z \in \mathcal{A}, \\ 0, & \text{otherwise.} \end{cases}$$

This means that the rearranged function $f^*(\tau)$ corresponds to the minimum value of y , such that τ is lower or equal than the total coverage of the subspaces of \mathcal{T} , wherein the function $f(\cdot)$ is lower than y . By extending this concept, it is possible to get the multivariate version of the rearrangement operator. In fact, if \mathcal{T} is a bounded multidimensional space, e.g. $[0, 1]^d$ with $d > 1$, the rearrangement operator takes the following form:

$$f^*(\mathbf{t}) = \inf \left\{ y \in \mathbb{R} : \int_0^1 \cdots \int_0^1 \mathbb{1}_{(-\infty, f(u_1, \dots, u_d)]}(y) du_1 \dots du_d \geq q(\mathbf{t}) \right\},$$

where $\mathbf{t} = (t_1, \dots, t_d)^\top$ and the quantity $q(\mathbf{t})$ is chosen such that

$$\int_{[0,1]^d} \mathbb{1}_{(-\infty, f(\mathbf{t})]}(y) d\mathbf{t} = \int_{[0,1]^d} \mathbb{1}_{(-\infty, f^*(\mathbf{t})]}(y) d\mathbf{t}$$

and it can be approximated as

$$q(\mathbf{t}) \approx \frac{1}{2} \prod_{i=1}^d t_i + \frac{1}{2} \left\{ 1 - \prod_{i=1}^d (1 - t_i) \right\}.$$

The former part of the summation in the approximation above refers to the coverage of the subspace wherein the rearranged function is surely less than $f^*(\mathbf{t})$. The latter part denotes the coverage of the subspace which excludes that part wherein the rearranged function is surely greater than $f^*(\mathbf{t})$.

The definition of the rearrangement operator in (3.4) is sufficient to obtain a monotonic quantile estimator. In practice, separate independent estimation of the conditional quantile function $Q_{Y|X=x}(\tau)$ in each point $\tau \in (0, 1)$ may provide curves which violate the monotonicity property. In these cases, since the conditional quantile function is estimated by points, then its empirical estimate $\hat{Q}_{Y|X=x}(\tau)$ can be transformed in $\hat{Q}_{Y|X=x}^*(\tau)$ through rearrangement as in (3.4).

Let $\hat{F}_{Y|X=x}(y)$ be the empirical conditional c.d.f., which can be defined through the *empirical conditional quantile function* $\hat{Q}_{Y|X=x}(\tau)$ as

$$\begin{aligned} \hat{F}_{Y|X=x}(y) &= \int_0^1 \mathbb{1}_{(-\infty, y]} \left\{ \hat{Q}_{Y|X=x}(\tau) \right\} d\tau, \text{ or equivalently as} \\ &= \int_0^1 \mathbb{1}_{[\hat{Q}_{Y|X=x}(\tau), +\infty)}(y) d\tau. \end{aligned} \quad (3.5)$$

In other words, the empirical c.d.f. can be viewed as the total length of the portions of domain wherein the empirical conditional quantile function is lower or equal than a threshold y . Since the empirical c.d.f. $\hat{F}_{Y|X=x}(y)$ is computed as in (3.5), it is always monotonically increasing for any quantile function used in the estimation procedure. This is the motivation which allows to define the *rearranged quantile estimator* as

$$\hat{Q}_{Y|X=x}^*(\tau) = \hat{F}_{Y|X=x}^{-1}(\tau) = \inf \left\{ y : \hat{F}_{Y|X=x}(y) \geq \tau \right\}, \quad (3.6)$$

which is monotone in τ (Chernozhukov *et al.*, 2009).

Chernozhukov *et al.* (2010) proved that this estimator has smaller estimation error than the original non rearranged. This implies that the rearranged conditional quantile function is closer to the true quantile curve $Q_{Y|X=x}(\tau)$, i.e. for any $p \in [1, +\infty)$ the following inequality is always satisfied:

$$\left\| \hat{Q}_{Y|X=x}^*(\tau) - Q_{Y|X=x}(\tau) \right\|_p \leq \left\| \hat{Q}_{Y|X=x}(\tau) - Q_{Y|X=x}(\tau) \right\|_p,$$

where $\| \cdot \|_p$ represents the L^p -norm of a measurable function.

They also proved that the coverage probability of the confidence interval for the rearrangement is greater than a given level $\alpha \in (0, 1)$; in fact, the confidence interval may cover non monotonic curves. By considering only monotonic function, the coverage interval reduces its length by maintaining

the same coverage probability. The proposal of Chernozhukov *et al.* (2009) is based on the application of the rearrangement technique to the confidence bands, in order to construct narrow confidence intervals covering only monotonic curves.

3.1.4 Multi-objective quantile estimators

Multi-objective optimisation techniques are the founding methods for a simultaneous quantile regression imposing non crossing constraints. In practice, instead of using the estimator in (3.2) to compute the estimates for each value of τ , these are obtained by optimising simultaneously a multi-objective function $\mathbf{h}(\theta_{\tau_1}, \dots, \theta_{\tau_m}): \mathbb{R}^{rm} \rightarrow \mathbb{R}_+^m$, where m is the number of quantile levels $\tau_1 < \tau_2 < \dots < \tau_m$ that must be computed.

The general multi-objective minimisation problem is designed to find an optimal point which is a compromise with respect to the result attained in the *utopia point*, wherein each loss function achieves its minimum. Since a unique global solution does not exist, the Pareto optimality is the main concept to define an optimal point (Pareto, 1906). Originally, this idea was expressed in terms of “*ophelimity maximum*” point, in which the maximum global benefit is attained when all the loss functions are close to their minimum, but some of them will improve while others will get worse results for any little change from that point. In other words, a Pareto optimal point achieves an equilibrium of benefits for each objective function. A complete and comprehensive review about multi-objective optimisation methods can be found in Marler and Arora (2004).

Simultaneous quantile regression minimises at the same time the sum of the asymmetric absolute deviations

$$h_j(\theta_{\tau_j}) = \sum_{i=1}^n \rho_{\tau_j} \{y_i - g(x_i; \theta_{\tau_j})\},$$

for each $j = 1, \dots, m$, subject to the non crossing constraints

$$g(x; \theta_{\tau_k}) < g(x; \theta_{\tau_{k+1}}), \quad (3.7)$$

which must be satisfied for $k = 1, \dots, m - 1$ and for all $x \in \mathcal{X}$.

According to Marler and Arora (2004), there exists three classes of possible methods to solve a general multi-objective optimisation problem. The distinction of these algorithms is based on the articulation of preferences, namely the criteria to combine all the functions to optimise into a unique one. Such preferences can be defined either *a priori* or *a posteriori*. If no preference is defined more objective procedures can be considered.

The quantile regression method proposed by Bondell *et al.* (2010) and the stepwise techniques developed by Wu and Liu (2009) enter in this class of multi-objective quantile estimators. They are both based on prior preferences, but the former method can be viewed as a procedure without preferences.

Many criteria can be chosen for combining the objective functions into a unique global one. Such a single global function can be viewed as a loss or utility function, which is optimised to reach a good compromise between the final solution and the utopia point $\theta_{\tau_j}^{\circ}$. In this context, by definition, the utopia point $\theta_{\tau_j}^{\circ}$ always satisfies the equality $h_j(\theta_{\tau_j}^{\circ}) = 0$, for any $j = 1, \dots, m$. The most common choices are presented in the following.

The weighted global criteria (Zeleny, 1973) combine the objective functions through a weighted exponential sum, so that the global loss can be described by the function $\ell(\theta)$, which is expressed as

$$\ell(\theta) = \sum_{j=1}^m h_j(\theta_{\tau_j})^p w_j,$$

or alternatively as

$$\ell(\theta) = \sum_{j=1}^m \{h_j(\theta_{\tau_j})w_j\}^p, \quad (3.8)$$

where $\theta = (\theta_{\tau_1}^{\top}, \dots, \theta_{\tau_m}^{\top})^{\top}$. The weights w_j may be all equal to $1/m$, if it is considered a setting without preferences. Alternatively, they may be defined by prior preferences such that $0 \leq w_j \leq 1$, for all $j = 1, \dots, m$, and $\sum_{j=1}^m w_j = 1$. The particular case of the *weighted sum criteria* refers to the previous loss function computed for $p = 1$.

The constrained version of quantile regression is a *simultaneous multiple quantile estimation* procedure based on this multi-objective optimisation technique by imposing the constraints in (3.7). Bondell *et al.* (2010) focuses on both linear and nonparametric models defined on a bounded domain. Their estimator is based on the minimisation of the loss function (3.8) with $p = 1$ and subject to the non crossing constraints. They restricted to the case in which $g(x; \theta_{\tau})$ is linear or, more generally, a linear combination of basis functions. The same idea can be extended with some difficulties also to get the estimates for nonlinear parametric models.

The lexicographic method (Fishburn, 1974) sorts the objective functions in order of importance and then they are optimised one at a time.

Of course, each function $h_j(\cdot)$ has to be univocally related to another $h_{(j)}(\cdot)$, where the index (j) denotes the position in the optimisation sequence. In order to guaranty the Pareto optimality, the optimisation is subject to further constraints when a variation of $\theta_{\tau_{(j)}}$ implies a change of value on $\theta_{\tau_{(k)}}$, for any $j > k$, with $j = 2, \dots, m$ and $k = 1, \dots, j - 1$. In practice, the minimisation of $h_{(j)}(\theta_{\tau_{(j)}})$ is subject to $h_{(k)}(\theta_{\tau_{(k)}}) \leq (1 + \delta_k) h_{(k)}(\theta_{\tau_{(k)}}^*)$, where $\theta_{\tau_{(k)}}^*$ is the minimum point of the function $h_{(k)}(\cdot)$ and $\delta_k \in [0, 1]$ represents a constant of constraint relaxation. Of course, $h_{(1)}(\cdot)$ is optimised without any constraint.

The *stepwise multiple quantile regression* (Wu and Liu, 2009) is very similar to this method. In fact, instead of estimating multiple quantile regression models at the same time, estimates are computed sequentially for each quantile level by imposing the non crossing constraints on the previous one. In practice, the most central curve is the first to be estimated. The other curves are sequentially calculated by imposing the constraints in (3.7) from the middle quantile level to the highest and, conversely, from the middle to the lowest.

The weighted min-max method is also called the *weighted Chebychev method* (Osyczka, 1978). It is a special case of (3.8) as $p \rightarrow \infty$, so that the loss function is computed as

$$\ell(\theta) = \max_j \{h_j(\theta_{\tau_j})w_j\}, \quad \forall j = 1, \dots, m.$$

When the minimisation of this loss function coincide with the minimisation of $h_j(\cdot)$ for a unique $j = 1, \dots, m$, then it is preferred to optimise the following function:

$$\ell(\theta) = \max_j \{|h_j(\theta_{\tau_j}) - h_j^\circ|w_j\}, \quad \forall j = 1, \dots, m,$$

where $h_j^\circ = \min_{\theta_{\tau_j}} h_j(\theta_{\tau_j})$ for any $j = 1, \dots, m$, namely the minimum value of $h_j(\cdot)$ obtained through non simultaneous optimisations. Since h_j° must be attained for all $j = 1, \dots, m$ before the global optimisation, the last loss function refers to the *a posteriori* articulation of preferences. This is done to improve the identifiability of all the parameters to estimate. Generally speaking, these methods do not guaranty the existence of a Pareto optimal point; however, if the loss function $\ell(\theta)$ admits a unique minimum, it is a Pareto optimal point.

The exponential weighted criteria were initially designed as an alternative to the weighted sum method in order to capture points on non

convex optimal surfaces (Athanasopoulos and Papalambros, 1996). They supply a loss function which is formulated as

$$\ell(\theta) = \sum_{j=1}^m \{\exp(p w_j) - 1\} \exp \{p h_j(\theta_{\tau_j})\},$$

where large values of p can lead to numerical instabilities during the minimisation phase.

The weighted product method (first introduced as *product of powers* by Bridgman, 1922) is mostly adopted when objective functions have different orders of magnitude. In the optimisation process the objective functions will have similar significance without applying any transformation. In this case, the loss function is expressed as

$$\ell(\theta) = \prod_{j=1}^m h_j(\theta_{\tau_j})^{w_j},$$

where the weights w_j indicates the relative significance of the function $h_j(\cdot)$. However, this method can introduce extra non linearity in the form of the loss function, which can potentially be cumbersome to solve from a computational viewpoint.

3.2 New estimation procedures with equality constraints

In many cases, some parameters may be considered constant for each quantile level. This means that the effects of these fixed parameters do not vary with τ , and thus, the changes across quantiles are imputable only to those parameters which vary. By considering the *random coefficients interpretation* of generic regression model $g(x; \theta_\tau)$, the vector of parameters θ_τ might be considered as the realisation of a multivariate stochastic process (see Koenker, 2005, section 2.6). From this point of view, if a parameter follows a Dirac distribution, it should be considered fixed across quantile levels during the estimation procedure. A particular case is given by the estimation of the parameters of a linear model defined on an one-dimensional unbounded domain, in fact only parallel lines are the unique solution for the non crossing curves problem (Koenker, 1984).

By considering the multi-objective optimisation techniques exposed in Section 3.1.4, new estimators can be proposed in order to obtain proper

estimates by imposing equality constraints on some parameters. Let $\theta = (\varphi^\top, \psi^\top)^\top$ be split in two vectors, where $\varphi = (\varphi_{\tau_1}^\top, \dots, \varphi_{\tau_m}^\top)^\top$, which denotes the vectors of *quantile varying coefficients*, and ψ , which represents the vector of *quantile fixed parameters*. This means that the τ -quantile regression model $g(x; \theta_\tau)$ can be written as $g(x; \varphi_\tau, \psi_\tau)$, where the equality $\psi_\tau = \psi$ must be satisfied for any $\tau \in (0, 1)$.

Since the lexicographic method, the exponential weighted criteria and the weighted product minimisation are computationally inefficient on both time demanding and numerical instabilities, the methods presented below are based mainly on the weighted sum and weighted min-max approach.

The parameters can be estimated by minimising a loss function computed through the weighted sum method, so that the resulting estimator is

$$\hat{\theta}^{(ws)} = \arg \min_{\varphi, \psi} \sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \{y_i - g(x_i; \varphi_{\tau_j}, \psi)\}. \quad (3.9)$$

In order to get better results, the non crossing constraints in (3.7) must be imposed as

$$g(x; \varphi_{\tau_k}, \psi) < g(x; \varphi_{\tau_{k+1}}, \psi),$$

for $k = 1, \dots, m - 1$ and for all $x \in \mathbb{R}$. On the other hand, the estimator based on the weighted min-max can be expressed as

$$\hat{\theta}^{(wm)} = \arg \min_{\varphi, \psi} \left(\max_j \left[\sum_{i=1}^n w_j \rho_{\tau_j} \{y_i - g(x_i; \varphi_{\tau_j}, \psi)\} \right] \right), \quad (3.10)$$

where $j \in \{1, \dots, m\}$. Also here the regression may be subject to the usual non crossing constraints as above.

Wagner (1959) pointed out that the usual linear programming techniques can be adopted for solving regression problems; in fact, the original minimisation problem is transformed into an equivalent. He studied both the minimisation of the sum of absolute deviations and the minimisation of the maximum absolute deviation. In practice, when the regression model is linear, the same optimisation technique can be applied to compute the estimates for (3.9), and the problems must easily be adjusted to be processed (see Section 3.2.1). Generally speaking, both proposed estimators refer to a simultaneous minimisation of all the objective functions computed for each quantile level; however, because of the nature of the parameters, it is possible to minimise the proposed loss functions by iterating two steps until convergence. In practice, the minimum of the global loss function is obtained when all the functions are as close as possible to their minimum. Especially for the

Algorithm 3.2.1 Quantile estimation with equality constraints

- 1: Select the quantile levels $\tau_j \in (0, 1)$, for $j = 1, \dots, m$.
- 2: Select an estimator between either (3.9) or (3.10).
- 3: Initialise all the parameter with a guess.
- 4: **repeat**
- 5: Minimise the loss function for φ by maintaining fixed ψ .
- 6: Minimise the loss function for ψ by maintaining fixed φ .
- 7: **until** The parameter estimates converge.

estimator in (3.10), it is possible to apply easily a two-step tuning procedure, which is performed as described in the algorithm 3.2.1. More details on the estimation by using (3.9) can be found in Section 3.2.1 and 3.2.2.

A noteworthy case can simplify computations, when only one parameter is allowed to vary according to the quantile levels. When the model or previous inferential results allow this setting, it is possible to avoid the non crossing constraints in (3.7). Reasonably, if all parameters are fixed across quantile levels, all quantile curves collapse into a single one. Nonetheless, by letting only one proper parameter vary with τ , it adjusts the location of the quantile curves automatically without imposing the constraints in (3.7). Of course, this is true if the resulting curves are unique for the selected values of τ . In fact, if there exist more optimal points of global minimum, the monotonicity of the estimated conditional quantile function is not guaranteed anymore. Only if the solution exists and it is unique, the algorithm 3.2.1 may be modified by using either the estimator in (3.9) or (3.10) to compute the quantile fixed parameters, while the quantile varying parameters are obtained with the estimator in (3.2).

A special case of the estimator in (3.9) was considered by Hogg (1975), wherein the function $g(\cdot)$ has the form of the classical linear location-shift regression model, where only the intercept is allowed to vary with τ . The regression method described by Koenker (2005, Section 6.8) does not consider the non crossing constraints.

From a computational point of view, the estimation technique with non crossing-constraints can be simplified by the adoption of an estimator, which leads to a penalised quantile regression. It can generally be defined as follows

$$\hat{\theta}^{(ps)} = \arg \min_{\varphi, \psi} \sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \{y_i - g(x_i; \varphi_{\tau_j}, \psi)\} + \kappa \sum_{k=1}^{m-1} P_{\tau_{k+1}, \tau_k}(g), \quad (3.11)$$

where κ is a constant controlling the trade-off between the deviance and the

penalty, and $P_{\tau_{k+1}, \tau_k}(g)$ is a penalty which depends on the condition imposed by the constraints in (3.7). Such a penalty can be defined as

$$P_{\tau_{k+1}, \tau_k}(g) = \frac{1}{2} \int_{\mathcal{X}} |g(x; \varphi_{\tau_{k+1}}, \psi) - g(x; \varphi_{\tau_k}, \psi)| dx - \frac{1}{2} \int_{\mathcal{X}} g(x; \varphi_{\tau_{k+1}}, \psi) - g(x; \varphi_{\tau_k}, \psi) dx,$$

and it represents the area between two consecutive quantile curves wherein the crossing constraints are not satisfied. However, such a formulation is not worthy for the penalties involving regression models with multiple covariates. For this reason, any suitable approximation can be considered, e.g.

$$P_{\tau_{k+1}, \tau_k}(g) = \frac{1}{2} \sum_{i=1}^n |g(x_i; \varphi_{\tau_{k+1}}, \psi) - g(x_i; \varphi_{\tau_k}, \psi)| - \frac{1}{2} \sum_{i=1}^n \{g(x_i; \varphi_{\tau_{k+1}}, \psi) - g(x_i; \varphi_{\tau_k}, \psi)\}.$$

In these cases, an unconstrained optimisation is performed through the estimator (3.11) and the final estimates are consistent with those obtained by a proper constrained optimisation based on the estimator (3.9).

Another proposal to avoid the imposition of non crossing constraints when two or more parameters may vary with τ . The idea comes from the rearrangement estimator (Chernozhukov *et al.*, 2010) described in Section 3.1.3. In practice, quantile varying coefficients can be view as a function of τ , i.e. $\varphi^\bullet: (0, 1) \rightarrow \mathbb{R}^q$, where q is the number of free parameters. In so doing, instead of applying the estimator in (3.6) for each x point, each function $\varphi_i^\bullet(\tau)$ will be rearranged into a monotonic function $\varphi_i^*(\tau)$, for $i = 1, \dots, q$. In other words, the estimates of φ are sorted by increasing order; namely, this procedure can be adopted only when the theoretical model satisfies the following condition:

$$g(x; \varphi_{\tau_k}, \psi) > g(x; \varphi_{\tau_{k+1}}, \psi) \iff \varphi_{\tau_k} > \varphi_{\tau_{k+1}}, \quad \forall k = 1, \dots, m-1.$$

When the chosen model does not satisfy such a condition, then it is still possible to obtain proper estimates by adopting a semi-parametric variation of the so-called *quantile sheet* introduced by Schnabel and Eilers (2013). In practice, functional parameters $\varphi^\bullet(\tau)$ are modelled for any value of $\tau \in [0, 1]$ and not only for some and fixed. In this case the conditional c.d.f. of the response variable can be described by a “*hyper-surface*”, which depends on both τ and the covariates. Any non-parametric model can be chosen for

the vector-valued function $\varphi^\bullet(\tau)$. This allows to introduce a certain degree of flexibility in the parametric nonlinear model. Among the non-parametric model for the parameters, the *Bernstein polynomials* (Bernstein, 1912) seem to be reasonable because of the easy interpretation of their coefficients. In so doing, the vector valued function $\varphi^\bullet(\tau)$ is defined by

$$\varphi^\bullet(\tau) = \sum_{j=0}^m \binom{m}{j} \phi_j \tau^j (1-\tau)^{m-j},$$

where $\phi_j \in \mathbb{R}^q$ is the j -th vector of coefficients to be estimated and it should approximate the true values of the function φ^\bullet evaluated in j/m . The idea behind the *semi-parametric quantile sheets* is based on the fact that the parameters of very close quantile curves have non significant variations; hence the difference $\phi_{j+1} - \phi_j$ tends to zero as $m \rightarrow \infty$. The following proposed estimator is based on a penalised quantile regression, which involves difference penalties P , i.e.

$$\hat{\theta}^{(qs)} = \arg \min_{\phi_1, \dots, \phi_m, \psi} \int_0^1 w(\tau) \sum_{i=1}^n \rho_\tau \left[y_i - g\{x_i; \varphi^\bullet(\tau), \psi\} \right] d\tau + P, \quad (3.12)$$

wherein $w(\tau)$ is a non negative weighting function for all $\tau \in [0, 1]$ and it satisfies the following property:

$$\int_0^1 w(\tau) d\tau = 1,$$

but it is usually considered constant, namely $w(\tau) = 1$. The penalty P is defined by

$$P = \kappa \|\Delta_d \Phi\|_F^2,$$

where $\|\cdot\|_F$ denotes the *Frobenius norm*, i.e. the square root of the sum of squares of all the components of a matrix (see Bernstein, 2009, Section 9.2). The matrix Φ has dimension $q \times m$ and its j -th row corresponds to the vector ϕ_j , while Δ_d is a $(q-d) \times q$ matrix which forms the differences of order d (Eilers and Marx, 1996), e.g.

$$\begin{aligned} \Delta_0 \Phi &= \Phi, \\ \Delta_1 \Phi &= \{\phi_j - \phi_{j+1}\} \quad \forall j = 1, \dots, q-1, \\ \Delta_2 \Phi &= \{\phi_j - 2\phi_{j+1} + \phi_{j+2}\} \quad \forall j = 1, \dots, q-2, \end{aligned}$$

and so on. Usually, d is greater than zero and it is often one or two. The integral involved in the estimator (3.12) is not analytically computable, therefore it must be approximated by numerical methods, such as the *trapezoidal quadrature* (Cruz-Urbe *et al.*, 2002).

3.2.1 Simultaneous linear quantile regression

A non crossing constrained quantile regression of a linear model is performed by combining the idea of the Hogg's estimator (1975) and the algorithm proposed by Koenker and Ng (2005) for inequality constrained quantile regression, as suggested in the paper of Bondell *et al.* (2010). Of course, any generalisation of this solution for other fixed and quantile varying parameters is always possible for the linear case (especially if a bounded domain is considered). The optimisation problem related to the estimator (3.9) can be expressed as a linear programming problem, when the function $g(\cdot)$ is linear and the non crossing constraints are imposed. Equivalently, the problem can be reformulated as

$$\min \sum_{j=1}^m w_j \{ \tau_j \mathbf{1}_n^\top \mathbf{u}_j + (1 - \tau_j) \mathbf{1}_n^\top \mathbf{v}_j \}, \quad (3.13)$$

subject to

$$\begin{bmatrix} \mathbf{X} & \mathbf{I}_{mn} & -\mathbf{I}_{mn} \\ -\mathbf{X} & -\mathbf{I}_{mn} & \mathbf{I}_{mn} \\ \mathbf{R} & \mathbf{O} & \mathbf{O} \end{bmatrix} \begin{bmatrix} \beta \\ \mathbf{u} \\ \mathbf{v} \end{bmatrix} \geq \begin{bmatrix} \mathbf{1}_{2m} \otimes \mathbf{y} \\ \mathbf{0}_{m-1} \end{bmatrix},$$

where $\mathbf{1}_n$ denotes an n -th dimensional vector of ones, $\mathbf{u} = (\mathbf{u}_1^\top, \dots, \mathbf{u}_m^\top)^\top$ and $\mathbf{v} = (\mathbf{v}_1^\top, \dots, \mathbf{v}_m^\top)^\top$, wherein \mathbf{u}_j and \mathbf{v}_j respectively represent the positive and negative parts of residuals of the correspondent τ_j , for $j = 1, \dots, m$. The design matrix \mathbf{X} is an $mn \times (m + 1)$ matrix defined as

$$\mathbf{X} = [\mathbf{1}_m \otimes \mathbf{x} \quad \mathbf{L} \otimes \mathbf{1}_n],$$

wherein the notation \otimes denotes the Kroneker product, $\mathbf{x} = (x_1, \dots, x_n)^\top$ and the lower triangular matrix

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

of dimension $m \times m$. The identity matrix \mathbf{I}_{mn} is a $mn \times mn$ diagonal matrix having ones as diagonal entries. The inequality constraints matrix \mathbf{R} is an $(m - 1) \times (m + 1)$ matrix defined as

$$\mathbf{R} = [\mathbf{0}_{m-1} \quad \mathbf{0}_{m-1} \quad \mathbf{I}_{m-1}],$$

wherein $\mathbf{0}_{m-1}$ represents a $(m - 1)$ -th dimensional vector of zeros. The matrix \mathbf{O} denotes an $(m - 1) \times mn$ matrix of zeros. The vector $\mathbf{y} = (y_1, \dots, y_n)^\top$ is

the vector of responses and the vector of parameters is such that

$$\beta = \left(\beta_1, \beta_{0\tau_1}, \beta_{0\tau_2} - \beta_{0\tau_1}, \dots, \beta_{0\tau_m} - \sum_{j=1}^{m-1} \beta_{0\tau_j} \right)^\top. \quad (3.14)$$

The problem in (3.13) is not written in a canonical form, therefore the dual problem can be expressed as

$$\max_{\mathbf{d}} \mathbf{d}^\top \begin{bmatrix} \mathbf{1}_m \otimes \mathbf{y} \\ \mathbf{0}_{m-1} \end{bmatrix}, \quad (3.15)$$

subject to

$$\begin{bmatrix} \mathbf{X}^\top & \mathbf{R}^\top \\ -\mathbf{X}^\top & -\mathbf{R}^\top \\ \mathbf{I}_{mn} & \mathbf{O}^\top \\ -\mathbf{I}_{mn} & \mathbf{O}^\top \\ \mathbf{O} & \mathbf{I}_{m-1} \end{bmatrix} \mathbf{d} \geq \begin{bmatrix} \mathbf{0}_{2mn} \\ \{\mathbf{w} \odot (\boldsymbol{\tau} - \mathbf{1}_m)\} \otimes \mathbf{1}_n \\ \{\mathbf{w} \odot (-\boldsymbol{\tau})\} \otimes \mathbf{1}_n \\ \mathbf{0}_{m-1} \end{bmatrix},$$

where the symbol \odot denotes the component-wise multiplication between two vectors or two matrices of the same dimension, the vector of quantile levels $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)^\top$ and the vector of weights $\mathbf{w} = (w_1, \dots, w_m)^\top$.

Even if the problem in (3.15) can be solved by standard linear programming algorithms, it is convenient to make a further adjustment. Let \mathbf{a} be a vector such that

$$\mathbf{a} = \mathbf{d} - \begin{bmatrix} \{\mathbf{w} \odot (\boldsymbol{\tau} - \mathbf{1}_m)\} \otimes \mathbf{1}_n \\ \mathbf{0}_{m-1} \end{bmatrix},$$

then the linear programming problem in (3.15) can be formulated in a canonical form as

$$\max_{\mathbf{a}} \mathbf{a}^\top \begin{bmatrix} \mathbf{1}_m \otimes \mathbf{y} \\ \mathbf{0}_{m-1} \end{bmatrix}, \quad (3.16)$$

subject to

$$\begin{bmatrix} \mathbf{X}^\top & \mathbf{R}^\top \\ -\mathbf{X}^\top & -\mathbf{R}^\top \\ \mathbf{I}_{mn} & \mathbf{O}^\top \\ -\mathbf{I}_{mn} & \mathbf{O}^\top \\ \mathbf{O} & \mathbf{I}_{m-1} \end{bmatrix} \mathbf{a} \geq \begin{bmatrix} \mathbf{X}^\top \{\mathbf{w} \odot (\boldsymbol{\tau} - \mathbf{1}_m)\} \otimes \mathbf{1}_n \\ -\mathbf{X}^\top \{\mathbf{w} \odot (\boldsymbol{\tau} - \mathbf{1}_m)\} \otimes \mathbf{1}_n \\ \mathbf{0}_{mn} \\ -\mathbf{1}_{mn} \\ \mathbf{0}_{m-1} \end{bmatrix}.$$

To get the estimates of the parameters of β as defined in (3.14), Gutenbrunner and Jurečková (1992) suggested an approach based on the regression rank score process. In practice, the first mn optimal values of the vector \mathbf{a} are used to compute the diagonal entries of the diagonal matrix \mathbf{A} , so that

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{A} (\mathbf{1}_m \otimes \mathbf{y}).$$

If the matrix $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ is singular, the usual least square estimator is adopted, i.e.

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{1}_m \otimes \mathbf{y});$$

however, as Gutenbrunner and Jurečková (1992) pointed out, the probability that $\mathbf{X}^\top \mathbf{A} \mathbf{X}$ is singular approaches to zero as $n \rightarrow \infty$.

3.2.2 Simultaneous nonlinear quantile regression

A non crossing constrained quantile regression of nonlinear models is performed by extending the method of Koenker and Park (1996). Such a method is based on numerical optimisation techniques in order to achieve suitable results for the estimator in (3.2). The technique can be summarised as an iterative search of the steepest descent direction and of the optimal step length. Here, in order to achieve a computational compromise, the proposed extension of this method can only be applied for a nonlinear constrained quantile regression without regarding the non crossing constraints. Even if the generalisation of Koenker and Park's method can not be achieved for any non crossing quantiles problem, it is always possible to study an *ad hoc* algorithm for each nonlinear model as a particular case by itself.

In order to compute the estimates by adopting the estimator in (3.9) for a generic nonlinear function $g(x; \varphi, \psi)$, it is necessary to consider the homogeneity property of the check function $\rho_\tau(\cdot)$:

Property 1. *Given a quantile level $\tau \in (0, 1)$ and a scalar $w \in \mathbb{R}_+$, the equation $\rho_\tau(wz) = w\rho_\tau(z)$ is always satisfied for any $z \in \mathbb{R}$, i.e. $\rho_\tau(\cdot)$ is a homogeneous function. (The proof is reported in Appendix A.1)*

El-Attar *et al.* (1979) stated the existence condition of the optimum values of the estimates. In fact, the solution is optimal if there exists a vector

$$\mathbf{d} \in [\tau_1 - 1, \tau_1]^n \times \dots \times [\tau_m - 1, \tau_m]^n, \quad (3.17)$$

such that

$$\mathbf{J}^\top \mathbf{d} = \mathbf{0} \quad (3.18)$$

and

$$\mathbf{g}_\theta^\top \mathbf{d} = \sum_{j=1}^m \sum_{i=1}^n \rho_{\tau_j} \left\{ w_j y_i - w_j g(x_i; \hat{\varphi}_{\tau_j}, \hat{\psi}) \right\}, \quad (3.19)$$

where the matrix $\mathbf{J} = \{\text{diag}(\mathbf{w}) \otimes \mathbf{1}_n\} \odot [\mathbf{G}, \Psi]$, with

$$\mathbf{G} = \begin{bmatrix} \left. \frac{\partial g(x_1; \varphi, \hat{\psi})}{\partial \varphi} \right|_{\varphi=\hat{\varphi}_{\tau_1}} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \left. \frac{\partial g(x_n; \varphi, \hat{\psi})}{\partial \varphi} \right|_{\varphi=\hat{\varphi}_{\tau_1}} & 0 & \dots & 0 \\ 0 & \left. \frac{\partial g(x_1; \varphi, \hat{\psi})}{\partial \varphi} \right|_{\varphi=\hat{\varphi}_{\tau_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \left. \frac{\partial g(x_n; \varphi, \hat{\psi})}{\partial \varphi} \right|_{\varphi=\hat{\varphi}_{\tau_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \left. \frac{\partial g(x_1; \varphi, \hat{\psi})}{\partial \varphi} \right|_{\varphi=\hat{\varphi}_{\tau_m}} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \left. \frac{\partial g(x_n; \varphi, \hat{\psi})}{\partial \varphi} \right|_{\varphi=\hat{\varphi}_{\tau_m}} \end{bmatrix},$$

and

$$\Psi = \begin{bmatrix} \left. \frac{\partial g(x_1; \hat{\varphi}_{\tau_1}, \psi)}{\partial \psi} \right|_{\psi=\hat{\psi}} \\ \vdots \\ \left. \frac{\partial g(x_n; \hat{\varphi}_{\tau_1}, \psi)}{\partial \psi} \right|_{\psi=\hat{\psi}} \\ \vdots \\ \left. \frac{\partial g(x_1; \hat{\varphi}_{\tau_m}, \psi)}{\partial \psi} \right|_{\psi=\hat{\psi}} \\ \vdots \\ \left. \frac{\partial g(x_n; \hat{\varphi}_{\tau_m}, \psi)}{\partial \psi} \right|_{\psi=\hat{\psi}} \end{bmatrix},$$

and the vector

$$\mathbf{g}_{\hat{\theta}} = (\mathbf{w} \otimes \mathbf{y}) - (\mathbf{w} \otimes \mathbf{1}_n) \odot \begin{bmatrix} g(x_1; \hat{\varphi}_{\tau_1}, \hat{\psi}) \\ \vdots \\ g(x_n; \hat{\varphi}_{\tau_1}, \hat{\psi}) \\ \vdots \\ g(x_1; \hat{\varphi}_{\tau_m}, \hat{\psi}) \\ \vdots \\ g(x_n; \hat{\varphi}_{\tau_m}, \hat{\psi}) \end{bmatrix}.$$

Since the approach adopted here is based on a steepest descent technique proposed by Osborne and Watson (1971), the estimation procedure becomes a sequence of the following minimisation problems:

$$\min_{\delta} \sum_{j=1}^m \sum_{i=1}^n \rho_{\tau_j} \left\{ w_j y_i - w_j g(x_i; \hat{\varphi}_{\tau_j}, \hat{\psi}) - \delta^{\top} \nabla_{i,j} \right\}$$

wherein the vector $\nabla_{i,j}$ has the same components of the row $n(j-1)+i$ of the matrix \mathbf{J} and δ is the vector representing the descent direction for updating the parameters. In order to find the optimal values of δ , some other steps are required. In fact, it is necessary to apply the Meketon's algorithm (1987) on a linear approximation of the function $g(\cdot)$ to compute the optimal values of the vector \mathbf{d} , and after the primal descent direction vector δ . In practice, by considering the vector form of the linear approximation, i.e.

$$\mathbf{g}_{\theta} \approx \mathbf{g}_{\hat{\theta}} - \mathbf{J}(\theta - \hat{\theta}),$$

the primal direction is computed as

$$\delta = (\mathbf{J}^{\top} \mathbf{D}^2 \mathbf{J})^{-1} \mathbf{J}^{\top} \mathbf{D}^2 \mathbf{g},$$

where the diagonal matrix $\mathbf{D} = \text{diag}\{\min(u_1, v_1), \dots, \min(u_{mn}, v_{mn})\}$. Here, the vectors \mathbf{u} and \mathbf{v} do not represent the positive and negative parts of the residuals as in the linear case, but they are computed as

$$\mathbf{u} = (\boldsymbol{\tau} \otimes \mathbf{1}_n) - \mathbf{d},$$

and

$$\mathbf{v} = (\boldsymbol{\tau} \otimes \mathbf{1}_n) + \mathbf{d}.$$

More accurately, the vector \mathbf{d} is initialised to be the null vector $\mathbf{0}_{mn}$ in order to satisfy the existence conditions for the optimal solution in (3.17),

(3.18) and (3.19). By applying the Meketon's algorithm, the vector \mathbf{d} is updated as

$$\mathbf{d} \leftarrow \mathbf{d} + \omega(\mathbf{s} \odot \boldsymbol{\alpha}),$$

where the scalar $\omega \in (0, 1)$ is a constant used to insure the feasibility of the solution, which is usually fixed at $\omega = 0.97$; the vector

$$\mathbf{s} = \mathbf{D}^2 \left\{ \mathbf{I}_{mn} - \mathbf{J} (\mathbf{J}^\top \mathbf{D}^2 \mathbf{J})^{-1} \mathbf{J}^\top \mathbf{D}^2 \right\} \mathbf{g}$$

and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)^\top \otimes \mathbf{1}_n$, wherein

$$\alpha_j^{-1} = \max_{i=1, \dots, n} \left\{ \max \left(\frac{s_{n(j-1)+i}}{\tau_j - d_{n(j-1)+i}}, -\frac{s_{n(j-1)+i}}{1 - \tau_j + d_{n(j-1)+i}} \right) \right\}$$

for any $j = 1, \dots, m$.

After executing twice the Meketon's algorithm and once the primal direction is computed, the estimates are updated as

$$\hat{\theta} \leftarrow \hat{\theta} + \tilde{\lambda} \delta,$$

wherein the scalar $\tilde{\lambda} \in [0, 1]$ is calculated as

$$\tilde{\lambda} = \arg \min_{\lambda} \sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \left\{ y_i - g(x_i; \hat{\theta} + \lambda \delta) \right\}$$

and it denotes the optimal step length according to the direction δ . After this, the vector $\mathbf{g}_{\hat{\theta}}$ and the matrix \mathbf{J} are updated according the new values of the estimates.

Before to iterate the entire procedure until convergence, it is necessary to adjust the vector \mathbf{d} to ensure the feasibility of the solution for the new values of \mathbf{J} . The adjustment refers to a projection of the vector \mathbf{d} onto the null space of the new matrix \mathbf{J} followed by shrinkage procedure, which guaranties the condition in (3.17), so that

$$\mathbf{d} \leftarrow \left\{ \mathbf{I}_{mn} - \mathbf{J} (\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top \right\} \mathbf{d},$$

and then the shrinking is performed as follows:

$$\mathbf{d} \leftarrow \mathbf{d} \odot (\mathbf{r} \otimes \mathbf{1}_n),$$

wherein components of the vector $\mathbf{r} = (r_1, \dots, r_m)^\top$ are computed as

$$r_j^{-1} = \begin{cases} (c_0 + \epsilon)/\tau_j & \text{if } c_0 > \tau_j, \\ (c_1 + \epsilon)/(1 - \tau_j) & \text{if } c_1 < \tau_j - 1, \\ 1 & \text{otherwise,} \end{cases}$$

for all $j = 1, \dots, m$, where

$$c_0 = \max_{i=1, \dots, n} \{ \max(0, d_{n(j-1)+i}) \},$$

$$c_1 = \min_{i=1, \dots, n} \{ \min(0, d_{n(j-1)+i}) \},$$

and $\epsilon > 0$ is a constant regarding the tolerance of the shrinkage.

Here, the stopping rules are based on several convergence criteria. In particular when the difference between the global loss at the previous step and that at the actual is less than a fixed tolerance threshold the algorithm stops. Also the step length $\tilde{\lambda}$ is evaluated if it is less than another tolerance threshold. If at least one of these conditions are satisfied or when the maximum number of iteration is reached, the algorithm stops.

3.2.3 Semi-parametric quantile sheets

The adoption of the Levenberg-Marquardt's algorithm (1944, 1963) is usually preferred for least square regression of nonlinear models. In this context, after some considerations, it is possible to apply this basic technique to improve convergence, time efficiency, and stability of the resulting estimates. Here, the proposed method to estimate the coefficients of a semi-parametric quantile sheet is based on a new perspective which can easily lead to other forms of asymmetric regression.

In order to obtain the estimates by the use of (3.12), the following least square problem is considered:

$$\hat{\theta} = \arg \min_{\theta} \frac{1}{2} \sum_{i \in \mathcal{I}} f_i(\theta)^2, \quad (3.20)$$

wherein \mathcal{I} is a generic set of indexes, f_i denotes the i -th generic loss function whose form will be specified below. According to the Levenberg-Marquardt's algorithm, the solution of this problem is iteratively found from a candidate point for θ by the following updating formula:

$$\theta \leftarrow \theta - \delta,$$

where the steepest descent direction δ is given by

$$\delta = \{ \mathbf{J}^\top \mathbf{J} + \lambda \text{diag}(\mathbf{J}^\top \mathbf{J}) \}^{-1} \mathbf{J}^\top \mathbf{f},$$

and λ is non negative scalar factor which is adjusted at each iteration in order to avoid slow convergence. The Jacobian matrix \mathbf{J} is obtained by numerical

differentiation of the functions involved in (3.20); in particular the i -th row of \mathbf{J} corresponds to the numerical approximation of the gradient of f_i . The vector \mathbf{f} consists of the evaluations of each function f_i in θ . As all the iterative algorithms, also the Levenberg-Marquardt's algorithm, stops according the usual convergence criteria or when the maximum iteration number is reached.

To calculate the coefficients of the models through the estimator (3.12), it is necessary to write the deviance as a sum of squares. If the integral is approximated by the trapezoidal quadrature, the estimator (3.12) can be rewritten as

$$\hat{\theta}^{(qs)} = \arg \min_{\phi_1, \dots, \phi_m, \psi} \sum_{k=0}^{K-1} \frac{1}{2K} \left\{ L_{\phi_1, \dots, \phi_m, \psi} \left(\frac{k}{K} \right) + L_{\phi_1, \dots, \phi_m, \psi} \left(\frac{k+1}{K} \right) \right\} + P,$$

where K is the number of equally spaced panels which are used to make discrete the reference domain of the integral, while

$$L_{\phi_1, \dots, \phi_m, \psi}(\tau) = w(\tau) \sum_{i=1}^n \rho_{\tau} \left[y_i - g\{x_i; \varphi^{\bullet}(\tau), \psi\} \right].$$

Therefore, the first K values of vector \mathbf{f} are defined by

$$f_i = \frac{1}{\sqrt{2K}} \left\{ L_{\phi_1, \dots, \phi_m, \psi} \left(\frac{i-1}{K} \right) + L_{\phi_1, \dots, \phi_m, \psi} \left(\frac{i}{K} \right) \right\}^{\frac{1}{2}}, \quad \forall i = 1, \dots, K,$$

and the remaining $(m-1)q$ values are given by the penalty P . In particular, if the first order difference is considered, then these last values of \mathbf{f} take the following form:

$$f_{K+(i-1)q+j} = \sqrt{\kappa}(\phi_{i,j} - \phi_{i+1,j}), \quad \forall i = 1, \dots, m-1, \text{ and } j = 1, \dots, q,$$

where the notation with a double subscript for $\phi_{i,j}$ denotes the component in the i -th row and j -th column of the matrix Φ .

This technique requires an initial common point for all the vectors ϕ_j , for $j = 1, \dots, m$. A more effective technique consists in substituting the check function $\rho_{\tau}(\cdot)$ with the following

$$\rho_{\tau}^*(z) = \begin{cases} \tau^2 z, & \text{if } z \geq 0, \\ -(1-\tau)^2 z, & \text{otherwise.} \end{cases} \quad (3.21)$$

In so doing, this check function forces the curves to spread from a central common starting location, but the resulting fitted curves do not identify the right quantile level (see the proof in Appendix A.2). To solve this problem,

it is necessary to associate the estimated curve to its proper quantile level. The transformation $\pi: [0, 1] \rightarrow [0, 1]$ is defined as

$$\pi(\tau) = \frac{\tau^2}{1 - 2\tau + 2\tau^2},$$

and it returns the true probability associated to the τ estimated curve. However, any conditional quantile can be computed through semi-parametric quantile sheets for any values of $\tau \in [0, 1]$. For this reason, when the check function $\rho_\tau^*(\cdot)$ is used in the estimation procedure, the inverse function of $\pi(\cdot)$ must be used to calculate the parameters of the nonlinear model through the Bernstein polynomials. In practice, the inverse function of $\pi(\cdot)$ is defined by

$$\pi^{-1}(\tau) = \begin{cases} \frac{1}{2}, & \text{if } \tau = \frac{1}{2}, \\ \frac{\tau - \sqrt{(1-\tau)\tau}}{2\tau - 1}, & \text{otherwise,} \end{cases} \quad (3.22)$$

so that the parameters are computed as

$$\hat{\varphi}^\bullet(\tau) = \sum_{j=0}^m \binom{m}{j} \hat{\phi}_j \{\pi^{-1}(\tau)\}^j \{1 - \pi^{-1}(\tau)\}^{m-j}. \quad (3.23)$$

If the resulting estimates do not satisfy the monotonicity property of quantiles, then the estimation procedure should be repeated with higher values of κ .

3.3 Testing multi-objective quantile estimators

Once the parameters are estimated, the model needs to be compared with other relevant and more flexible models for selecting the optimal approximation of the real τ -quantile function. In reality, such a function is not known, so there is the necessity of establish which criteria can be adopted for model selection when multiple quantiles are simultaneously estimated.

The analysis of the residuals is also important as well as the model selection. The most common techniques introduced in the literature for least squares methods were also extended in order to make deeper analyses also for a quantile regression settings.

Nonetheless, the construction of confidence regions can be performed to make inference on the parameters and this can lead to a better interpretation of the estimates. This technique can also supply an alternative method for model selection when two estimated models are nested.

3.3.1 Goodness of fit tests

The test statistic R^1 is quite common in robust statistics for checking the lack of fit of a median regression model. Such a statistic is the analogous to the usual R^2 in least squares regression and it was introduced to evaluate the goodness of fit by McKean and Sievers (1987). Its asymmetric version was studied some years later by Koenker and Machado (1999), who proposed other tests for the inferential process. Some of these tests are more common among statisticians, such as the likelihood ratio test based on the Laplace distribution, while the others are based on asymptotic approximations, e.g. the Wald test. However, none of these tests evaluates the performance of the regression models simultaneously; in fact, all them and their related tests for nested models are separately performed for each quantile level.

Let $\hat{g}_0(x)$ be the estimates of a simple model, which is also called the *reduced model* and it is often chosen to be constant and independent on any covariates. A more complex model (or “*full model*”), $\hat{g}_1(x)$, is then compared with $\hat{g}_0(x)$ by the use of the statistic $R_{\tau_j}^1$. When separate quantile levels are estimated, such a statistic is computed for each $j = 1, \dots, m$ as

$$R_{\tau_j}^1 = 1 - \frac{V_{\tau_j}^\bullet}{V_{\tau_j}^\circ}, \quad (3.24)$$

where

$$V_{\tau_j}^\bullet = \sum_{i=1}^n \rho_{\tau_j} \{y_i - \hat{g}_1(x_i)\}$$

represents the deviance of the full model and

$$V_{\tau_j}^\circ = \sum_{i=1}^n \rho_{\tau_j} \{y_i - \hat{g}_0(x_i)\},$$

denotes the deviance of the reduced one. Since $V_{\tau_j}^\circ > V_{\tau_j}^\bullet$, the results of $R_{\tau_j}^1$ always belong to the interval $[0, 1]$. In particular, the statistic denotes a lack of fit for values closer to 0, while higher values corresponds to an improvement of the goodness of fit.

When multiple quantile levels are simultaneously estimated, a global measure of goodness of fit should be considered instead. For this purpose, two quantities can be considered:

- \tilde{R}^1 , which can be calculated almost as (3.24);
- \bar{R}^1 , which summaries the individual statistics $R_{\tau_j}^1$, for all $j = 1, \dots, m$.

In practice, \tilde{R}^1 can be viewed as a global goodness of fit statistics, and the way it is computed depends on the global loss function used during the estimation procedure. E.g. if the estimator in (3.9) is selected, then the resulting statistic will be computed as

$$\tilde{R}^1 = 1 - \frac{\sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \{y_i - \hat{g}_{1,\tau_j}(x_i)\}}{\sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \{y_i - \hat{g}_{0,\tau_j}(x_i)\}}.$$

Such a statistic, even if it is computed with the same criteria adopted for the estimation, it may suffer of a “quantile reference” problem. In fact, this problem arises when the deviances are mixed together or, in the worst cases, when the ratio is obtained with the deviances of two different quantiles, especially if the estimator in (3.10) is adopted.

The latter proposal, \bar{R}^1 , is based on a unique index which might be computed either as the weighted average, the weighted median or the minimum of those quantities obtained as in (3.24). The minimum statistics is calculated as

$$\bar{R}^1 = \min\{R_{\tau_1}^1, \dots, R_{\tau_m}^1\}$$

and it seems to be reasonable, because when it exceeds an optimal level all the other are greater. This means that the full model is optimal for all the quantile levels.

Other tests presented in the literature are devoted to the selection of regression models for a specific quantile level. E.g. the test proposed by He and Zhu (2003) checks the model inadequacy (both for linear and nonlinear models) and it is based on the *cusum process* of the residuals (a standard technique in statistical quality control, see Montgomery, 2000). To be more precise, the proposed statistic was defined as follows

$$C_{\tau_j} = \max_{\|\mathbf{a}\|=1} n^{-1} \sum_{i=1}^n (\mathbf{a}^\top W_{\tau_j}(x_i))^2, \quad (3.25)$$

which is equivalent to the largest eigenvalue of the matrix

$$n^{-1} \sum_{i=1}^n W_{\tau_j}(x_i) W_{\tau_j}^\top(x_i), \quad (3.26)$$

where the function $W: \mathbb{R}^K \rightarrow \mathbb{R}^r$ is defined as

$$W_{\tau_j}(\mathbf{t}) = n^{-1/2} \sum_{i=1}^n \rho_{\tau_j} \left\{ y_i - g(x_i; \hat{\theta}_{\tau_j}) \right\} \nabla_{\theta} g(x_i; \hat{\theta}_{\tau_k}) \mathbb{1}_{(-\infty, t_1] \times \dots \times (-\infty, t_K]}(x_i),$$

where $\nabla_{\theta}g(x_i; \hat{\theta}_{\tau_k})$ denotes the gradient of the regression model computed with respect to the parameters. Since a cusum process is a cumulative sum process, in this case the function $W_{\tau_j}(\mathbf{t})$ is a vector-weighted cumulative sum involving the gradients $\nabla_{\theta}g(x_i; \hat{\theta}_{\tau_k})$, which is computed only for those observations that are lower than a specific value \mathbf{t} , and the weights, which are given by the application of check function ρ_{τ_j} on the residuals. The statistic in (3.25) is equivalent to *spectral norm* (Bernstein, 2009, proposition 9.2.3) of the matrix in (3.26), in fact it may be interpreted as the norm of the total deviance of the vector-weighted cusum process $W_{\tau_j}(\mathbf{t})$. This perspective is justified by the fact that $W_{\tau_j}(\mathbf{t})$ weakly converges to a Gaussian process (He and Zhu, 2003). In order to accept the goodness of fit hypothesis, the statistic C_{τ_j} must be lower than the critical value c_{α} which guaranties a level of significance α , i.e.

$$\Pr(C_{\tau_j} < c_{\alpha}) = 1 - \alpha.$$

He and Zhu (2003) assert that the test is consistent for any fixed alternative, and it is asymptotically distribution-free. However, they suggested a computer simulation based procedure to get an approximated critical value. Wilcox (2008) introduced a simplified approach based on this method by ranking the gradients $\nabla_{\theta}g(x_i; \hat{\theta}_{\tau_k})$, so that the improvement in computational efficiency of the original algorithm is obtained during the calculation of $W_{\tau_j}(\mathbf{t})$.

Even if each separate statistic gives a specific information on the goodness of fit for its quantile level, a global technique to evaluate simultaneous quantile regression models is not studied yet. A possible approach to extend the previously described technique for simultaneous regressions consists in evaluating the statistics C_{τ_j} for any $j = 1, \dots, m$ and to establish their empirical joint distribution by Monte Carlo methods. As before, each critical values $c_{j,\alpha}$ must satisfy

$$\Pr\{(C_{\tau_1} < c_{1,\alpha}) \cap \dots \cap (C_{\tau_m} < c_{m,\alpha})\} = 1 - \alpha.$$

Other techniques are based on the information criteria (Koenker, 2005, Section 4.9.1), e.g. the *Akaike information criterion*

$$AIC = \log \left[\frac{1}{m n} \sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \left\{ y_i - g(x_i; \hat{\varphi}_{\tau_j}, \hat{\psi}) \right\} \right] + p$$

or the *Schwarz information criterion*

$$SIC = \log \left[\frac{1}{m n} \sum_{j=1}^m \sum_{i=1}^n w_j \rho_{\tau_j} \left\{ y_i - g(x_i; \hat{\varphi}_{\tau_j}, \hat{\psi}) \right\} \right] + \frac{p}{2} \log(m n),$$

where p is the mean number of parameters that are effectively estimated for each separate model, i.e.

$$p = p_f + \frac{p_c}{m},$$

where p_f denotes the number of free parameters, while p_c represents the number of fixed parameters which remain constant across quantiles. A more sophisticated method based on information criteria is the *focused model selection* (Behl *et al.*, 2012). Other alternatives to these techniques are based on the idea of a regression via “*lasso*” shrinkage (Tibshirani, 1996), e.g. *regularised simultaneous model selection* (Zou and Yuan, 2008).

3.3.2 Analyses of residuals

Usually, the analysis of residuals is a powerful tool to check the basic properties of the residuals based on the assumption of normality (or other distributions). These methods are further options to establish the goodness of fit in terms of outliers detection, homoscedasticity of the residuals and model adequacy. On the other hand, most of these analyses are not useful when a quantile regression setting is adopted. Since quantile regression is most robust than least squares methods, the outlier detection can be used for different purposes than excluding or re-weighting observations for a further more robust estimation. The analysis of residuals might be useful for checking the heteroscedasticity of the residuals, but this does not affect the results since simultaneous multiple quantile regression is able to catch this feature of the data.

To perform these analyses from a graphical point of view, the most common plots are the *residuals vs. fitted values* and the *normal quantile-quantile* (Faraway, 2002). The former plot is useful to access to the randomness of residuals and homoscedasticity, while the latter is applied to check the residuals normality. Of course, by adjusting these techniques for the quantile regression context, these analyses must be performed separately for each quantile level.

In a quantile regression setting, residuals vs. fitted values plot is adopted to check graphically some odd patterns, which imply a lack of fit. By the nature of the technique, the τ -quantile of the residuals is zero and this fact can be used to diagnose premature algorithm stops. To establish the lack of fit, however, it is necessary to see if the residuals are locally random by maintaining the quantile proportions, namely there should be τ negative and $1 - \tau$ non negative residuals. By taking inspiration from the methods proposed by Redden *et al.* (2004) and Zou and Yuan (2008), it is possible to summarise the graphical information with a statistic for testing the goodness

of fit. This new technique for the analysis of residuals can be described as follows:

1. introduce an indicator variable $D_\tau = \mathbb{1}_{\mathbb{R}_+} \left\{ y - g(x; \hat{\theta}_\tau) \right\}$, which classifies the observations according to the position of their response variable with respect to the regression model;
2. perform a non parametric logistic regression on D_τ by using the response variable y as covariate and shrinking to zero the estimates of the parameters related to the smoother. This means that the logistic regression model will be

$$\Pr(D_\tau = 1) = \text{logit}^{-1} \{ \kappa_\tau + s_\tau(y) \}, \quad (3.27)$$

where

$$\text{logit}^{-1}(z) = \frac{1}{1 + \exp(-z)},$$

κ_τ is a constant parameter which plays the role of the intercept, and the function $s_\tau(\cdot)$ is a generic smoother obtained as a weighted sum of basis functions. This estimation procedure refers to the *generalised additive models* (Hastie and Tibshirani, 1990) and, in this case, to a further extension of this method as a regularised logistic regression (as described in Hastie *et al.*, 2001, Section 18.4);

3. perform a log-likelihood ratio test to determine if the smoother estimates are significant or not. In other words, the test should discern among $H_0: s(y) = 0$, for all $y \in \mathbb{R}$, versus $H_1: s(y) \neq 0$, for some $y \in \mathbb{R}$. The test statistic based on the log-likelihood ratio test is asymptotically distributed as χ^2 with degrees of freedom computed as the difference of the effective degrees of freedom between two models: one including the smoother and another one excluding it. If H_0 is accepted, then the quantile regression model is accurate enough to characterise the conditional quantile of $Y|X = x$.

Other tests might be performed through regularised logistic regression as in (3.27) by substituting the covariates x to the response y .

The normal quantile-quantile plot (or Q-Q plot), it is not a proper tool in this context. However, if the error distribution is assumed to be the *asymmetric Laplace* (Yu and Zhang, 2005), it might be adopted to determine if a monotone non decreasing transformation applied to the data might improve the results (Santos and Elian, 2012). In fact, if $h(\cdot)$ is a monotone

non decreasing function, then by the equivalence property of quantiles, the following relationship holds:

$$Q_{h(Y)|X=x}(\tau) = h \{Q_{Y|X=x}(\tau)\}.$$

The Q-Q plot can be drawn by following the definition of quantile residuals given by Dunn and Smyth (1996), i.e.

$$\hat{\varepsilon}_{\text{Dunn}}(\tau) = \Phi^{-1} \left\{ F_{\text{La}}(y, g(x; \hat{\theta}_\tau), \hat{\sigma}_\tau, \tau) \right\},$$

where

$$\hat{\sigma}_\tau = \frac{1}{n} \sum_{i=1}^n \rho_\tau \left\{ y_i - g(x_i; \hat{\theta}_\tau) \right\},$$

$\Phi(\cdot)$ is the standard normal c.d.f. and F_{La} denotes the c.d.f. of an asymmetric Laplace r.v., which is defined as

$$F_{\text{La}}(y, g(x; \theta_\tau), \sigma_\tau, \tau) = \begin{cases} \tau \exp \left[\frac{1-\tau}{\sigma_\tau} \{y - g(x; \theta_\tau)\} \right], & \text{if } y \leq g(x; \theta_\tau), \\ 1 - \frac{1-\tau}{\exp \left[\frac{\tau}{\sigma_\tau} \{y - g(x; \theta_\tau)\} \right]}, & \text{otherwise.} \end{cases}$$

3.3.3 Confidence regions and intervals

After the model selection, the *standard error* (s.e.) of each parameter is computed for making other tests on the obtained estimates. Bootstrap methods are mostly applied in this context to get an approximate estimate of the s.e. of the coefficients. In particular, these resampling techniques were studied by Koenker (1994) to construct confidence intervals. These methods, however, can also be used as a further alternative to check the goodness of fit when nested models are compared.

There exist other ways to estimate the s.e. of the parameters, but all of them are based on asymptotic properties of the limiting distribution of the parameters as the sample size $n \rightarrow \infty$. In many real cases, such properties are not valid; in fact they are based on some conditions which might not be satisfied. A simple example is given by the *independent and identical distributed* (i.i.d.) errors. If data are affected by an heteroscedastic behaviour, the residuals highlight the fact that the i.i.d. condition is no more satisfied; therefore the asymptotic properties cannot be used.

The sparsity test is the most direct method to compute confidence interval, because it is based on the asymptotic normality assumption of the

estimated parameters distribution. In fact, under the assumptions of i.i.d. errors with c.d.f. F and density f , such that $f\{F^{-1}(\tau_j)\} > 0$ in a neighbourhood of τ_j for each $j = 1, \dots, m$,

$$\sqrt{n}(\hat{\theta}_{\tau_j} - \theta_{\tau_j}) \xrightarrow{d} \mathbf{N}(\mathbf{0}, \mathbf{\Omega}_j),$$

where the matrix $\mathbf{\Omega}_j$ is defined analogously as in Koenker and Bassett Jr. (1978), so that it takes the following form for the nonlinear quantile regression:

$$\mathbf{\Omega}_j = \lim_{n \rightarrow \infty} \frac{\tau_j(1 - \tau_j)}{f\{F^{-1}(\tau_j)\}^2} \left[\frac{1}{n} \sum_{i=1}^n \nabla g_{i,j} \nabla g_{i,j}^\top \right]^{-1},$$

wherein $\nabla g_{i,j}$ is the gradient of $g(x_i; \theta_{\tau_j})$ calculated with respect to θ_{τ_j} , for any $i = 1, \dots, n$ and $j = 1, \dots, m$. The sparsity function, namely the reciprocal of the density function, has to be estimated and, since it is unknown, it is a good practice to estimate it with the empirical quantile function, i.e.

$$\begin{aligned} \hat{s}(\tau_j) &= \hat{f}\{\hat{F}^{-1}(\tau_j)\}^{-1}, \\ &= \{\hat{F}^{-1}(\tau_j + h_n) - \hat{F}^{-1}(\tau_j - h_n)\} / (2h_n), \end{aligned}$$

where \hat{f} and \hat{F}^{-1} are respectively the estimated density and quantile functions of the error term and bandwidth h_n approaches to zero as $n \rightarrow \infty$. It was shown that the performances of this method with a finite sample size vary a lot depending on the choice of the bandwidth. This provides a direct method also for independent but not identically distributed data, when the asymptotic covariance matrix takes form

$$\mathbf{\Omega}_j = \lim_{n \rightarrow \infty} \frac{\tau_j(1 - \tau_j)}{n} \left[\sum_{i=1}^n \mathbf{F}_i \right]^{-1} \left[\sum_{i=1}^n \nabla g_{i,j} \nabla g_{i,j}^\top \right] \left[\sum_{i=1}^n \mathbf{F}_i \right]^{-1},$$

wherein the matrix $\mathbf{F}_i = f_i\{F_i^{-1}(\tau_j)\} \nabla g_{i,j} \nabla g_{i,j}^\top$, with error specific c.d.f. F_i and density f_i . An extension of this technique can easily be developed also for simultaneous nonlinear quantile regression.

Koenker (1994) compared the coverage frequencies and the lengths of the confidence intervals computed with several bootstrap techniques in a simulation study. After that, he stated that the xy -pairs bootstrap and the method developed by Parzen *et al.* (1994) perform better than other bootstrap methods. Mainly, these techniques are able to maintain the heteroscedasticity in the data. In fact, in the former each observed pair $\{x_i, y_i\}$, for $i = 1, \dots, n$, can be sampled with replacement via Monte-Carlo procedures with probability $1/n$. The latter generates only one extra data point to include in a new

regression setting, so that it can obtain a realisation of the quantile regression process.

Some years later, He and Hu (2002) introduced the *Markov Chain Marginal Bootstrap* (MCMB) method to reduce the computational time of the previous resampling techniques and it can be suitable for high dimensional models. Kocherginsky *et al.* (2005) extended this method for the construction of confidence intervals in a regression quantile setting and they compared it with the previous techniques. At each iteration, the algorithm draws for each parameter a new sample with replacement from the original data. Then, it estimates a single parameter by maintaining fixed the others. In the end, since the resulting sequence is a Markov chain, it might suffer of the autocorrelation problem. Kocherginsky *et al.* (2005) also provided a further method (MCMB-A) based on affine transformation of the parameter space to reduce the correlations in the sequences.

Other two extension of these methods for nonlinear regression can be found in Kocherginsky and He (2007). The former consist in a linear transformation of the estimating equations (MCMB-B), while the latter (MCMB-AB) is based on a combined approach of the MCMB-A and MCMB-B.

Generally speaking, these bootstrap procedures and resampling techniques destroy the original dependence structure of the data, therefore they should properly be used when observations are independent. In these particular circumstances, the confidence intervals can easily be computed on independent sequences of parameters through order statistics or other empirical quantile techniques (e.g. see Hyndman and Fan, 1996). The covariance matrix of the parameters can be estimated as

$$\mathbf{\Omega} = \frac{1}{n} \sum_{i=0}^n (\theta_i^* - \tilde{\theta})(\theta_i^* - \tilde{\theta})^\top,$$

where θ_i^* denotes the i -th parameter vector computed via Monte-Carlo methods, while $\tilde{\theta}$ is a central vector of parameters commonly calculated with the averages of the bootstrapped parameters.

In order to avoid the use of resampling methods, it is possible to adopt a quantity which measures the ‘‘proximity’’ of the observations to the fitted model as θ varies. Seber and Wild (2003, see Section 3.3.1) stated that it seems appropriate to calculate the confidence regions for θ according to the contours of the loss function $\ell(\theta)$. In so doing, a confidence region is a set of parameters which satisfy a condition, i.e.

$$\left\{ \theta : \ell(\theta) \leq \ell(\hat{\theta}) b_\alpha \right\}, \quad (3.28)$$

where $b_\alpha > 1$ is a multiplicative critical level such that the coverage probability of the region is $1 - \alpha$. Equivalently, a region can also be expressed as

$$\left\{ \theta: \ell(\theta) - \ell(\hat{\theta}) \leq c_\alpha \right\},$$

where $c_\alpha > 0$ is an additive critical level. Since these confidence regions are not based on any approximation they are called “exact”, but generally the value of b_α or c_α is unknown.

Through the use of the concentration inequalities, it is possible to approximate the critical values of these regions. For the region in (3.28), one can obtain a test statistic which has a bounded support in $[0, 1]$; in fact, since $\ell(\hat{\theta})$ is the minimum attained loss, the quantity $T = \ell(\hat{\theta})/\ell(\theta)$ is always included in the interval $[0, 1]$ for any value of θ . In practice, it is required that

$$\Pr \{ \lambda_0 \leq T \leq \lambda_1 \} \geq 1 - \alpha, \quad (3.29)$$

where λ_0 and λ_1 are respectively the upper and lower bounds which guaranty the equality in the previous formula.

It expects under the null hypothesis that the value of T is closed to one, namely it seems reasonable to consider $\mathbf{E}[T] \approx 1$ and $\lambda_1 = 1$. By the application of the Hoeffding’s inequality (Hoeffding, 1963), one can obtain the following threshold:

$$\lambda_0 = 1 - \sqrt{\frac{\log(\alpha)}{-2n}}. \quad (3.30)$$

Another threshold can be computed from the Uspensky’s inequality (Uspensky, 1937) so that

$$\lambda_0 = 1 - c \sqrt{\frac{1 - \alpha}{n\alpha}}, \quad (3.31)$$

where $c \in [0, 0.25]$ is a constant related to the variance of T . It is reasonable to set $c = 12^{-1/2}$ (see Appendix A.3). Both thresholds approximately guaranty a coverage with probability $1 - \alpha$. More details and the proofs are presented in Appendix A.3.

Here, the critical value b_α corresponds to λ_0^{-1} , and since the two different kind of “exact” region must be identical, the correspondent critical value c_α is given by

$$c_\alpha = \ell(\hat{\theta})(\lambda_0^{-1} - 1).$$

By expanding $\ell(\theta)$ in $\hat{\theta}$ through Taylor series, i.e.

$$\ell(\theta) \approx \ell(\hat{\theta}) + \nabla \ell(\hat{\theta})^\top (\theta - \hat{\theta}) + \frac{1}{2} (\theta - \hat{\theta})^\top \nabla^2 \ell(\hat{\theta}) (\theta - \hat{\theta}),$$

where $\nabla\ell(\hat{\theta})$ and $\nabla^2\ell(\hat{\theta})$ are respectively the gradient and the hessian matrix of the loss function, it is possible to obtain an approximated confidence region. Since the loss function in quantile regression is not differentiable everywhere, the gradient computed in the minimum point of loss can be considered to be the null vector (even if its value is not computable). For the same reason, also the hessian matrix must be approximated numerically, so that the resulting confidence region takes the following form:

$$\left\{ \theta: (\theta - \hat{\theta})^\top \tilde{\nabla}^2\ell(\hat{\theta})(\theta - \hat{\theta}) \leq 2c_\alpha \right\}. \quad (3.32)$$

3.4 Conclusion

After a meticulous description of established general methods for the estimation of quantile curves, which do not violate the monotonicity property of quantiles, new suitable estimation techniques are introduced in this chapter for nonlinear models. The equality constraints are considered when dealing with particular cases, wherein the data variability is uniquely explained by the variation of a single parameter. More general methods are inspired by penalised estimators, where the penalty refers to a distance which quantifies the dissatisfaction of the constraints. The most useful estimation technique is based on the idea of quantile sheets, wherein the parameters of the nonlinear model are functional and dependent on the probability levels.

Several tests presented in the literature are adjusted in order to be able to perform the model selection and the residuals analysis for simultaneous estimated curves. The construction of confidence regions is also considered in order to assess the equivalence of similar models for a data set. More evidence of the capabilities of these methods is found in the results obtained via simulations, which are discussed in Appendix A.4 and exposed in Appendix A.5.

In order to apply some of these methods, it is necessary to study and understand the main issues and characteristics of the data collected. The preliminary analyses of the data are performed in order to isolate a reduced area and this is described in detail in the next chapter. The parameter estimations and results discussion will therefore be focused on this chosen area.

Chapter 4

Preliminary Data Analysis

In order to understand better how to analyse the data and what kind of techniques are more appropriate to interpret the phenomena related to the precipitation process, a cluster analysis is initially performed to classify different locations in some common areas. Through this method it is possible to identify the spatial distribution of the process and to investigate seasonal effects. In so doing, better model involving both statical and physical description can be developed for more accurate prediction.

The procedures of data collection, the description of the analysed data and the adjustments performed are shortly described in Section 4.1. The methodology adopted for the cluster analysis and cluster validation is explained in Section 4.2. The quantity selected to perform the analysis and results are discussed in Section 4.3. Conclusions are presented in Section 4.4.

4.1 Precipitation data

The National Oceanic and Atmospheric Administration (NOAA) system is aided by satellite constellations around the globe provided by the National Environmental Satellite Data and Information Service (NESDIS). This remote-sensing system is a useful tool to measure the surface temperatures, biomass burning, pollutants and cloud cover. Many other meteorological variables are collected by the use of particular sensors in the satellites, e.g. the temperature and the humidity comes from infrared and microwave radiometers, while the upper winds are based on a tracking system that follows the movement of clouds, aerosols and water molecules. However, observations about temperature and humidity are more precise over oceans, but they are more noisy over land depending on the land-use and surface variations. Through specific procedures described in Levizzani *et al.* (2007), precipita-

tion data are estimated with the information obtained from satellites.

Rainfall data are also collected in other ways; in particular, Michaelides (2008) described the instruments and the methods for continuous ground data acquisition. All the materials refer to sensors, scanners and cameras for precipitation measurements to improve the estimates from satellites. In fact, since ground measurements are more precise than the satellites estimates, they are used to calibrate the satellites' outcomes. By the use of this methodology, it is possible to get data over a wide area with a reduced bias.

4.1.1 General Information

The analysed data are related to the rainfall estimated by the American National Weather Service (NWS) and River Forecast Centers (RFCs). Data are collected daily from 1st January 2005 to 24th January 2013 on grid cells. The projection grid system is defined by the Hydrological Rainfall Analysis Project (HRAP), even if the latitude and longitude coordinates are provided as defined by the usual polar stereographic projection. Daily precipitation data sets are available from the NOAA web site¹ which also provides monthly and yearly data.

Each data set consists of a varying number of observations with the recorded values for the following variables:

id defines a unique integer value that identifies each grid cell;

hrapx represents the column number of the HRAP grid cell, in particular higher numbers refers to eastern cells;

hrapy denotes the row number of the HRAP grid cell, where higher numbers are northern cells;

latitude and longitude refer to the polar stereographic coordinate system, in particular to the latitude and longitude of the HRAP grid point (i.e. the cell centre);

globvalue 24-hours, 1-month or 1-year precipitation values in inches. The “-2” values correspond to “Missing Data”, and the cells with no precipitation (i.e. 0 inches) are systematically removed from the grid.

¹<http://www.noaa.gov/>

4.1.2 Data adjustments

If each data set is considered separately from the others, it is only possible to study the precipitation with regard to the spatial information within a specific time window (of the length of one day, month or year). In order to deal only with the temporal information or with both spatial and temporal information, these data need to be merged in a unique large data set.

Since the grid cells are labelled with an ID-number from 0 to 864362, it is considered a dual relationship between the two data sets, one storing the spatial coordinates of each grid cell and the other with the recorded precipitation data and their temporal occurrence. The grid cells presenting either missing observations or zero precipitation during the entire monitored period were removed from the final data sets. In so doing, some computational advantages are obviously related to the amount of stored data as well as a slight decrease of computational time.

In practice, the first data set with the spatial information consists mainly of grid cells coordinates, in fact the variables for each cell are the ID, the HRAP coordinates and the corresponding latitude and longitude. The second data set collects the precipitation data according to their time occurrence, namely each grid cell denotes a sample unit with its ID and the precipitation values for each day, month or year as variables.

4.2 Methodology

Basic statistical tools and more advanced data mining techniques are used to discover the main features of large data sets. These descriptive techniques can be summarised by

data inspection, which is needed to analyse phenomena from a descriptive point of view;

cluster analysis, for finding groups in the data;

discriminant analysis, to construct classifiers in order to choose the best group for new observations;

knowledge conversion, to represent and summarise the results for better interpretations.

After each phase, validation techniques must be applied to ensure the achievement of the most accurate findings.

Cluster analysis is adopted for many application fields and, in particular, it is used for data reduction, prediction based on groups and generation or

testing of hypotheses. Here, the information in the data can be compressed by the identification of an optimal number of data partitions. Since the groups summarise the main features of the data, the most central object will be processed for other analyses, such as performing a regression.

Initially, this kind of analysis is performed in order to select groups of cells with the same or most similar temporal pattern. This can reduce the number of sample units for successive analyses, regarding to both the spatial and temporal components of the rainfall process. The results can also be useful not only to reduce the original spatial lattice, but also to build a proximity matrix among groups in order to provide a spatial network.

Several classification techniques are available and usually they are divided in supervised and unsupervised methods. The main goal of the former is related to the discriminant analysis, when there is previous knowledge about different groups. In practice, a new observation has to be associated to one known group. The latter performs a search of optimal clusters in order to improve the description of the data. To adopt an objective approach, no assumption about the process generating the data is made. Therefore, in this case, unsupervised procedures are preferred to the supervised.

The choice of the clustering algorithm is not only based on previous knowledge; in fact the selection of a suitable procedure can be driven by the type of input data, the definition of the dissimilarities between points and by the theory related to the algorithm itself. In practice, when the sample size is very large, it is unfeasible to store the distances between each couple of sample units. For this reason, algorithms such as K-Means are preferred because they will perform much better from a computational viewpoint. The algorithm selection must be done by considering also the shape of clusters, the ability of outliers detection and complexity of the procedure.

Grid-based clustering is mainly adopted for the quantification of spatial data to a finite number of cells, these algorithms summarise each feature through some statistical parameters and then generate hierarchies at different levels. On the other hand, self-organising maps are based on neural networks, wherein each neuron is a centroid.

The self-organising maps are a good choice for discovering clusters that are not known a priori; in fact they show invisible structures of the data and it is good for outliers detection. The complexity of this algorithm is $O(n)$, where n denotes the sample size. This algorithm combines aspects of partitioning techniques with some idea about hierarchical methods; in fact the resulting partition is characterised by an underlain ordered structure.

All clustering methods should search for well separated groups whose components are the closest to each other. However, if the optimal number of clusters is not known a priori, it may be found through cluster validation

techniques. Thus, the resulting groups must be evaluated using quantitative methods in order to establish whether the partition fits the data. All validating methods use quantities which measure the compactness and separation of the groups.

In practice, there are three approaches for cluster validation and they are based on external, internal or relative criteria. The first evaluates the results on a pre-specified structure, the second involves quantities related directly to data (e.g. dissimilarity matrix) and the third criterion compares the resulting clusters with others given by the same algorithm but obtained with different parameters (e.g. by changing the initial value of centroids).

4.2.1 Data transformations

Transformations are applied to original data to improve both the performance of clustering methods and the results reliability. On the other hand, transformations reduce the comparability between original and transformed data in terms of results.

Each variable in the data-set may be treated with its most appropriate transformation. However, the impact of some transformations on final cluster results can be very high as it happens when nonlinear transformations are applied (e.g. the rank data transformation). Only linear ones do not change the resulting final groups and, for this reason, the mean-variance standardisation is often used to improve the efficiency of the algorithm.

The logarithm is a classical nonlinear transformation that is used to deal with positive data. It is often used to reduce heteroscedasticity and the distribution skewness of the original data. This transformation, after some adjustments, can be used in order to standardise the first three moments of the transformed variable. In practice, if the following transformation is adopted

$$X^* = \frac{\log(X + \alpha) - \mu}{\sigma}, \quad (4.1)$$

the values of $\alpha \in \mathbb{R}_+$, $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}_+$ satisfy the conditions $E[X^*] = 0$, $\text{Var}[X^*] = 1$ and $\text{Skew}[X^*] = 0$. In practice, α stabilises the skewness of the transformed variable and its value is found by numerical method, while the values of μ and σ respectively control the usual standardisation of the mean and the variance.

Other standardisations can be applied in order to obtain values between the interval $[0, 1]$. One method consists in a location and scale transformation, wherein the location is related to the minimum observed value and the

scale is based on the difference between the maximum and the minimum, i.e.

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)}.$$

By the use of particular functions, it is possible to get the similar transformations also for non negative data, e.g.

$$X^* = 1 - \exp(-\gamma X), \quad (4.2)$$

wherein $\gamma \in \mathbb{R}_+$ is chosen to mitigate the exponential effect for extreme values.

More sophisticated variable transformations are based on ranks, or by the use of nonparametric methods in order to perform the cluster analysis on functional transformation (e.g. see Abraham *et al.*, 2003, for the classification through B-splines). Cluster analysis can be performed also on indexes that were previously identified or given by known standards. This approach is preferred than others if these indexes are the estimates of a fixed number of the parameters of a suitable model.

Even if Gordon (1999) suggested that the selection of the dissimilarity measure should be data-driven, non negative data are related to quantitative variables which are treated independently on their support. This means that each type of variable has a specific set of measures to apply and, if this approach is followed, then no transformation is needed.

If the seasonal properties of the data are the main characteristics under study, they must be studied through Fourier series in order to approximate any periodic function of the time

$$f(t) = \sum_{i=0}^{\infty} \{\alpha_i \sin(i\omega t) + \beta_i \cos(i\omega t)\},$$

where α_i and β_i are the coefficients of the series, t denotes the time and ω represents the frequency of the oscillations. The approximation of the function is computed by truncating the series, e.g.

$$f(t) \approx \beta_0 + \alpha_1 \sin(\omega t) + \beta_1 \cos(\omega t). \quad (4.3)$$

As Dunstan *et al.* (1982) pointed out, the cosinor model (Guercin, 1973) is an alternative to the previous approximation, i.e.

$$f(t) \approx \theta_0 + \theta_1 \cos(\omega t - \varphi), \quad (4.4)$$

wherein φ denotes the phase, namely the position of peak in the interval $(0, 2\pi\omega^{-1}]$, θ_0 is the mean level of the function and θ_1 represents the amplitude

of the cosine, which is related to the intensity of the seasonal cyclicity. The estimates of these four parameters, which are obtained through the usual nonlinear regression techniques, summarise all the information needed for the cluster analysis. In practice, it is preferred to get the estimates for the parameters of the approximation in (4.3) rather than for the cosinor in (4.4). This is done because the parameters ω and φ are very sensitive to the value of θ_1 and, in many cases, they are not easily identifiable. However, the parameter ω in (4.3) suffers of the same inferential problem and it might be fixed *a priori* by adopting conventional settings (such as the year length).

4.2.2 Kohonen networks (self-organising maps)

The fundamental idea of self-organising maps was introduced in the early 1980s by Teuvo Kohonen. This kind of maps are similar to neural networks, because they are based on network architectures inspired by the nervous system. This method was successfully employed in various pattern recognition problems involving very noisy data. Recently, Gorricha and Lobo (2011) applied this technique to geo-referenced data and also to extreme precipitation data (see particularly Gorricha *et al.*, 2012).

Kohonen networks are defined as set of neighbouring neurons (or computational units) in a network. These neurons compete in their activities through interactions with their neighbours in order to recognise adaptively specific data-clusters characterised by different patterns. This technique is a type of centroid-based clustering, where each neuron is univocally associated to its respective centroid (which is the most informative object for the correspondent group of observations). The main goal is to find those centroids that provide the best approximation for a specific subset of observations.

For this kind of network, the learning procedure (i.e. the adaptation of the neural network) is said to be unsupervised, sometimes competitive or self-organising. The simplest version of this technique can be explained from both a neurological and a mathematical-statistical viewpoint. Actually, these networks work like a brain where close neuronal cells share similar information, which is located in dedicated brain zones in order to perform specific tasks. The neurological point of view consists in only one neuron or local group of neurons that gives an active response to input stimuli. Each neuron is located in a meaningful neuronal coordinate system (or neuronal grid), so that responses tend to become ordered according to the position of the neuron in the network. In particular, each coordinate corresponds to a specific domain of input signals and this provides a way to interpret the input information. The latter refers to neurons as centroids that are updated sequentially in order to be enough close to a set of most similar data. The network impose

a topographic ordering on neurons, so that neighbouring centroids are also updated. In this way, this technique enforces the resulting cluster centroids to be locally similar, so that hierarchical relationships can easily be studied.

Let $\mathbf{m}_j \in \mathbb{R}^r$ be the j -th neuron of the Kohonen network, where the index $j = 1, \dots, K$ and r denotes the dimension of the vectors related to each neuron. In the basic version of these networks, each neuron is parametrised with respect to a two-dimensional grid, which consists of integer coordinate pairs defined as $\mathbf{q}_j \in Q_1 \times Q_2$, wherein the set $Q_1 = \{1, 2, \dots, q_1\}$ and similarly $Q_2 = \{1, 2, \dots, q_2\}$. Such a grid can be formed by neurons represented in a higher dimensional space, namely $\mathbf{q}_j \in Q_1 \times Q_2 \times \dots \times Q_s$ so that the total amount of cluster centroids is given by $K = q_1 q_2 \dots q_s$. Thus, there exists an one-to-one relationship between the index j and the neuronal coordinates of the grid \mathbf{q}_j . From the grid, it is possible to determine the neurons with similar patterns. Without considering a grid, this can be done also in a more cumbersome manner involving a proximity matrix which provides the similarity rates among neurons.

The fitting algorithm tries to approximate, as far as possible, the data points by processing each observed object $\mathbf{x}_i \in \mathbb{R}^r$ at a time, where the index $i = 1, \dots, n$ and n represents the total amount of observations. A distance measure $d: \mathbb{R}^r \times \mathbb{R}^r \rightarrow \{0\} \cup \mathbb{R}_+$ is often chosen between the Manhattan or the Euclidean distances, in order to compare sample units and centroids.

The procedure starts with the centroids initialisation by random selection, because unordered centroids will be sorted during the fitting procedure. However, this policy is not the fastest way to converge to the best final results. In practice, the procedure is much faster if the choice of initial centroids is not completely randomised. Any ordered initialisation, when it is computationally feasible, is always better than a random one (Kohonen, 2001). One way to choose the centroids can be done by selecting the most representative sample units through the “build phase” of the *partitioning around medoids* (PAM) algorithm, but this choice is inappropriate for large data-sets. The linear initialisation (Kohonen, 2001) is a further method to get the starting values of the centroids, but it cannot be applied for large amounts of data. This kind of initialisation implies the computation of the autocorrelation matrix of $\mathbf{x} \in \mathbb{R}^r$ followed by a dimension reduction performed through *principal component analysis* (PCA) in order to obtain \mathbf{x}^* by selecting the eigenvectors related to the s largest eigenvalues. After the projection of the observations x_i in a s dimensional sub-space, for any $i = 1, \dots, n$, the principal components \mathbf{x}_i^* are partitioned in K sets with respect to a rectangular array, whose dimension depends on the neuronal grid. The initial values of the centroids \mathbf{m}_j , for $j = 1, \dots, K$, are then identified with the mean of \mathbf{x}_i belonging to the j -th partition. After, the algorithm will perform with the

Algorithm 4.2.1 Kohonen networks (self-organising maps)

-
- 1: Initialise the centroids either by random or guided sample units selection.
 - 2: **repeat**
 - 3: **for** $i = 1$ to n **do**
 - 4: Select the i -th sample unit.
 - 5: Determine which centroid is the closest to the selected sample unit.
 - 6: Update this centroid its grid-neighbours.
 - 7: **end for**
 - 8: **until** The centroids converge or the maximum number of iterations is reached.
 - 9: Define clusters by assigning each sample unit to its closest centroid.
-

original observations.

Once centroids are initialised, it is necessary to find for each sample units \mathbf{x}_i its closest centroid \mathbf{m}_j and, after, to update it with its neighbours \mathbf{m}_k according to the grid topology, where $k \in \{1, \dots, K\}$. This procedure is repeated until centroids convergence, as described in the algorithm 4.2.1. In other words, \mathbf{m}_j is the centroid which minimises the quantity $d(\mathbf{x}_i, \mathbf{m}_j)$. All centroids \mathbf{m}_k are considered as neighbours of \mathbf{m}_j if and only if the distance between \mathbf{q}_j and \mathbf{q}_k is less than a fixed threshold $\varrho > 0$. The fitting iterations are based on the recursive formulation of the arithmetic average. This means that at each iteration the centroids become closer to their central position. Therefore, the centroids update phase moves \mathbf{m}_k centroids toward \mathbf{x}_i as follows:

$$\mathbf{m}_k \leftarrow \mathbf{m}_k + \lambda h(\mathbf{q}_j, \mathbf{q}_k) (\mathbf{x}_i - \mathbf{m}_k),$$

where λ is positive scalar value which denotes the learning rate that can varies with the number of iteration needed for the convergence (λ is usually decreasing sigmoidally from 1 to 0). Particular interest can be found in the function $h(\cdot, \cdot)$, which can give more weight to neighbouring centroids closer to \mathbf{m}_j ; e.g. it can be a constant function, i.e. $h(\mathbf{q}_j, \mathbf{q}_k) = 1$, or a radial function depending on a distance, i.e.

$$h(\mathbf{q}_j, \mathbf{q}_k) = \exp\left(-\frac{d(\mathbf{q}_j, \mathbf{q}_k)}{\sigma_j}\right),$$

where σ_j controls the topological radius of the neighbourhood of \mathbf{m}_j . Baykal and Erkmen (1999) showed how this function, if properly specified, can improve the cluster results by avoiding false convergences or premature stops. It also increases the convergence speed and efficiency of the algorithm. In

practice, they suggested to use the following formulation:

$$h(\mathbf{q}_j, \mathbf{q}_k) = \frac{\sigma_j - d(\mathbf{q}_j, \mathbf{q}_k)}{\sigma_j},$$

where the value of σ_j is chosen in order to satisfy $\delta_{\max} = \sigma_j - 1$, wherein δ_{\max} denotes the maximum distance between \mathbf{q}_j and \mathbf{q}_k .

When centroids converged, it is possible to associate each sample unit with its closest centroid, so that the groups can easily be recognised. Other analytical possibilities can be found in the study of centroids characteristics, which can underline salient pattern features for each group.

There exists another version of self-organisation and its name is due to the fitting procedure. The batch-maps (Kohonen, 2001) perform a batch-learning algorithm, which differs only for centroids update step. In this case each \mathbf{m}_j is updated as follows:

$$\mathbf{m}_j = \frac{\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} h(\mathbf{q}_j, \mathbf{q}_k) \mathbf{x}_i}{\sum_{k \in \mathcal{K}} h(\mathbf{q}_j, \mathbf{q}_k) |\mathcal{I}_k|}, \quad (4.5)$$

wherein \mathcal{K} is the set neighbouring centroids indexes (it always includes j), similarly \mathcal{I}_k is the set of indexes for those points which are closer to \mathbf{m}_k , while the notation $|\mathcal{I}_k|$ denotes the cardinality of the set \mathcal{I}_k . After some iterations the cardinality of the set \mathcal{K} must decrease, so that the number of neighbours reduces. In so doing, the first iterations sort the centroids in the grid according their information, while the last iterations adjust their values. This algorithm improves the original self-organising maps because it is computationally faster and the resulting centroids are asymptotically stable. A noteworthy case must be considered when the function

$$h(\mathbf{q}_j, \mathbf{q}_k) = \begin{cases} 0, & \text{if } j \neq k, \\ 1, & \text{otherwise,} \end{cases}$$

because the batch-maps corresponds with the K-Means. In practice we can refer to the batch-maps as a sorted extension of the K-means method.

4.2.3 Cluster validation

The evaluation of clusters can be conducted through the calculation of specific compactness and separation indexes. The concept of compactness refers to the distance between each observation and its cluster centroid, while the separation is related to the intersection among the highest density zones of the support of the clusters. Although the variance is commonly used to

measure the variability within and between the groups, other validation indexes were developed to deal both quantitative and qualitative data. These other criteria are based on the single linkage, to measure the distance between two groups, the complete linkage to measure the distance between the farthest members in the same group. The comparison of centroids is less demanding from a computational point of view and it is also useful to measure the average distance between clusters. A complete review about clustering validation methods can be found in Halkidi *et al.* (2001).

With respect to the criteria cited in the introduction of Section 4.2, the external ones are based on hypotheses tests, wherein the null hypothesis states that the structure of the data is completely random. Since the distribution of the test statistics cannot be found easily, the use of Monte Carlo techniques lead to complex procedure and to unfeasible computational waiting-time. Internal criteria are considered only when it is possible to obtain quantities inherited to the data, in order to evaluate either the hierarchy of clusters or each single clustering scheme. Also these techniques suffer of computational problems in particular for large memory demands.

The so-called relative criteria are much better if they are compared with the others. These methods can be divided in two categories. The former refers clustering techniques which do not need an initial value for the number of clusters. Usually, these procedures try to find the optimal number of clusters that underlie the data structure, so that cluster validation can be avoided. The latter refers to those algorithms that need the number of groups to be specified in order to initialise the procedure. A validation technique for such algorithms consists in selecting a suitable index and computing it after running the cluster algorithm. The execution of the algorithm must be done iteratively from a predefined minimum to a predefined maximum number of groups. In the end, the resulting partition corresponds to the best obtained value computed for the chosen index. Those indexes that are mostly used for checking cluster results are the modified Hubert Γ statistic, the Dunn index and the Davies-Bouldin index.

The *modified Hubert* Γ is computed as

$$\Gamma = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^n v_{i,j} d(\mathbf{m}_{[i]}, \mathbf{m}_{[j]}),$$

where n is the sample size, $v_{i,j}$ is the component in the i -th row and j -th column of the proximity matrix $\mathbf{V} \in \{0, 1\}^{n \times n}$ and $d(\mathbf{m}_{[i]}, \mathbf{m}_{[j]})$ is the distance between centroids of the clusters where the sample units $\mathbf{x}_i, \mathbf{x}_j$ belong. This statistic is not computable when the number of cluster is equal either to one or n ; however, in the other cases high values of Γ denote high evidence for

the presence of compact clusters.

The *Dunn index* is define as

$$D = \frac{\min_{i=1,\dots,K-1} \left\{ \min_{j=i+1,\dots,K} d(\mathcal{C}_i, \mathcal{C}_j) \right\}}{\max_{k=1,\dots,K} \text{diam}(\mathcal{C}_k)},$$

where $d(\mathcal{C}_i, \mathcal{C}_j)$ is the distance between two clusters defined as

$$d(\mathcal{C}_i, \mathcal{C}_j) = \min_{\mathbf{x} \in \mathcal{C}_i, \mathbf{y} \in \mathcal{C}_j} d(\mathbf{x}, \mathbf{y}), \quad (4.6)$$

and $\text{diam}(\mathcal{C}_k)$ is the diameter of the k -th cluster \mathcal{C}_k and it is computed as

$$\text{diam}(\mathcal{C}_k) = \max_{\mathbf{x}, \mathbf{y} \in \mathcal{C}_k} d(\mathbf{x}, \mathbf{y}). \quad (4.7)$$

Such an index is used with the purpose to identify compact and well separated groups; in fact, the optimal result is achieved when the distance between clusters is large and the diameter of each group is small. Thus, high values indicates strong evidence for compact and well separated clusters.

The *Davies-Bouldin index* is calculated as

$$DB = \frac{1}{K} \sum_{i=1}^K \max_{j=1,\dots,K, j \neq i} \frac{s_i + s_j}{d_{i,j}},$$

wherein the ratio is a similarity measure which is based on a measure of cluster dispersion s_i and on a measure of dissimilarity between clusters $d_{i,j}$. Usually, the measure defined in (4.7) is assigned to s_i and that in (4.6) is assigned to $d_{i,j}$. A desirable property for the final groups is to be less similar as possible, therefore lower values of this index refers to better results.

Other methods are based on the distance between two clusters, in particular the *centroid hierarchical clustering* considers the distances between couples of representative objects. After the execution of the clustering algorithm, the resulting centroids become the new objects for further cluster analysis by adopting the usual hierarchical (agglomerative or divisive) techniques.

The *total separation index* is another quantity which globally evaluates the separation between clusters by considering the distances between representative objects. It is defined as

$$TS = \sum_{k=1}^K \left(\sum_{l=1}^K d(\mathbf{m}_k, \mathbf{m}_l) \right)^{-1} \frac{\max_{i=1,\dots,K-1, j=i+1,\dots,K} d(\mathbf{m}_i, \mathbf{m}_j)}{\min_{i=1,\dots,K-1, j=i+1,\dots,K} d(\mathbf{m}_i, \mathbf{m}_j)}$$

and, of course, the higher values indicate strong evidence of well-separated clusters.

All the methods relate to the relative criteria can also be used to evaluate the results from a self-organising map. However, those referring only to the centroids are preferred, because they require less computational efforts. In practice, when the optimal number of cluster is lower than the number of fitted neurons, it is necessary to merge some groups. After this, we can adopt one of two procedures: the former is based on the reduction of the number of centroids in the grid followed by the update of the remaining centroids; the latter is based on groups with multiple representative objects (namely each cluster is represented by one or more centroids).

4.3 Application

As described in Section 4.1, the data are available with three different time resolutions. For the moment, only monthly precipitation data are considered in this chapter, in order to handle less time points per location cell and get a reasonable compromise between computational efficiency and information-loss (due to the aggregation of daily values). After merging the data as described in Section 4.1.2 and removing all non-informative points (i.e. missing data or non-zero precipitation during all the recorded period, see Figure 4.1), a cluster analysis methods is performed on the 505682 remaining points.

In order to maintain an ordered among the clusters, the self-organising maps introduced in Section 4.2.2 are applied. In particular, by the use of 3D self-organising maps, the relationship among neighbouring centroids are explained by a three dimensional grid as depicted in Figure 4.2. The colour of each neuron in the grid is defined by the grid position through the *red*, *green*, *blue* (RGB) standard (Gorricha and Lobo, 2011). This standard is defined by three additive primary colour lights such that their weighted sum can reproduce all the colours needed (Gauch and Hsia, 1992). This can help when spatial groups need to be visualised in a map. In fact, this allows to improve the representation of the clustering results, because the similarities among clusters are more evident and the each cluster is more distinguishable from the others.

The data are initially analysed without transformation by considering a $4 \times 4 \times 3$ neuronal grid. In so doing, the number of groups is set to be 48, which is the default value required by the software to initialise the algorithm. Even if this number is not the optimal or real number of clusters, it is always possible to find a better value by the application of cluster validation techniques.

The centroids initialisation is done by selecting randomly, because other initialisation techniques are computationally unfeasible for both insufficient memory size and waiting time. For this reason, the batch algorithm is preferred to the original one, so that the values of centroids are updated as described in (4.5), where the weighting function $h(\cdot, \cdot)$ gives one for all the neighbouring centroids and zero otherwise.

The Euclidean distance is chosen to compare observations and centroids, and this measure is also used to find neighbouring centroids. The neighbourhood is defined by a radius σ , which reduces (almost linearly) from 4 to 0 during the network training. In practice, this is the optimal setting, which gives good results with less computational efforts.

The data are then analysed by considering the transformations explained in Section 4.2.1, in particular the mean-variance standardisation, the skewness adjustment in (4.1) and the range normalisation in (4.2), wherein γ is here the inverse of the arithmetic mean of the variable.

The resulting clusters are located in homogeneous areas (see Figure 4.3), but they have different shapes if they are compared with those obtained through variables transformations. This is essentially due to the transformation itself and to the centroids initialisation, which is done randomly. In order to maintain spatial relationships among these areas, it is possible to identify those that are similar to each other. In practice, each colour represents an

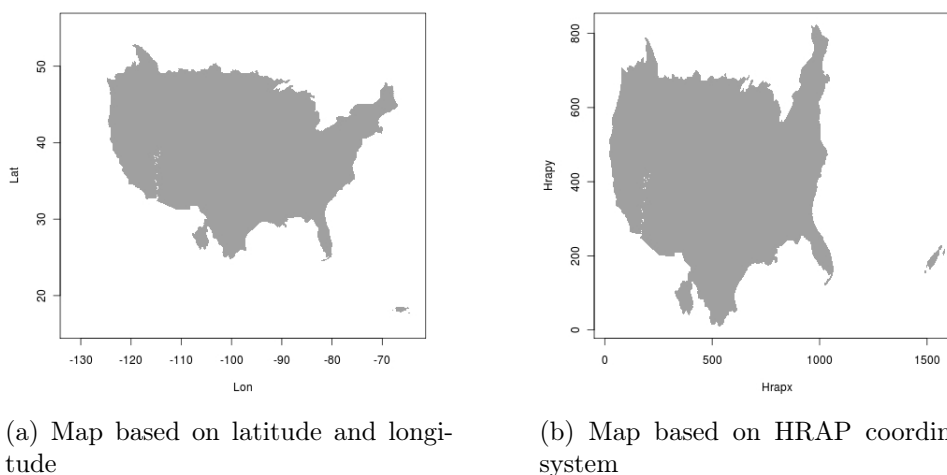


Figure 4.1: The grey area denotes the location points that are informative for the analysis of the process, while the white zone represents locations with missing data or non-zero precipitation during all the recorded period.

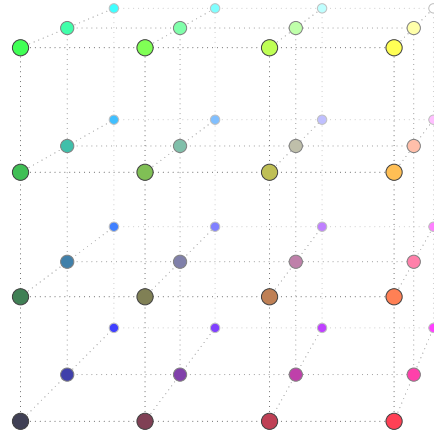


Figure 4.2: Three dimensional grid to define relationships among centroids.

area, so that similar colours denotes similar precipitation patterns in those zones; in fact, if some colours are close each other in the topology of the grid (as in Figure 4.2), the related centroids are close each other according to the chosen metric (see Figure B.4 and B.5).

Cluster validation techniques based on the modified Hubert Γ statistic, the Dunn index and the Davies-Bouldin index cannot be applied, because the computation of the quantities to obtain the value of each index is prohibitive. Since these methods evaluate a statistic for a different number of groups, it is necessary to change iteratively the structure of the network in order to select the optimal one. For these reasons, it is preferred to compute only one time these statistics for the different transformation by fixing the number of clusters.

The results shown in Table 4.1 illustrate the clusters separation and compactness; however, it is not possible to compare the adopted transformations because these indexes are not standardised and they are computed with distances calculated in different topological spaces.

A general view of the cluster quality can be seen in Figure B.2 and B.3, where each neurons are marked with the maximum distance between inner-cluster points (which is related to the compactness of a group) and the minimum distance between clusters (which is related to the cluster separation). All the values associated to the correspondent neurons are not good enough

according to the optimal clustering properties; in fact the distance between the clusters is very small (i.e. groups are not well separated), while the diameter of the clusters is very large (i.e. groups are not compact).

By the use of the centroid hierarchical clustering, it is possible to reduce the number of clusters. Those groups, whose centroid are closer according to some criteria, are merged in a unique cluster. The Euclidean distance is selected to compare centroids, since it was previously adopted for the network fitting. In this case, the agglomerative hierarchical algorithm with the Ward's linkage was preferred in order to get the optimal partition. Figure B.6 and B.7

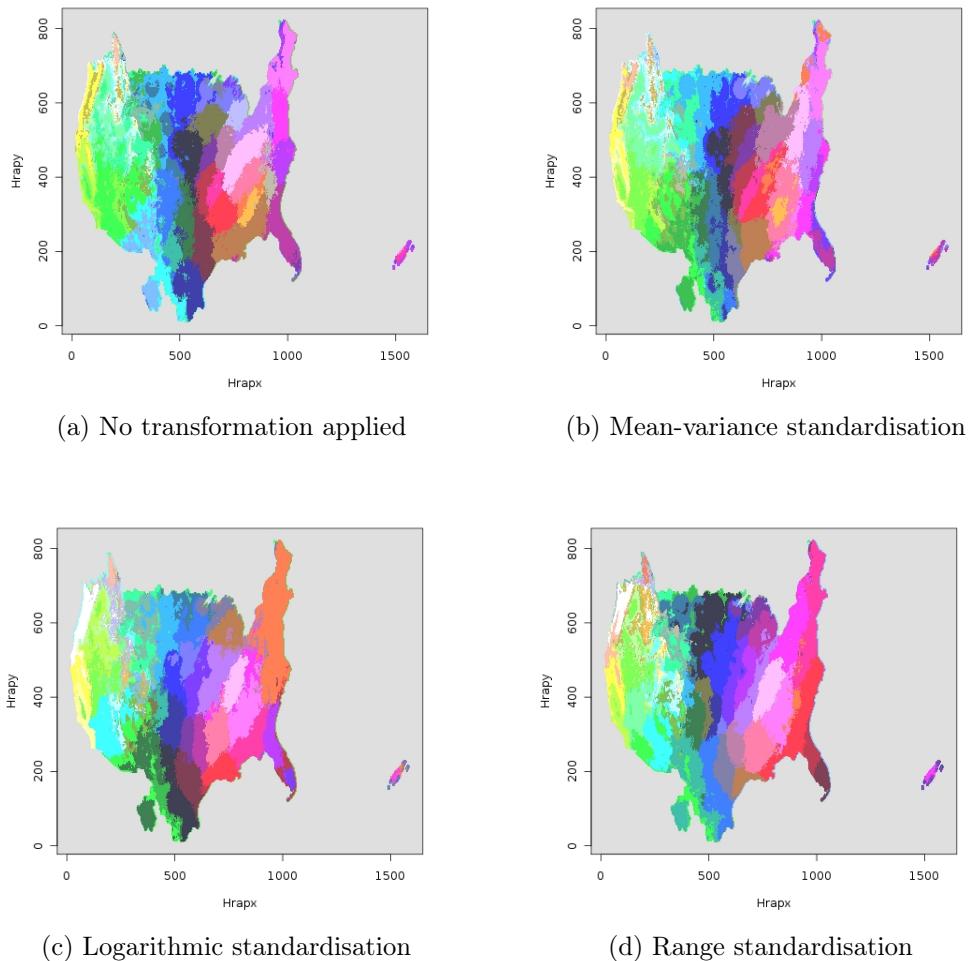


Figure 4.3: Different precipitation areas obtained with random initialisation (colours here are used to identify the clusters).

Table 4.1: Cluster validation statistics

	Hubert mod. Γ	Dunn	Davies- Bouldin	Total sep.
No transformation	1.40	0.00069	131.10	0.69
No transformation (PCA)	1.44	0.00077	135.10	0.87
Mean-variance standardization	0.55	0.00082	136.29	1.48
Mean-variance standard. (PCA)	0.59	0.00057	142.17	2.12
Logarithmic standardization	0.57	0.00655	83.76	0.76
Logarithmic standard. (PCA)	0.59	0.00319	104.00	0.93
Range standardization	0.18	0.00294	104.18	2.66
Range standardization (PCA)	0.18	0.00701	74.06	2.62

show the results of this cluster validation technique. Figure B.8 and B.9 show the merging costs. A cutting point at four clusters seems reasonable also from a spatial point of view (see Figure B.1).

The same kind of cluster analysis is then performed with the same settings, exception made for the linear initialisation of the centroid by the use of the PCA (as explained in Section 4.2.2) and the increased the number of iterations in the fitting procedure. The clustered areas are shown in Figure 4.4 and, even if the results are almost similar to the previous, the areas showed in Figure 4.4a are more homogeneous than those showed in Figure 4.3a. However, the clusters suffer from the same problems of the previous analysis. The validation statistics in Table 4.1 does not show improvements in the Dunn and Davies-Bouldin statistics, while the improvements achieved in the modified Hubert Γ and in the total separation length are slight.

Further analyses can be conducted through a model-based approach, which allows to reduce the dimensionality of the data set. As described in Section 4.2.1, the seasonal components of rainfall time series can be modelled by the use of the Fourier orthogonal series approximation in (4.3). By assuming the year as the frequency of the cycles, the parameter ω is set to be $\pi/6$. To improve the information related to the rainfall process, other indexes are considered for the cluster analysis, i.e. the minimum and the maximum precipitation observed during the entire monitored period and the standard deviation of the residuals after the regression. The cluster analysis is then performed through the same approach adopted above. In practice, the Kohonen network is initialised by the use of the PCA method, where the indexes mentioned before are considered as observations. The other setting characteristics are maintained unaltered.

The Figure 4.5 shows the resulting area obtained through this approach. They are more heterogeneous with respect to the previous, because the indexes summarise the observed data patterns. Table 4.2 is reported for completeness reasons and it shows the cluster validation statistics that are not comparable with each other and with those reported in Table 4.1.

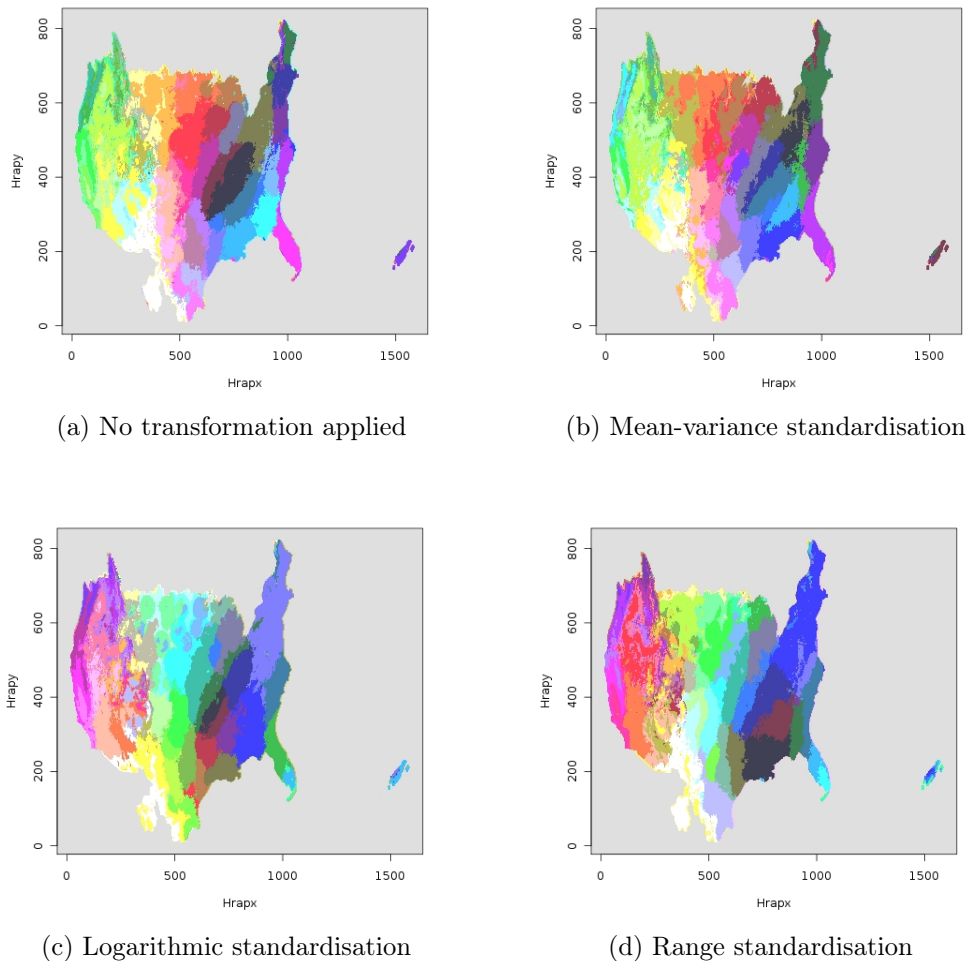


Figure 4.4: Different precipitation areas obtained with linear initialisation (colours here are used to identify the clusters).

	Hubert mod. Γ	Dunn	Davies- Bouldin	Total sep.
Informative indexes	0.22	0.00017	300.14	12.15

Table 4.2: Cluster validation statistics for the model-based approach

4.4 Conclusion

The cluster analysis highlighted those groups that are found through the self-organising maps. Such groups are not compact and not separated. This is due to the high correlation among neighbouring sample cells. There is no improvement by merging all groups in four clusters; in fact, the distances among the final groups are as small as the distances among the 48 initial groups. The same results emerged also by changing the type of network initialisation. By adopting a model-based approach and the use of other indexes, the final clusters are less homogeneous than those found with all the observations.

The evidence of similar pattern is very high for near locations. There is also a common pattern among all these analyses, which is evident on a large scale. In fact, the most different patterns are found between the east and west coasts, rather than from the north and the south of the United States.

These results can be useful for further analyses, such as the study of the spatial correlation (in particular for the presence or the absence of anisotropy) and other more sophisticated studies of seasonality cycles and extreme events.

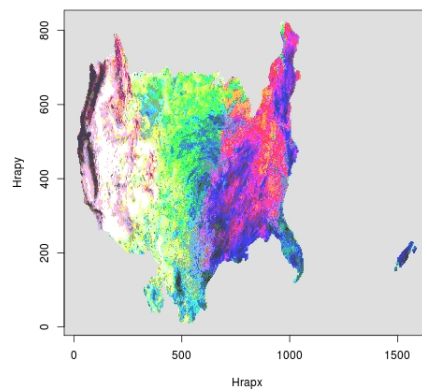


Figure 4.5: Different precipitation areas classified through informative indexes (colours here are used to identify the clusters).

However, these results are utilised in the next chapter to reduce the number of location cells, and select those where to perform regression analyses.

Chapter 5

Least squares regression analyses

The least squares regression is a useful technique to access the conditional expectation of a response variable for some given covariates without introducing any distributive assumption on the data. The Levenberg-Marquardt algorithm (1944; 1963) is the most used and accurate method to obtain the solutions of nonlinear least squares problems. The most recent implementation is due to Markwardt (2009), which introduced several innovations with respect to previously existing software. This tool is adopted here to study the phenomenon from a standard regression perspective. According to the results obtained in Chapter 4, an initial quantitative evaluation of the process can be given and a deeper investigation can lead to more suitable models.

Section 5.1 describes how to handle the data in order to perform the regression analyses. The description of the results obtained with the Bass model is given in Section 5.2. The extension of the generalised Bass model, the identification rationale of the intervention function, and the related results obtained via nonlinear least squares are discussed in Section 5.3. The main findings are summarised in Section 5.4.

5.1 Data reduction

After preliminary analyses (see Chapter 4), several homogeneous areas were discovered by performing a cluster analysis. By looking at the most salient features of the obtained centroids (e.g. cyclical patterns and rainfall magnitude), it was decided to consider only a part of the observed areas. In particular, the centroids related to the north-west side of the country exhibit stable rainfall cycles. However, since the clusters obtained with the previous analyses are not well-separated, the selection of each location cell was performed by checking if the correlation between the optimal centroid

(see Figure 4.4a) and each cell was greater than 0.8. By ignoring highly correlated cells located very far from the previously recognised are, the west part of the Washington State and a small part in the north-west of Oregon State were selected (see Figure 5.1).

The final area consists of 6582 cells and, for each of them, it was decided to shift the starting position of the year to a common starting point. The 15th August was chosen mostly for practical reasons. Since the simplified approach based on the innovation diffusion requires to estimate the parameters of the models for each year, it is reasonable to think that the time origin of the process must be in a day with no relevant rainfall. The longest period of the year, when either no rainfall or a negligible quantity of daily observed precipitation is recorded, is positioned around the 15th August. In

Further temporal adjustments were taken into account to aggregate the data by one week periods. This time-frame is considered in order to deal with less zeroes and for having still a considerable number of time-points during the year. This avoids the problem of dealing with zero inflated data, as well as not compromising the accuracy of the estimates. In addition, since each year starts by convention at the 15th August, the total amount of days within this “*hydrological year*” must be 364. This concept is necessary in order to refer to a standardised year of 52 weeks, instead of 52.14 (or 52.29) weeks, so that the considered sequence of years 2006 – 2012 starts on the 15th August 2005 and ends on the 14th August 2012.

In order to aggregate the data by week, it is essential to standardise the number of days in the year. Therefore, the observed rainfall in the 29th February of each leap year was split and added to the amount observed in the previous and successive days. For example, the resulting quantity assigned to the 28th February is given by the summation of its observed value and half of the value observed the day after; similarly, the same computation is performed for the 1st March. Subsequently, the value observed at the 14th August is added to the value of the 13th August. This choice is consistently justified by the nature of the process; in fact, no precipitation or a non significant quantity is usually measured during the last day of the year.

5.2 The Bass model

The regression analyses are here based on the innovation diffusion simplified view of the rainfall process. Originally, the Bass model was formulated as a Riccati’s differential equation (Bass, 1969), which can be used to get an approximate description of the the temporal evolution of the rainfall within the year.

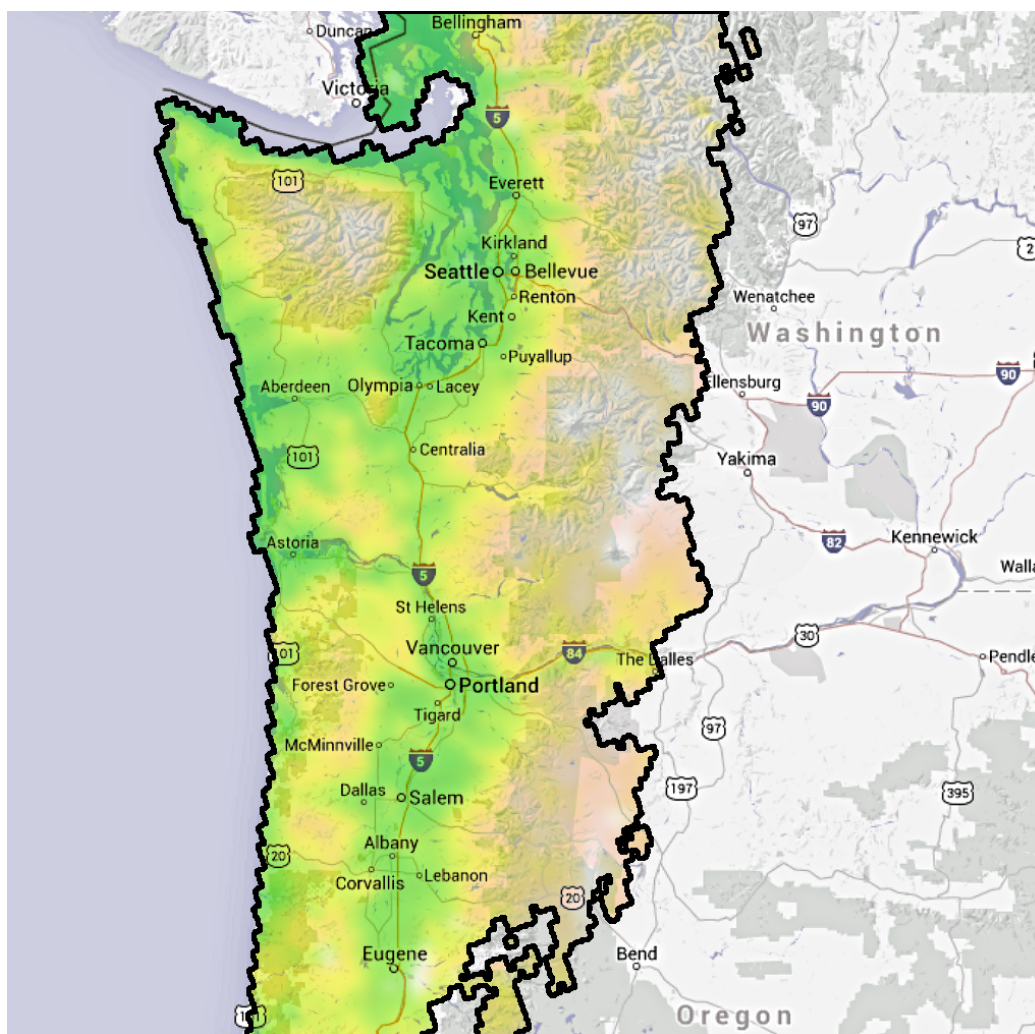


Figure 5.1: Selected region to analyse (west part of the Washington state and north-west of Oregon state). Map provided by Google (<https://maps.google.com/>) and adjusted with GIMP (<http://www.gimp.org/>).

Seber and Wild (2003) identified two approaches to estimate the parameters of sigmoidal growth models. The former is based on a stochastic viewpoint, wherein it is necessary to consider a stochastic extension of the differential equation and the propagation of errors. The latter is related to a deterministic perspective, which requires a closed form solution of differential equation, so that the standard nonlinear least squares techniques can easily be applied. In this case, the deterministic approach is followed because there exists a closed form solution of the Bass model in (2.1), the resulting estimates are more stable and reliable, and finally, the errors are easier to handle.

Conventionally, the parameters of the Bass model are estimated with the accumulated rates of adoption; therefore, the resulting estimates are more accurate, because the response is less noisy than the instantaneous rates of adoption. Nevertheless, it is always possible to estimate the parameters with the instantaneous rates by fitting the model

$$\frac{\partial y(t)}{\partial t} = \mu\zeta \frac{(\zeta + \xi)^2 \exp\{-t(\zeta + \xi)\}}{[\zeta + \xi \exp\{-t(\zeta + \xi)\}]^2} \quad (5.1)$$

and, in some particular cases (Guseo, 2004), it is even required, although it is preferred to avoid this way to operate. For these reasons, it was decided to fit the Bass model formulated for the accumulated rates of adoption in (2.1).

The parameters of the cumulative Bass model are estimated for each separate location cell and year. Usually, each parameter estimate is represented with a meaningful colour assigned to its location cell in each yearly map. However, the maps produced in this fashion show the main spatial patterns and local variations with noisy behaviours, because the estimates are independently calculated without considering the information in the neighbourhood. On the other hand, the estimation of a global model, wherein the parameters are allowed to vary spatially, is prohibitive in terms of computational time efficiency and usage of memory, and it does not guarantee the same level of accuracy.

A simpler method to produce spatially related estimates is based on the idea of a Bass model with random parameters. Since the estimates can still be considered spatially dependent, it is reasonable to apply a moving spatial average based on the eighth nearest neighbouring cells in order to identify non stochastic spatial patterns of the model parameters. More precisely,

$$\bar{\mu}_i = \frac{1}{1 + |\mathcal{N}_i|} \left(\hat{\mu}_i + \sum_{j \in \mathcal{N}_i} \hat{\mu}_j \right), \quad (5.2)$$

where $\bar{\mu}_i$ denotes the local average of the parameter μ with respect to the i -th cell, $\hat{\mu}_j$ indicate the least squares estimate of the parameter μ obtained for

j -th cell, and the notation $|\mathcal{N}_i|$ represents the cardinality of the set \mathcal{N}_i , whose elements corresponds to the indexes of the neighbouring cells. In particular, since the index $i \notin \mathcal{N}_i$, such a set coincides with the Moore neighbourhood for most of the central cells. Of course, the same expression can be used to draw smooth maps for all parameters.

The results are illustrated in Figure 5.2, 5.3, and 5.4. Each Figure consists of seven maps (one for each year), which are related to the same area. Next to each map, the empirical distribution density of the random effects, e.g. $\hat{\mu}_i - \bar{\mu}_i$, is shown. Such a density was globally evaluated for each year, and it is mainly used to access the global variability of the local smooth.

Figure 5.2 shows the maps of the smoothed estimates of the parameter μ . These maps show a very strong stability, which is connected with the total amount of precipitation for each year. It is evident that the zones with higher precipitation persist in the same spatial position. A closer look of these areas in the map exposed in Figure 5.1 reveals that the orography has a dominant role in explaining the intensity of the rainfall. In addition, the west part of a mountain is characterised by estimates having higher values than the east part of the same mountain. This is mainly due to a prevalent air-drift of humid masses. In fact, most of the water molecules forming raindrops evaporate from the Pacific ocean and move towards the continent. The *orographic precipitation* is formed when these masses reach the mountains (as explained in Section 2.1). A better evidence of this description can be given by reproducing the same results on three-dimensional maps based on latitude, longitude and elevation.

Figure 5.3 shows the maps of the smoothed estimates of the parameter ζ . The smoothed estimates of this parameter present different patterns year after year. The same spatial persistence is not maintained as is the case for the parameter μ . However, it is possible to note that higher values are located in the north, while lower values are obtained in the south of the considered area. In comparison with Figure 5.1, higher values corresponds mainly to areas with higher population densities, and thus characterised by a larger amount of aerosol particles of human origins. Nevertheless, cells having dense vegetation also provide a large amount of aerosol particles by transpiration. For example, the maps of the years 2009 and 2010 show high values in areas with low population densities. Approximately, from a temporal point of view, the estimated values exhibit an increasing trend until the year 2010, but they exhibit an inversion of the trend during the last two years.

The smoothed estimates of the parameter ξ are shown in Figure 5.4. These estimates are spatially more stable than the estimates obtained for the parameter ζ but less than those for μ . In fact, the higher values obtained within each year are mostly located in mountainous areas. Since the

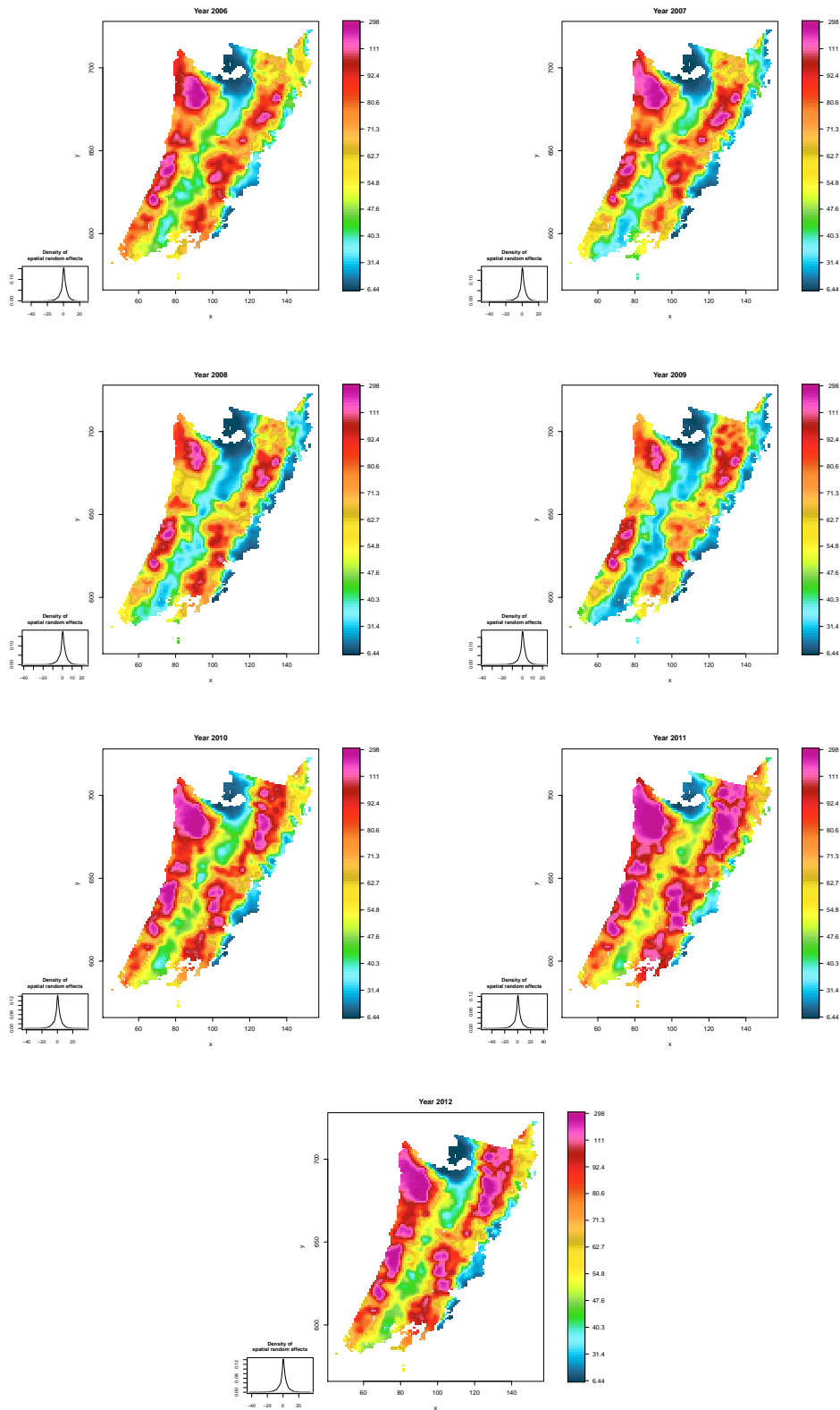


Figure 5.2: Smoothed estimates of the parameter μ of the Bass model, 2006 – 2012.

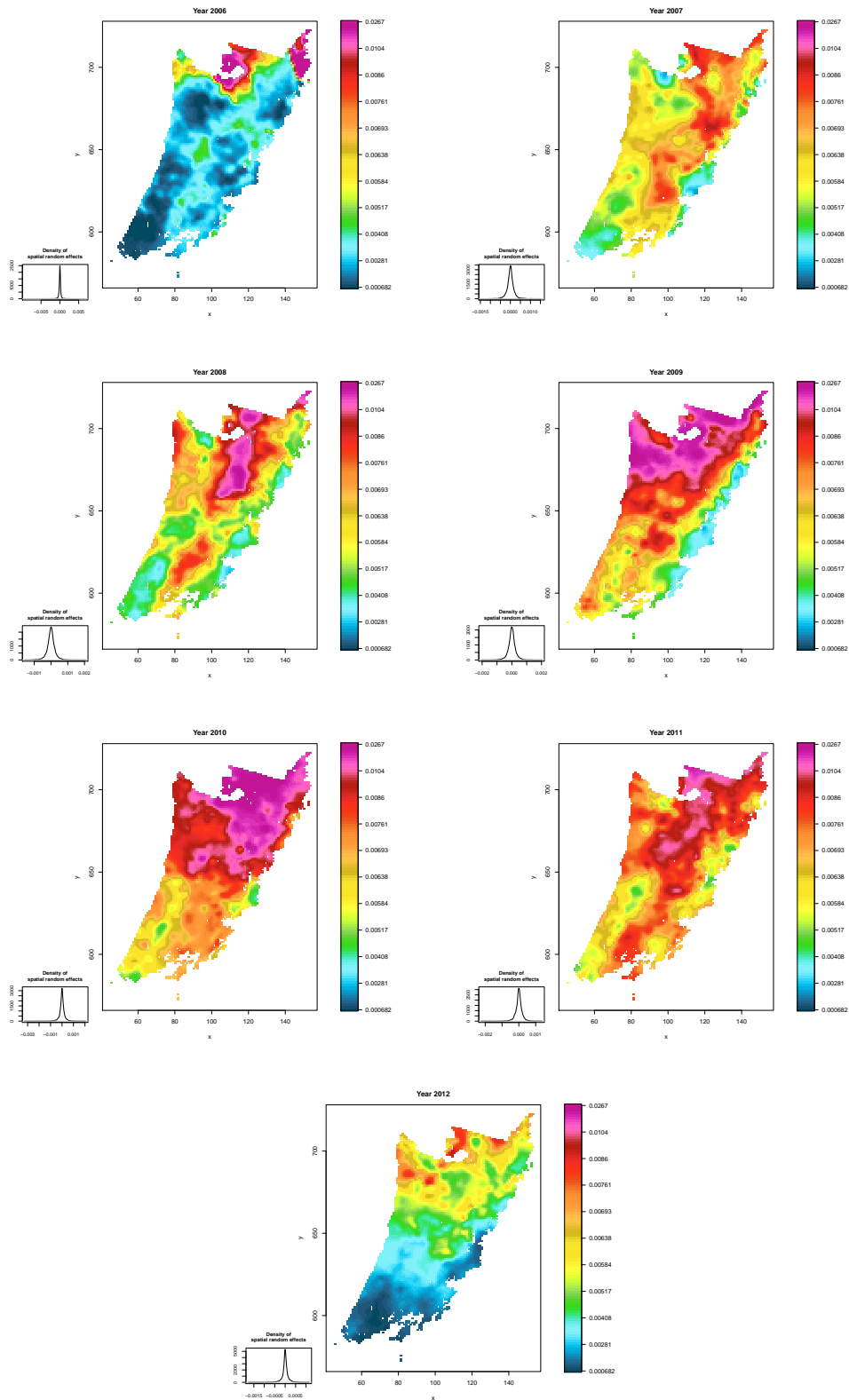


Figure 5.3: Smoothed estimates of the parameter ζ of the Bass model, 2006 – 2012.

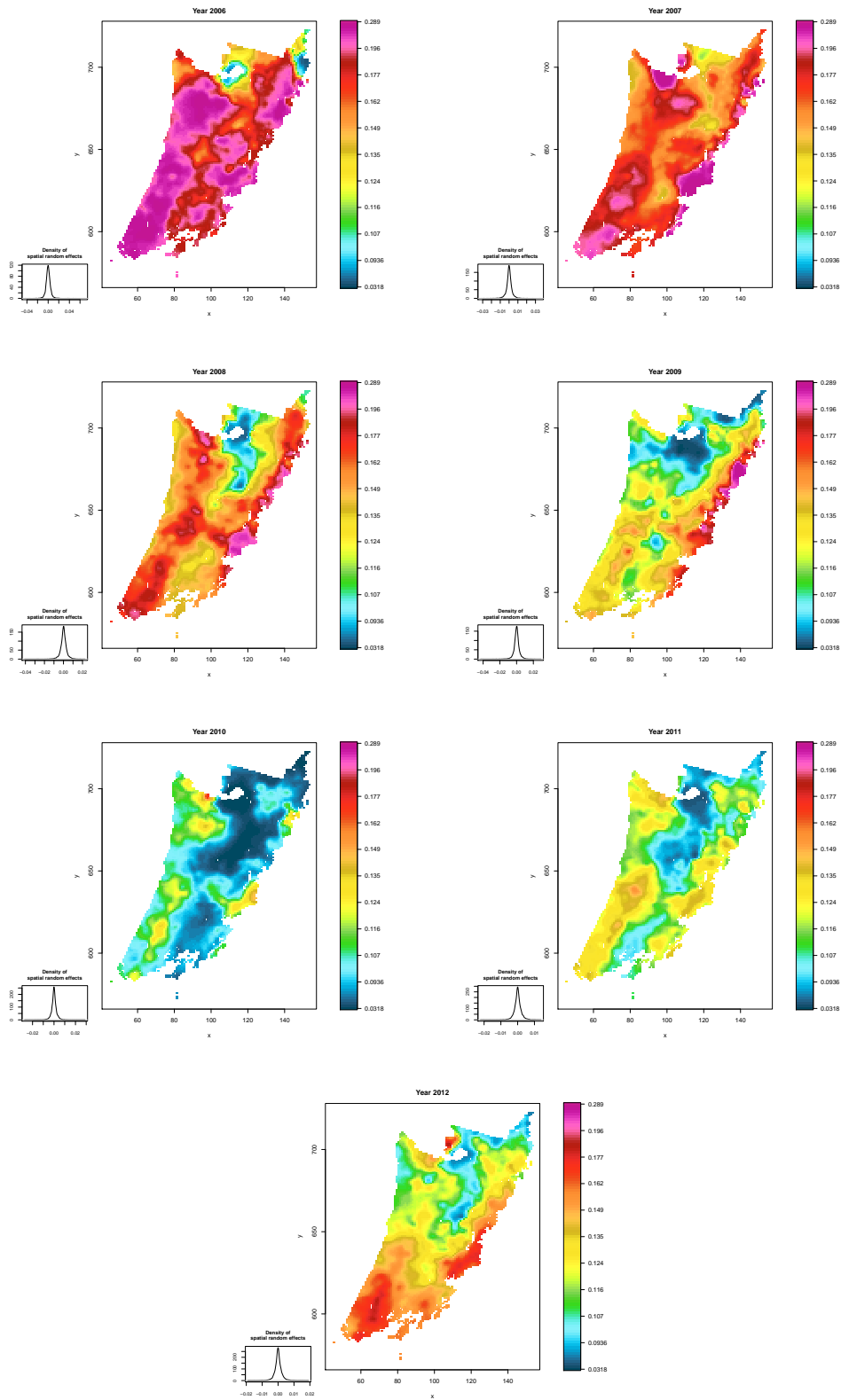


Figure 5.4: Smoothed estimates of the parameter ξ of the Bass model, 2006 – 2012.

parameter ξ is related to the coalescence effect, it is reasonable to relate it to variations of temperature due to the altitude, or to a strong presence of powerful air-stream that intensifies such an effect by advection. Also in this case, the parameter estimates approximately manifest a temporal trend, which is decreasing until the year 2010 and increasing afterwards.

5.3 The generalised Bass model

The first attempt in the literature to extend the Bass model was proposed by Bass *et al.* (1994). Such an extension is based on a simple idea, which introduces a flexible formulation through a non autonomous Riccati's differential equation. Instead of using non parametric techniques (such as local polynomials or splines), further flexibility is achieved by a parametric intervention function $a(t)$, as in the formulation in (2.2). However, a much simpler model was initially considered before fitting the generalised Bass model. Such a model was introduced by Bemmaor (1994), and it was adopted here for its ability in controlling asymmetrical trends of the trajectory. Its formulation for accumulative rates of adoption is the following:

$$y(t) = \mu\zeta^\omega \frac{1 - \exp\{-t(\zeta + \xi)\}}{[\zeta + \xi \exp\{-t(\zeta + \xi)\}]^\omega},$$

where ω is the parameter that governs the asymmetrical shape of the trajectory. Almost every analysed cell presented non stable behaviours: premature convergence and untrustworthy estimates were the main encountered problems. After a graphical inspection of some time-series related to the non convergent cells, it was more evident and necessary to adopt a model that could better adjust to bimodal configurations.

As explained above, the generalised Bass model is capable of adjusting the modelled trajectory in many ways (see Guseo, 2004, Section 3.7). The main issue of the generalised model is the identification of a proper intervention function, which can adequately change the time-flow of the process. The following steps were performed for a graphical assistance:

1. the selection of few cells, wherein the bimodal trends were graphically identified, was necessary to achieve a proper result with moderate efforts;
2. a smoothed version of the time-series was calculated via moving averages, but only its cumulative version was examined. In so doing, the noise reduction around the signal was attained in order to increase the estimation stability of the cumulative Bass model;

3. the correspondent estimates obtained on the smoothed series with the cumulative Bass model in (2.1) were considered, even if they are always and really different to the correspondent least squares solution under the generalised Bass model. In fact, since all parameters are estimated simultaneously, the parameters involved in the intervention function $a(t)$ produce different estimates of μ , ζ and ξ . In addition, the estimates of the parameter μ were adjusted in order to apply the inverse function of the cumulative Bass model, i.e.

$$t = -(\xi + \zeta)^{-1} \log \left\{ \frac{\zeta(\mu - y)}{\xi y + \zeta \mu} \right\}, \quad (5.3)$$

which can be written for the generalised Bass model as

$$\int_0^t a(z) dz = -(\xi + \zeta)^{-1} \log \left\{ \frac{\zeta(\mu - y)}{\xi y + \zeta \mu} \right\},$$

where the quantity $\mu - y$ must always be positive. For this reason, it was decided to utilise the following adjustment

$$\hat{\mu}^* = \begin{cases} \hat{\mu}, & \text{if } \hat{\mu} > \max(y), \\ \max(y) + \epsilon, & \text{otherwise,} \end{cases}$$

where $\epsilon = 0.4$ is set to calibrate the last part of the curve.

4. once the inverse function in (5.3) was computed, it was possible to reproduce approximately the shape of the true underlying intervention function $a(t)$. The graphs exposed in Figure 5.5 show both the approximation of the integrated function $A(t)$, namely

$$\tilde{A}(t) \approx A(t) = \int_0^t a(z) dz,$$

and the approximation of $a(t)$, which is based on the the following formulation:

$$a(t) \approx \tilde{A}(t + 1) - \tilde{A}(t).$$

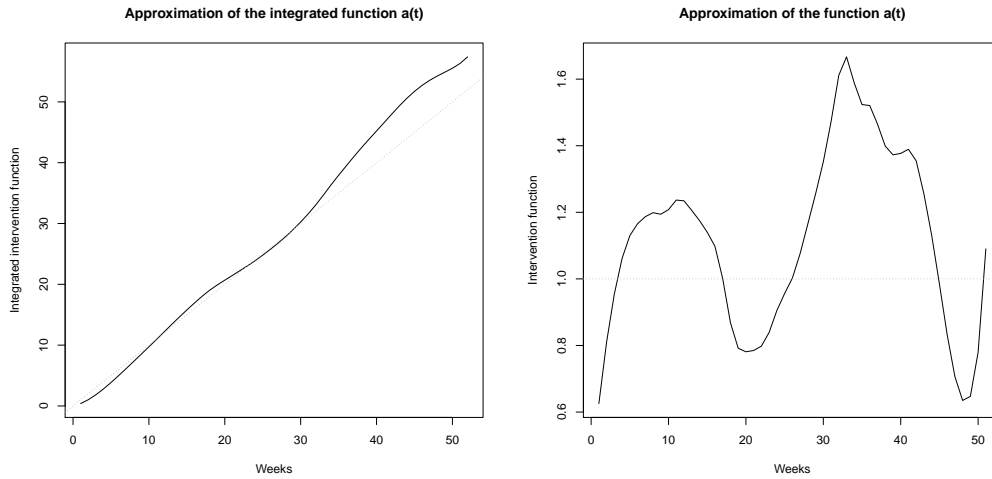


Figure 5.5: Approximations of the intervention function (on the right) and its integral (on the left).

Since the curve on the right graph in Figure 5.5 shows an oscillating behaviour, this can be modelled through harmonic functions. In particular, the simple formulation introduced in (2.3) is a reasonable choice to catch the principal characteristics of seasonal patterns (such as magnitude and frequency).

The parameters of the cumulative generalised Bass model are estimated through the Levenberg-Marquardt algorithm (1944; 1963), and the regression results are illustrated in Figure 5.6, 5.7, 5.8, 5.9, and 5.10.

The maps of the estimates obtained for the parameter μ are shown in Figure 5.6. They are not very different than those produced for the standard Bass model (see Figure 5.2 for comparison), and this means that the total amount of rainfall for each year can accurately be estimated by the use of both models.

The maps of the estimated values for the parameter ζ are exposed in Figure 5.7. The same conclusion for the parameter ζ of the Bass model are still valid. In fact, even if the generalised model produced slightly different estimates, they are still capable to highlight the differences of aerosol presence among geographical cells.

The introduction of an intervention strongly affects the estimates of the parameter ξ . However, the maps in Figure 5.8 show almost the same spatial patterns of those obtained for the Bass model, therefore it is reasonable to consider a strong connection between the effect explained by ξ (the coalescence) and the effects due to the orography (such as local variations of

temperature or air-stream).

Figure 5.9 shows the estimates of the parameter α , which are globally less stable than those obtained for the parameters described above. There is not any particular temporal trend between two consecutive years. This means that the intensity of the process is not related to the orography and further inspections based on other environmental variables are required.

In the end, the maps exposed in Figure 5.10 show the estimates of the parameter β . Such estimates are the most unstable among all the presented parameters, because they depend on the estimates of α ; in fact, if $\alpha \approx 0$, then the value the parameter β is not identifiable. Spatially, the mountainous areas (especially the most elevated) present cells with less homogeneous values than the neighbours. The year 2008 is an exception to this general thought, because the value of the parameter α has a main role during the estimation procedure (see Figure 5.9, year 2008, the red zones are almost zero).

Other models based on the mixture in (2.4) were estimated; however, the results lack spatial stability and are inferentially inaccurate. For these reasons, they are not reported here.

5.4 Conclusion

Standard statistical techniques are used in this chapter in order to obtain useful hydrological information behind the data. After exposing how the observations are handled and aggregated for better and more stable regression settings, two models presented in Chapter 2 are considered.

The results show how the micro-physical effects and characteristics of the selected zone can be estimated through macro-physical models. From a graphical perspective, it is interesting to see how the estimates are locally stable; in fact, they are mostly related to particular features of the zones (such as the land use and elevation). More specifically, the heaviest rainfall are recorded in those location cells where the estimates are strongly connected with high altitudes, intense wind-advection, and substantial presence of liquescent aerosol particles.

Even if these analyses are useful for the inference of the expected yearly evolution of the process, it is not possible to assess the empirical distribution of the rainfall over time. For this reason, the new semi-parametric quantile regression is applied in the next chapter to extract more information about the distribution of the data.

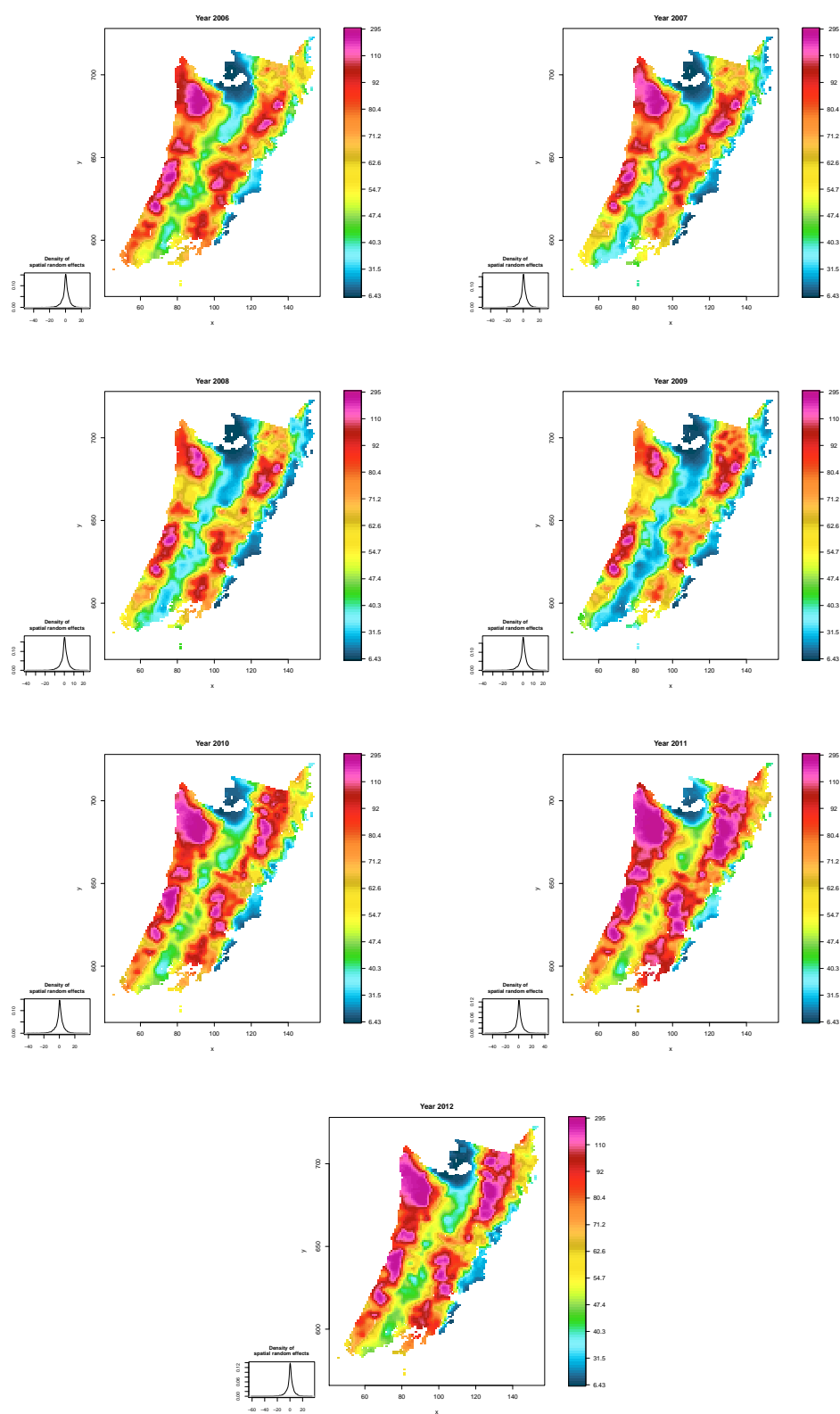


Figure 5.6: Smoothed estimates of the parameter μ of the generalised Bass model, 2006 – 2012.

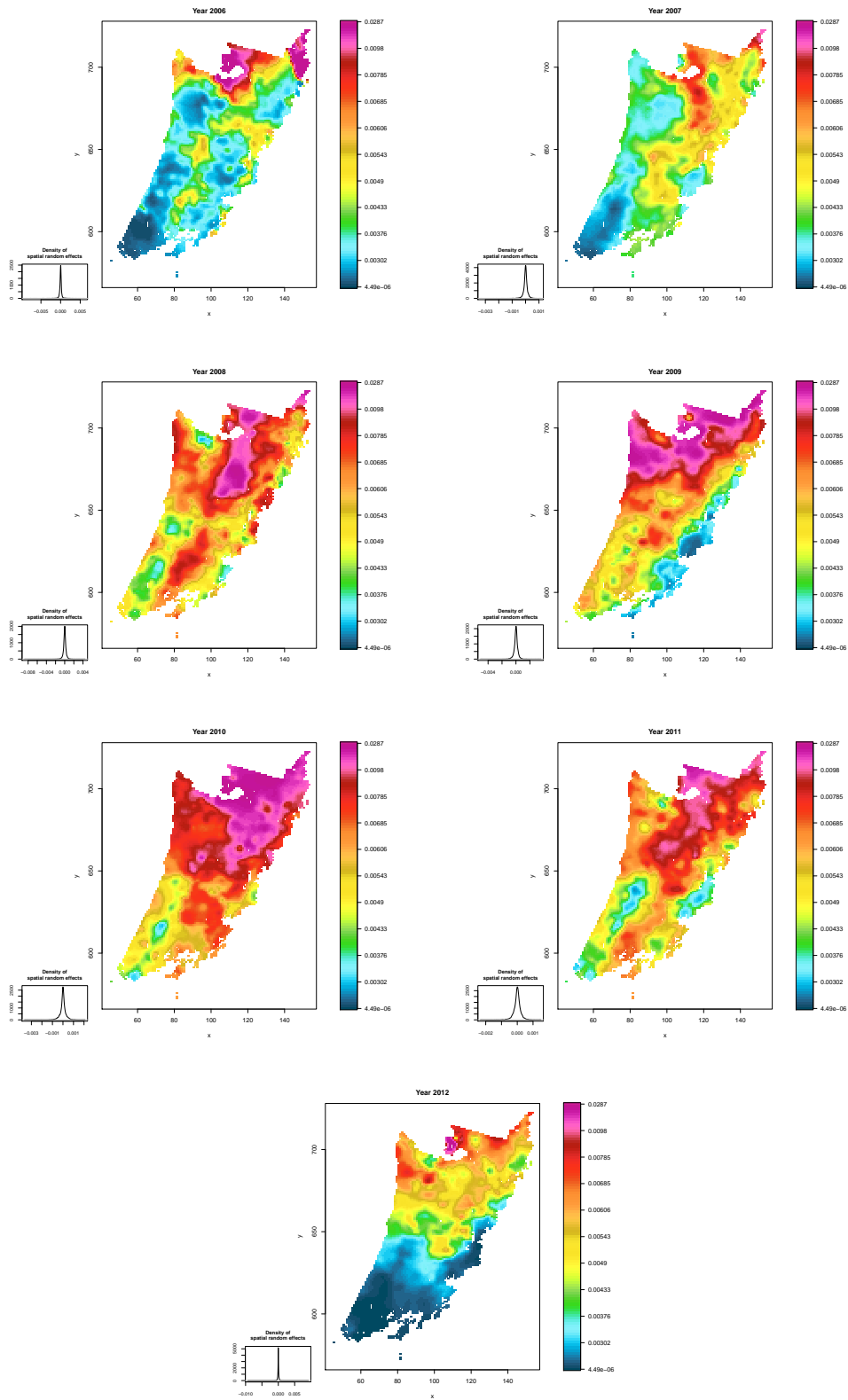


Figure 5.7: Smoothed estimates of the parameter ζ of the generalised Bass model, 2006 – 2012.

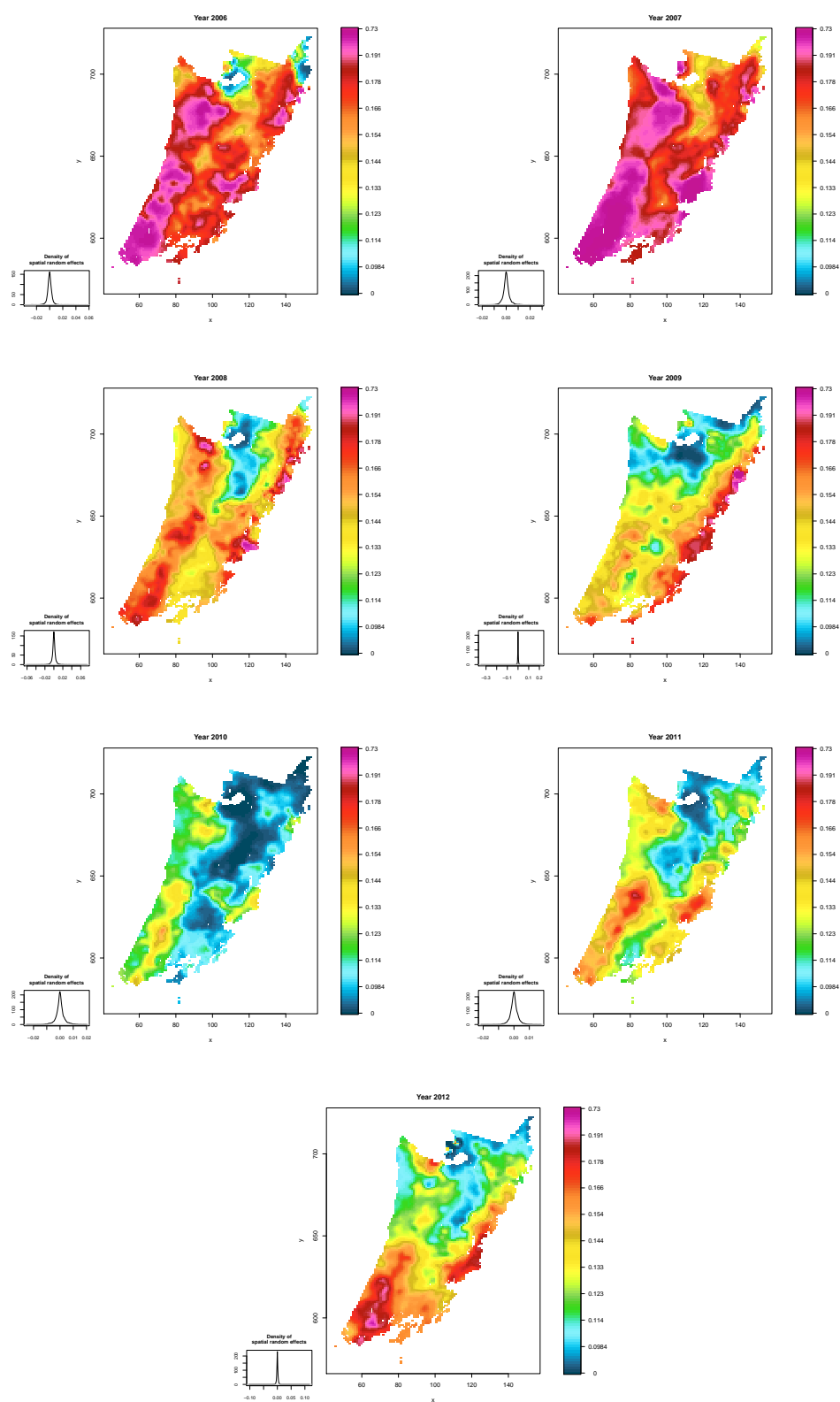


Figure 5.8: Smoothed estimates of the parameter ξ of the generalised Bass model, 2006 – 2012.

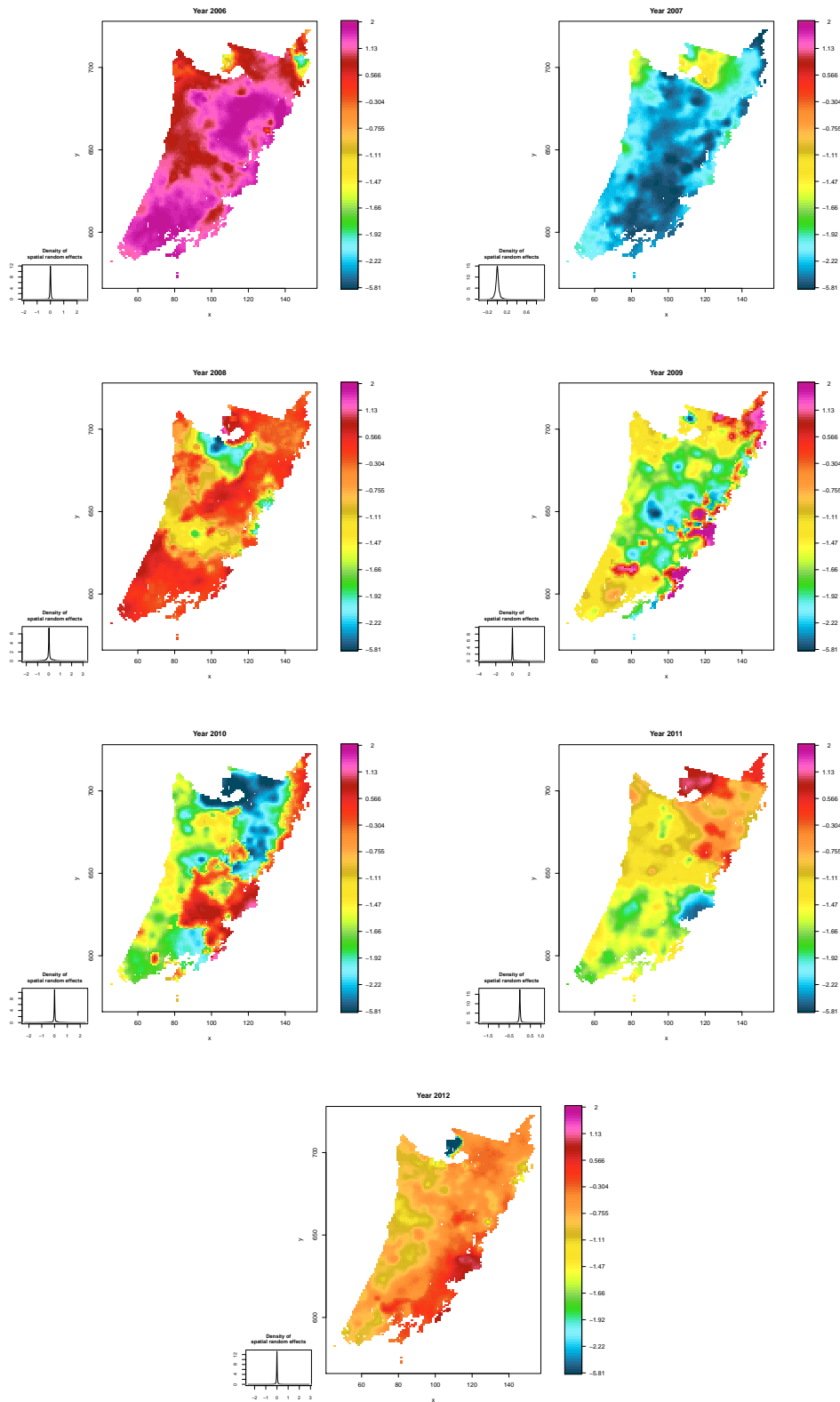


Figure 5.9: Smoothed estimates of the parameter α of the generalised Bass model, 2006 – 2012.

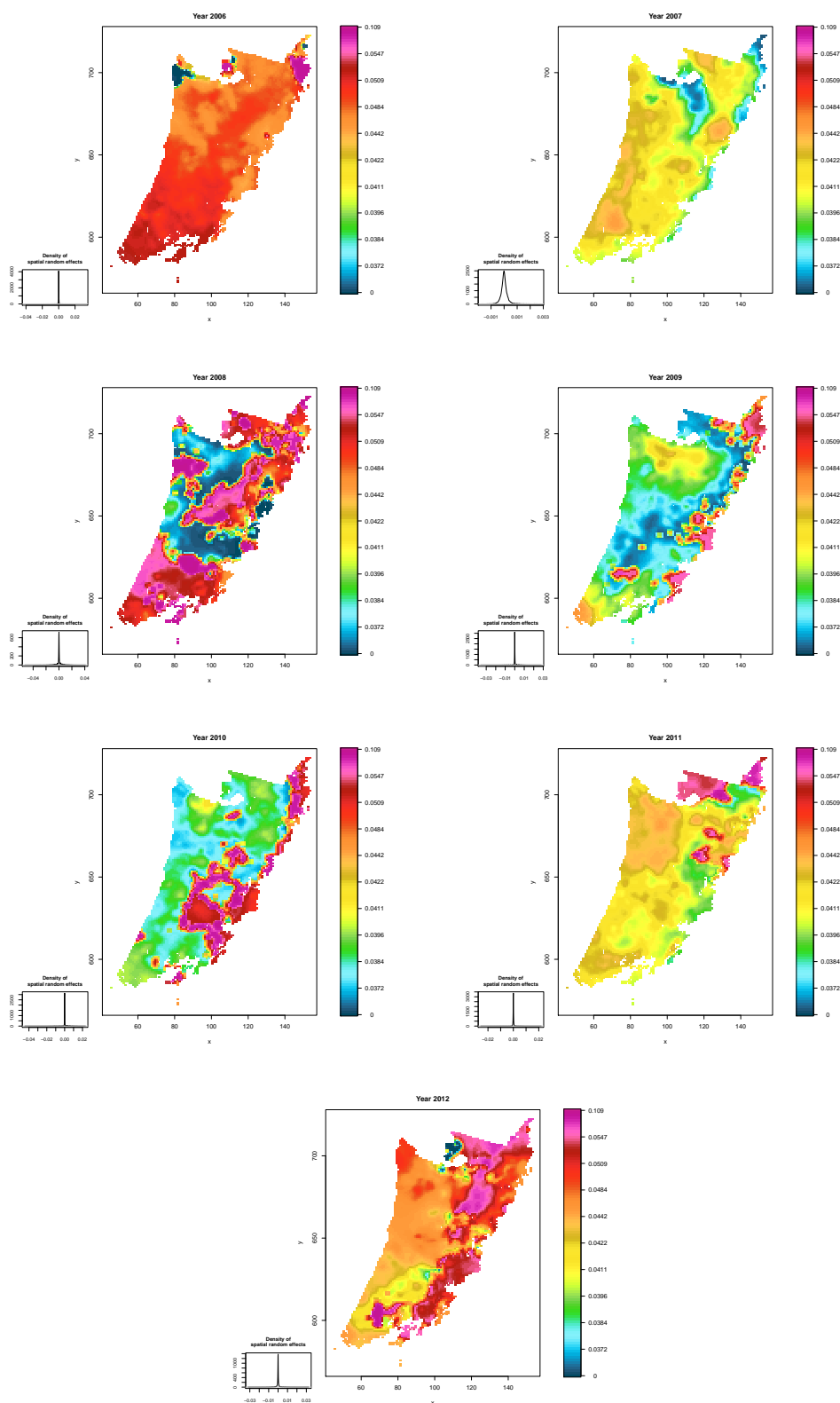


Figure 5.10: Smoothed estimates of the parameter β of the generalised Bass model, 2006 – 2012.

Chapter 6

Quantile regression analyses

Quantile regression is a powerful technique which can identify the conditional quantile of a response variable, so that it is possible to make inference on the distribution of $Y|X = x$ without introducing assumptions on the shape of the residual density. The estimation method of Koenker and Park (1996) is usually applied to perform nonlinear quantile regression for each separate probability level. However, the semi-parametric quantile sheet regression procedure described in Section 3.2.3 is used for the estimation of simultaneous quantile curves. In particular, non crossing curves for any probability level can be obtained by the use of this estimator. By the introduction of nonparametric functional parameters and a penalty term in the objective function, the whole conditional distribution of the response variable can be characterised by the chosen nonlinear model.

Section 6.1 discusses the Bass model extension for semi-parametric quantile regression, practical regression details (e.g. the initialisation), and the presentation of the results. A similar extension for the generalised Bass model and the description of results are considered in Section 6.2. Brief conclusions are reported in Section 6.3.

6.1 The Bass model

The Bass model, formulated for the accumulated rates of adoption, was considered in Chapter 5 in order to stabilise and improve the accuracy of the parameter estimates. Unfortunately, this way to operate strongly affects the results of quantile regression; in fact, as previously mentioned, the accumulated rates of adoption are less noisy than the instantaneous rates. Of course, if the aim of the analysis is concerned with the cumulative quantities, then quantile regression must be performed for the regression model in (2.1). On

the other hand, if the analysis is based on explaining the instantaneous rates of adoption, then the model in (5.1) must be considered. From a mathematical point of view, the instantaneous rates can be decomposed in two additive terms, i.e.

$$y'(t) = g(t) + e(t),$$

where $g(t)$ is a generic deterministic signal to identify and $e(t)$ is a generic noise term, whose distribution is centred in zero. It is quite evident that the cumulative quantities can be expressed as

$$y(t) = G(t) + \int_0^t e(z) dz,$$

where $G(t)$ is given by the definite integral of $g(t)$, while the last additive term is a stochastic diffusion process. From these two expressions, it is possible to establish that the two additive noise terms have two different distributions. Making inference via quantile regression on the distribution for one of these two formulations, it is not immediate to find a solution for the other, whether it exists. Here, since the focus is on the instantaneous precipitation rates, the model in (5.1) is considered.

In order to obtain a semi-parametric quantile sheet model, all the parameters considered here are varying with the probability level τ . The functional parameters, i.e. $\mu(\tau)$, $\xi(\tau)$ and $\zeta(\tau)$, are formulated through the modified Bernstein polynomials as defined in (3.23). Also for quantile regression, the local spatial variations are ignored during regression in order to achieve a compromise between the computational efficiency and estimates accuracy. In practise, the initialisation is based on the following setup:

- $\mu_i(\tau) = \hat{\mu}_i$,
- $\zeta_i(\tau) = \hat{\zeta}_i$,
- $\xi_i(\tau) = \hat{\xi}_i$,

where $\hat{\mu}_i$, $\hat{\zeta}_i$ and $\hat{\xi}_i$ are the least squares estimates related to the i -th cell.

The extension of the Bass model with random parameters adopted to produce spatial dependent estimates must further be generalised in this context. In particular, crossing quantile curves could be obtained when the application of the local average in (5.2) to the coefficients is used to compute each functional parameter. Therefore, it is preferred to consider random functional parameters, so that $\hat{\mu}_i(\tau) = \bar{\mu}_i(\tau) + \varrho(\tau)$, where the non stochastic spatial patterns $\bar{\mu}_i(\tau)$ are identified by

$$\bar{\mu}_i(\tau) = \frac{1}{1 + |\mathcal{N}_i|} \left\{ \hat{\mu}_i(\tau) + \sum_{j \in \mathcal{N}_i} \hat{\mu}_j(\tau) \right\},$$

and the function $\varrho(\tau)$ is a noise term which satisfies $E[\varrho(\tau)] = 0$ for all $\tau \in [0, 1]$.

The Figure 6.1, 6.2 and 6.3 illustrate the maps of parameters for nine probability levels and their variations through the years. Although other graphical representation can be drawn to express the difference among quantile curves, it seems better to provide the maps for the parameter values in order to assess which effects have more influence on the conditional distribution of the response variable. Furthermore, a more informative perception of the physical process can be adopted to describe extreme events.

Figure 6.1 shows the smoothed maps for the functional parameters $\mu(\tau)$ evaluated for $\tau = 0.1, \dots, 0.9$. Across the years, the maps are very stable as it was observed for the estimates obtained by least squares methods. It is also evident that the largest amount of rainfall are located in the most elevated cells. However, the functional parameter rises as the probability level increases, and its largest variations are related to the cells where the overall amount of precipitation is the largest.

Figure 6.2 shows the smoothed maps for the functional parameters $\zeta(\tau)$ evaluated for $\tau = 0.1, \dots, 0.9$. As for the least squares estimates, the results obtained through quantile regression do not present the same spatial stability of $\mu(\tau)$. Nonetheless, it is still possible to see how the human activities and wild vegetation areas exhibit the largest estimated values. Furthermore, most of the estimated functional parameter $\zeta(\tau)$ is stable as τ varies. Essentially, it may be considered almost constant, but it is characterised by slight increments as τ increases. Thus, this parameter has not so strong effects on the conditional distribution of the response variable, and it is essential to control the shape of the estimated quantile trajectories.

The smoothed maps of the functional parameter $\xi(\tau)$ are shown in Figure 6.3. These results are more stable than those obtained for $\zeta(\tau)$, but not as much as for those calculated for $\mu(\tau)$. In particular, as for the least squares estimates, the mountainous areas have higher parameters values. However, this functional parameter has relevant effects on the conditional distribution of the response variable, as well as on the estimated quantile trajectories. In fact, even if the maps drawn from 2006 to 2009 exhibit few variations according to different values of τ , the three successive years are characterised by more evident changes.

6.2 The generalised Bass model

The generalisation formulated in (2.2) is adopted in order to extend the Bass model for accumulated rates of adoption. Here, for the same reasons

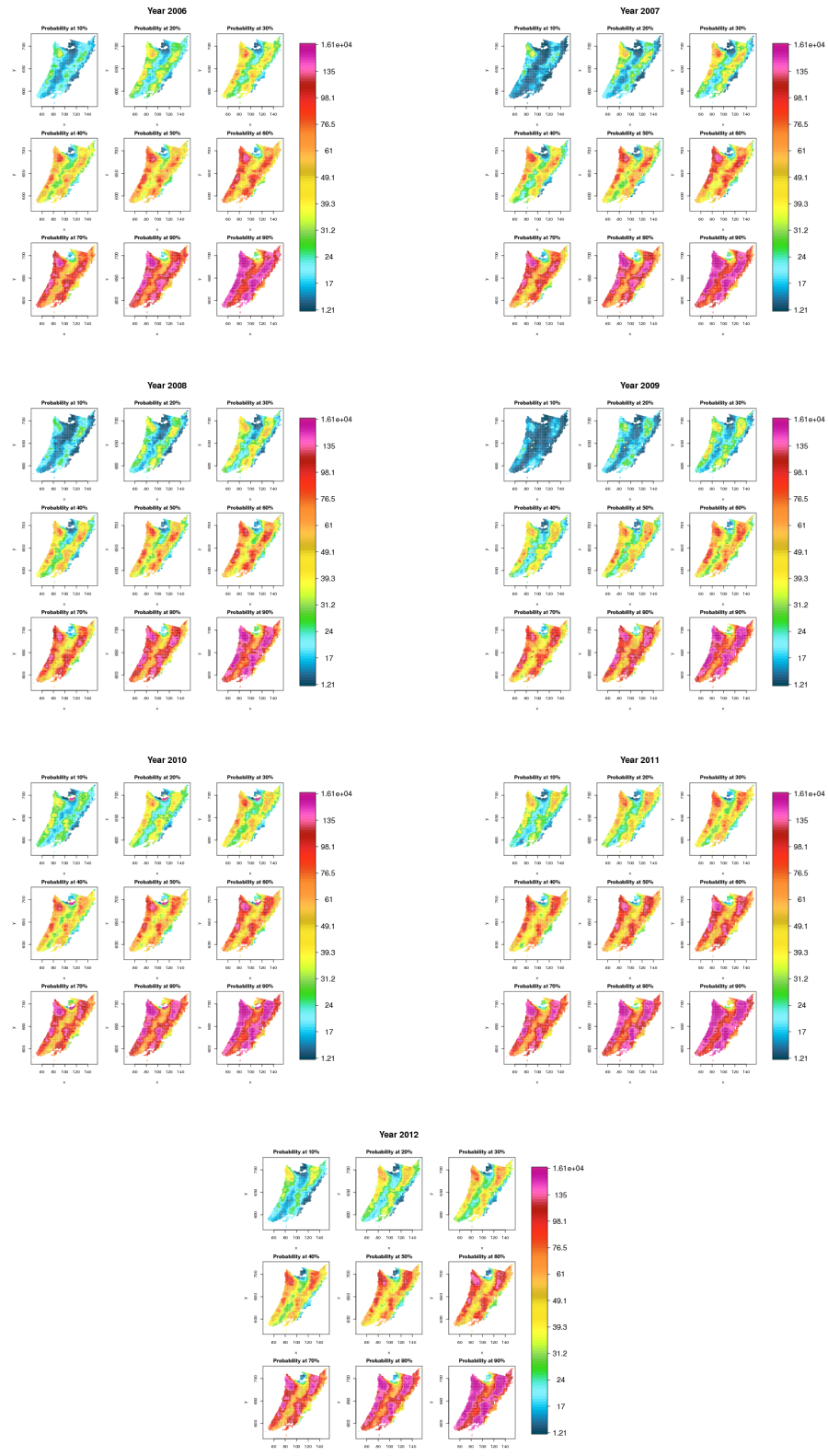


Figure 6.1: Smoothed quantile estimates of the parameter μ of the Bass model, 2006 – 2012.

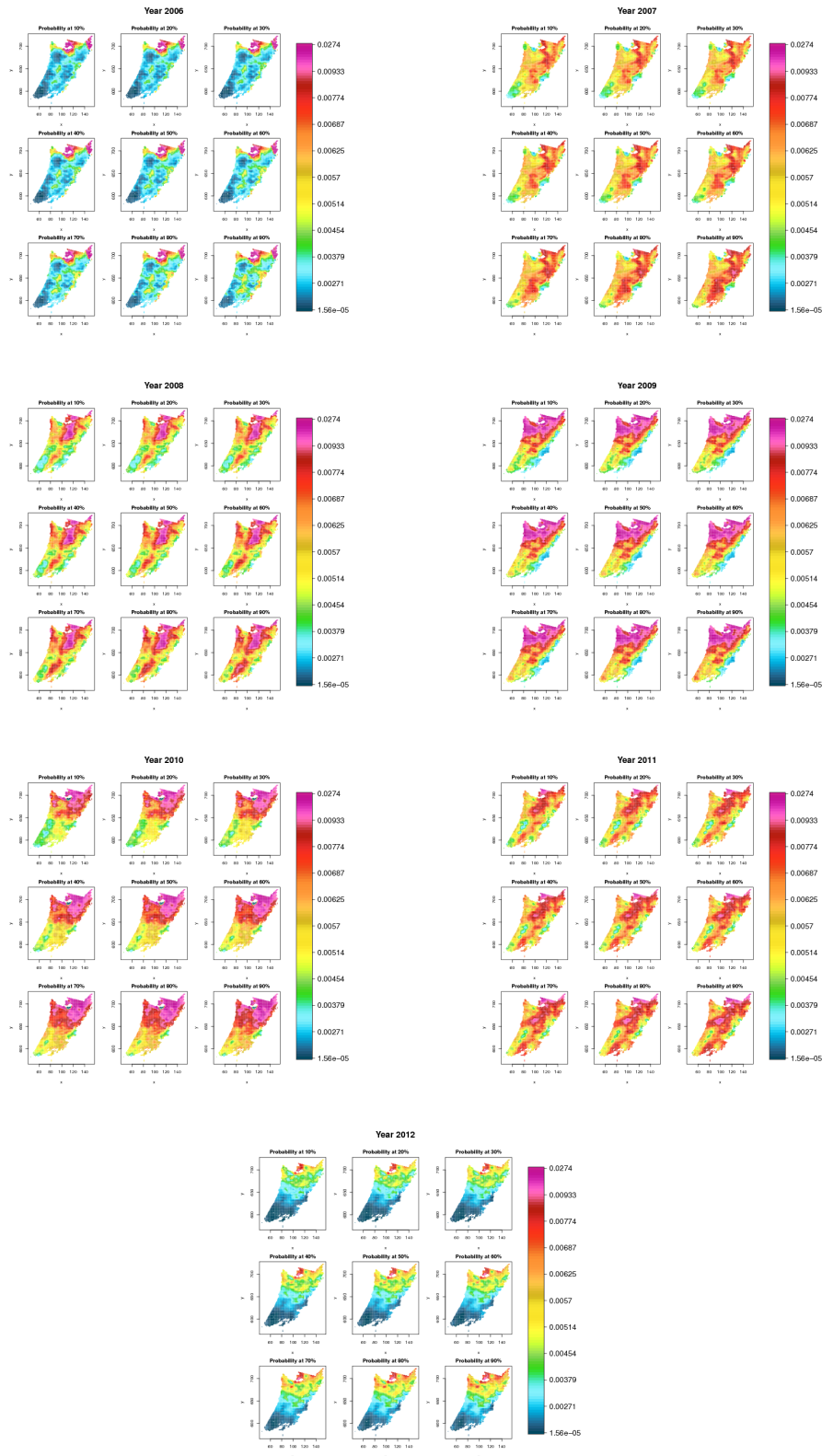


Figure 6.2: Smoothed quantile estimates of the parameter ζ of the Bass model, 2006 – 2012.

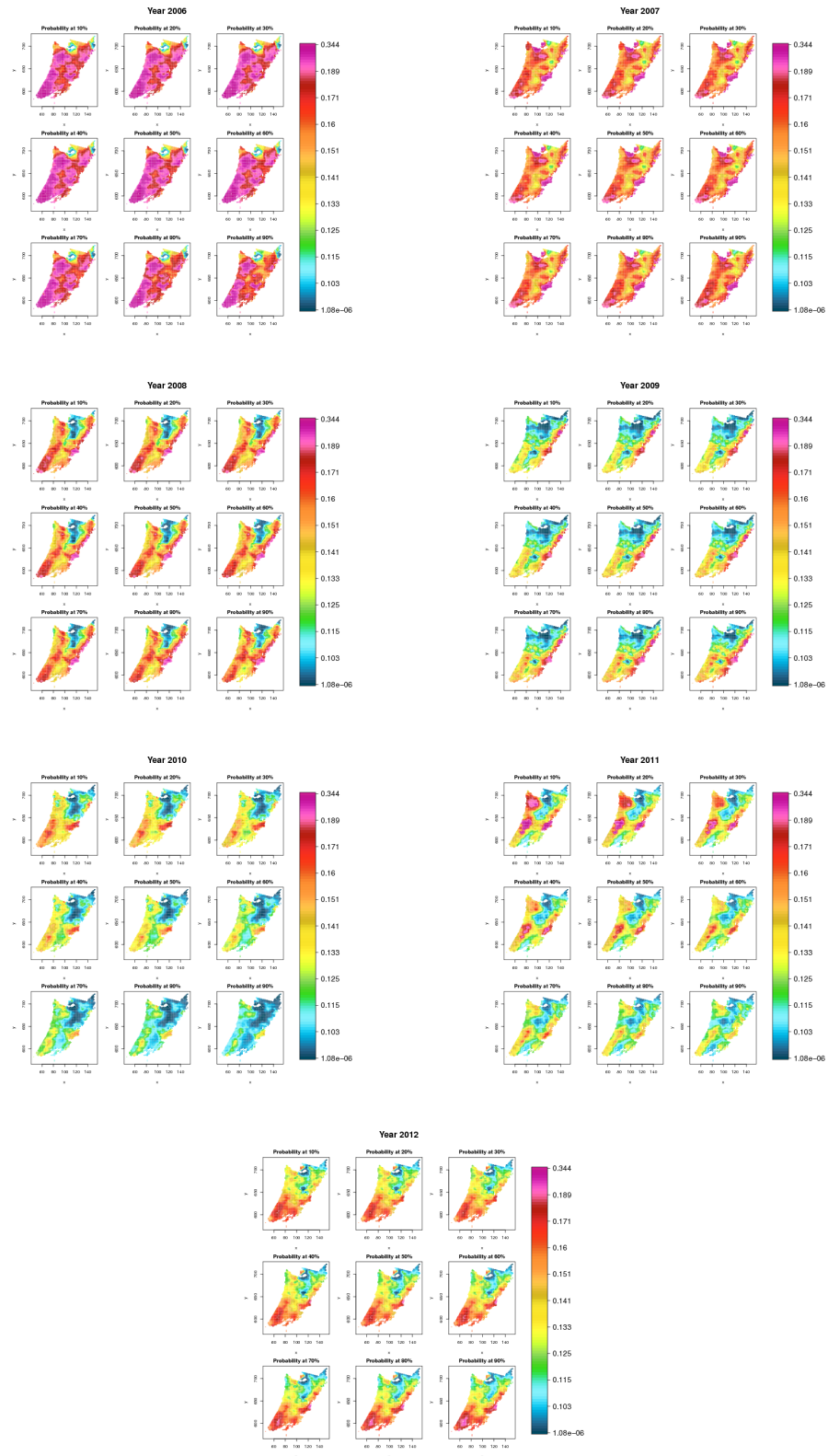


Figure 6.3: Smoothed quantile estimates of the parameter ξ of the Bass model, 2006 – 2012.

mentioned in Section 6.1, the generalised Bass model for the instantaneous precipitation rates is considered in order to apply quantile regression. By taking the first derivative with respect to the time t the resulting model is the following:

$$y'(t) = \frac{\mu \zeta (\zeta + \xi)^2 \{1 + 2\pi\alpha\beta \cos(2\pi\beta t)\} \exp \left[(\zeta + \xi) \{t + \alpha \sin(2\pi\beta t)\} \right]}{\left(\zeta \exp \left[(\zeta + \xi) \{t + \alpha \sin(2\pi\beta t)\} \right] + \xi \right)^2}.$$

As for the Bass model, a suitable model for semi-parametric quantile sheet regression can be formulated by replacing each parameter with its functional form based on the modified Bernstein polynomials in (3.23), and the regression is initialised with the following setup:

- $\mu_i(\tau) = \hat{\mu}_i$,
- $\zeta_i(\tau) = \hat{\zeta}_i$,
- $\xi_i(\tau) = \hat{\xi}_i$,
- $\alpha_i(\tau) = \hat{\alpha}_i$,
- $\beta_i(\tau) = \hat{\beta}_i$,

where $\hat{\mu}_i$, $\hat{\zeta}_i$, $\hat{\xi}_i$, $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the least squares estimates for the generalised Bass model related to the i -th cell.

The smoothed maps of the functional parameter of the generalised Bass model are reported in Figure 6.4, 6.5, 6.6, 6.7 and 6.8. They are obtained by applying the same techniques performed for the Bass model in the previous section.

The results for the functional parameter $\mu(\tau)$ are shown in Figure 6.4. These maps are very similar to those produced for the Bass model (see Figure 6.1 for comparison), even if very few slight differences are also recognised in some small areas. Essentially, the process is spatially stable, and this denotes a strong relationship with the orography as noted previously. In addition, the variations of this parameter through quantiles have a strong impact on the conditional distribution of the rainfall, because they quantify the variability of the response. From a long-term perspective, the functional parameter are not affected by strong fluctuations; however, it is possible to observe interesting growth during the last three years.

The smoothed maps for the functional parameter $\zeta(\tau)$ are shown in Figure 6.5. These maps present the same spatial characteristics of those produced for the Bass model, and this can be interpreted as a localised effect

which is connected with the human activities. As it was found in the previous chapter, the northern areas have the highest estimated values; in fact, in those zones deliquescent aerosol particles are produced and, successively, released in the atmosphere. Moreover, by looking at the variations across quantile levels, it is possible to notice a slightly significant increasing trend.

Figure 6.6 shows the smoothed maps for the functional parameter $\xi(\tau)$. As for the Bass model, these maps have a strong spatial stability, which is mostly related to the physical interpretation of the parameter ξ ; in fact, the coalescence effect is strictly affected by thermodynamical aspects, which are more intense near the elevated positions (e.g. wider atmospheric temperature ranges). More evident patterns are highly visible across probability levels, which permit to quantify the contribution of the coalescence effect. In other words, since the considered generalised Bass model take into account seasonal effects, high values of $\xi(\tau)$ increase the magnitude of initial oscillations, while low vales give more emphasis to the final fluctuations. However, these effects on the seasonality are strictly linked with the values assumed by $\zeta(\tau)$. By looking at the maps, and considering the role of the nonlinear interaction between the coalescence and the seasonality effects, it is quite evident that the conditional distribution of the observed phenomenon manifests long-term changes.

Figure 6.7 and 6.8 expose the smoothed maps for the functional parameters $\alpha(\tau)$ and $\beta(\tau)$ respectively. These functional parameters are less stable than the previous. In fact, these maps do not show any spatial persistence neither evident long-term evolutions. Nonetheless, it is interesting to note that these functional parameter are stable across quantiles. This issue is mainly due to the usage of the penalty for the estimation of the functional parameters.

6.3 Conclusion

In this chapter, regression analyses are also performed by considered the main models adopted in the previous chapter. However, the new procedure based on the semi-parametric quantile sheets is utilised to assess the distribution of the data. Practical details are described in order to understand how the algorithm should be initialised for a stable regression setting in order to obtain reliable results.

The results show how the empirical distribution of each cell differs from the other according to the values of the parameters and their variations with respect to the probability level. By performing these analyses, it is also possible to study the trajectory shapes of the extreme phenomena and

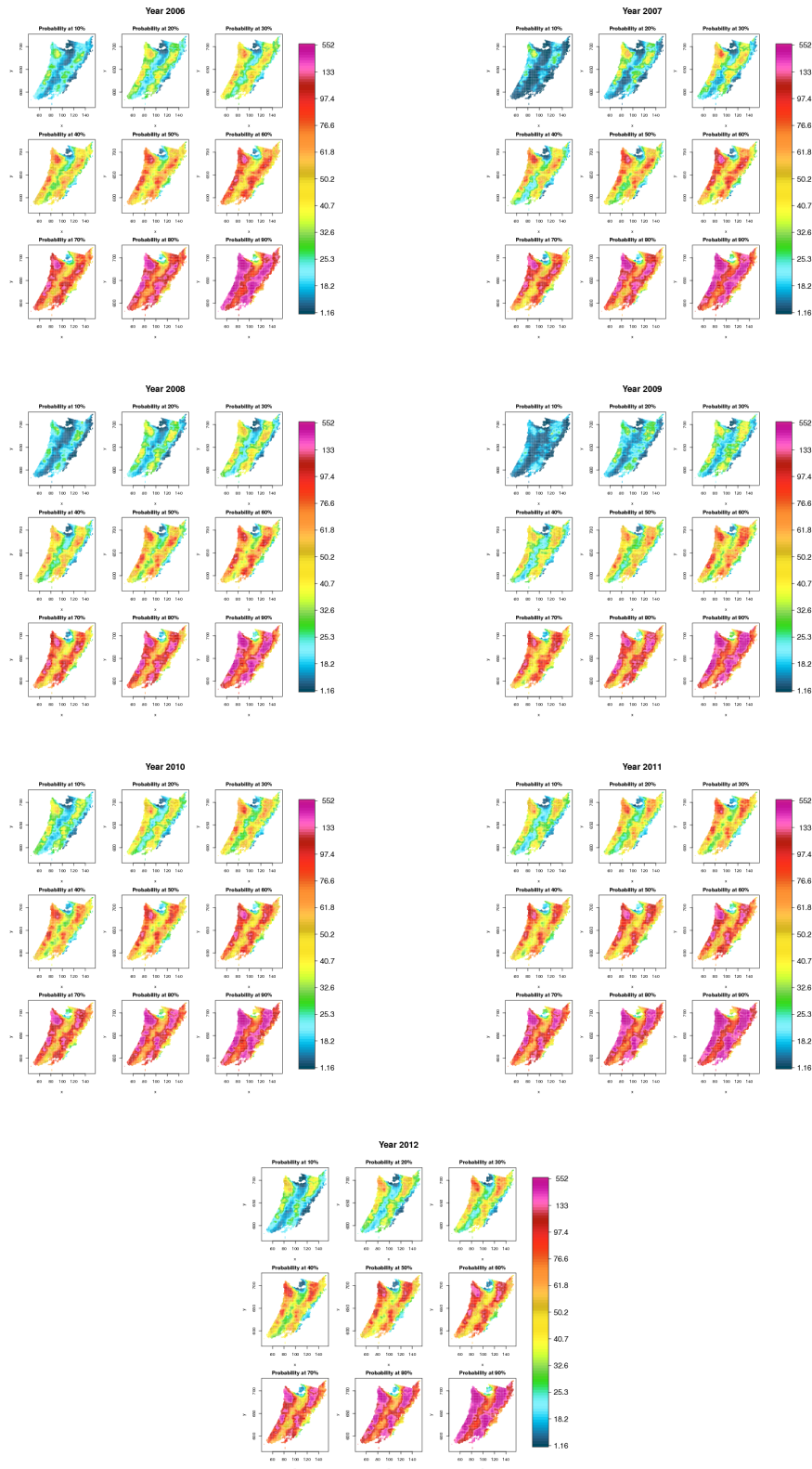


Figure 6.4: Smoothed quantile estimates of the parameter μ of the generalised Bass model, 2006 – 2012.

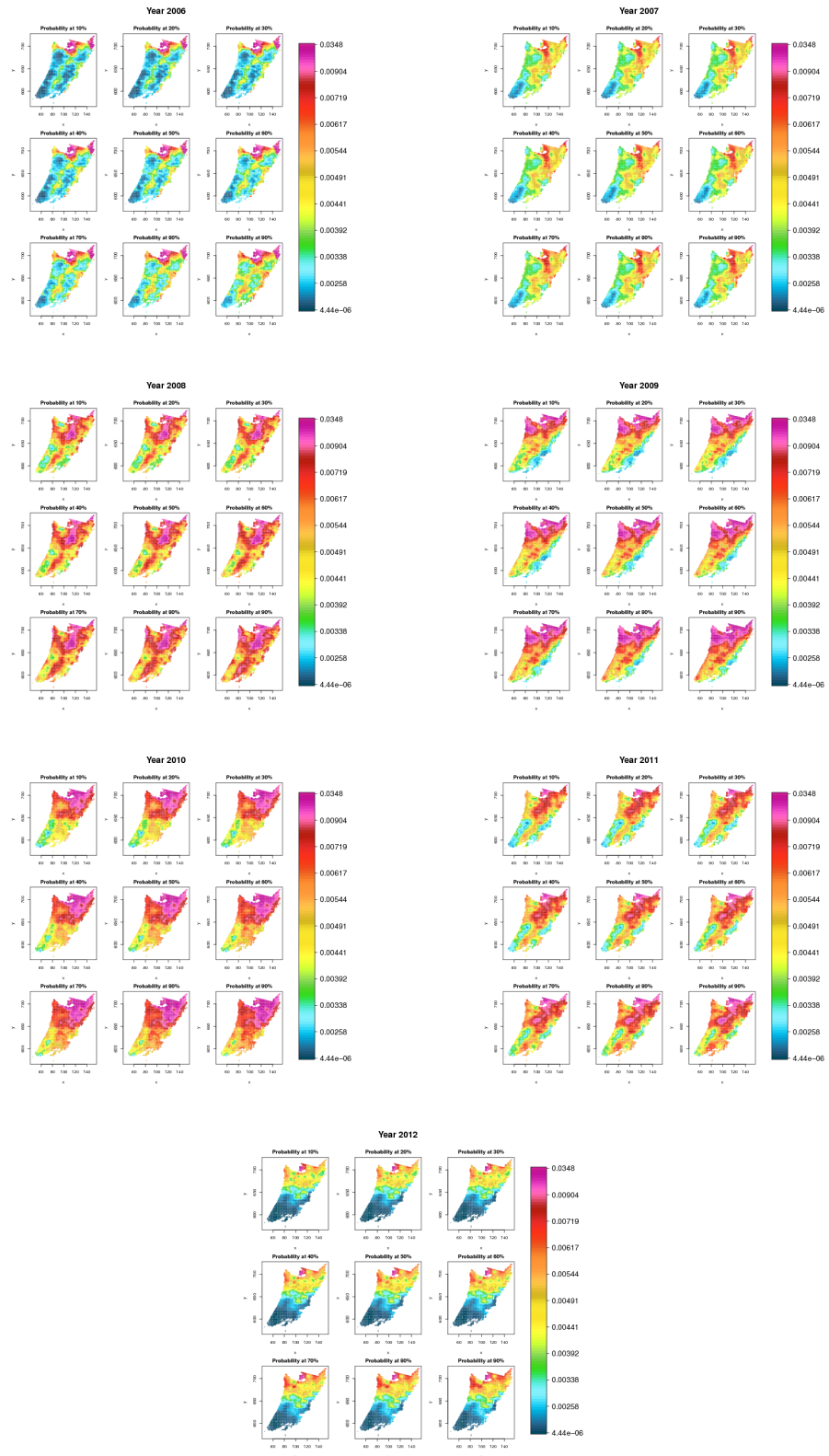


Figure 6.5: Smoothed quantile estimates of the parameter ζ of the generalised Bass model, 2006 – 2012.

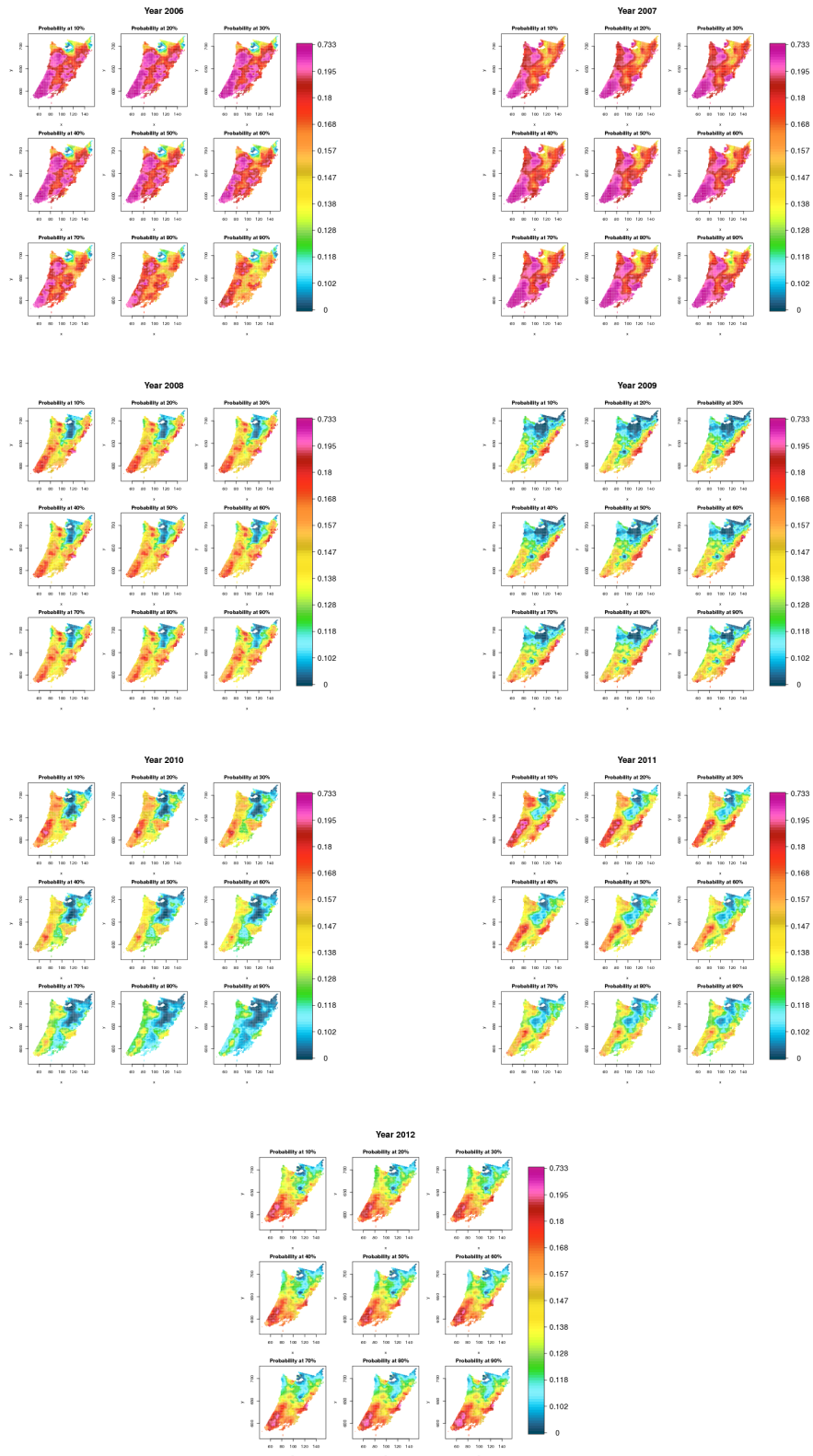


Figure 6.6: Smoothed quantile estimates of the parameter ξ of the generalised Bass model, 2006 – 2012.

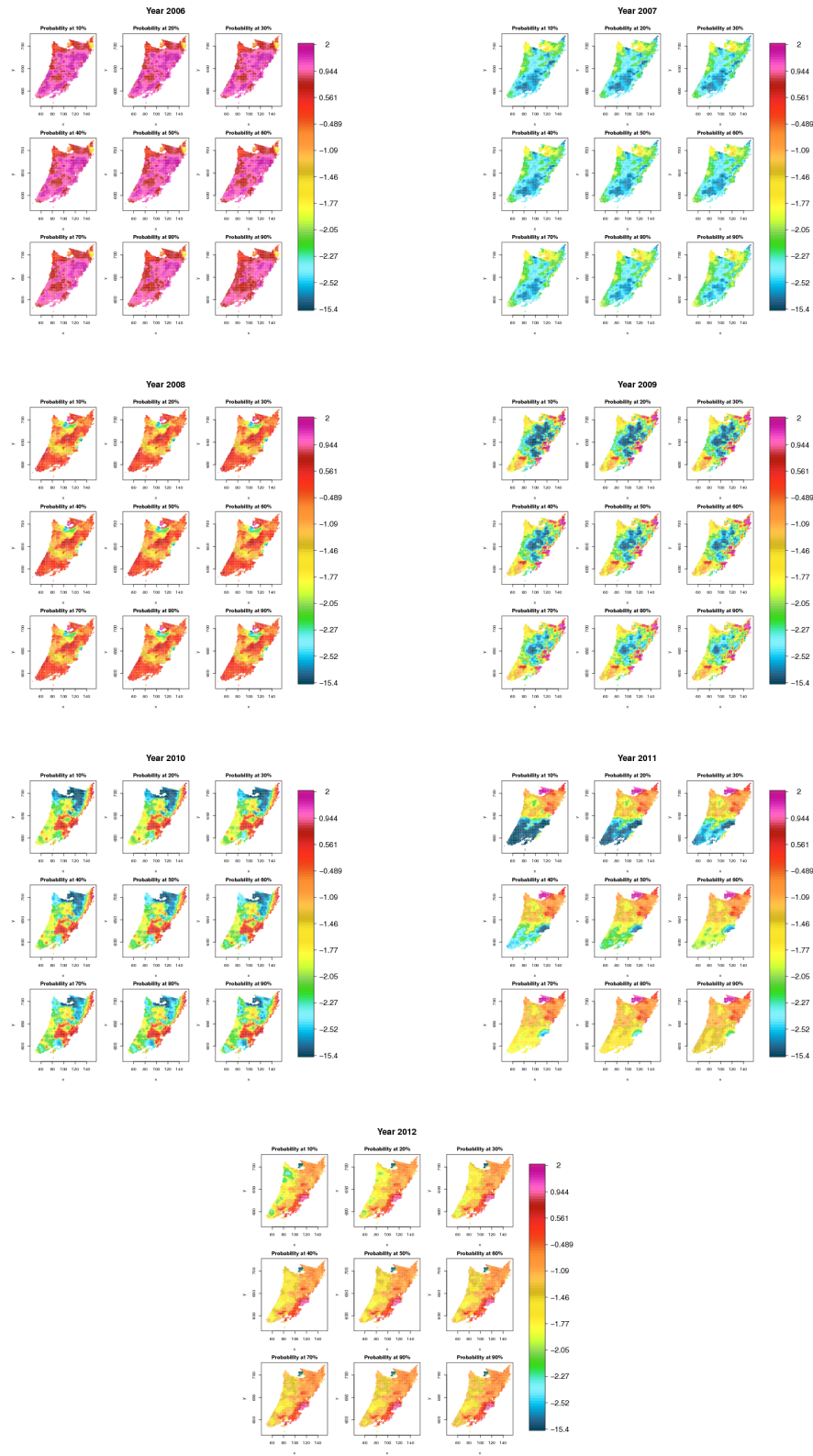


Figure 6.7: Smoothed quantile estimates of the parameter α of the generalised Bass model, 2006 – 2012.

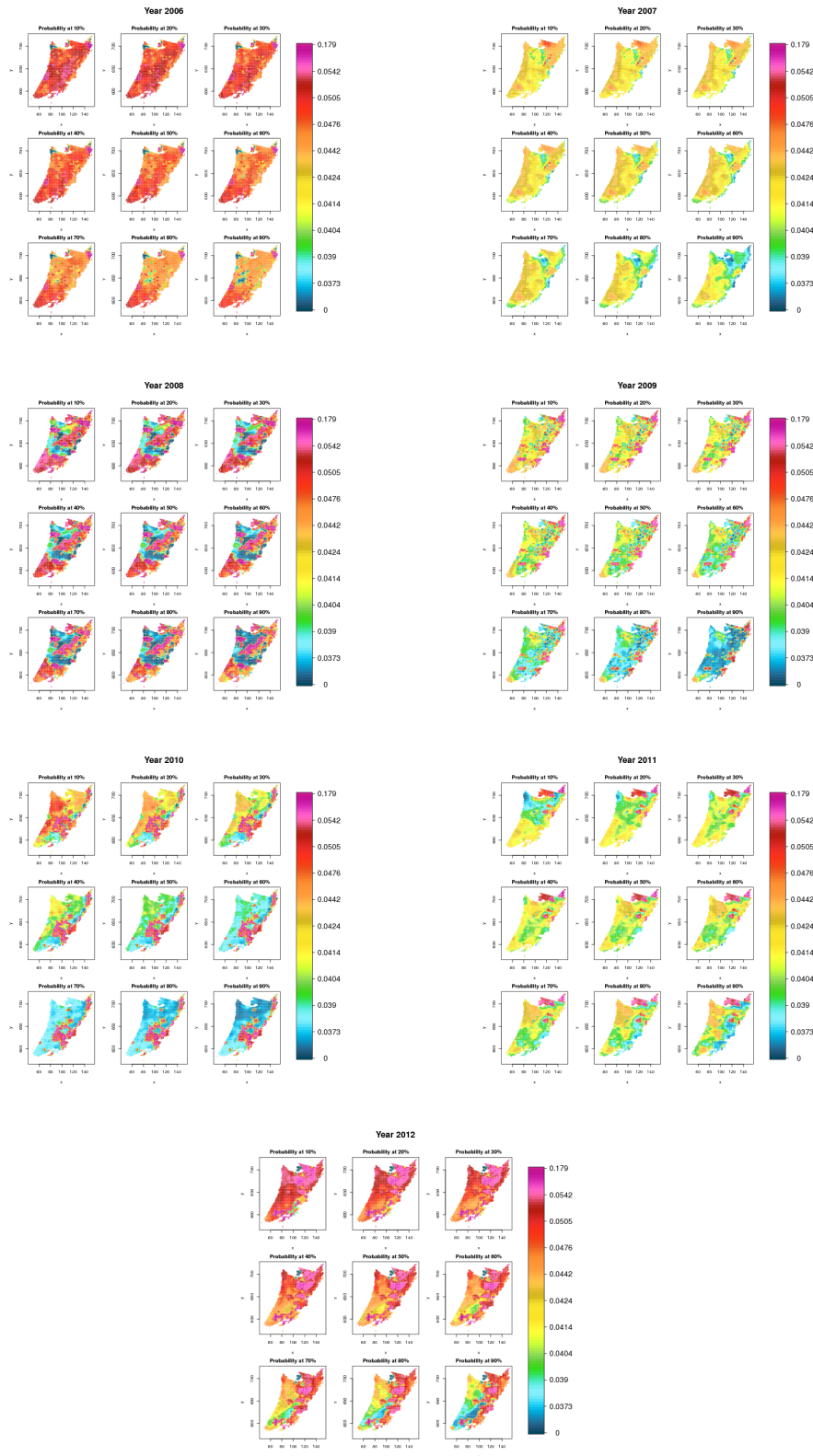


Figure 6.8: Smoothed quantile estimates of the parameter β of the generalised Bass model, 2006 – 2012.

establish which parameter is more sensitive to the distributive variability. Although the features of the estimates remain almost the same as in the previous chapter, a more detailed description of the rainfall process and its probabilistic characterisation is given here by the use of quantile regression.

Chapter 7

Conclusions

Spatiotemporal phenomena related to the rainfall measurements can be characterised by physical models, this can be an alternative to the identification of spatiotemporal patterns based on standard analytical tools, which mainly focus on explaining spatiotemporal correlations via nonparametric regression models or variograms (or correlograms) functions for stochastic processes. The physical processes generating the precipitation are considered within a simplified physical framework, where the mathematical formulation of a regression model plays a central role. In so doing, the spatiotemporal dependence of the data is strictly linked to the chosen model, which explains the complexities of the phenomena. From this perspective, the temporal evolution of the spatial patterns is given by the model, which takes into account the potential amount of rainfall, the presence of the deliquescent particles, the coalescence which forms the raindrops and other effects (such as seasonality). All these aspects are well-known in the hydrological literature and, in the recent years, most of the work related to the formation of the precipitation focus on the clouds microphysics, wherein the active-nuclei of the aerosols and the thermodynamical nature of the coalescence are central concepts. In particular, the principal aim of the performed analyses arrives from the necessity of understanding if the relations observed among neighbouring cells and consecutive years are attributable to explicable physical mechanisms.

Most of the data used to assess this issue are collected daily through satellites; therefore, they can be organised as time series with a location index regarding a specific spatial cell. Such a multitude of data requires analytical tools which must have high performance features, e.g. parallel computing and effective estimation algorithm. Nonetheless, it is still necessary to focus on a reduced area and lose some of the temporal resolution. The data organisation in homogeneous areas is obtained via cluster analysis, which is also useful to

identify the most suitable zones according to their temporal pattern, while further adjustments are attained by weekly aggregation. This allows an affordable amount of data to be considered in order to achieve optimal results in reasonable time.

The purpose of these analyses mainly concerns simplified physical mechanisms which are essential for the explanation of observed correlations. Originally, the hydrological theory developed by Koren and Feingold (2011) is based on the concept of self-organisation, which characterises several natural, social and economical systems. The regression models inspired by the diffusion of innovations are introduced to approximate the evolution of the rainfall process within a year through a more straightforward perspective. Such models and the adopted estimation techniques allow to quantify the basic effects associated with the complexities of the process.

The phenomenon in the considered area is mostly originated by condensation of water vapour mainly due to the orography of the territory. However, this thermodynamical point of view can be studied from a more challenging perspective, which also takes into account the microphysics of clouds. In practical terms, although the innovation diffusion models are a theoretical approximation of these microphysical processes, they can quantify and separate two main effects: the initialisation due to the aerosol moisture in clouds, and the reproduction of the coalescence mechanisms. A graphical representation of the estimates highlights the areas where these effects are more relevant. In particular, such regions have meaningful connections with highly populated zones and elevation measurements. In addition, the innovation diffusion models can also identify the total rainfall observed during the year. Other features of the data (e.g. inner year seasonality) are also considered by applying some generalisation of the standard Bass' model (1969; Bass *et al.*, 1994).

The adopted inferential tools are based on regression techniques, in particular the nonlinear least squares and innovative nonlinear quantile regression techniques are here considered. The former is a standard tool to identify and assess the expectation of the conditional response variable, $E[Y|X = x]$, without forcing assumptions on the error distribution. The latter regression method is introduced to increase the computational efficiency and make inference on the conditional distribution of the response given the covariates. While the former method focuses on the conditional expectation, the latter concerns with the conditional quantile of the response variable given a specific probability level. In so doing, it is possible to study the distribution of the rainfall in a specific space-time position, which in turn allows for the study of the local variations of the observations, and the behaviour of the model parameters in correspondence with probability levels. For instance,

the model estimated for a high probability (e.g. 0.95) characterises extreme values of heavy rainfall, while the same model with parameters estimated for a low probability level (e.g. 0.05) identifies trajectories which are adequate for lower precipitation levels. However, since the standard techniques for non-linear quantile regression developed by Koenker and Park (1996) does not take into account the position other curves to fit, suitable regression methods are investigated. The semi-parametric quantile sheet algorithm based on the concept introduced by Schnabel and Eilers (2013) can be applied in order to achieve optimal results. In fact, the resulting curves are not intersecting, so that they respect the monotonicity property of quantiles. Therefore, it is possible to reconstruct the conditional cumulative distribution of $Y|X = x$ for any probability level by estimating only few coefficients.

In practical terms, the regression techniques applied to the data highlight strong connections with the salient features of measurement area. The evidence of this fact is mainly visible on the maps of the parameters, e.g. the expected total amount of precipitation is greater near the most elevated zones. From these analyses, the effects of aerosol particles appear to be more evident and highly linked with urban areas, as well for the coalescence which is stronger where the wind-advection is more turbulent. The introduction of further parameters for the inner year seasonality stabilises the estimated values of the basic parameters (i.e. the total amount of precipitation, the aerosol effects and the coalescence).

From a graphical perspective, the major distinction between the results obtained through the nonlinear least squares and quantile regression is clearer when the estimates of the parameters are significantly varying according to the probability levels. This fact is more noticeable for the parameter controlling the amount of yearly rainfall, while the other parameters are quite constant across the probabilities, even if little variations are necessary and statistically significant for controlling the evolution of the trajectories. The resulting parameters maps confirm the results obtained via nonlinear least squares; however, quantile regression methods can quantify the intrinsic variability of the observed data and describe more details of the underlying dynamics.

More sophisticated analyses can be performed in order to increase the precision of the model in short-term forecasts. By the introduction of the extension of SARMAX models (Guseo and Dalla Valle, 2005), it is possible to get more accurate estimates of the conditional expectations, where the residuals are modelled via the usual linear time series methods. However, even if this concept can be applied to quantile regression techniques, it requires more methodological efforts because it is not so evident how to deal with the residuals dynamics. In particular, while quantile autoregression is

not a novelty, the MA processes in a quantile regression context are still an unexplored research field. Other interesting research opportunities consists of the study of long-term parameters evolution. While proper methods are well-established for the analyses of repeated functional time series, it is more challenging to extend these techniques for quantile-sheets, where the unknown future parameter estimates must satisfy the monotonicity property of quantiles. Other methodological research perspectives are concerned with the evaluation of the variability of the estimates and trajectories, especially under space-time dependence assumption of the data.

Appendix A

Notes on quantile regression

A.1 Proof of the homogeneity property

The proof of the Property 1 is quite obvious. Since the scalar w is a positive real number, the inequality $wz < 0$ is maintained for all $z < 0$. Conversely, it is also true that $wz \geq 0$, if and only if $z \geq 0$. Therefore, the resulting values of the function $\rho_\tau(wz)$ can be computed as

$$\rho_\tau(wz) = \begin{cases} wz\tau, & \text{if } z \geq 0, \\ wz(\tau - 1), & \text{if } z < 0, \end{cases}$$

wherein the conditions are expressed only for z and not for the product wz . It is evident that, by factorising the scalar w , it is possible to get the equality

$$\rho_\tau(wz) = w\rho_\tau(z)$$

and this completes the proof.

Also the check function $\rho_\tau^*(\cdot)$ satisfies the homogeneity property and the proof it is obvious.

A.2 Proof of quantile level identification

When the check function $\rho_\tau^*(\cdot)$ in (3.21) is adopted to estimate the model coefficients, it means that the solution of the decision problem

$$\min_{\nu(x;\tau)} \mathbf{E} [\rho_\tau^*\{(Y|X) - \nu(X; \tau)\} | X = x],$$

does not coincide with the τ quantile level. As it was done before, a model not depending on x can be considered without any loss of generality, hence

an equivalent problem can be formulated as

$$\min_{\nu_\tau} \mathbf{E} \{ \rho_\tau^*(Y - \nu_\tau) \},$$

where the expectation is defined as

$$\begin{aligned} \mathbf{E} \{ \rho_\tau^*(Y - \nu_\tau) \} &= \int_{-\infty}^{+\infty} \rho_\tau^*(Y - \nu_\tau) dF_Y, \\ &= (1 - \tau)^2 \int_{-\infty}^{\nu_\tau} (y - \nu_\tau) dF_Y + \tau^2 \int_{\nu_\tau}^{+\infty} (y - \nu_\tau) dF_Y. \end{aligned}$$

By differentiating with respect to ν_τ , the solution of the equation

$$(1 - \tau)^2 \int_{-\infty}^{\nu_\tau} dF_Y - \tau^2 \int_{\nu_\tau}^{+\infty} dF_Y = 0$$

correspond to the desired minimum, when it is unique. By rewriting the previous equation as follows

$$(1 - \tau)^2 \int_{-\infty}^{\nu_\tau} dF_Y = \tau^2 \left(1 - \int_{-\infty}^{\nu_\tau} dF_Y \right),$$

the right quantile level is given by

$$\int_{-\infty}^{\nu_\tau} dF_Y = \frac{\tau^2}{1 - 2\tau + 2\tau^2},$$

and not by τ . Suppose that the quantity

$$\pi_\tau = \int_{-\infty}^{\nu_\tau} dF_Y$$

denotes the c.d.f. of Y computed in ν_τ . By solving the following equation

$$(2\pi_\tau - 1)\tau^2 - 2\pi_\tau\tau + \pi_\tau = 0$$

with respect to τ , it is possible to obtain the inverse function in (3.22), and this completes the proof.

A.3 Proofs of critical values consistency

Let the random variables X_1, \dots, X_n be independent and defined on a bounded support, such that $\Pr(X_i \in [0, 1]) = 1$, for any $i = 1, \dots, n$, the Hoeffding's inequality states that

$$\Pr \left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbf{E}[X_i] \geq n^2 k \right) \leq \exp(-2n^3 k^2),$$

wherein k is an arbitrary positive constant (compare Theorem 2.6 in Hoeffding, 1963).

Since the test statistic considered in region (3.29) is a random variable T with bounded support $[0, 1]$, it may assume that any realisation of T comes from an average process, in this specific case

$$T = 1 - \frac{1}{n} \sum_{i=1}^n X_i.$$

Furthermore, since the expectation of T under the null hypothesis is approximately one, it is easy to note from the previous formula that the expectation of the summation must approximately be zero. For this reason, it may assume that $E[X_i] \approx 0$ for all $i = 1, \dots, n$.

The probability of the region in (3.29) with $\lambda_1 = 1$, can be written as

$$\begin{aligned} \Pr(\lambda_0 \leq T \leq \lambda_1) &= 1 - \Pr(T \leq \lambda_0), \text{ or equivalently as} \\ &= 1 - \Pr(1 - T \geq 1 - \lambda_0), \end{aligned}$$

since T exceed one with probability zero. From the Hoeffding's inequality, the following relationships are true:

$$\begin{aligned} \Pr(1 - T \geq 1 - \lambda_0) &= \Pr(1 - T \geq nk), \\ &\leq \exp(-2n^3k^2), \\ &= \alpha. \end{aligned}$$

From these statements, since $\alpha = \exp(-2n^3k^2)$,

$$k = \sqrt{-\frac{\log(\alpha)}{2n^3}},$$

and since $1 - \lambda_0 = nk$, it obtains the following result

$$\lambda_0 = 1 - \sqrt{-\frac{\log(\alpha)}{2n}},$$

which completes the proof for the threshold in (3.30).

In order to obtain the threshold in (3.31) further assumptions on the variance of T need to be introduced. Let $\sigma^2 \propto n^{-1}$ be the variance of T , such that

$$\sigma^2 = \text{Var}[T] = \text{Var}[1 - T] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i],$$

which is due to the independence of the variables X_i , for $i = 1, \dots, n$. Assuming the homoscedasticity of the hidden variables, i.e. $\text{Var}[X_i] = \text{Var}[X_j]$, for any $i, j = 1, \dots, n$ and, since $\sigma^2 \propto n^{-1}$, it follows that $\sigma^2 = n^{-1} \text{Var}[X_i]$. In other words, the variance of X_i is constant for all $i = 1, \dots, n$.

Under all the previous assumptions, the Uspensky's inequality states that

$$\Pr\left(\frac{1}{n} \sum_{i=1}^n X_i \geq \sigma k\right) \leq \frac{1}{1+k^2},$$

which leads to

$$\Pr(1 - T \geq 1 - \lambda_0) = \Pr(1 - T \geq \sigma k) \leq (1 + k^2)^{-1} = \alpha.$$

From the last equality one solves it for k , so that

$$k = \sqrt{\frac{1}{\alpha} - 1},$$

and, since $1 - \lambda_0 = \sigma k$, by assuming that σ^2 takes the form c^2/n , where c is a positive constant, it obtains the threshold in (3.31) which completes the proof.

Special values of c can be considered. In particular, from the previous assumptions, it follows that $\text{Var}[X_i] = c^2$ and, since $X_i \in [0, 1]$, then $\text{Var}[X_i] \leq 0.25$ for any $i = 1, \dots, n$. Assuming that the distribution of X_i is unimodal and $\text{E}[X_i^3] \approx 0$, then $\text{Var}[X_i] \leq 1/12$, for all $i = 1, \dots, n$ (Dharmadhikari and Joag-Dev, 1989). Thus it is reasonable to set $c^2 = 1/12$.

A.4 Simulation studies

In order to understand the main differences among the estimators presented in the previous sections, Monte-Carlo techniques are used. Data from a linear and a nonlinear model are simulated with different sample sizes and error distributions. In so doing, it is possible to show the main characteristics of each situation. More details on the adopted simulation techniques are explained in the the Appendix A.5.

The obtained results are organised in tables exposed in Appendix A.6. Four indexes are selected to access the properties of the estimators, in particular

the estimates mean, which is computed through the arithmetic average of the estimates, i.e.

$$M_\tau = \frac{1}{N} \sum_{i=1}^N \hat{\theta}_{\tau,i};$$

the mean squared error, which is calculated as the average of the squared deviances between the estimates and true parameter θ (which is known by the simulation setting), i.e.

$$S_\tau = \frac{1}{N} \sum_{i=1}^N \left(\hat{\theta}_{\tau,i} - \theta \right)^2;$$

the squared bias, namely the squared difference between the estimates mean and the true value, i.e.

$$B_\tau^2 = (\theta - M_\tau)^2;$$

the estimator variance, which is defined as the difference between the mean squared error and the squared bias, i.e.

$$V_\tau = S_\tau - B_\tau^2.$$

Three estimators are compared for the linear settings, while only two for the nonlinear. In the former case, the Koenker and Bassett Jr. (1978) and the Bondell *et al.* (2010) estimators (coded in the tables as KO and BO respectively as subscript for each parameter) are compared with that proposed in (3.9) (coded as SS), which is calculated with the technique explained in Section 3.2.1. In the latter, the nonlinear quantile regression is applied as described in Koenker and Park (1996) and the estimator (3.9) is computed as shown in Section 3.2.2. These methods are selected because they are easily comparable by looking at the performances of the parameters without checking the discrepancies between the τ -quantile regression model and the true τ -quantile function. In particular, since the result of the restricted quantile regression method (He, 1997) consists in the sum of two estimated functions as in (3.3), the conditional estimated quantile model is not directly comparable through an evaluation of the parameters. Similarly, the rearrangement quantile estimator (Chernozhukov *et al.*, 2010) is not directly comparable with other methods, because it exchanges the values of the estimated conditional quantiles. This reveals to be cumbersome to determine the proximity between estimates and the true quantile function in a continuum. Other multi-objective regression methods are not considered for a lack of *ad hoc* algorithms and related software.

In order to compare the performances of the proposed confidence regions, only the results from the nonlinear model simulations are used. Since the best model is known by the simulation setting, the proposed tests for the goodness of fit are not performed here also for reducing the computational

efforts. However, the minimum of the statistics in (3.24) is computed (i.e. \bar{R}^1), in order to establish the behaviour of a global test of performance under the null hypothesis.

A.4.1 A linear regression model

The experiments are designed such that the values assigned to $x \in [a, b]$ are unique. In particular, $x_1 = a$ and $x_{i+1} - x_i = (b-a)(n-1)^{-1}$, with $a = -2$ and $b = 2$, for any $i = 1, \dots, n-1$. The observations from a linear model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ are simulated with pseudo-independent errors ε_i . The parameters are set to be $\beta_0 = 5$ and $\beta_1 = 3$. Here, in each simulation setting, the error terms follow four different distributions, i.e. the standard Normal $\mathbf{N}(0, 1)$, the standard Laplace $\mathbf{La}(0, 1)$, the standard Cauchy $\mathbf{Ca}(0, 1)$ and a special case of the Beta-Normal distribution $\mathbf{BeN}(0.05, 0.05, 0, 1)$ (introduced by Eugene *et al.*, 2002). All these univariate distributions are symmetric and centred in zero, so that $\varepsilon \in [-c, c]$ with 95% of probability. The first three are unimodal and the last one is bimodal. In particular, $c \approx 1.96$ for the standard Normal distribution, $c \approx 3$ for the Laplace, $c \approx 12.71$ for the Cauchy and $c \approx 10.65$ for the Beta-Normal distribution. In order to recognise more easily the theoretical properties of the estimators on the simulation, four different sample sizes are considered for each error distribution, in particular for $n \in \{25, 100, 400, 1600\}$. Here, the estimates are obtained for $\tau \in \{0.1, 0.2, \dots, 0.9\}$. A quite large number of simulated data sets for each setup is chosen to be $N = 10,000$.

The results show that the variance of each estimator decreases and its bias tend to zero as $n \rightarrow \infty$, consequently the mean squared error reduces as the sample size increases. Mostly, the proposed estimator performs better (in terms of squared bias and estimator variance) than the other already presented in the literature. Only in few cases the estimator of Bondell *et al.* (2010) has less variance and hence it is more efficient; in fact, during the simulation and estimation phases it was considered a bounded domain for the models.

When the error term is Normal distributed or as a Cauchy, it is not possible to recognise from the results which is the less biased estimator, because the squared bias of an estimator is the lowest for some quantile levels τ , while the opposite happen for other values of τ . However, the results obtained with the Beta-Normal distributed error (exception made for the computed averages) are not very reliable; in fact, the theoretical values of the parameters are affected by numerical accuracy of the quantile function of the Beta distribution.

A.4.2 A nonlinear regression model

The nonlinear model is related to the theory of innovation diffusion processes, especially according to the formulation of Bass (1969). The original model was expressed in terms of the following differential equation:

$$\dot{z} = \mu \left(\zeta + \xi \frac{z}{\mu} \right) \left(1 - \frac{z}{\mu} \right),$$

wherein \dot{z} denotes instantaneous rates, z the cumulative adoptions and the parameters ζ , ξ and μ respectively represent the innovation effect, the imitation effect and the saturation level of the process. Since a closed form solution for both the cumulative and the instantaneous adoptions exists as a function of the time, it is possible to fit these two models with a single data set. In fact, once the instantaneous rates are randomly generated, the respective cumulative adoptions can easily be calculated.

The model for the cumulative adoption can be written as

$$z = \mu \frac{1 - \exp\{-(\zeta + \xi) t\}}{1 + \xi/\zeta \exp\{-(\zeta + \xi) t\}} + \varepsilon, \quad (\text{A.1})$$

where the time $t \geq 0$ plays the same role of the covariate x in the previous section, while the error term ε depends on the random fluctuation of the instantaneous rates of adoption. Since the deterministic model for \dot{z} is by definition the first derivative with respect to t of the deterministic part in the model (A.1), the stochastic representation of instantaneous rates is written as follows:

$$\dot{z} = \eta \mu \frac{(\zeta + \xi)^2 \exp\{-(\zeta + \xi) t\}}{\zeta [1 + \xi/\zeta \exp\{-(\zeta + \xi) t\}]^2},$$

where the multiplicative error term can be expressed as

$$\eta = 1 + \frac{\varepsilon^* \zeta [1 + \xi/\zeta \exp\{-(\zeta + \xi) t\}]^2}{\mu (\zeta + \xi)^2 \exp\{-(\zeta + \xi) t\}},$$

wherein ε^* is an additive error term derived from (A.1). Since \dot{z} must be non negative, then η is a non negative random fluctuation centred in one (more precisely $E[\eta] = 1$). The parameters are set to be $\zeta = 0.0005$, $\xi = 0.005$ and $\mu = 1000$. Also this case, the values assigned to $t \in [a, b]$ are unique, because repeated measurements at the same time are usually not available in real contexts. In practice, the values of t satisfied the equalities $t_1 = a$ and $t_{i+1} - t_i = (b - a)(n - 1)^{-1}$, with $a = 0$ and $b = 150$, for any $i = 1, \dots, n - 1$, where the sample sizes considered are $n \in \{25, 50, 100, 200\}$. Here, the distributions for η are the Gamma $\text{Ga}(16, 16)$, the Log-Normal

LN($-0.025, 0.05$), the Uniform $U(0.5, 1.5)$ and the Kumaraswamy-Gamma distribution (Cordeiro and de Castro, 2011) $KwGa(1.88, 2, 16, 16)$. In particular $\eta \in [0, c]$ with 95% of probability, where $c \approx 1.44$ for the Gamma distribution, $c \approx 1.41$ for the Log-Normal distribution and $c \approx 1.29$ for the Kumaraswamy-Gamma distribution. Also in this case $N = 10,000$ data sets are simulated. The estimates are obtained for $\tau \in \{0.25, 0.50, 0.75\}$.

The results show how the theoretical properties of quantile regression can be extended to the simultaneous nonlinear quantile regression. The tables in Appendix A.6.2 show that most of the computed variances of the estimator in (3.9) are lower than those calculated for the estimator of Koenker and Park (1996). Also in this context, the behaviour of the estimator variance and the mean squared error is similar to the linear case. The bias can be a problem when it deals with nonlinear model fitting as it happens with least-squares estimates. Also here, it is easy to see this problem by comparing the parameter values of the simulation setting and the averages of the estimates computed for the scenarios with Uniform noise. In many scenarios the estimator in (3.9) performs better in the estimation of the parameters ζ and ξ rather than μ .

To access the main characteristics of the \bar{R}^1 statistic for a goodness of fit test, the Table A.33 shows the minimum, the 5% quantile, the first quartile, the median, the mean, the third quartile and the maximum of the computed statistics for each distribution. As it happens for the evaluation of the R^2 , here it seems more reasonable to set a threshold around 0.65 instead of 0.75. From the results in Table A.33, it is possible to state that the precision of test statistic increases as $n \rightarrow \infty$. This fact suggests that the threshold should slightly tend to 1 as the sample size increases.

The comparisons of the tests are based on coverage probabilities of the confidence regions in (3.28) and (3.32). In practice, the critical values obtained from the Hoeffding's inequality are closer to the empirical quantile of the test statistic under the null hypothesis than those computed with the Uspensky's inequality. The simulation results show that the coverage probabilities for the approximate confidence regions are not good as well those for the "exact" confidence regions. Both critical values tends to the same limit as $n \rightarrow \infty$, and the improvement of coverage probabilities for the approximate confidence regions rises as n increases (an effect due to the assumption on the expectation of the test statistic under the null hypothesis). Even if the Hoeffding's threshold is more accurate, both inequalities provides good coverage probabilities for the "exact" confidence regions. These considerations are valid in all the considered case because the calculated thresholds are not dependent on the distribution of the errors.

Algorithm A.5.1 Simulation of a pseudo-random number distributed as a Beta-Normal $\text{BeN}(\alpha, \beta, 0, 1)$ with $\alpha = \beta$

```

1: Generate a pseudo-random number  $B \sim \text{Be}(\alpha, \alpha)$ .
2: Generate another pseudo-random number  $U \sim \text{U}(0, 1)$ .
3: if  $U > 0.5$  then
4:   if  $B > 0.5$  then
5:     Calculate  $Y = \Phi^{-1}(1 - B)$ .
6:   else
7:     Calculate  $Y = \Phi^{-1}(B)$ .
8:   end if
9: else
10:  if  $B > 0.5$  then
11:    Calculate  $Y = -\Phi^{-1}(1 - B)$ .
12:  else
13:    Calculate  $Y = -\Phi^{-1}(B)$ .
14:  end if
15: end if

```

A.5 Notes on the simulation accuracy

If the inverse of the c.d.f. F_Y is known, the easiest way to generate a pseudo-random number from F_Y is the so-called inversion technique. In practice, a pseudo-random number U is generated from a standard Uniform $\text{U}(0, 1)$ as explained by Matsumoto and Nishimura (1998), so that $Y = F_Y^{-1}(U)$. This is the most efficient solution for generating Normal, Log-Normal, Laplace or Cauchy pseudo-random numbers.

If the inversion technique is chosen for the generation from a Beta-Normal distribution, it requires more efforts because it is not efficient to calculate Y from the quantile function. In practice, if $Y \sim \text{BeN}(\alpha, \beta, 0, 1)$, the r.v. $\Phi(Y)$ is distributed as a Beta $\text{Be}(\alpha, \beta)$, where $\Phi(\cdot)$ is the c.d.f. of a r.v. distributed as a standard Normal. In a straightforward way, a pseudo-random number V is generated from a Beta $\text{Be}(\alpha, \beta)$, so that $Y = \Phi^{-1}(V) \sim \text{BeN}(\alpha, \beta, 0, 1)$.

When the parameter α and β satisfy the equality $\alpha = \beta$, the resulting Beta-Normal distribution is symmetric (Eugene *et al.*, 2002). In this case, for improving the numerical accuracy of the simulation, it is preferred to adopt a different technique described in Algorithm A.5.1. This improvement is essential, when both parameters are less than one.

A different technique is applied to simulate from a Beta $\text{Be}(\alpha, \beta)$, because the application of its quantile function is computationally inefficient. Thus, the acceptance-rejection method is applied with some slight changes

Algorithm A.5.2 Simulation of a pseudo-random number distributed as a Kumaraswamy-Gamma $\text{KwGa}(\alpha, \beta, \gamma, \delta)$ with $(\alpha, \beta)^\top \in [1, +\infty)^2 \setminus \{(1, 1)^\top\}$

- 1: Calculate $M = \beta\alpha^\beta\{(\alpha - 1)/(\alpha\beta - 1)\}^{1-1/\alpha}\{(\beta - 1)/(\alpha\beta - 1)\}^{\beta-1}$
 - 2: **repeat**
 - 3: Generate a pseudo-random number $Y \sim \text{Ga}(\gamma, \delta)$.
 - 4: Compute $U_1 = F_Y(Y)$, where F_Y is the c.d.f. of a Gamma $\text{Ga}(\gamma, \delta)$.
 - 5: Generate another pseudo-random number $U_2 \sim \text{U}(0, 1)$.
 - 6: **until** $U_2M \leq \alpha\beta U_1^{\alpha-1}(1 - U_1^\alpha)^{\beta-1}$
 - 7: Accept the last pseudo-random number Y that was generated.
-

as described in Cheng (1978). Also a Gamma $\text{Ga}(\alpha, \beta)$ must be simulated through the acceptance-rejection method to reduce computational time. The algorithms proposed by Ahrens and Dieter (1974, when $0 < \alpha < 1$) or Ahrens and Dieter (1982, when $\alpha \geq 1$) are the most efficient in terms of time and numerical accuracy.

The generation of a Kumaraswamy-Gamma pseudo-random number can be done similarly as for the Beta-Normal distribution. In fact, Y can be generated by drawing a pseudo-random number from a Kumaraswamy distribution through the inversion technique and after, the Gamma quantile function is applied to it. As it was mentioned above, the computation of the Gamma quantile function should be avoided, because it is almost four times slower than the computation of its c.d.f..

The algorithm A.5.2 is proposed to simulate from a Kumaraswamy-Gamma $\text{KwGa}(\alpha, \beta, \gamma, \delta)$. Such an algorithm is based on the acceptance-rejection method and can be adopted only when $\alpha \geq 1$ and $\beta \geq 1$, but not for $\alpha = \beta = 1$. Such conditions guaranty the uniqueness of the mode for a Kumaraswamy $\text{Kw}(\alpha, \beta)$ distribution. In practice, a pseudo-random number Y , drawn from Gamma $\text{Ga}(\gamma, \delta)$, will be accepted with probability

$$\frac{f_{\text{Kw}}\{F_{\text{Ga}}(Y)\}}{f_{\text{Kw}}\left\{\left(\frac{\alpha - 1}{\alpha\beta - 1}\right)^{1/\alpha}\right\}},$$

where the numerator is the Kumaraswamy density function computed on the Gamma c.d.f. of Y and the quantity on the denominator is the maximum value of the Kumaraswamy density (obtained in the mode).

A.6.2 Tables related to the non linear regression

Table A.17: Average of the estimates for the Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.96287e-03	5.00388e-03	5.06523e-03
ξ_{KO}	5.14294e-02	5.05173e-02	4.99598e-02
μ_{KO}	8.32415e+02	9.85083e+02	1.15622e+03
ζ_{SS}	5.00749e-03	5.00749e-03	5.00749e-03
ξ_{SS}	5.03572e-02	5.03572e-02	5.03572e-02
μ_{SS}	8.41000e+02	9.85980e+02	1.14246e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.98493e-03	5.00828e-03	5.03056e-03
ξ_{KO}	5.06306e-02	5.02133e-02	5.00005e-02
μ_{KO}	8.27424e+02	9.81945e+02	1.15577e+03
ζ_{SS}	5.00585e-03	5.00585e-03	5.00585e-03
ξ_{SS}	5.01684e-02	5.01684e-02	5.01684e-02
μ_{SS}	8.31055e+02	9.82741e+02	1.15026e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.00237e-03	5.00927e-03	5.01501e-03
ξ_{KO}	5.02670e-02	5.00857e-02	5.00160e-02
μ_{KO}	8.25023e+02	9.80355e+02	1.15594e+03
ζ_{SS}	5.00657e-03	5.00657e-03	5.00657e-03
ξ_{SS}	5.00745e-02	5.00745e-02	5.00745e-02
μ_{SS}	8.26658e+02	9.80557e+02	1.15340e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.99722e-03	5.00663e-03	5.00888e-03
ξ_{KO}	5.01425e-02	5.00341e-02	4.99922e-02
μ_{KO}	8.23173e+02	9.79884e+02	1.15551e+03
ζ_{SS}	5.00349e-03	5.00349e-03	5.00349e-03
ξ_{SS}	5.00297e-02	5.00297e-02	5.00297e-02
μ_{SS}	8.24168e+02	9.79899e+02	1.15394e+03

Table A.18: Estimator variance for the Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.27128e-06	8.94871e-07	9.47074e-07
ξ_{KO}	4.30502e-05	2.59176e-05	2.68889e-05
μ_{KO}	5.98621e+03	6.21972e+03	9.12698e+03
ζ_{SS}	6.70397e-07	6.70397e-07	6.70397e-07
ξ_{SS}	1.90296e-05	1.90296e-05	1.90296e-05
μ_{SS}	6.28795e+03	6.39473e+03	9.24352e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	6.32598e-07	4.63612e-07	4.81081e-07
ξ_{KO}	1.85627e-05	1.26056e-05	1.28609e-05
μ_{KO}	2.91339e+03	3.06347e+03	4.58701e+03
ζ_{SS}	3.44324e-07	3.44324e-07	3.44324e-07
ξ_{SS}	9.24487e-06	9.24487e-06	9.24487e-06
μ_{SS}	2.99079e+03	3.15467e+03	4.60735e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	3.08246e-07	2.31225e-07	2.47604e-07
ξ_{KO}	8.50585e-06	5.94340e-06	6.44025e-06
μ_{KO}	1.45555e+03	1.56654e+03	2.27774e+03
ζ_{SS}	1.70998e-07	1.70998e-07	1.70998e-07
ξ_{SS}	4.41419e-06	4.41419e-06	4.41419e-06
μ_{SS}	1.46904e+03	1.57727e+03	2.28782e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.57187e-07	1.17964e-07	1.21857e-07
ξ_{KO}	4.05621e-06	2.99893e-06	3.08857e-06
μ_{KO}	7.13785e+02	7.85090e+02	1.12863e+03
ζ_{SS}	8.64254e-08	8.64254e-08	8.64254e-08
ξ_{SS}	2.19320e-06	2.19320e-06	2.19320e-06
μ_{SS}	7.22726e+02	7.85508e+02	1.13212e+03

Table A.19: Squared bias for the Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.37853e-09	1.50533e-11	4.25554e-09
ξ_{KO}	2.04314e-06	2.67639e-07	1.61430e-09
μ_{KO}	1.08398e+02	3.40811e+01	6.66748e-01
ζ_{SS}	5.61152e-11	5.61152e-11	5.61152e-11
ξ_{SS}	1.27575e-07	1.27575e-07	1.27575e-07
μ_{SS}	3.60878e+02	4.53466e+01	1.67539e+02
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.27002e-10	6.85319e-11	9.33856e-10
ξ_{KO}	3.97621e-07	4.54845e-08	2.64980e-13
μ_{KO}	2.93809e+01	7.28906e+00	1.31768e-01
ζ_{SS}	3.41981e-11	3.41981e-11	3.41981e-11
ξ_{SS}	2.83558e-08	2.83558e-08	2.83558e-08
μ_{SS}	8.19305e+01	1.22158e+01	2.64708e+01
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.60424e-12	8.58878e-11	2.25219e-10
ξ_{KO}	7.13119e-08	7.34353e-09	2.55362e-10
μ_{KO}	9.11698e+00	1.23013e+00	2.83958e-01
ζ_{SS}	4.31611e-11	4.31611e-11	4.31611e-11
ξ_{SS}	5.55708e-09	5.55708e-09	5.55708e-09
μ_{SS}	2.16695e+01	1.72037e+00	4.03310e+00
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	7.74923e-12	4.40040e-11	7.87678e-11
ξ_{KO}	2.03050e-08	1.16608e-09	6.15925e-11
μ_{KO}	1.36723e+00	4.07763e-01	1.17512e-02
ζ_{SS}	1.22133e-11	1.22133e-11	1.22133e-11
ξ_{SS}	8.79159e-10	8.79159e-10	8.79159e-10
μ_{SS}	4.68778e+00	4.26874e-01	2.16207e+00

Table A.20: Mean squared error for the Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.27266e-06	8.94886e-07	9.51330e-07
ξ_{KO}	4.50933e-05	2.61852e-05	2.68905e-05
μ_{KO}	6.09460e+03	6.25381e+03	9.12765e+03
ζ_{SS}	6.70453e-07	6.70453e-07	6.70453e-07
ξ_{SS}	1.91572e-05	1.91572e-05	1.91572e-05
μ_{SS}	6.64883e+03	6.44007e+03	9.41106e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	6.32825e-07	4.63681e-07	4.82015e-07
ξ_{KO}	1.89603e-05	1.26511e-05	1.28609e-05
μ_{KO}	2.94277e+03	3.07076e+03	4.58714e+03
ζ_{SS}	3.44358e-07	3.44358e-07	3.44358e-07
ξ_{SS}	9.27322e-06	9.27322e-06	9.27322e-06
μ_{SS}	3.07272e+03	3.16689e+03	4.63382e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	3.08252e-07	2.31311e-07	2.47829e-07
ξ_{KO}	8.57716e-06	5.95074e-06	6.44051e-06
μ_{KO}	1.46467e+03	1.56777e+03	2.27803e+03
ζ_{SS}	1.71041e-07	1.71041e-07	1.71041e-07
ξ_{SS}	4.41975e-06	4.41975e-06	4.41975e-06
μ_{SS}	1.49070e+03	1.57899e+03	2.29185e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.57195e-07	1.18008e-07	1.21936e-07
ξ_{KO}	4.07652e-06	3.00010e-06	3.08863e-06
μ_{KO}	7.15152e+02	7.85498e+02	1.12864e+03
ζ_{SS}	8.64376e-08	8.64376e-08	8.64376e-08
ξ_{SS}	2.19408e-06	2.19408e-06	2.19408e-06
μ_{SS}	7.27413e+02	7.85935e+02	1.13428e+03

Table A.21: Average of the estimates for the Log-Normal distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.95772e-03	5.01996e-03	5.07413e-03
ξ_{KO}	5.11807e-02	5.04003e-02	4.99292e-02
μ_{KO}	8.48863e+02	9.81643e+02	1.13804e+03
ζ_{SS}	5.00991e-03	5.00991e-03	5.00991e-03
ξ_{SS}	5.03321e-02	5.03321e-02	5.03321e-02
μ_{SS}	8.55802e+02	9.82820e+02	1.12500e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.98595e-03	5.01040e-03	5.03720e-03
ξ_{KO}	5.04811e-02	5.01684e-02	4.99361e-02
μ_{KO}	8.43547e+02	9.78238e+02	1.13535e+03
ζ_{SS}	5.00767e-03	5.00767e-03	5.00767e-03
ξ_{SS}	5.01306e-02	5.01306e-02	5.01306e-02
μ_{SS}	8.46430e+02	9.78794e+02	1.13028e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.98922e-03	4.99769e-03	5.01774e-03
ξ_{KO}	5.02576e-02	5.01372e-02	4.99970e-02
μ_{KO}	8.41967e+02	9.78114e+02	1.13543e+03
ζ_{SS}	5.00070e-03	5.00070e-03	5.00070e-03
ξ_{SS}	5.00953e-02	5.00953e-02	5.00953e-02
μ_{SS}	8.43401e+02	9.78026e+02	1.13285e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.00161e-03	5.00859e-03	5.01167e-03
ξ_{KO}	5.00796e-02	5.00061e-02	4.99766e-02
μ_{KO}	8.39783e+02	9.76185e+02	1.13453e+03
ζ_{SS}	5.00609e-03	5.00609e-03	5.00609e-03
ξ_{SS}	5.00046e-02	5.00046e-02	5.00046e-02
μ_{SS}	8.40311e+02	9.76097e+02	1.13309e+03

Table A.22: Estimator variance for the Log-Normal distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	9.13940e-07	7.05524e-07	8.31802e-07
ξ_{KO}	3.10329e-05	2.03202e-05	2.33635e-05
μ_{KO}	4.25085e+03	4.82781e+03	7.91228e+03
ζ_{SS}	5.38086e-07	5.38086e-07	5.38086e-07
ξ_{SS}	1.48834e-05	1.48834e-05	1.48834e-05
μ_{SS}	4.59154e+03	5.05982e+03	7.79203e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.48677e-07	3.70748e-07	4.27600e-07
ξ_{KO}	1.28671e-05	9.98822e-06	1.11834e-05
μ_{KO}	2.14455e+03	2.40555e+03	3.87453e+03
ζ_{SS}	2.77223e-07	2.77223e-07	2.77223e-07
ξ_{SS}	7.35458e-06	7.35458e-06	7.35458e-06
μ_{SS}	2.16895e+03	2.46880e+03	3.86717e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.24309e-07	1.84834e-07	2.20533e-07
ξ_{KO}	5.98961e-06	4.72126e-06	5.64897e-06
μ_{KO}	1.06603e+03	1.21862e+03	1.88325e+03
ζ_{SS}	1.41387e-07	1.41387e-07	1.41387e-07
ξ_{SS}	3.56985e-06	3.56985e-06	3.56985e-06
μ_{SS}	1.08269e+03	1.22836e+03	1.89398e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.09848e-07	9.17424e-08	1.07206e-07
ξ_{KO}	2.84701e-06	2.34709e-06	2.77311e-06
μ_{KO}	5.25122e+02	6.05676e+02	9.73668e+02
ζ_{SS}	6.86597e-08	6.86597e-08	6.86597e-08
ξ_{SS}	1.77410e-06	1.77410e-06	1.77410e-06
μ_{SS}	5.30948e+02	6.08502e+02	9.79162e+02

Table A.23: Squared bias for the Log-Normal distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.78780e-09	3.98251e-10	5.49600e-09
ξ_{KO}	1.39411e-06	1.60229e-07	5.01635e-09
μ_{KO}	1.01900e+02	4.01023e+01	1.56793e+01
ζ_{SS}	9.82974e-11	9.82974e-11	9.82974e-11
ξ_{SS}	1.10266e-07	1.10266e-07	1.10266e-07
μ_{SS}	2.90126e+02	5.64040e+01	8.24946e+01
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.97340e-10	1.08260e-10	1.38358e-09
ξ_{KO}	2.31421e-07	2.83445e-08	4.08875e-09
μ_{KO}	2.28368e+01	8.57379e+00	1.62693e+00
ζ_{SS}	5.88790e-11	5.88790e-11	5.88790e-11
ξ_{SS}	1.70687e-08	1.70687e-08	1.70687e-08
μ_{SS}	5.87001e+01	1.21366e+01	1.44343e+01
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.16158e-10	5.32495e-12	3.14554e-10
ξ_{KO}	6.63385e-08	1.88272e-08	8.96309e-12
μ_{KO}	1.02270e+01	7.86278e+00	1.82033e+00
ζ_{SS}	4.83793e-13	4.83793e-13	4.83793e-13
ξ_{SS}	9.08702e-09	9.08702e-09	9.08702e-09
μ_{SS}	2.14588e+01	7.37658e+00	1.51623e+00
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.57671e-12	7.37760e-11	1.36124e-10
ξ_{KO}	6.33535e-09	3.73218e-11	5.47496e-10
μ_{KO}	1.02872e+00	7.65381e-01	2.00534e-01
ζ_{SS}	3.70759e-11	3.70759e-11	3.70759e-11
ξ_{SS}	2.10028e-11	2.10028e-11	2.10028e-11
μ_{SS}	2.38024e+00	6.19642e-01	9.73898e-01

Table A.24: Mean squared error for the Log-Normal distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	9.15728e-07	7.05923e-07	8.37298e-07
ξ_{KO}	3.24270e-05	2.04804e-05	2.33685e-05
μ_{KO}	4.35275e+03	4.86791e+03	7.92796e+03
ζ_{SS}	5.38184e-07	5.38184e-07	5.38184e-07
ξ_{SS}	1.49937e-05	1.49937e-05	1.49937e-05
μ_{SS}	4.88167e+03	5.11622e+03	7.87452e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.48874e-07	3.70856e-07	4.28984e-07
ξ_{KO}	1.30985e-05	1.00166e-05	1.11874e-05
μ_{KO}	2.16739e+03	2.41412e+03	3.87616e+03
ζ_{SS}	2.77282e-07	2.77282e-07	2.77282e-07
ξ_{SS}	7.37165e-06	7.37165e-06	7.37165e-06
μ_{SS}	2.22765e+03	2.48094e+03	3.88160e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.24425e-07	1.84840e-07	2.20847e-07
ξ_{KO}	6.05594e-06	4.74009e-06	5.64898e-06
μ_{KO}	1.07626e+03	1.22648e+03	1.88507e+03
ζ_{SS}	1.41387e-07	1.41387e-07	1.41387e-07
ξ_{SS}	3.57894e-06	3.57894e-06	3.57894e-06
μ_{SS}	1.10415e+03	1.23574e+03	1.89549e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.09850e-07	9.18162e-08	1.07342e-07
ξ_{KO}	2.85334e-06	2.34713e-06	2.77365e-06
μ_{KO}	5.26150e+02	6.06441e+02	9.73869e+02
ζ_{SS}	6.86967e-08	6.86967e-08	6.86967e-08
ξ_{SS}	1.77412e-06	1.77412e-06	1.77412e-06
μ_{SS}	5.33328e+02	6.09121e+02	9.80136e+02

Table A.25: Average of the estimates for the Uniform distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.99528e-03	4.98690e-03	4.99804e-03
ξ_{KO}	5.21599e-02	5.11936e-02	5.04010e-02
μ_{KO}	7.80580e+02	9.98058e+02	1.21941e+03
ζ_{SS}	4.97279e-03	4.97279e-03	4.97279e-03
ξ_{SS}	5.08056e-02	5.08056e-02	5.08056e-02
μ_{SS}	7.87584e+02	1.00309e+03	1.20270e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.00743e-03	4.99720e-03	4.98806e-03
ξ_{KO}	5.10182e-02	5.06736e-02	5.02460e-02
μ_{KO}	7.68041e+02	1.00123e+03	1.23321e+03
ζ_{SS}	4.98788e-03	4.98788e-03	4.98788e-03
ξ_{SS}	5.04004e-02	5.04004e-02	5.04004e-02
μ_{SS}	7.69367e+02	1.00433e+03	1.22583e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.01536e-03	4.99352e-03	4.98459e-03
ξ_{KO}	5.04902e-02	5.04046e-02	5.01866e-02
μ_{KO}	7.59175e+02	1.00049e+03	1.24106e+03
ζ_{SS}	4.98817e-03	4.98817e-03	4.98817e-03
ξ_{SS}	5.02547e-02	5.02547e-02	5.02547e-02
μ_{SS}	7.59858e+02	1.00192e+03	1.23718e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.00717e-03	4.99225e-03	4.99114e-03
ξ_{KO}	5.01969e-02	5.01723e-02	5.00884e-02
μ_{KO}	7.54006e+02	9.99736e+02	1.24541e+03
ζ_{SS}	4.99100e-03	4.99100e-03	4.99100e-03
ξ_{SS}	5.01119e-02	5.01119e-02	5.01119e-02
μ_{SS}	7.53992e+02	1.00048e+03	1.24352e+03

Table A.26: Estimator variance for the Uniform distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.45644e-06	1.70227e-06	9.96379e-07
ξ_{KO}	8.95051e-05	5.22566e-05	2.99533e-05
μ_{KO}	1.04264e+04	1.31866e+04	1.05756e+04
ζ_{SS}	1.04829e-06	1.04829e-06	1.04829e-06
ξ_{SS}	3.11124e-05	3.11124e-05	3.11124e-05
μ_{SS}	1.19388e+04	1.40883e+04	1.10196e+04
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.40011e-06	9.74395e-07	5.00570e-07
ξ_{KO}	4.13100e-05	2.72343e-05	1.35276e-05
μ_{KO}	5.71713e+03	7.29635e+03	5.65830e+03
ζ_{SS}	5.43297e-07	5.43297e-07	5.43297e-07
ξ_{SS}	1.47381e-05	1.47381e-05	1.47381e-05
μ_{SS}	6.10336e+03	7.59097e+03	5.65789e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	7.64212e-07	5.49406e-07	2.70423e-07
ξ_{KO}	2.03701e-05	1.45518e-05	6.99519e-06
μ_{KO}	2.91523e+03	3.78314e+03	2.88926e+03
ζ_{SS}	2.94354e-07	2.94354e-07	2.94354e-07
ξ_{SS}	7.71003e-06	7.71003e-06	7.71003e-06
μ_{SS}	3.02782e+03	3.84593e+03	2.88691e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	3.82921e-07	2.85247e-07	1.37649e-07
ξ_{KO}	9.85646e-06	7.35317e-06	3.55402e-06
μ_{KO}	1.47906e+03	1.97404e+03	1.49499e+03
ζ_{SS}	1.50225e-07	1.50225e-07	1.50225e-07
ξ_{SS}	3.88709e-06	3.88709e-06	3.88709e-06
μ_{SS}	1.51472e+03	2.00325e+03	1.50855e+03

Table A.27: Squared bias for the Uniform distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.22795e-11	1.71622e-10	3.85714e-12
ξ_{KO}	4.66499e-06	1.42476e-06	1.60788e-07
μ_{KO}	9.35162e+02	3.77164e+00	9.35630e+02
ζ_{SS}	7.40521e-10	7.40521e-10	7.40521e-10
ξ_{SS}	6.48926e-07	6.48926e-07	6.48926e-07
μ_{SS}	1.41255e+03	9.56585e+00	2.23723e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.51320e-11	7.82300e-12	1.42536e-10
ξ_{KO}	1.03671e-06	4.53729e-07	6.05028e-08
μ_{KO}	3.25489e+02	1.52136e+00	2.82057e+02
ζ_{SS}	1.46884e-10	1.46884e-10	1.46884e-10
ξ_{SS}	1.60329e-07	1.60329e-07	1.60329e-07
μ_{SS}	3.75089e+02	1.87608e+01	5.84414e+02
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.35966e-10	4.20536e-11	2.37563e-10
ξ_{KO}	2.40337e-07	1.63691e-07	3.48079e-08
μ_{KO}	8.41875e+01	2.40270e-01	7.99911e+01
ζ_{SS}	1.40028e-10	1.40028e-10	1.40028e-10
ξ_{SS}	6.48765e-08	6.48765e-08	6.48765e-08
μ_{SS}	9.71778e+01	3.70475e+00	1.64274e+02
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.14321e-11	5.99952e-11	7.85246e-11
ξ_{KO}	3.87680e-08	2.96732e-08	7.80704e-09
μ_{KO}	1.60480e+01	6.97092e-02	2.10804e+01
ζ_{SS}	8.10437e-11	8.10437e-11	8.10437e-11
ξ_{SS}	1.25281e-08	1.25281e-08	1.25281e-08
μ_{SS}	1.59395e+01	2.27139e-01	4.20455e+01

Table A.28: Mean squared error for the Uniform distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.45646e-06	1.70244e-06	9.96383e-07
ξ_{KO}	9.41700e-05	5.36813e-05	3.01141e-05
μ_{KO}	1.13615e+04	1.31903e+04	1.15112e+04
ζ_{SS}	1.04903e-06	1.04903e-06	1.04903e-06
ξ_{SS}	3.17614e-05	3.17614e-05	3.17614e-05
μ_{SS}	1.33513e+04	1.40979e+04	1.32568e+04
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.40017e-06	9.74403e-07	5.00712e-07
ξ_{KO}	4.23467e-05	2.76880e-05	1.35881e-05
μ_{KO}	6.04262e+03	7.29787e+03	5.94036e+03
ζ_{SS}	5.43443e-07	5.43443e-07	5.43443e-07
ξ_{SS}	1.48985e-05	1.48985e-05	1.48985e-05
μ_{SS}	6.47844e+03	7.60973e+03	6.24231e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	7.64448e-07	5.49448e-07	2.70661e-07
ξ_{KO}	2.06105e-05	1.47155e-05	7.02999e-06
μ_{KO}	2.99942e+03	3.78339e+03	2.96926e+03
ζ_{SS}	2.94495e-07	2.94495e-07	2.94495e-07
ξ_{SS}	7.77490e-06	7.77490e-06	7.77490e-06
μ_{SS}	3.12500e+03	3.84964e+03	3.05119e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	3.82972e-07	2.85307e-07	1.37728e-07
ξ_{KO}	9.89523e-06	7.38284e-06	3.56183e-06
μ_{KO}	1.49511e+03	1.97411e+03	1.51607e+03
ζ_{SS}	1.50306e-07	1.50306e-07	1.50306e-07
ξ_{SS}	3.89961e-06	3.89961e-06	3.89961e-06
μ_{SS}	1.53066e+03	2.00348e+03	1.55060e+03

Table A.29: Average of the estimates for the Kumaraswamy-Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.96110e-03	5.00901e-03	5.05037e-03
ξ_{KO}	5.07817e-02	5.02272e-02	4.98564e-02
μ_{KO}	8.87874e+02	9.93806e+02	1.10715e+03
ζ_{SS}	5.01025e-03	5.01025e-03	5.01025e-03
ξ_{SS}	5.01390e-02	5.01390e-02	5.01390e-02
μ_{SS}	8.93971e+02	9.94001e+02	1.09849e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.97924e-03	5.00168e-03	5.01717e-03
ξ_{KO}	5.03527e-02	5.01008e-02	4.99524e-02
μ_{KO}	8.85682e+02	9.93247e+02	1.10735e+03
ζ_{SS}	4.99724e-03	4.99724e-03	4.99724e-03
ξ_{SS}	5.00881e-02	5.00881e-02	5.00881e-02
μ_{SS}	8.88160e+02	9.93763e+02	1.10391e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.99879e-03	5.00532e-03	5.00788e-03
ξ_{KO}	5.01432e-02	5.00415e-02	4.99938e-02
μ_{KO}	8.84305e+02	9.92273e+02	1.10783e+03
ζ_{SS}	5.00284e-03	5.00284e-03	5.00284e-03
ξ_{SS}	5.00374e-02	5.00374e-02	5.00374e-02
μ_{SS}	8.85541e+02	9.92401e+02	1.10600e+03
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.99891e-03	5.00063e-03	5.00442e-03
ξ_{KO}	5.00557e-02	5.00175e-02	4.99869e-02
μ_{KO}	8.83428e+02	9.92206e+02	1.10798e+03
ζ_{SS}	5.00180e-03	5.00180e-03	5.00180e-03
ξ_{SS}	5.00054e-02	5.00054e-02	5.00054e-02
μ_{SS}	8.84138e+02	9.92283e+02	1.10711e+03

Table A.30: Estimator variance for the Kumaraswamy-Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.57544e-07	3.97279e-07	4.35763e-07
ξ_{KO}	1.77517e-05	1.13604e-05	1.24814e-05
μ_{KO}	2.84035e+03	2.77734e+03	3.77231e+03
ζ_{SS}	3.01117e-07	3.01117e-07	3.01117e-07
ξ_{SS}	8.35528e-06	8.35528e-06	8.35528e-06
μ_{SS}	2.99653e+03	2.91399e+03	3.78891e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.74345e-07	2.07967e-07	2.23813e-07
ξ_{KO}	7.70450e-06	5.53984e-06	5.87745e-06
μ_{KO}	1.49210e+03	1.43953e+03	1.92417e+03
ζ_{SS}	1.55257e-07	1.55257e-07	1.55257e-07
ξ_{SS}	4.06722e-06	4.06722e-06	4.06722e-06
μ_{SS}	1.51189e+03	1.45471e+03	1.93209e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.38788e-07	1.03705e-07	1.11696e-07
ξ_{KO}	3.73343e-06	2.65311e-06	2.91592e-06
μ_{KO}	7.56036e+02	7.14678e+02	9.80137e+02
ζ_{SS}	7.82160e-08	7.82160e-08	7.82160e-08
ξ_{SS}	2.01157e-06	2.01157e-06	2.01157e-06
μ_{SS}	7.62282e+02	7.14868e+02	9.79486e+02
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	6.92620e-08	5.19675e-08	5.65703e-08
ξ_{KO}	1.75786e-06	1.31415e-06	1.44921e-06
μ_{KO}	3.58231e+02	3.44837e+02	4.75750e+02
ζ_{SS}	3.92118e-08	3.92118e-08	3.92118e-08
ξ_{SS}	9.85038e-07	9.85038e-07	9.85038e-07
μ_{SS}	3.60297e+02	3.48112e+02	4.77938e+02

Table A.31: Squared bias for the Kumaraswamy-Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.51300e-09	8.12020e-11	2.53710e-09
ξ_{KO}	6.11132e-07	5.16125e-08	2.06124e-08
μ_{KO}	2.46388e+01	3.66148e+00	8.40375e-01
ζ_{SS}	1.05057e-10	1.05057e-10	1.05057e-10
ξ_{SS}	1.93073e-08	1.93073e-08	1.93073e-08
μ_{SS}	1.22361e+02	4.44580e+00	9.18680e+01
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	4.31046e-10	2.81525e-12	2.94717e-10
ξ_{KO}	1.24411e-07	1.01507e-08	2.26923e-09
μ_{KO}	7.68487e+00	1.83568e+00	5.16205e-01
ζ_{SS}	7.59535e-12	7.59535e-12	7.59535e-12
ξ_{SS}	7.76305e-09	7.76305e-09	7.76305e-09
μ_{SS}	2.75659e+01	3.50033e+00	1.73305e+01
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.47475e-12	2.82499e-11	6.20478e-11
ξ_{KO}	2.05085e-08	1.72634e-09	3.81479e-11
μ_{KO}	1.94745e+00	1.45174e-01	5.90828e-02
ζ_{SS}	8.06880e-12	8.06880e-12	8.06880e-12
ξ_{SS}	1.40093e-09	1.40093e-09	1.40093e-09
μ_{SS}	6.92120e+00	2.58881e-01	4.30053e+00
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.18048e-12	3.90810e-13	1.95515e-11
ξ_{KO}	3.10058e-09	3.07866e-10	1.70800e-10
μ_{KO}	2.68374e-01	9.81158e-02	8.60939e-03
ζ_{SS}	3.25669e-12	3.25669e-12	3.25669e-12
ξ_{SS}	2.90134e-11	2.90134e-11	2.90134e-11
μ_{SS}	1.50919e+00	1.52739e-01	9.27758e-01

Table A.32: Mean squared error for the Kumaraswamy-Gamma distribution

$n = 25$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	5.59057e-07	3.97360e-07	4.38300e-07
ξ_{KO}	1.83629e-05	1.14120e-05	1.25020e-05
μ_{KO}	2.86499e+03	2.78100e+03	3.77315e+03
ζ_{SS}	3.01222e-07	3.01222e-07	3.01222e-07
ξ_{SS}	8.37459e-06	8.37459e-06	8.37459e-06
μ_{SS}	3.11889e+03	2.91843e+03	3.88078e+03
$n = 50$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	2.74776e-07	2.07970e-07	2.24108e-07
ξ_{KO}	7.82891e-06	5.54999e-06	5.87972e-06
μ_{KO}	1.49978e+03	1.44137e+03	1.92469e+03
ζ_{SS}	1.55265e-07	1.55265e-07	1.55265e-07
ξ_{SS}	4.07499e-06	4.07499e-06	4.07499e-06
μ_{SS}	1.53946e+03	1.45821e+03	1.94942e+03
$n = 100$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	1.38790e-07	1.03734e-07	1.11758e-07
ξ_{KO}	3.75394e-06	2.65484e-06	2.91596e-06
μ_{KO}	7.57983e+02	7.14823e+02	9.80196e+02
ζ_{SS}	7.82241e-08	7.82241e-08	7.82241e-08
ξ_{SS}	2.01297e-06	2.01297e-06	2.01297e-06
μ_{SS}	7.69204e+02	7.15126e+02	9.83787e+02
$n = 200$	$\tau = 25\%$	$\tau = 50\%$	$\tau = 75\%$
ζ_{KO}	6.92631e-08	5.19679e-08	5.65898e-08
ξ_{KO}	1.76096e-06	1.31446e-06	1.44938e-06
μ_{KO}	3.58499e+02	3.44935e+02	4.75759e+02
ζ_{SS}	3.92151e-08	3.92151e-08	3.92151e-08
ξ_{SS}	9.85067e-07	9.85067e-07	9.85067e-07
μ_{SS}	3.61806e+02	3.48264e+02	4.78866e+02

A.6.3 Tables related to the tests

Goodness of fit tests

Table A.33: Quantities related to the distribution of \bar{R}^1 test statistic

$n = 25$	Min.	5% Qu.	25% Qu.	Median	Mean	75% Qu.	Max.
Ga	0.456	0.607	0.669	0.709	0.707	0.747	0.888
LN	0.497	0.650	0.704	0.741	0.738	0.774	0.912
U	0.452	0.546	0.598	0.636	0.638	0.677	0.865
KwGa	0.597	0.724	0.769	0.799	0.797	0.827	0.925
$n = 50$	Min.	5% Qu.	25% Qu.	Median	Mean	75% Qu.	Max.
Ga	0.511	0.622	0.666	0.694	0.693	0.721	0.826
LN	0.576	0.668	0.704	0.727	0.727	0.751	0.853
U	0.476	0.555	0.591	0.615	0.617	0.643	0.777
KwGa	0.656	0.735	0.768	0.788	0.787	0.808	0.892
$n = 100$	Min.	5% Qu.	25% Qu.	Median	Mean	75% Qu.	Max.
Ga	0.571	0.637	0.666	0.687	0.686	0.705	0.780
LN	0.621	0.679	0.705	0.722	0.722	0.739	0.819
U	0.523	0.564	0.588	0.606	0.606	0.624	0.707
KwGa	0.681	0.746	0.768	0.782	0.782	0.797	0.854
$n = 200$	Min.	5% Qu.	25% Qu.	Median	Mean	75% Qu.	Max.
Ga	0.595	0.648	0.669	0.683	0.683	0.697	0.756
LN	0.650	0.691	0.708	0.720	0.720	0.732	0.782
U	0.535	0.570	0.588	0.600	0.601	0.613	0.689
KwGa	0.708	0.754	0.769	0.780	0.779	0.790	0.832

Coverage of confidence regions

Table A.34: Coverage probabilities for testing data from a Gamma distribution

	Nominal level $1 - \alpha$	Hoeffding's coverage	Uspensky's coverage	Hoeffding's approx. cover.	Uspensky's approx. cover.
$n = 25$	0.9000	0.9069	0.8289	0.1128	0.0763
	0.9500	0.9393	0.9456	0.1438	0.1500
	0.9900	0.9790	1.0000	0.2099	0.5937
	0.9990	0.9934	1.0000	0.2983	0.0000
$n = 50$	0.9000	0.9750	0.9377	0.3657	0.2807
	0.9500	0.9877	0.9888	0.4188	0.4299
	0.9900	0.9969	1.0000	0.5189	0.8151
	0.9990	0.9995	1.0000	0.6166	0.0000
$n = 100$	0.9000	0.9954	0.9863	0.6912	0.6000
	0.9500	0.9981	0.9984	0.7428	0.7523
	0.9900	0.9994	1.0000	0.8190	0.9628
	0.9990	0.9999	1.0000	0.8775	1.0000
$n = 200$	0.9000	0.9999	0.9983	0.9167	0.8657
	0.9500	1.0000	1.0000	0.9382	0.9424
	0.9900	1.0000	1.0000	0.9668	0.9973
	0.9990	1.0000	1.0000	0.9845	1.0000

Table A.35: Critical values for testing data from a Gamma distribution

	Nominal level $1 - \alpha$	Empirical threshold	Hoeffding's threshold	Uspensky's threshold
$n = 25$	0.9000	0.7893	0.7854	0.8268
	0.9500	0.7443	0.7552	0.7483
	0.9900	0.6487	0.6965	0.4255
	0.9990	0.5461	0.6283	-0.8248
$n = 50$	0.9000	0.8944	0.8483	0.8775
	0.9500	0.8707	0.8269	0.8220
	0.9900	0.8190	0.7854	0.5938
	0.9990	0.7553	0.7372	-0.2903
$n = 100$	0.9000	0.9490	0.8927	0.9134
	0.9500	0.9361	0.8776	0.8742
	0.9900	0.9060	0.8483	0.7128
	0.9990	0.8652	0.8142	0.0876
$n = 200$	0.9000	0.9746	0.9241	0.9388
	0.9500	0.9683	0.9135	0.9110
	0.9900	0.9548	0.8927	0.7969
	0.9990	0.9336	0.8686	0.3548

Table A.36: Coverage probabilities for testing data from a Log-Normal distribution

	Nominal level $1 - \alpha$	Hoeffding's coverage	Uspensky's coverage	Hoeffding's approx. cover.	Uspensky's approx. cover.
$n = 25$	0.9000	0.9083	0.8292	0.1255	0.0856
	0.9500	0.9440	0.9500	0.1636	0.1715
	0.9900	0.9771	1.0000	0.2393	0.6305
	0.9990	0.9924	1.0000	0.3303	0.0000
$n = 50$	0.9000	0.9766	0.9452	0.3956	0.3093
	0.9500	0.9877	0.9895	0.4560	0.4672
	0.9900	0.9967	1.0000	0.5582	0.8494
	0.9990	0.9996	1.0000	0.6563	0.0000
$n = 100$	0.9000	0.9959	0.9866	0.7255	0.6392
	0.9500	0.9981	0.9983	0.7750	0.7856
	0.9900	0.9997	1.0000	0.8468	0.9700
	0.9990	0.9999	1.0000	0.8979	1.0000
$n = 200$	0.9000	0.9998	0.9978	0.9332	0.8870
	0.9500	1.0000	1.0000	0.9533	0.9565
	0.9900	1.0000	1.0000	0.9761	0.9977
	0.9990	1.0000	1.0000	0.9872	1.0000

Table A.37: Critical values for testing data from a Log-Normal distribution

	Nominal level $1 - \alpha$	Empirical threshold	Hoeffding's threshold	Uspensky's threshold
$n = 25$	0.9000	0.7919	0.7854	0.8268
	0.9500	0.7484	0.7552	0.7483
	0.9900	0.6509	0.6965	0.4255
	0.9990	0.5336	0.6283	-0.8248
$n = 50$	0.9000	0.8990	0.8483	0.8775
	0.9500	0.8747	0.8269	0.8220
	0.9900	0.8206	0.7854	0.5938
	0.9990	0.7615	0.7372	-0.2903
$n = 100$	0.9000	0.9496	0.8927	0.9134
	0.9500	0.9366	0.8776	0.8742
	0.9900	0.9080	0.8483	0.7128
	0.9990	0.8654	0.8142	0.0876
$n = 200$	0.9000	0.9743	0.9241	0.9388
	0.9500	0.9675	0.9135	0.9110
	0.9900	0.9522	0.8927	0.7969
	0.9990	0.9314	0.8686	0.3548

Table A.38: Coverage probabilities for testing data from a Uniform distribution

	Nominal level $1 - \alpha$	Hoeffding's coverage	Uspensky's coverage	Hoeffding's approx. cover.	Uspensky's approx. cover.
$n = 25$	0.9000	0.8734	0.7857	0.0664	0.0437
	0.9500	0.9172	0.9243	0.0877	0.0930
	0.9900	0.9611	1.0000	0.1374	0.4555
	0.9990	0.9856	1.0000	0.2033	0.0000
$n = 50$	0.9000	0.9584	0.9093	0.2349	0.1736
	0.9500	0.9749	0.9781	0.2818	0.2917
	0.9900	0.9917	0.9999	0.3703	0.7115
	0.9990	0.9978	1.0000	0.4712	0.0000
$n = 100$	0.9000	0.9918	0.9734	0.5424	0.4526
	0.9500	0.9969	0.9975	0.6037	0.6149
	0.9900	0.9993	1.0000	0.6946	0.9144
	0.9990	1.0000	1.0000	0.7783	1.0000
$n = 200$	0.9000	0.9989	0.9957	0.8462	0.7792
	0.9500	0.9994	0.9995	0.8802	0.8858
	0.9900	1.0000	1.0000	0.9260	0.9910
	0.9990	1.0000	1.0000	0.9550	1.0000

Table A.39: Critical values for testing data drawn a Uniform distribution

	Nominal level $1 - \alpha$	Empirical threshold	Hoeffding's threshold	Uspensky's threshold
$n = 25$	0.9000	0.7701	0.7854	0.8268
	0.9500	0.7163	0.7552	0.7483
	0.9900	0.6088	0.6965	0.4255
	0.9990	0.4920	0.6283	-0.8248
$n = 50$	0.9000	0.8814	0.8483	0.8775
	0.9500	0.8546	0.8269	0.8220
	0.9900	0.7936	0.7854	0.5938
	0.9990	0.7155	0.7372	-0.2903
$n = 100$	0.9000	0.9426	0.8927	0.9134
	0.9500	0.9277	0.8776	0.8742
	0.9900	0.8965	0.8483	0.7128
	0.9990	0.8634	0.8142	0.0876
$n = 200$	0.9000	0.9713	0.9241	0.9388
	0.9500	0.9640	0.9135	0.9110
	0.9900	0.9473	0.8927	0.7969
	0.9990	0.9238	0.8686	0.3548

Table A.40: Coverage probabilities for testing data from a Kumaraswamy-Gamma distribution

	Nominal level $1 - \alpha$	Hoeffding's coverage	Uspensky's coverage	Hoeffding's approx. cover.	Uspensky's approx. cover.
$n = 25$	0.9000	0.9037	0.8274	0.1690	0.1167
	0.9500	0.9377	0.9442	0.2104	0.2222
	0.9900	0.9778	0.9999	0.3005	0.7054
	0.9990	0.9949	1.0000	0.4031	0.0000
$n = 50$	0.9000	0.9738	0.9415	0.4661	0.3718
	0.9500	0.9864	0.9885	0.5307	0.5436
	0.9900	0.9953	1.0000	0.6317	0.8966
	0.9990	0.9995	1.0000	0.7298	0.0000
$n = 100$	0.9000	0.9951	0.9860	0.7843	0.6995
	0.9500	0.9976	0.9981	0.8273	0.8346
	0.9900	0.9994	1.0000	0.8866	0.9847
	0.9990	0.9999	1.0000	0.9320	1.0000
$n = 200$	0.9000	0.9991	0.9973	0.9542	0.9192
	0.9500	0.9998	1.0000	0.9706	0.9727
	0.9900	1.0000	1.0000	0.9860	0.9997
	0.9990	1.0000	1.0000	0.9938	1.0000

Table A.41: Critical values for testing data from a Kumaraswamy-Gamma distribution

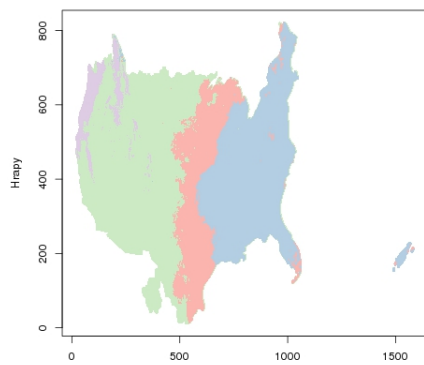
	Nominal level $1 - \alpha$	Empirical threshold	Hoeffding's threshold	Uspensky's threshold
$n = 25$	0.9000	0.7872	0.7854	0.8268
	0.9500	0.7416	0.7552	0.7483
	0.9900	0.6573	0.6965	0.4255
	0.9990	0.5421	0.6283	-0.8248
$n = 50$	0.9000	0.8949	0.8483	0.8775
	0.9500	0.8705	0.8269	0.8220
	0.9900	0.8179	0.7854	0.5938
	0.9990	0.7593	0.7372	-0.2903
$n = 100$	0.9000	0.9486	0.8927	0.9134
	0.9500	0.9362	0.8776	0.8742
	0.9900	0.9085	0.8483	0.7128
	0.9990	0.8616	0.8142	0.0876
$n = 200$	0.9000	0.9744	0.9241	0.9388
	0.9500	0.9680	0.9135	0.9110
	0.9900	0.9525	0.8927	0.7969
	0.9990	0.9258	0.8686	0.3548

Appendix B

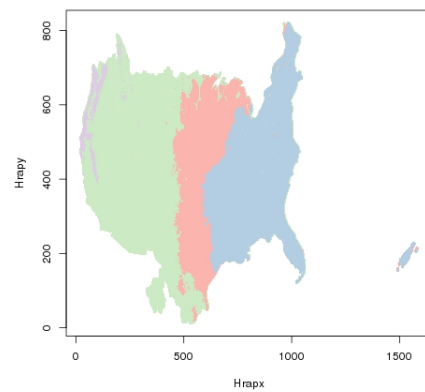
Cluster Analysis Results

B.1 Clustering with random initialisation

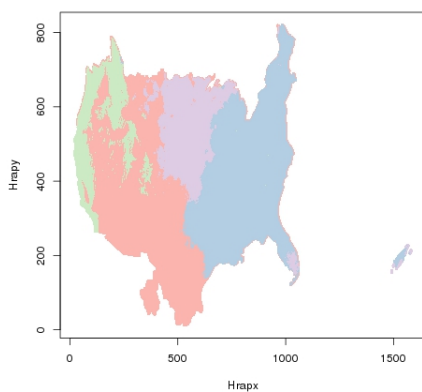
B.1.1 Resulting clusters



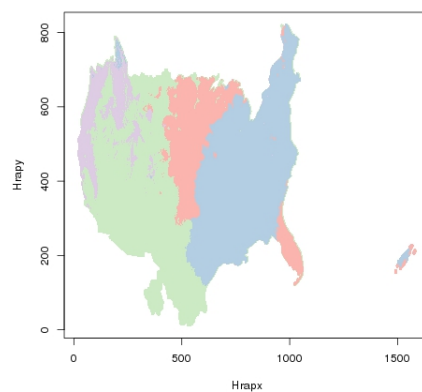
(a) No transformation applied



(b) Mean-variance standardisation



(c) Logarithmic standardisation



(d) Range standardisation

Figure B.1: Different precipitation areas after cluster validation.

B.1.2 Clustering validation graphs

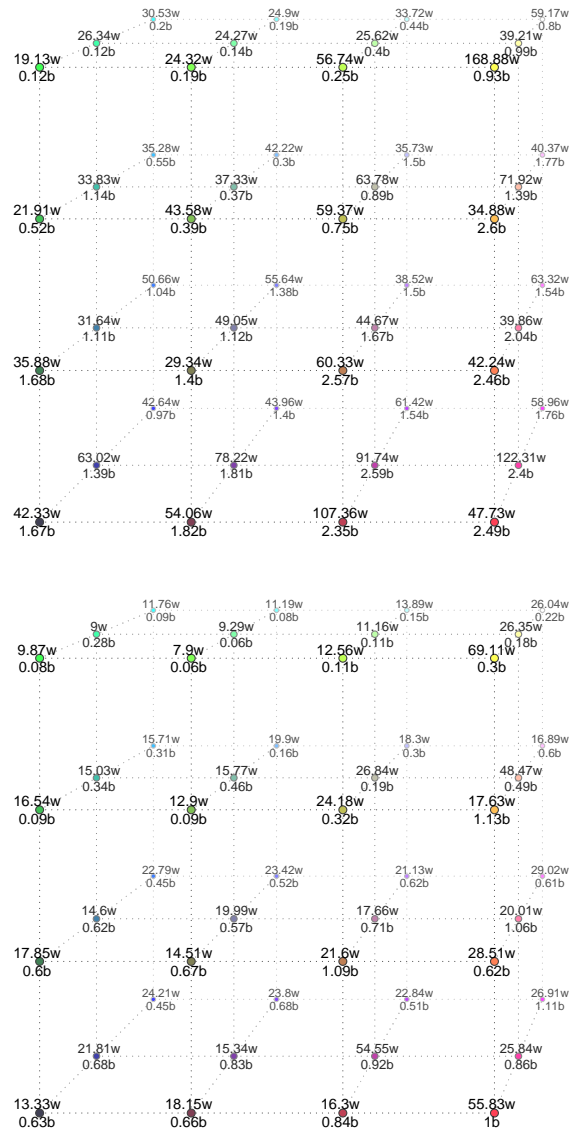


Figure B.2: Each map shows for each neuron the between clusters distance (b) and the clusters diameters (w). The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

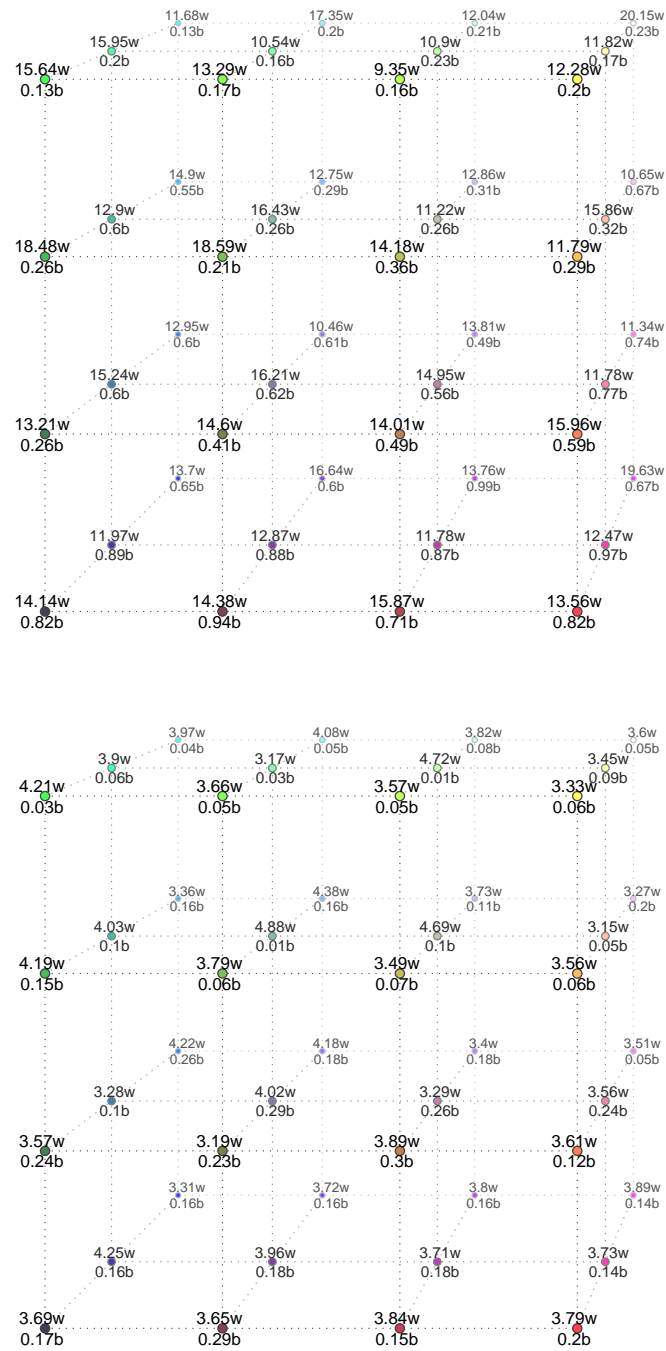


Figure B.3: Each map shows for each neuron the between clusters distance (b) and the clusters diameters (w). The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.1.3 Resulting centroids

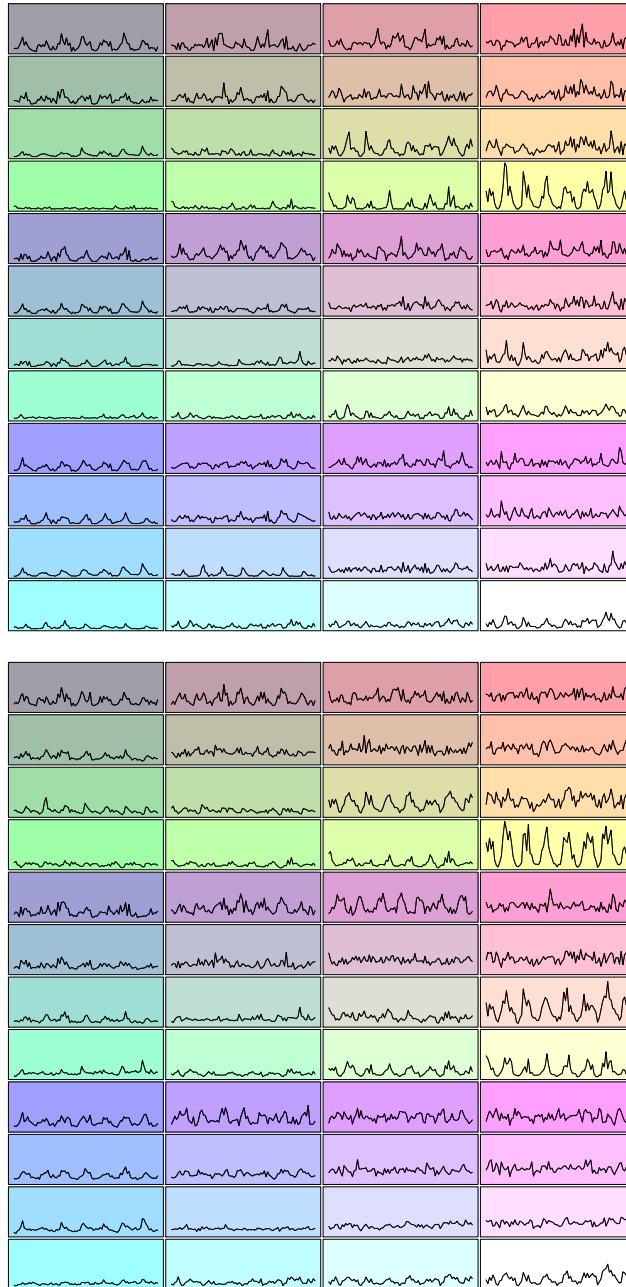


Figure B.4: Figures for centroids pattern with time in months on the x-axis. The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

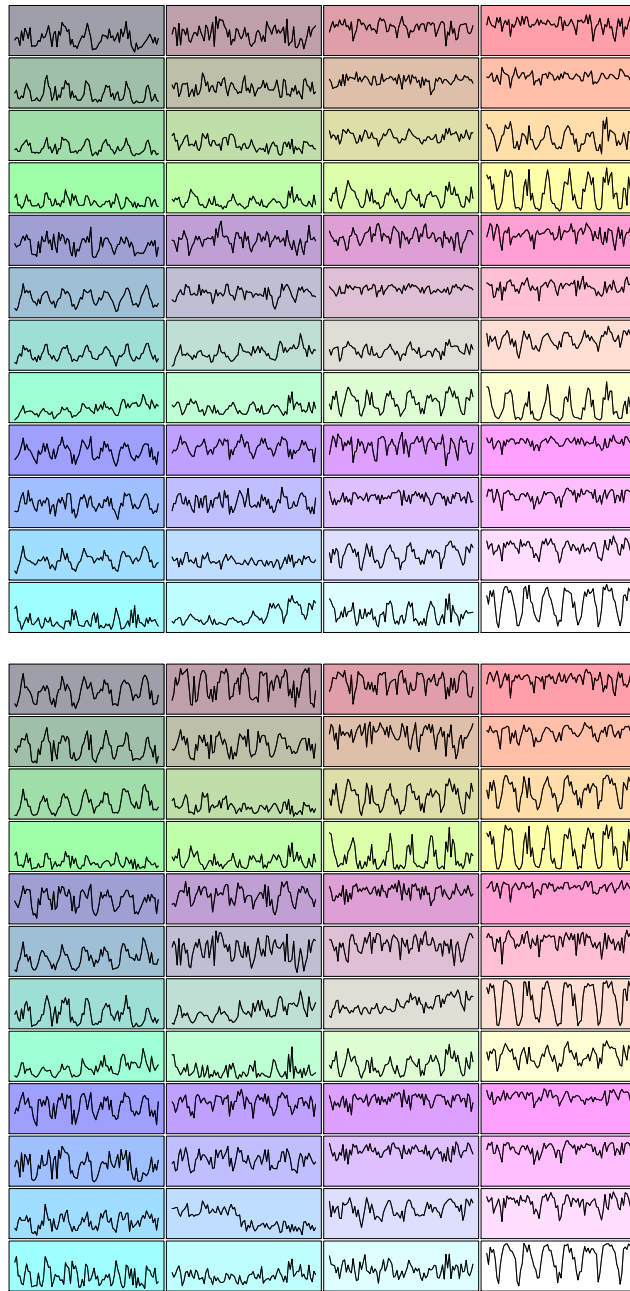


Figure B.5: Figures for centroids pattern with time in months on the x-axis. The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.1.4 Centroid hierarchical clustering

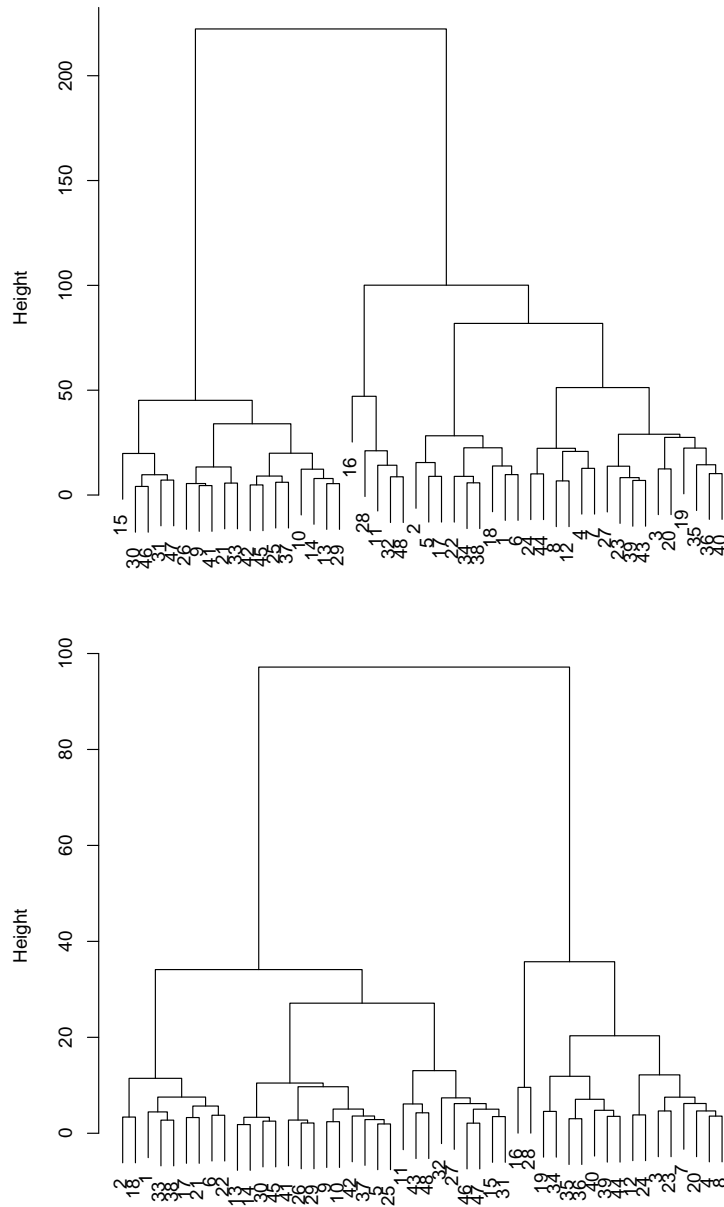


Figure B.6: Each figure shows the dendrogram computed with the Ward's linkage method. The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

B.1.5 Merging Costs

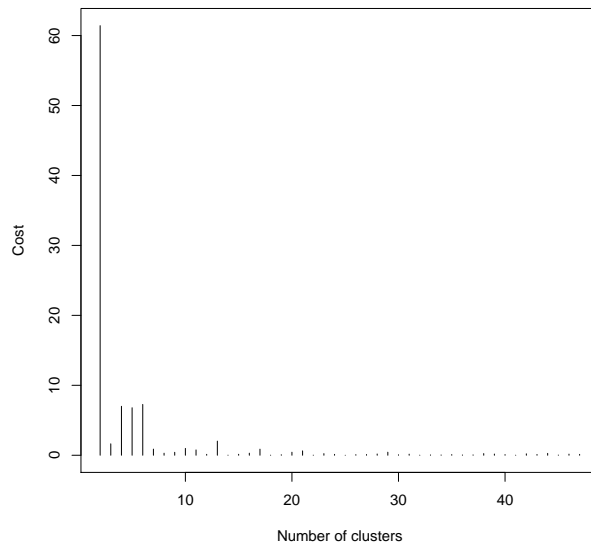
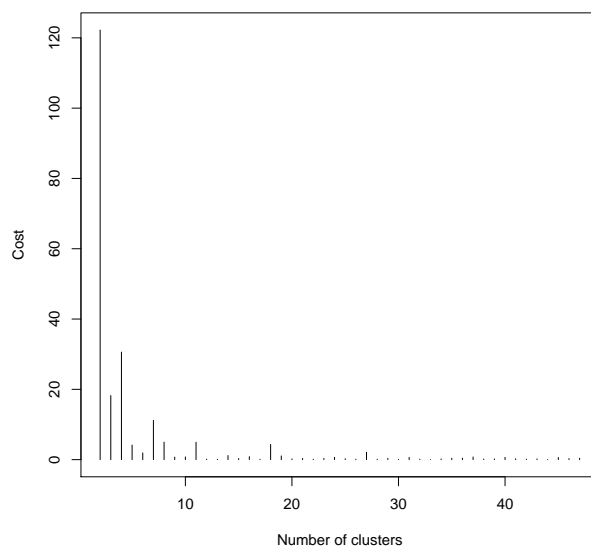


Figure B.8: Each figure shows the costs by merging clusters. The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

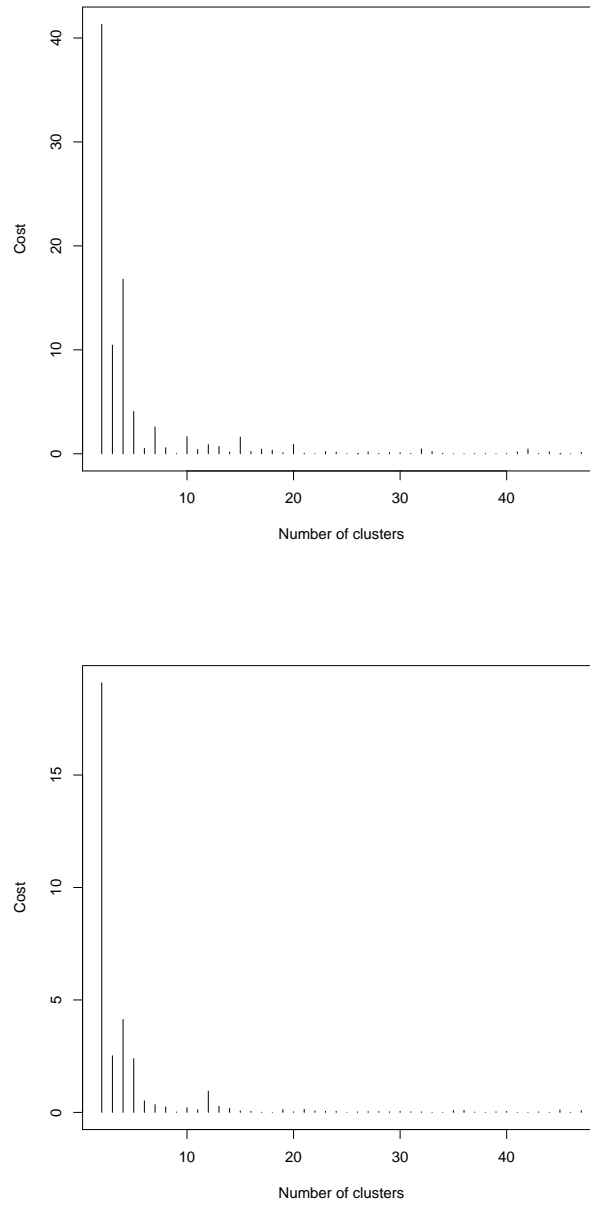
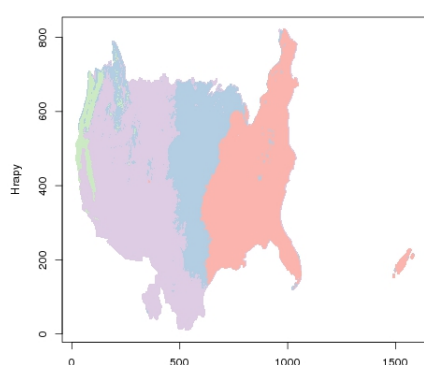


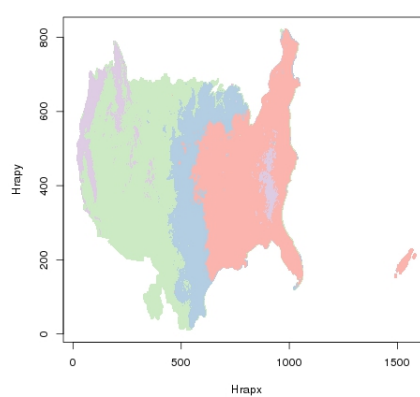
Figure B.9: Each figure shows the costs by merging clusters. The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.2 Clustering with linear initialisation

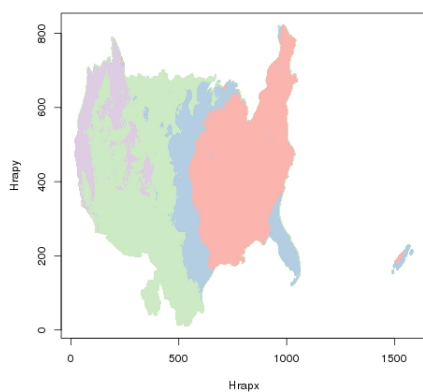
B.2.1 Resulting clusters (PCA)



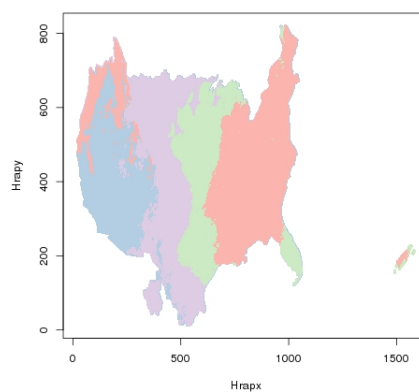
(a) No transformation applied



(b) Mean-variance standardisation



(c) Logarithmic standardisation



(d) Range standardisation

Figure B.10: Different precipitation areas after cluster validation.

B.2.2 Clustering validation graphs (PCA)

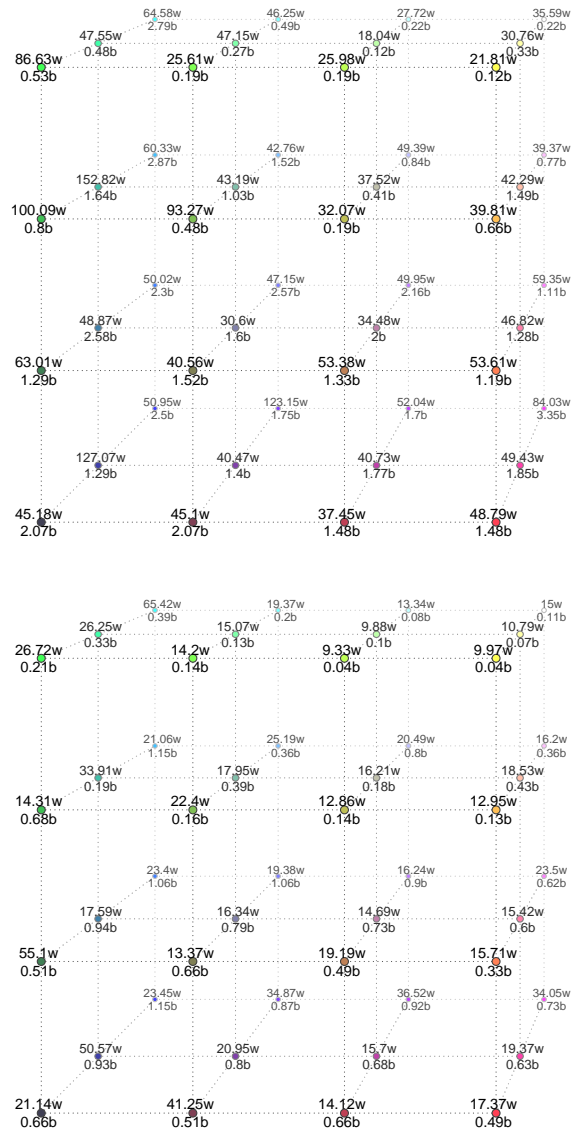


Figure B.11: Each map shows for each neuron the between clusters distance (b) and the clusters diameters (w). The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

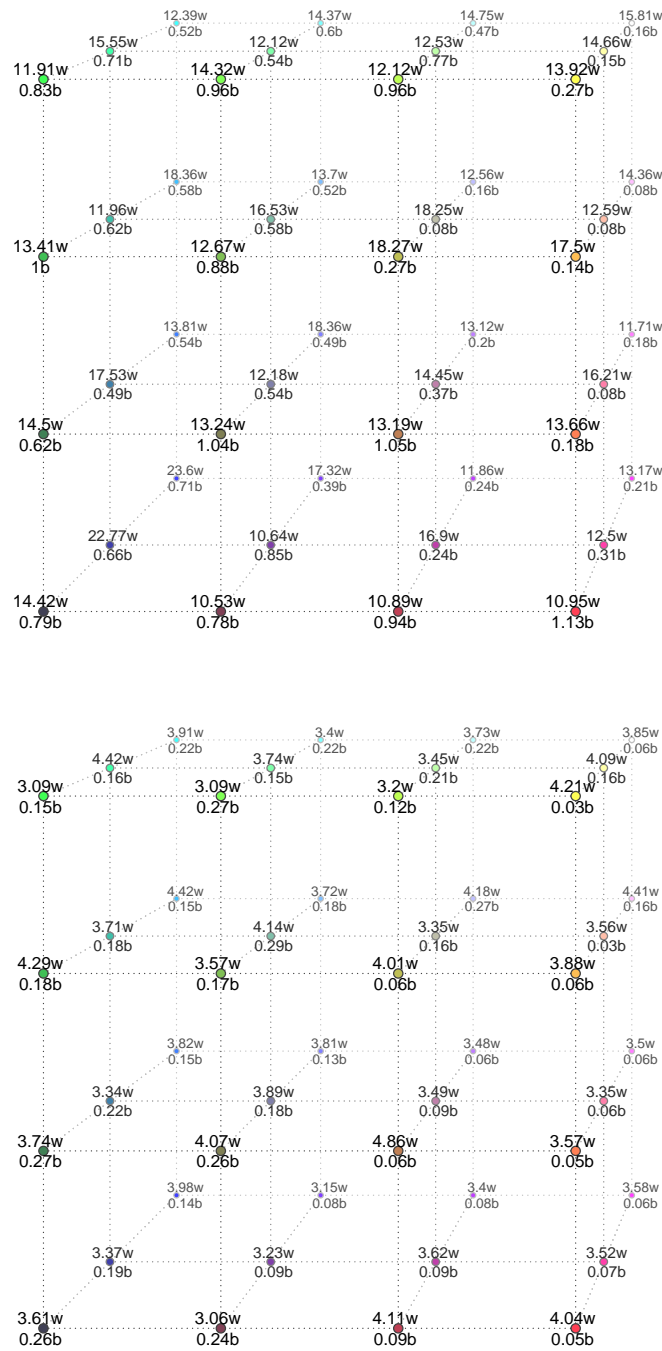


Figure B.12: Each map shows for each neuron the between clusters distance (b) and the clusters diameters (w). The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.2.3 Resulting centroids (PCA)

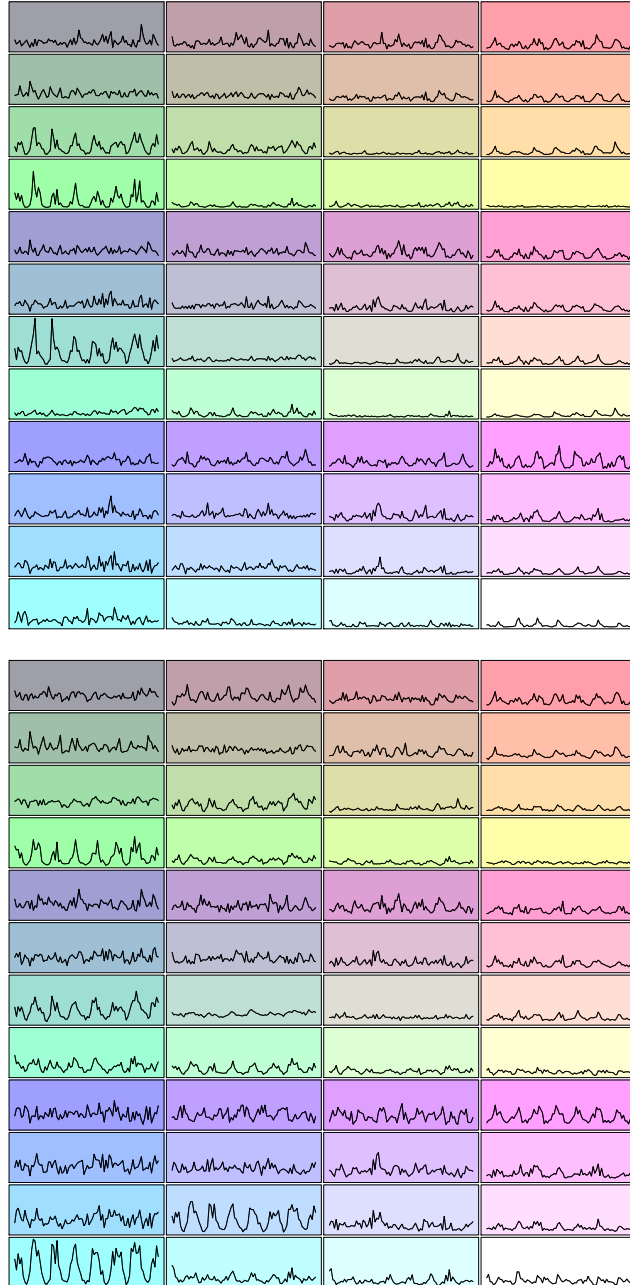


Figure B.13: Figures for centroids pattern with time in months on the x-axis. The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

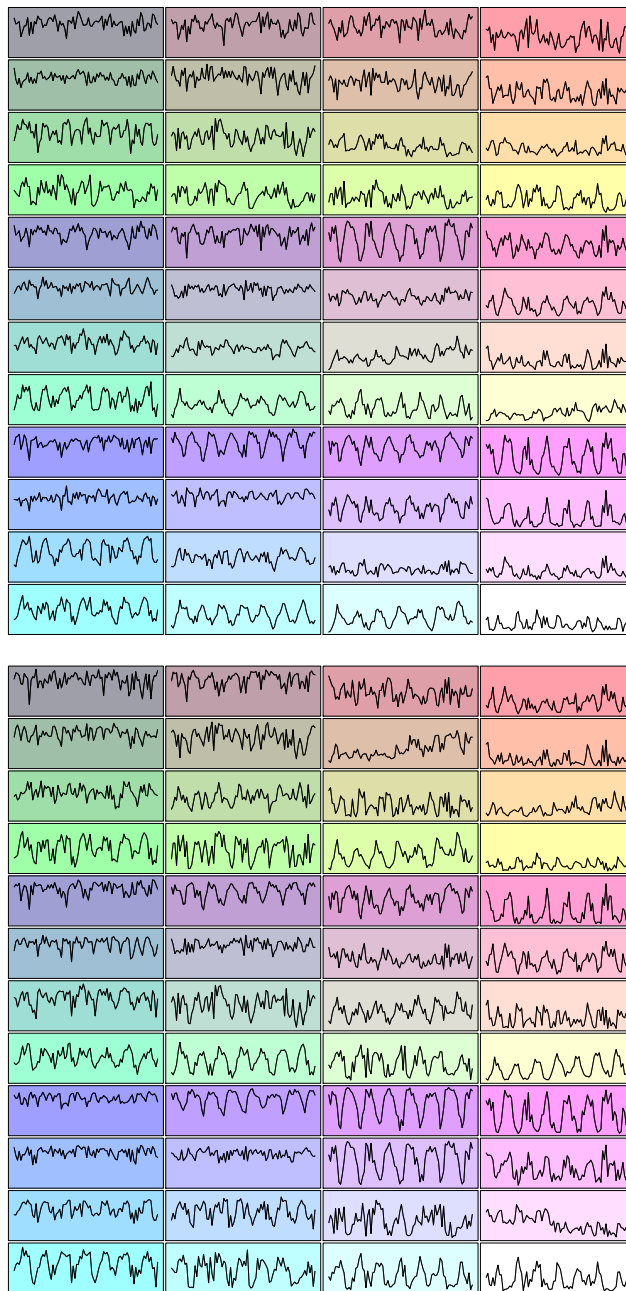


Figure B.14: Figures for centroids pattern with time in months on the x-axis. The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.2.4 Centroid hierarchical clustering (PCA)

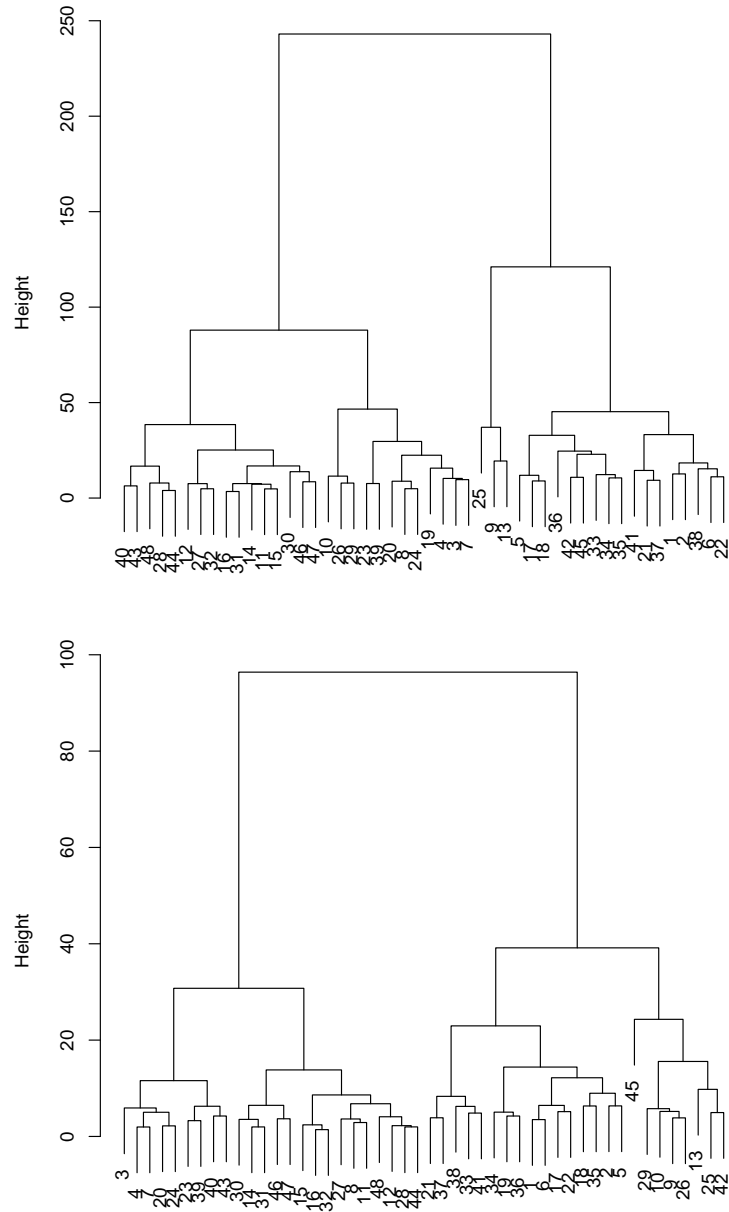


Figure B.15: Each figure shows the dendrogram computed with the Ward's linkage method. The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

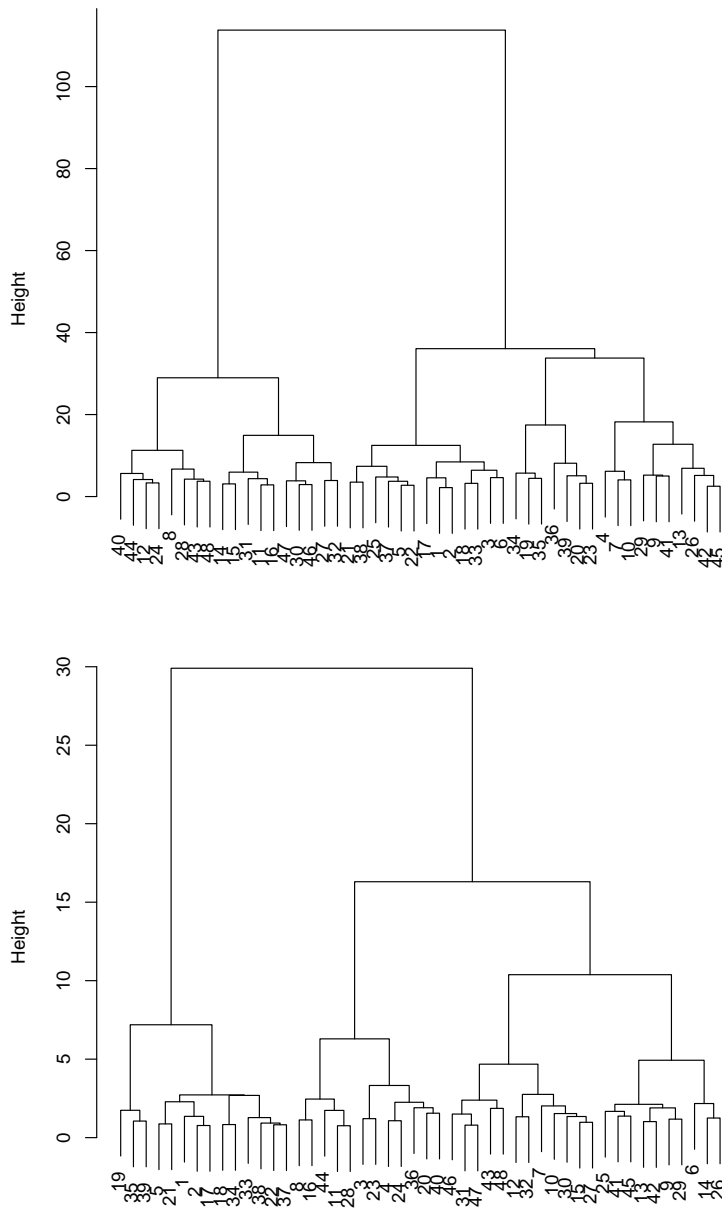


Figure B.16: Each figure shows the dendrogram computed with the Ward's linkage method. The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.2.5 Merging Costs (PCA)

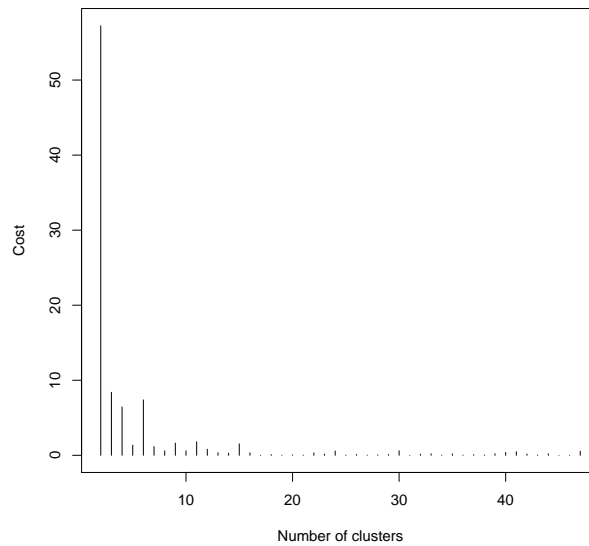
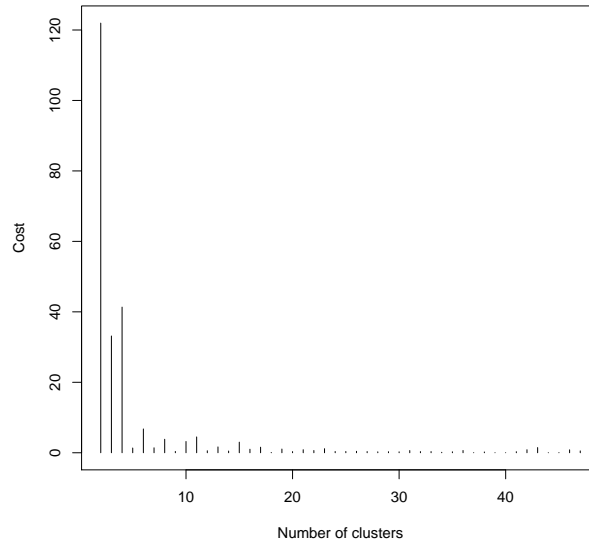


Figure B.17: Each figure shows the costs by merging clusters. The top figure is for the non transformed data and the bottom figure is for the mean-variance standardised data.

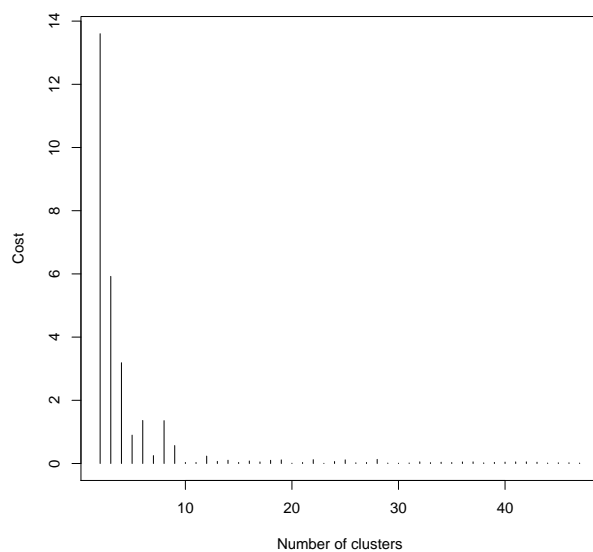
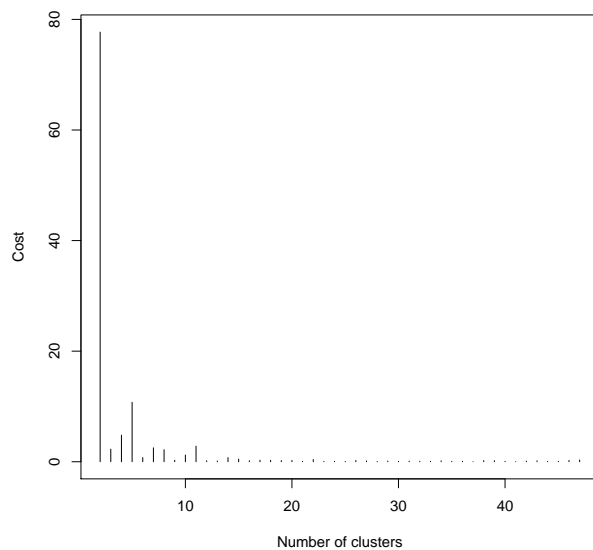


Figure B.18: Each figure shows the costs by merging clusters. The top figure is for the logarithmic standardised data and the bottom figure is for the range standardised data.

B.3 Clustering with informative indexes

B.3.1 Clustering validation graphs (indexes)

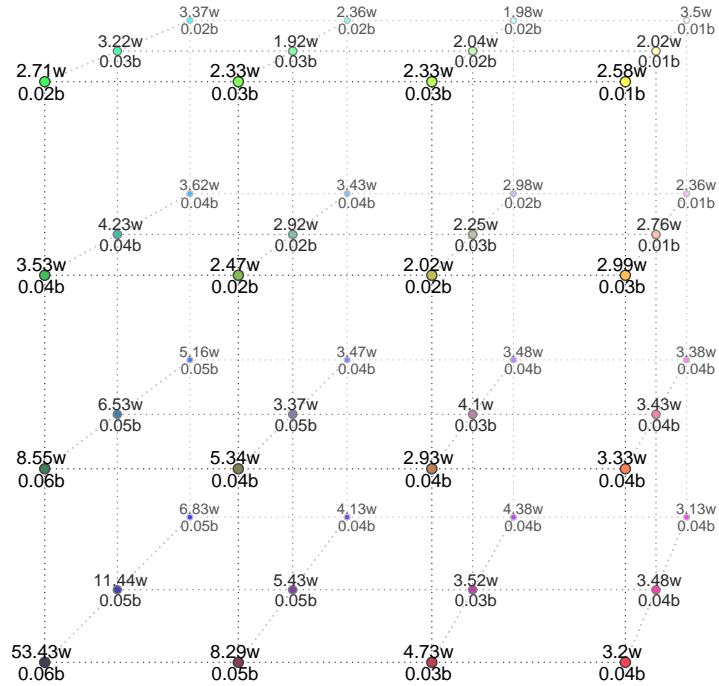


Figure B.19: The map shows for each neuron the between clusters distance (b) and the clusters diameters (w).

B.3.2 Resulting centroids (indexes)

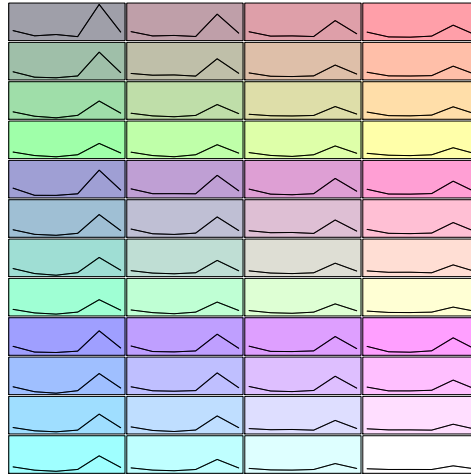


Figure B.20: Figure for centroids pattern with the first three point (from the left to the right) the mean estimates of β_0 , α_1 and β_1 for the group, the mean the minimum and the maximum observed precipitation and the mean standard deviation of the residuals.

B.3.3 Centroid hierarchical clustering (indexes)

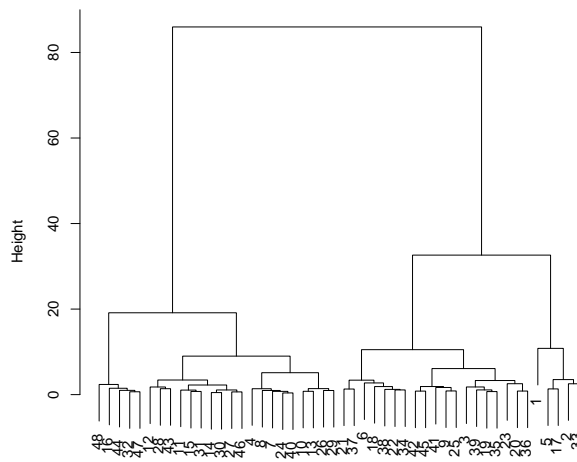


Figure B.21: The figure shows the dendrogram computed with the Ward's linkage method.

B.3.4 Merging Costs (indexes)

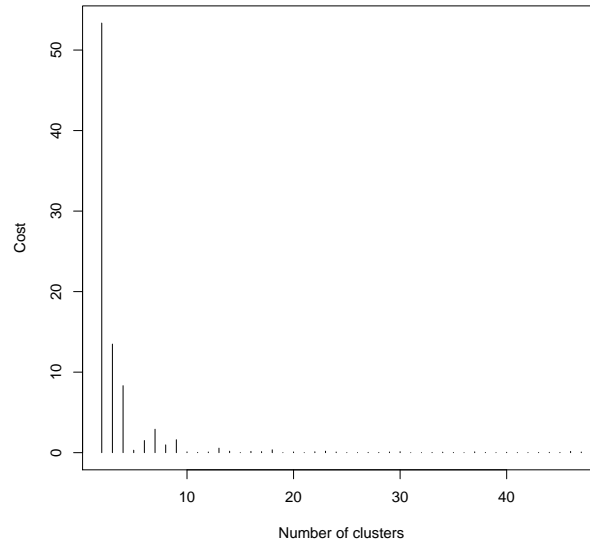


Figure B.22: The figure shows the costs by merging clusters.

B.3.5 Resulting clusters (indexes)

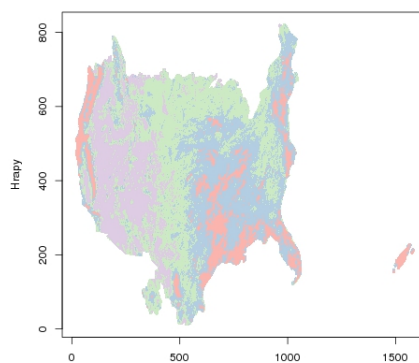


Figure B.23: Different precipitation areas after cluster validation.

Bibliography

- Abraham, C., Cornillon, P. A., Matzner-Løber, E., and Molinari, N. 2003. Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics*, **30**(3), 581–595.
- Ahrens, J. H., and Dieter, U. 1974. Computer methods for sampling from gamma, beta, poisson and binomial distributions. *Computing*, **12**(3), 223–246.
- Ahrens, J. H., and Dieter, U. 1982. Generating gamma variates by a modified rejection technique. *Communications of the ACM*, **25**(1), 47–54.
- Andreae, M. O., and Rosenfeld, D. 2008. Aerosol–cloud–precipitation interactions. Part 1. The nature and sources of cloud-active aerosols. *Earth-Science Reviews*, **89**(1), 13–41.
- Athan, T. W., and Papalambros, P. Y. 1996. A note on weighted criteria methods for compromise solutions in multi-objective optimization. *Engineering Optimization*, **27**(2), 155–176.
- Bass, F. M. 1969. A new product growth for model consumer durables. *Management Science*, **15**(5), 215–227.
- Bass, F. M., Krishnan, T. V., and Jain, D. C. 1994. Why the Bass model fits without decision variables. *Marketing Science*, 203–223.
- Baykal, N., and Erkmen, A. M. 1999. Extended self organizing feature map: a tagged potential field approach. *Neural Processing Letters*, **10**(1), 57–72.
- Behl, P., Claeskens, G., and Dette, H. 2012. Focused model selection in quantile regression.
- Bemmaor, A. C. 1994. *Research traditions in marketing*. Boston, MA: Kluwer Academic Publishers. Chap. Modeling t, pages 201–223.

- Bernstein, D. S. 2009. *Matrix mathematics: theory, facts, and formulas*. Princeton University Press.
- Bernstein, S. N. 1912. Démonstration du théoreme de Weierstrass fondée sur le calcul des probabilités. *Comm. Soc. Math. Kharkov*, **13**(1).
- Bondell, H. D., Reich, B. J., and Wang, H. 2010. Noncrossing quantile regression curve estimation. *Biometrika*, **97**(4), 825–838.
- Boscovich, R. J. 1779. *Les éclipses: poëme (latin) en six chants*.
- Bridgman, P. W. 1922. *Dimensional analysis*. Yale University Press.
- Busovaca, S. 1985. Handling degeneracy in a nonlinear l_1 algorithm.
- Chen, C., and Wei, Y. 2005. Computational issues for quantile regression. *Sankhyā: The Indian Journal of Statistics*, 399–417.
- Cheng, R. C. H. 1978. Generating beta variates with nonintegral shape parameters. *Communications of the ACM*, **21**(4), 317–322.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. 2009. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, **96**(3), 559–575.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. 2010. Quantile and probability curves without crossing. *Econometrica*, **78**(3), 1093–1125.
- Cordeiro, G. M., and de Castro, M. 2011. A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, **81**(7), 883–898.
- Cruz-Uribe, D., Neugebauer, C. J., and Lafayette, W. 2002. Sharp error bounds for the trapezoidal rule and Simpson’s rule. *Journal of Inequalities in Pure and Applied Mathematics*, **3**(4), 1–22.
- Dharmadhikari, S. W., and Joag-Dev, K. 1989. Upper bounds for the variances of certain random variables. *Communications in Statistics-Theory and Methods*, **18**(9), 3235–3247.
- Dunn, P. K., and Smyth, G. K. 1996. Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**(3), 236–244.

- Dunstan, F. D. J., Barham, S., Kemp, K. W., Nix, A. B. J., Rowlands, R. J., Wilson, D. W., Phillips, M. J., and Griffiths, K. 1982. A feasibility study for early detection of breast cancer using breast skin temperature rhythms. *The Statistician*, **31**(1), 37–52.
- Edgeworth, F. Y. 1888. XXII. On a new method of reducing observations relating to several quantities. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **25**(154), 184–191.
- Eilers, P. H. C., and Marx, B. D. 1996. Flexible smoothing with B-splines and penalties. *Statistical science*, **11**(2), 89–102.
- El-Attar, R. A., Vidyasagar, M., and Dutta, S. R. K. 1979. An algorithm for l_1 -norm minimization with application to nonlinear l_1 -approximation. *SIAM Journal on Numerical Analysis*, **16**(1), 70–86.
- Eugene, N., Lee, C., and Famoye, F. 2002. Beta-normal distribution and its applications. *Communications in Statistics. Theory and Methods*, **31**(4), 497–512.
- Faraway, J. J. 2002. *Practical regression and ANOVA using R*.
- Feingold, G., Koren, I., Wang, H., Xue, H., and Brewer, W. A. 2010. Precipitation-generated oscillations in open cellular cloud fields. *Nature*, **466**(7308), 849–852.
- Fishburn, P. C. 1974. Lexicographic orders, utilities and decision rules: a survey. *Management Science*, 1442–1471.
- Gauch, J. M., and Hsia, C. W. 1992. Comparison of three-color image segmentation algorithms in four color spaces. *Pages 1168–1181 of: Proceedings of SPIE*, vol. 1818.
- Gauss, C. F. 1809. *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*.
- Gordon, A. D. 1999. *Classification*. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC.
- Gorricha, J., and Lobo, V. 2011. Improvements on the visualization of clusters in geo-referenced data using self-organising maps. *Computers & Geosciences*.

- Gorricha, J., Costa, A., and Lobo, V. 2012. Spatial characterization of extreme precipitation in Madeira island using geostatistical procedures and a 3D SOM. *Pages 98–104 of: GEOProcessing 2012, The Fourth International Conference on Advanced Geographic Information Systems, Applications, and Services.*
- Guercin, J. P. 1973. Les variations périodiques du nombre d'accidents de la circulation. *Journal de la Société de statistique de Paris*, **114**, 64–67.
- Guidolin, M., and Guseo, R. 2013. Modelling seasonality in innovation diffusion. *Technological Forecasting and Social Change.*
- Guseo, R. 2004. *Interventi strategici e aspetti competitivi nel ciclo di vita di innovazioni.* Tech. rept. Working Paper Series.
- Guseo, R., and Dalla Valle, A. 2005. Oil and gas depletion: diffusion models and forecasting under strategic intervention. *Statistical Methods and Applications*, **14**(3), 375–387.
- Guseo, R., and Guidolin, M. 2010. Cellular automata with network incubation in information technology diffusion. *Physica A: Statistical Mechanics and its Applications*, **389**(12), 2422–2433.
- Gutenbrunner, C., and Jurečková, J. 1992. Regression rank scores and regression quantiles. *The Annals of Statistics*, **20**(1), 305–330.
- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. 2001. On clustering validation techniques. *Journal of Intelligent Information Systems*, **17**(2), 107–145.
- Han, D. 2011. *Concise hydrology.*
- Hardy, G. H., Littlewood, J. E., and Polya, G. 1934. Inequalities.
- Hastie, T., Tibshirani, R., and Friedman, J. 2001. *The elements of statistical learnin.*
- Hastie, T. J., and Tibshirani, R. J. 1990. *Generalized additive models.* Monographs on Statistics and Applied Probability, vol. 43. Chapman and Hall London.
- He, X. 1997. Quantile curves without crossing. *American Statistician*, 186–192.

- He, X., and Hu, F. 2002. Markov chain marginal bootstrap. *Journal of the American Statistical Association*, **97**(459), 783–795.
- He, X., and Zhu, L. X. 2003. A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, **98**(464), 1013–1022.
- Heylighen, F. 2001. The science of self-organization and adaptivity. *The Encyclopedia of Life Support Systems*, **5**(3), 253–280.
- Hoeffding, W. 1963. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, **58**(301), 13–30.
- Hogg, R. V. 1975. Estimates of percentile regression lines using salary data. *Journal of the American Statistical Association*, 56–59.
- Hyndman, R. J., and Fan, Y. 1996. Sample quantiles in statistical packages. *The American Statistician*, **50**(4), 361–365.
- Kocherginsky, M., and He, X. 2007. Extensions of the Markov chain marginal bootstrap. *Statistics & Probability Letters*, **77**(12), 1258–1268.
- Kocherginsky, M., He, X., and Mu, Y. 2005. Practical confidence intervals for regression quantiles. *Journal of Computational and Graphical Statistics*, **14**(1), 41–55.
- Koenker, R. 1984. A note on L-estimates for linear models. *Statistics & Probability Letters*, **2**(6), 323–325.
- Koenker, R. 1994. Confidence intervals for regression quantiles. *Asymptotic Statistics*, 349–359.
- Koenker, R. 2005. *Quantile regression*. Vol. 38. Cambridge University Press.
- Koenker, R., and Bassett Jr., G. 1978. Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33–50.
- Koenker, R., and Machado, J. A. F. 1999. Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 1296–1310.
- Koenker, R., and Ng, P. 2005. Inequality constrained quantile regression. *Sankhyā: The Indian Journal of Statistics*, 418–440.
- Koenker, R., and Park, B. J. 1996. An interior point algorithm for nonlinear quantile regression. *Journal of Econometrics*, **71**(1), 265–283.

- Kohonen, T. 1990. The self-organizing map. *Proceedings of the IEEE*, **78**(9), 1464–1480.
- Kohonen, T. 2001. *Self-organizing maps*. Springer Series in Information Sciences. Springer.
- Koren, I., and Feingold, G. 2011. Aerosol–cloud–precipitation system as a predator-prey problem. *Proceedings of the National Academy of Sciences*, **108**(30), 12227.
- Laplace, P. S. 1846. *Oeuvres de Laplace: Traité de mécanique céleste*. Oeuvres de Laplace. Imprimerie Royale.
- Legendre, A. M. 1806. Nouvelle formule pour réduire en distances vraies les distances apparentes de la Lune au Soleil ou à une étoile. *Pages 30–54 of: Mémoires de l'Académie des sciences de l'Institut de France*. Didot frères, fils et cie.
- Levenberg, K. 1944. A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, **2**, 164–168.
- Levin, Z., and Cotton, W. R. 2008. *Aerosol pollution impact on precipitation: a scientific review*. Springer Verlag.
- Levizzani, V., Bauer, P., and Turk, F. J. 2007. *Measuring precipitation from space: EURAINSAT and the future*. Advances in global change research. Springer.
- Lotka, A. J. 1910. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, **14**(3), 271–274.
- Markwardt, C. B. 2009. Non-linear least-squares fitting in IDL with MPFIT. *Pages 251–254 of: Bohlender, D. A., Durand, D., and Dowler, P. (eds), Astronomical Data Analysis Software and Systems XVIII*. Astronomical Society of the Pacific Conference Series, vol. 411.
- Marler, R. T., and Arora, J. S. 2004. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, **26**(6), 369–395.
- Marquardt, D. W. 1963. An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics*, **11**(2), 431–441.

- Matsumoto, M., and Nishimura, T. 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans. Model. Comput. Simul.*, **8**(1), 3–30.
- McKean, J. W., and Sievers, G. L. 1987. Coefficients of determination for least absolute deviation analysis. *Statistics & Probability Letters*, **5**(1), 49–54.
- Mehrotra, S. 1992. On the implementation of a primal-dual interior point method. *SIAM Journal on Optimization*, **2**(4), 575–601.
- Meketon, M. S. 1987. Least absolute value regression. *Manuscript, AT&T Bell Laboratories, Holmdel, New Jersey*.
- Michaelides, S. 2008. *Precipitation: advances in measurement, estimation, and prediction*. Springer.
- Montgomery, D. C. 2000. *Introduction to statistical quality*. John Wiley. New York.
- Nagel, K., and Raschke, E. 1992. Self-organizing criticality in cloud formation? *Physica A: Statistical Mechanics and its Applications*, **182**(4), 519–531.
- Osborne, M. R., and Watson, G. A. 1971. On an algorithm for discrete nonlinear L1 approximation. *The Computer Journal*, **14**(2), 184–188.
- Osyczka, A. 1978. An approach to multicriterion optimization problems for engineering design. *Computer Methods in Applied Mechanics and Engineering*, **15**(3), 309–333.
- Pareto, V. 1906. *Manuale di economia politica*. Società Editrice.
- Parzen, M. I., Wei, L. J., and Ying, Z. 1994. A resampling method based on pivotal estimating functions. *Biometrika*, **81**(2), 341–350.
- Petters, M. D., and Kreidenweis, S. M. 2007. A single parameter representation of hygroscopic growth and cloud condensation nucleus activity. *Atmospheric Chemistry and Physics*, **7**(8), 1961–1971.
- Powell, M. J. D. 1969. *A hybrid method for equations*. Tech. rept. T.P. 364. A.E.R.E. Didcot, Berkshire, England.

- Rayleigh, L. 1916. LIX. On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **32**(192), 529–546.
- Redden, D. T., Fernández, J. R., and Allison, D. B. 2004. A simple significance test for quantile regression. *Statistics in medicine*, **23**(16), 2587–2597.
- Riccati, J. 1758. *Opere*. Treviso: appresso Jacopo Giusti.
- Santos, B. R., and Elian, S. N. 2012. Analysis of residuals in quantile regression: an application to income data in Brazil. *Pages 723–728 of: Proceeding of the 27-th International Workshop on Statistical Modelling*. Charles University in Prague.
- Schnabel, S. K., and Eilers, P. H. C. 2013. Simultaneous estimation of quantile curves using quantile sheets. *AStA Advances in Statistical Analysis*, **97**(1), 77–87.
- Seber, G. A. F., and Wild, C. J. 2003. *Nonlinear regression*. Vol. 503. LibreDigital.
- Stevens, B., and Feingold, G. 2009. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature*, **461**(7264), 607–613.
- Tang, C., and Bak, P. 1988. Mean field theory of self-organized critical phenomena. *Journal of Statistical Physics*, **51**(5-6), 797–802.
- Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.
- Uspensky, J. V. 1937. *Introduction to mathematical probability*. McGraw-Hill Book Company, Inc.
- Volterra, V. 1927. *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*. C. Ferrari.
- Wagner, H. M. 1959. Linear programming techniques for regression analysis. *Journal of the American Statistical Association*, 206–212.
- Wilcox, R. R. 2008. Quantile regression: a simplified approach to a goodness-of-fit test. *Journal of Data Science*, **6**, 547–556.

-
- Wu, Y., and Liu, Y. 2009. Stepwise multiple quantile regression estimation using non-crossing constraints. *Statistics and Its Interface*, **2**, 299–310.
- Yu, K., and Zhang, J. 2005. A three-parameter asymmetric Laplace distribution and its extension. *Communications in Statistics—Theory and Methods*, **34**(9-10), 1867–1879.
- Zeleny, M. 1973. Compromise programming. *Multiple Criteria decision Making*, 262–301.
- Zou, H., and Yuan, M. 2008. Regularized simultaneous model selection in multiple quantiles regression. *Computational Statistics & Data Analysis*, **52**(12), 5296–5304.

Luca Sartore

CURRICULUM VITAE

Personal Details

Date of Birth: February 18, 1985
Place of Birth: Cittadella (PD), Italy
Nationality: Italian

Contact Information

University of Padova
Department of Statistics
via Cesare Battisti, 241-243
35121 Padova. Italy.

Tel. +39 049 827 4111
e-mail: sartore@stat.unipd.it

Current Position

Since January 2011; (expected completion: January 2014)
PhD Student in Statistical Sciences, University of Padova.
Thesis title: Quantile Regression and Bass Models in Hydrology
Supervisor: Prof. Renato Guseo
Co-supervisor: Dr. Claudia Furlan.

Research interests

- High performance computational statistics
- Spatiotemporal statistics
- Regression methods and multivariate statistics
- Signal processing

Education

September 2008 – July 2010
Master (laurea specialistica/magistrale) degree in Statistics for Business Management.
University Ca' Foscari of Venice, Faculty of Economy
Title of dissertation: "Geostatistical models for 3-D data"
Supervisor: Prof. Carlo Gaetan
Final mark: 110/110 with honors

September 2004 – April 2008
Bachelor degree (laurea triennale) in Statistics and Computer Science for Business Management.
University Ca' Foscari of Venice, Faculty of Economy
Title of dissertation: "Verosimiglianza composita per dati geostatistici: metodi di stima ed un'applicazione"
Supervisor: Prof. Carlo Gaetan
Final mark: 108/110.

Work experience

August 2007 – October 2007

Pegaso IT Ltd.

Software Programmer (*training*).

Awards and Scholarship

2011

PhD scholarship (University of Padua).

Computer skills

- Operative system: Linux, Windows, OSX
- Programming: R, C, Qt, Visual Basic .Net, VBScripts, JavaScripts, Fortran
- Markup Languages: L^AT_EX, HTML, XML, CSS, ASP
- Information Systems: Oracle Database, SQL, MySQL, GIS
- Other skills: Parallel computing (OpenMP and CUDA)

Language skills

Venetian / Italian: native; English: fluent (written/spoken); German: intermediate (written/spoken).

Publications

Articles in journals

Sartore L. (2013). spMC: modelling spatial random fields with continuous lag Markov chains. *The R Journal* Vol. 5 No. 2, 16–28.

Conference presentations

Sartore, L. (2013). Constrained simultaneous nonlinear quantile regression. (contributed oral presentation) *DAGStat 2013*, Freiburg, Germany, March 18-22.

Other Interests

Organ music (playing and composing)

Choral music (http://www2.cpd1.org/wiki/index.php/Luca_Sartore)

References

Prof. Renato Guseo

Department of Statistical Science
University of Padua
Via C. Battisti, 241
35121 Padova (ITALY)
Phone: +39 049 827 4146
e-mail: renato.guseo@unipd.it

Prof. Paolo Fabbri

Department of Geosciences
University of Padua
Via Gradenigo, 6
35121 Padova (ITALY)
Phone: +39 049 827 9124
e-mail: paolo.fabbri@unipd.it

Prof. Carlo Gaetan

Department of Environmental Sciences, Informatics and Statistics
Ca' Foscari University of Venice
San Giobbe, Cannaregio 873
30121 Venezia (ITALY)
Phone: +39 041 234 7437
e-mail: gaetan@unive.it