STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN

Mercè Castellà Sánchez

Dipòsit Legal: T 962-2014

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN

PhD Thesis

Mercè Castellà Sánchez

Universitat Rovira i Virgili

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Mercè Castellà Sánchez

Statistical modelling and analysis of summer very hot events in mainland Spain

PhD Thesis

Thesis Advisor: Prof. Manola Brunet India

Geography Department

Centre for Climate Change

UNIVERSITAT ROVIRA I VIRGILI

Tortosa

2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI

DEPARTAMENT DE GEOGRAFIA-URV

C/Joanot Martorell, 15
43480 Vila-seca
Tel. 977 55 87 51

FAIG CONSTAR que aquest treball, titulat “Statistical modelling and analysis of summer very hot events in mainland Spain”, que presenta Mercè Castellà Sánchez per a l’obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament de Geografia d’aquesta universitat.

Tortosa, 7 de gener de 2014

La directora de la tesi doctoral

Dra. Manola Brunet India

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

In loving memory of my mother

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Content

Acknowledgements i

Abstract iii

1 Introduction 1

2 Statistical theory of extreme values analysis 9

2.1 Classical Extreme Value Theory (EVT) 11

2.2 The r Largest Order Statistic model 15

2.3 Peaks Over Threshold (POT) approach 17

2.4 Poisson-GPD model for excesses 19

2.5 Point Process approach 20

2.5.1 Parameter estimation 22

2.5.2 Incorporating non-stationarity into the model 24

2.5.3 Uncertainty and confidence intervals 24

2.5.4 Model diagnostics 28

2.5.5 Model selection 30

2.5.6 Effective return levels 30

3 Data and methodology 33

3.1 The area in study 33

3.2 Data description 34

3.2.1 The Spanish Daily Adjusted Temperature Series (SDATS) 34

3.2.2 Large-scale datasets 37

3.3 The Methodology applied 40

3.3.1 Definition of VHD and VHN 42

3.3.2 Threshold selection 43

3.3.3 Declustering 44

3.3.4 Model: Non-stationary point process 46
3.3.5 Suitability of the covariates into the model 48
3.3.6 Effective return levels 50
3.3.7 Composite maps 50
4 Results and discussion 51
4.1 Very Hot Days (VHD) 51
4.1.1 Influence of SLP anomalies on summer VHD 51
4.1.2 Influence of SST anomalies on summer VHD 61
4.1.3 Influence of SM anomalies on summer VHD 69
4.2 Very Hot Nights 75
4.2.1 Influence of SLP anomalies on summer VHN 75
4.2.2 Influence of SST anomalies on summer VHN 83
4.2.3 Influence of SM anomalies on summer VHN 88
4.3 Observed time trends in extreme temperatures 92
4.3.1 Observed trend in the location parameter (1940-1972) 93
4.3.2 Observed trends in the location parameter (1973-2010) 97
4.4 Observed trends in the scale parameter 103
4.5 Effective return levels of temperatures extremes (1973-2010) 109
5 Summary, conclusions and outlook 115
5.1 Influence of large-scale variables on VHD and VHN 115
5.2 Observed changes in extreme distributions and return levels 118
5.3 Conclusions and outlook 120
6 References 123
Appendix 131
List of Figures 157
List of Tables 163

Acknowledgements

I would like to express my gratitude to all who gave me the possibility to complete this thesis. First, I would like to express my gratitude to my PhD advisor Prof. Manola Brunet India for her continuous support and guidance during my study period. Thanks for proposing me this challenging topic of extreme value analysis, for the confidence in my capacity to carry out this task and for the encouragements along this path. Also, for the fruitful discussion we had when writing this thesis.

The PhD fellowship was granted by the Centre for Climate Change (C3). It has been a pleasure to be one of the first PhD students at the C3. This fellowship gave me the opportunity to work independently on my own research ideas. Writing this dissertation was one of the most challenging academic tasks I have ever had.

During the PhD period, I had the opportunity to participate in an Advanced Study Program during the Summer Colloquium on "Statistical Assessment of Extreme Weather Phenomena under Climate Change", which took place in the National Center for Atmospheric Research (NCAR), Colorado, USA. This activity was crucial to my career, since I had the opportunity to learn a lot about the statistical techniques which I have then applied in this work to model and analyse the extreme temperatures. I am very grateful to Eric Gilleland and Rick Katz for introducing me to the use of R package extRemes (freely available at http://www.r-project.org), which was used as a main statistical tool in my analysis and, also, for the scientific support and expert advice they gave me during these years.

In addition, I would like to thank my colleagues of the Centre for Climate Change (C3) to encourage me and share the office with a friendly atmosphere. I am especially grateful to Dr. Constanta Boroneant for her scientific discussions and advices, which helped me with the interpretation of the results. Also thanks for her patience, positive energy and encouragements and her willingness to review my thesis. I am also very grateful to Dr. Dimitrios Efthymiadis for his help and his valuable comments and suggestions.

I acknowledge all the useful discussions with the scientists I have met during the conferences, meetings and workshops.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

I am also grateful to people that I knew at the Campus Terres de l'Ebre to share good moments with me (coffee-breaks, lunches, talks, ...) and also for encourage me to finish my PhD.

My sincere thanks go to my brother Xavi for giving me warm family days. For the good times we have had, for example, good lunches and dinners in my home, which helped me a lot. Also thanks to him for mounting my domestic office, where I have written much of these lines and I finished the thesis.

Special thanks to my nieces, Mar and Emma, for their sincerely affection to me and the enjoyable moments shared, although, I saw them less than I would have wanted.

I would also like to thank my aunt Lolita for always being there, for listening to me and for giving me her support in difficult times.

Last but not least, I would like to give my most gratefulness to my best friend Graci Gómez for her unconditionally support, continuous encouragement and patience that helped me very much. Also, for believing in me and for the great times lived during this adventure. Thanks for everything!

Abstract

Extreme temperature events are of particular importance due to their severe impact on the environment, the economy and the society. Focused on the uppermost (>95th) percentiles of summer daily adjusted maximum (Tx) and minimum (Tn) temperature from the Spanish Daily Adjusted Temperature Series (SDATS), in this thesis a modelling and analysis of summer very hot days (VHD) and nights (VHN) over mainland Spain has been carried out by applying the methodology of Point Process (PP) Approach based on Extreme Value Theory. Through PP approach it has been investigated whether large-scale variables of Sea Level Pressure (SLP), Sea Surface Temperature (SST) and Soil Moisture (SM) are associated to the occurrence and intensity of these exceptional events. Furthermore, observed changes and trends in Tx and Tn extreme distribution have been analysed for two different periods 1940-1972 and 1973-2010 and 5, 10, 20, 50 and 100-year effective returns levels have been estimated for the most recent period.

Three large-scale atmospheric circulation patterns associated to the occurrence and intensity of VHD and VHN have been identified, showing stronger SLP anomalies during VHD events. The Southerly Flow Pattern which enforces a southerly component inflow of warm and dry air masses from Saharan Africa to affect the IP. The Weak South-westerly Airflow Pattern, which gives a weak warm westerly or south-westerly airflow over the IP and the North-westerly Airflow Pattern, which returns a warmed north-westerly airflow when it passes across the IP, although much weaker for VHN episodes. SST anomalies preceding an extreme temperature event have, in general, an important role in the intensity and frequency of VHD and VHN in mainland Spain, although the effect is not homogenous in space. Another finding is related to the influence of SM anomaly prior an extreme temperature event occurs, since it has been found that SM deficit during the previous days of an extreme event has an important contribution to its occurrence and intensity over all the locations analysed except over the northern coast.

Changes in extreme temperatures were observed at all stations, but they are not uniformly distributed in space and time. In general, these changes have been attributed to changes in the location parameter μ rather than in the scale parameter σ of the GEV distribution. Results reveal different behaviour of changes in both extreme distributions

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

for each period analysed. A shift toward colder values of both Tx and Tn extreme distributions have been detected over the period 1940-1972, while for the period 1973-2010 a meaningful shift toward warmer values have been observed especially for Tn extreme distribution. For the last warm period and for Tx, coastal and north-western locations exhibit the highest trends in μ. In the case of Tn extremes, the highest trends were found mainly in southern Spain. The estimations of the 20-year return level suggest increases of extreme temperatures for all the analysed series, although the largest increases in daily Tx extremes have been found in the northern coast of Spain and in daily Tn extremes in north-eastern Spain.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

1 Introduction

Hot extreme temperature events are of great scientific and societal interest and are receiving increasingly attention in the recent years due to their strong environmental and socioeconomic impacts. Drought coupled with extreme hot temperatures and low humidity can increase the risk of wildfire (IPCC, 2012). The agriculture sector is also influenced by extreme hot temperatures, since different crop species are very sensitive to extreme temperatures (Hatfield et al., 2011), particularly the grain yields (Prasad et al., 2006) and cereals (Rodríguez-Puebla et al., 2007). When extreme hot temperatures persist, water resources are affected as well due to increasing water demand and also power lines sag in because of the high demand of electricity (Colombo et al., 1999). This cascade of hot temperatures also includes affectations in human health, comfort and mortality (García-Herrera et al., 2005; Tobías et al., 2010; Tobías et al., 2012) such as the high toll paid during the summer 2003 and 2010 heat waves in Europe, hence advocating for the need to improve our knowledge on their changes in occurrence, intensity and duration in a changing climate.

Changes in extreme temperature are considerably different to changes in the mean temperature (e.g. Brown et al., 2008; Klein Tank et al. 2009). Also, climate model projections suggest considerably different changes in extreme temperatures than in their mean (Kharin and Zwiers, 2005; Clark et al., 2006). In addition, changes in extreme temperatures have an important contribution on climate change impacts (Meehl and Tebaldi, 2004).

Due to the interest of analysing extreme temperature episodes during recent years, an increased number of scientific publications assessing extreme temperature events are available, emphasising the scientific importance of their study. In this regard and in addition to many scientific papers, one of the most recent handbooks published in the field of climate extremes analysis is written by AghaKouchak et al. (2013), which provides a collection of the state-of-the-art methodologies and approaches suggested for detecting extremes, trend analysis, accounting for non-stationarities, and uncertainties associated with extreme value analysis in a changing climate. This is a clear indication of scientific interest for better knowing climate extremes triggers and impacts.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

A number of scientific articles have been dedicated to study changes in extreme temperatures at different spatial and temporal scales. At global scale (Alexander et al., 2006; Brown et al., 2008; Kharin and Zwiers, 2005; Donat et al., 2013), at the European (Moberg and Jones, 2005; Moberg et al., 2006; Klein Tank and Können, 2003) and national and local scales (Brunet et al., 2007a and 2007b; Furió and Meneu, 2011; Rodríguez-Puebla et al., 2010) for the Iberian Peninsula (IP) or some of its regions (Abaurrea and Cebrián, 2002; Abaurrea et al., 2007; El Kenawy et al., 2011; Lana et al., 2009; Serra et al., 2010). Their results may differ to some extent due to the methodological approaches adopted, data used, spatial and temporal scales analysed and their scopes. In this regard, Alexander et al. (2006) and recently Donat et al. (2013) computed and analysed seasonal and annual extreme indices derived from daily temperature and precipitation station data at global scale, finding significant changes since the mid-20th century, particularly strong from 1979 onwards. Both assessments reported significant shift in the probability distribution function of temperature indices associated with warming in a very large area of the Northern Hemisphere midlatitudes, especially for those indices derived from daily minimum temperature (Tn), although the indices derived from daily maximum temperature (Tx) showed similar changes but to a smaller extent. Brown et al. 2008 analysed the observed daily temperature anomalies with regard to the normal climate (1961-1990) and concluded that since 1950 extreme daily maximum and minimum temperatures warmed over most regions showing a significant positive trend in extreme daily temperature anomalies for both upper and lower tails of their distributions. For most regions, positive trend magnitudes were larger and covered a larger area for daily Tn than for Tx. In Europe, Moberg et al. (2006) also observed an overall warming during the entire 20th century. They found increasing trends both in daily Tx and Tn extreme indices averaged over the whole Europe. Over mainland Spain, Brunet et al. (2007a) found significant long-term (1850-2005) trends for summer extreme warm days and nights defined as Tx>90th and Tn>90th, respectively, with the highest coefficients estimated for the period 1973-2005. Also, at regional scale over Northeastern Spain, El Kenawy et al. (2011) found an increase in the frequency and intensity of hot extremes during the period 1960-2006, and this upward trend in hot extremes was more pronounced over the last two decades.

Hence, there is clear evidence from observations that warm extreme temperatures have considerably increased during recent years. Furthermore, climate projections

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

suggest that extreme temperature events will become more frequent and more severe in the future (Kharin and Zwiers, 2005; Meehl and Tebaldi, 2004).

The occurrence of extremes is usually the result of multiple factors, which can act either on the large-scale or on the regional (and local) scale (IPCC, 2012). For example, large-scale anomalous atmospheric circulation patterns can determine de main air flow over the study area and regional feedbacks linked to land-atmosphere interactions with, for example, the Soil Moisture (SM), can modulate overall changes in extremes (IPCC, 2012).

At large-scale, anomalous atmospheric conditions have been linked with extreme temperature events in Europe (Andrade et al., 2012; Cassou et al., 2005; Della-Marta et al., 2007). For example, the configuration of anomalously low sea level pressure (SLP) over central North Atlantic, anomalously high SLP over western Europe and negative anomalies over eastern Europe was associated with summer heat waves over western Europe and, especially, over the IP (Della-Marta et al., 2007). Anomalous atmospheric circulation also has been related to the European 2003 summer heat wave (Beniston and Diaz, 2004; Garcia-Herrera et al., 2010). In the IP, increases in warm days were linked to an increase in geopotential height at 500 hPa over the North Atlantic and to a decrease in the Scandinavian teleconnexion index which is associated with positive height anomalies over Scandinavia and weaker centers of opposite sign over Western Europe and Eastern Russia (Rodríguez-Puebla et al., 2010). Strong anomalies at different levels, from the surface and lower troposphere up to the mid troposphere, have been also related with very warm days over north-eastern Spain (El Kenawy et al., 2012a).

In addition to specific atmospheric conditions, warm extreme temperatures have been linked to sea surface temperatures (SST). The ocean can absorb and dissipate heat, influencing the climate due to a constant exchange of heat, momentum and water between the ocean and the atmosphere. Previous studies reported a strong relationship between hot temperatures and SST, as for example, anomalously warm SST in spring and winter have been associated with the occurrence of western European summer heat waves (Della-Marta et al., 2007). Also, extreme temperatures over north-western Europe, the Euro-Mediterranean region and Eurasia have been associated with well-defined anomalous patterns in both atmospheric circulation and SSTs (Carril et al., 2008). In particular, during the 2003 summer heat wave, the north-western part of the

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Mediterranean Sea exhibited strong anomalies of SST and, also, over the North Sea and the surrounding parts of the North Atlantic large SST anomalies were observed (Feudale and Shukla, 2011).

At a local scale, it is also known SM plays an important role in intensifying extreme events. When the soil is moist, most of the incident solar radiation goes to evaporating water rather than heating the air. In contrast, if there is a high incoming radiation and high vapour pressure deficit, the SM deficit starts due to the increase of evapotranspiration and there is a higher relative heating of the air form sensible heat flux. Thus, extremely high air temperatures are more likely to occur during the days when the SM is low (Brabson et al., 2005).

A deficit of SM has been linked with summer hot extremes in Europe (Jaeger and Seneviratne, 2011) and, particularly, in the south-eastern Europe (Hirschi et al., 2010). Persistence of European heat waves were also related to SM anomalies according to the results reported by Brabson et al. (2005) for eastern England, where not only the occurrence of extreme temperatures resulted from periods with low moisture but also long spells of extreme temperatures were related to moisture deficit. In southern Europe, drier soils favour warm and dry northward flow, increasing the probability of strong heat wave episodes in the middle or the end of the summer (Zampieri et al., 2009).

As it has been briefly discussed above, changes in daily extreme temperatures have been identified in many studies conducted at local, regional or global scales. Taking into account the results reported for the observed and expected increases along with the negative effects associated with these severe events, there is a need to better understand them. Therefore, to know the observed temporal changes and trends of the extreme distribution would be interesting, but also to understand the dynamical factors and physical processes responsible for the occurrence and intensity of these events would be essential. In the IPCC Special Report (IPCC, 2012) which focuses on the relationship between climate change and extreme weather and climate events, the impacts of such events, and the strategies to manage the associated risks, the importance of understanding climate extreme characteristics in order to improve and advance toward better climate change adaptation strategies is highlighted.

The study of extreme temperatures is a challenging task because of the rare occurrence of extreme events. Climate extremes can be evaluated by using both extreme indicators (indices) and Extreme Value Theory (EVT). The extreme indicators are based

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

on high order statistics on the tails of the probability distribution, which describe particular characteristics of extremes, including frequency, amplitude and persistence of events that occurs several times in a year, such as daily Tx beyond the 90th percentile. These events are often referred in literature as “moderate extremes” (Brunet et al., 2007a; Zhang et al., 2011). Besides, the EVT is the branch of probability and statistics dedicated to characterize the behaviour of extreme observations and complements the descriptive extreme indices in order to evaluate the statistical characteristics of rare events that lies farer in the tails of the probability distribution, which only occur infrequently and are not expected to be observed each year (Zwiers et al., 2013). In this case, the extremes in study are very high quantiles, such as the 95th or higher percentiles of daily Tx. Extreme value methods are powerful statistical methods for studying extremes and provide a class of models to enable extrapolation from observed data and to quantify uncertainties of such extrapolations. In addition, it is possible to account for non-stationary conditions in extreme value analysis.

Over Spain, few studies exploring large-scale forcing factors, such as SLP, SST or SM, associated with extreme temperatures have been found in the peer-reviewed literature. In this regard, progress has been recently achieved at the regional scale by El Kenawy et al. (2011, 2012a, 2012b) for the north-eastern Spain. But to my knowledge, there is no study yet on SM influence on extreme temperatures in mainland Spain.

This thesis is aimed at giving new insights on large-scale factors influencing summer (JJA) very hot days (VHD) and nights (VHN) over mainland Spain by using an adequate and robust methodology namely Point Process (PP) approach based on EVT. This methodology has been comprehensively described and discussed in the handbook of Coles, (2001) and will be briefly introduced in chapter 2 of this thesis. The PP approach also provides an interpretation of extreme value behaviour that unifies all the asymptotic extreme value models and leads directly to a likelihood that enables a more natural formulation of non-stationarity in threshold excesses compared with results reached from the generalized Pareto model (Coles, 2001).

The analysis of the present work is focused on the uppermost percentiles (>95th) of JJA daily adjusted Tx and Tn series. Following Coles (2001), Brown et al. (2008) and Sillmann et al. (2011), large-scale variables have been included as covariates in the statistical model of extreme values. Specifically, it has been investigated whether SLP, SST and SM anomalies can influence the frequency and the intensity of VHD and VHN

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

over mainland Spain. This also encompasses the investigation of possible anomaly patterns related to extreme events for the first time in Spain using this methodology. Working with the PP model provides an opportunity to incorporate the effects of large-scale variables into the extremal analysis. In addition, non-stationary PP approach has been used as an indirect method of fitting data to the GEV distribution to explore changes and trends in extreme temperatures, since changes in extremes can be linked to changes in daily Tx and Tn extreme distribution. Figure 1.1 shows a schematic example of how changes in the location of GEV distribution can affect extreme temperatures.

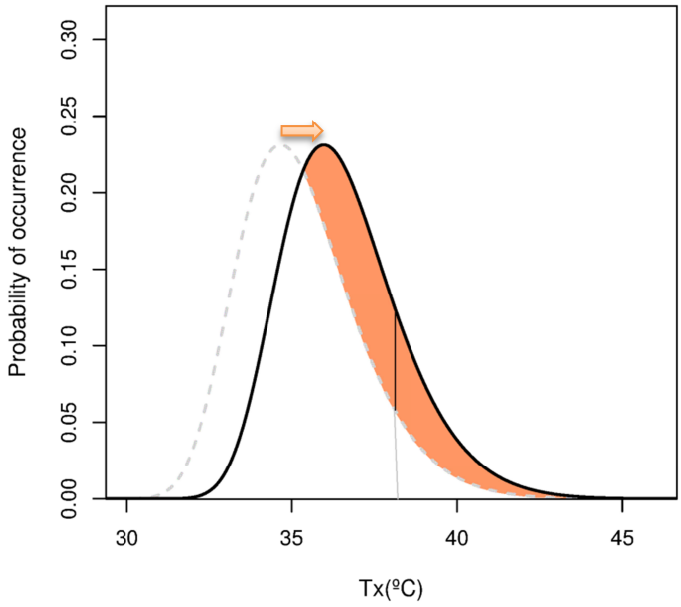


Figure 1.1. Schematic diagram showing a change in the location of the GEV distribution.

Modelling threshold excesses is well established in the literature. The original idea was developed by hydrologists to statistically modelling floods (Todorovic and Zelenhasic, 1970). Davison and Smith (1990) adopted a Generalized Pareto Distribution (GPD) with a Poisson Process into a single two dimensional process to model the frequency and intensity of floods and financial events respectively. Similar methodology has been used to analyse droughts (Abaurrea and Cebrián, 2002), modelling and forecasting extreme temperature events (Abaurrea et al., 2007; Dalelane and Deutschländer, 2013), analyse changes in extreme daily temperatures (Brown et al., 2008) and to investigate hot spells characteristics (Katsoulis and Hatzianastassiou, 2005). The methodology of PP approach have been applied, for example, to the

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

statistical modelling of hot spells and heat waves in Arizona and Colorado (USA) and France (Furrer, 2010); however, in Spain this robust methodology is rarely used to study extreme temperature events, hence, the importance of the present work.

This thesis represents a contribution to better understand large-scale circulation patterns, anomalous SST and SM deficits influencing summer extreme temperatures in mainland Spain. Furthermore, improving current knowledge on extreme temperature statistical characteristics and trends would give us a better perspective on their expected changes in the context of climate change and their potential impacts. Such information provides a better understanding of the exceptional occurrence of these events and could be very useful to undertake more reliably future projections.

The goals of this thesis are summarized through the following research questions, which are intended to be answered in detail in the next chapters.

Research questions:

- Which are the coherent large-scale anomaly patterns of SLP, SST and SM associated with summer very hot days and very hot nights at specific stations in Spain?
- Which are the statistically significant relationships between the frequency and intensity of extreme temperature events and SLP, SST and SM anomalies, identified by applying the PP approach?
- To what extent the characteristics of the observed Tx and Tn GEV distributions have changed since 1940 onwards?
- Which are the expected values of Tx and Tn extreme temperatures likely to occur in the future in mainland Spain locations?
- Which are the stations where are expected the largest increasing temperature rates in the 20-year return period analysis?

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

This thesis is organized as follows: in Section 2 the statistical theory of extreme values analysis is presented, the data and methodology used is described in Section 3, the findings are presented, put in the context of previous studies and discussed in Section 4 and, in Section 5, the most important findings are summarized and an outlook for further work is provided.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

2 Statistical theory of extreme values analysis

This section is intended at presenting an overview of the various statistical theories assessing extreme values based on observational data and to justify why the Point Process (PP) approach has been chosen to deal with the analysis of this thesis. In the literature there are a wide variety of excellent handbooks that provide mathematical descriptions of the extreme value theory. A useful handbook providing a comprehensive mathematical background of statistical modelling and focused on practical application and data analysis is written by Coles (2001). The book of Beirlant (2004) presents new probability models, inference and data analysis techniques oriented towards practical cases of extreme values. The handbook of Haan and Ferreira (2006) presents an excellent introduction to extreme value theory with complete theoretical treatments, while the handbook of Reiss and Thomas (2007) constitutes a compendium of extreme value analysis in the field of applied statistics. In this thesis the handbook of Coles (2001) has been used as a primer reference, due to the comprehensive theoretical framework of extreme value analysis, including contemporary technique based on PP model which has been chosen as the main approach in the analysis of summer extreme temperature events over mainland Spain. Also, the practical examples for solving the real problems presented in this handbook have been very useful for guiding the applications presented in this thesis.

Standard models of extreme values are derived from asymptotic arguments using limit laws as approximation to the distribution function first identified by Fisher and Tippett (1928), in which each model is characterized by its distribution. In order to better understand the theory and the modelling of extreme values, first, the basic and the general concepts for a simple case are described:

If $X_1,...,X_n$ is a sequence of independent and identically distributed random variables, then the maximum of the sequence over a "n-observation" period is:

$$M_n = \max(X_1,...,X_n)$$

The statistical behaviour X_i is unknown and then the corresponding behaviour of M_n cannot be exactly calculated. However, under suitable assumptions, it may be possible to approximate the true distribution by a simpler distribution returned by a limiting argument. The approximate behaviour of M_n for large values of *n* follows from detailed

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

limit arguments by letting $n \to \infty$, leading to a family of models that can be calibrated by the observed values of M_n.

There are different approaches to estimate the unknown parameters of the model, namely: *Probability Weighted Moments* (PWM), *L-moments* or *likelihood-based techniques*, but the last ones are unique in their adaptability to model-change. Although the estimated equations change if a model is modified, the underlying methodology is essentially unchanged.

In this thesis, the *maximum likelihood estimation* (MLE) is adopted, because its application is straightforward in the presence of covariates. The principle of MLE is to adopt the model with the largest likelihood, since for all the models under consideration this is the one constraint that assigns the highest probability to the observed data (See more details in the description of MLE in the subsection 2.5.1).

Once the parameters have been estimated, it is important to quantify the uncertainty due to sampling variability, especially in extreme value modelling, where quite small model changes can be largely magnified on extrapolation.

To reach robust conclusions about statistical features of the population and also to make good extrapolation, an accuracy fitted model is required, which implies it is necessary to assess the goodness-of-fit. Normally, the assessment of the accuracy of a model in terms of its agreement is done with the data used to estimate it, due to the lack of additional data sources against which the model can be judged. To explore the suitability of the models, Coles (2001: 36) suggests analysing the diagnostic plots, where comparison between the estimated distribution function and the empirical distribution function is examined.

In climate processes, seasonal effects or trends are usually apparent, due to different climate patterns or long-term climate changes. This information can be included in the model, which will have a non-stationary distribution (i.e., changing systematically throughout the time).

More complex models, which are non-stationary and use more data, can reduce uncertainties of the statistical modelling. The introduction of the non-stationarity into the model is carried out considering some parameters of the theoretical distribution function depending on the covariate. The covariates could incorporate trends, cycles or physical variables (e.g. measures of large-scale atmosphere-ocean circulation patterns) (Katz et al., 2002). Therefore, the important issue is to select the appropriate model,

which should be the simplest model possible that explains as much of the variation in the data as possible. To deal with this, it is necessary to test the improvements to the model gained by introducing the covariates. The methodology used in the present study is described in details in section 2.5.5.

2.1 Classical Extreme Value Theory (EVT)

The Extreme Value Theory (EVT) is a statistical discipline dedicated to characterising and quantifying the stochastic behaviour of extreme observations. It was formulated around the mid-20th century. This theory is mainly focused on describing the behaviour of the upper or lower tails of the statistical distribution data. Emil Gumbel was a pioneer in the application of the statistics of extremes to modelling extremal behaviour of observed physical processes, particularly in the fields of climatology, hydrology and oceanography (Gumbel, 1958). Initially, the theory was mainly applied in hydrology, due to the need to assess the return periods of floods, although later it was used by other disciplines, such as climatology, economics and engineering.

The asymptotic model characterization, which represents the starting point of EVT, is briefly described in here. The model works with groups of data into blocks of equal length (block maxima or minima) and it fits the data to the maximums of each block, then the model focuses on the statistical behaviour of:

$$M_n = \max(X_1,...,X_n)$$

where $X_1,...,X_n,$, is a sequence of independent random variables having a common distribution function F. X_i stands for the values of a process measured on a regular time-scale and M_n represents the maximum of the process over n time units of observations, that means that if n is the number of observations in a year, then M_n corresponds to the annual maximum.

An example of Block Maxima with the maximum observed daily maximum temperature (Tx) over each year (annual maximum) is shown in Figure 2.1. The plot represents the temporal evolution of the annual absolute values of daily Tx at Albacete station during 1940-2010.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

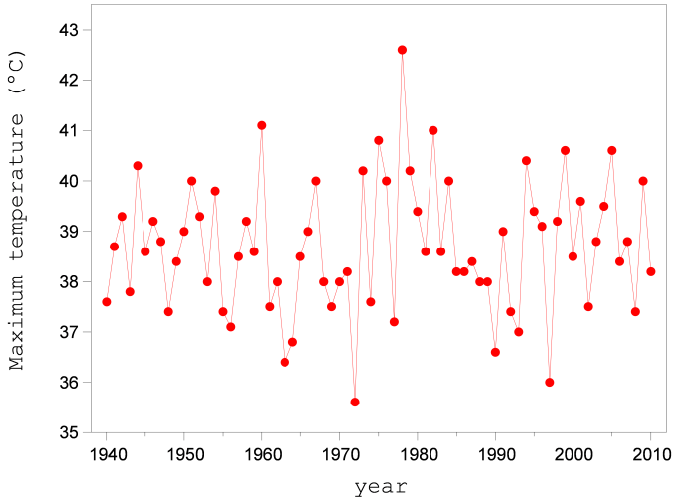


Figure 2.1. Annual maximum of daily maximum temperature (Tx) series at Albacete station for the 1940-2010 period.

In theory, the distribution of M_n can be derived exactly for all values of n:

$$\Pr\{M_n \le z\} = \Pr\{X_1 \le z, ..., X_n \le z\} = \Pr\{X_1 \le z\} \times ... \times \Pr\{X_n \le z\} = \{F(z)\}^n$$

but unfortunately the distribution function F is unknown as it started. It is necessary to look for approximate families of models for F^n, which can only be estimated on the basis of extreme data. The arguments are essentially an extreme value analogue of the central limit theory.

Procedure starts by looking at the behaviour of F^n as $n \to \infty$. But this alone is not enough: for any $z < z_+$, where z_+ is the upper end-point of F, then z_+ is the smallest value of z such as $F(z) = 1$, $F^n(z) \to 0$ as $n \to \infty$. So, the distribution of M_n degenerates of a point mass on z_+. This difficulty is avoided by allowing a linear renormalization of the variable M_n:

$$M_n^* = \frac{M_n - b_n}{a_n},$$

for sequence of constants $\{a_n > 0\}$ and $\{b_n\}$. Appropriate choices of the $\{a_n\}$ and $\{b_n\}$ stabilize the location and scale of M_n^* as n increases, avoiding the difficulties that arise with the variable M_n.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

The entire range of possible limit distributions for M_n^* is given by the extremal types theorem 3.1 (Coles, 2001:46). Theorem 3.1 denotes if there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such as

$$\Pr\{(M_n - b_n)/a_n \le z\} \to G(z) \qquad \text{as } n \to \infty$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$I: G(z) = \exp\left\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\right\}, \quad -\infty < z < \infty \quad \text{Gumbel}$$

$$II: G(z) = \begin{cases} 0, & z \le b \\ \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z > b \end{cases} \quad \text{Fréchet}$$

$$III: G(z) = \begin{cases} \exp\left\{-\left(\frac{z-b}{a}\right)^{-\alpha}\right\}, & z < b \\ 1, & z \ge b \end{cases} \quad \text{Weibul}$$

where α>0 and a>0.

The rescaled sample maxima $(M_n - b_n)/a_n$ converge in distribution to a variable having a distribution within one of the three families. The three classes of extremal distributions are the Gumbel, the Fréchet and the Weibull and each family has a location and scale parameter, b and a, respectively. Additionally, the Fréchet and Weibull families have a shape parameter α.

The three types of extreme value distributions are the only possible limits for the distribution of the M_n^* regardless of the distribution F for the population, and each one describes very different limiting behaviour in the tail of the distribution, which implies quite different representations of extreme value behaviour.

Note the choice of block size can be crucial because too small can lead to bias in estimation and extrapolation since the approximation by the limit model is likely to be poor and too large blocks generate too few block maxima, which leads to large estimation variance, see (Coles, 2001:54).

In the early stage of the EVT applications, it was usual to choose one of the three distribution families, and then to estimate the parameters of that distribution. But this

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

technique have disadvantages because, firstly, the choice of the most appropriate distribution is subjective and, secondly, once such a decision is made, subsequent inferences assume this choice to be correct and do not allow for uncertainty that such a selection involves, even though this uncertainty may be considerable.

A better analysis is offered by a combination of the Gumbel, Fréchet and Weibull families into a single parametric family; namely, the generalized extreme value (GEV) family of distributions, and their distribution functions are:

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

defined on the set $\{z : 1+\xi(z-\mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty, \sigma > 0$ and $-\infty < \xi < \infty$. The model has three parameters: a location parameter, μ; a scale parameter, σ; and a shape parameter, ξ.

If the random variable X_i has a GEV distribution, then the standardized variable $(X-\mu)/\sigma$ has a distribution that does not depend on either μ or σ, only on ξ.

The location parameter specifies where the distribution is ‘centred’, the scale parameter is ’spread’ and the shape governs the tail behaviour of the distribution and assumes three possible types:

i. $\xi = 0$, a light-tailed (or Gumbel) distribution;
ii. $\xi > 0$, a heavy-tailed (or Fréchet) distribution;
iii. $\xi < 0$, a bounded (or Weibull) distribution.

The type (i) distribution has an unbounded upper tail which decreases at a relatively rapid (i.e., exponential) rate, the type (ii) distribution also has an unbounded upper tail but it decreases at such a slow (i.e., power law) rate and the type (iii) distribution has a finite upper bound at $x = \mu - (\sigma/\xi)$.

The three types of the GEV family distributions with different behaviour of their tail are presented in Figure 2.2.

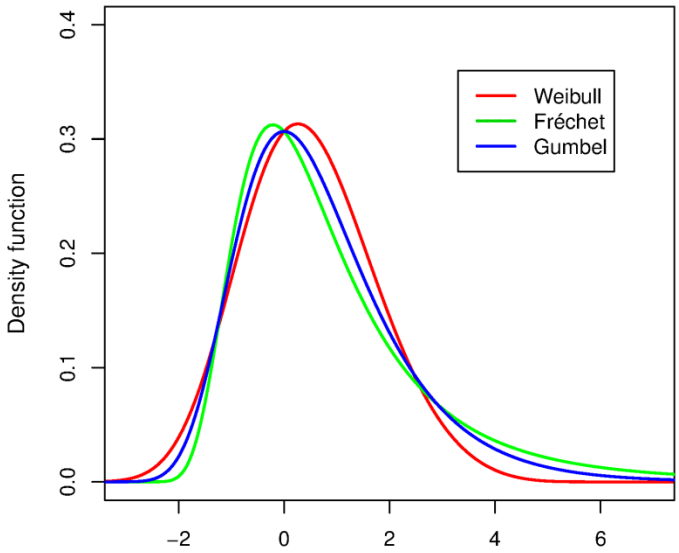


Figure 2.2. Plot of the GEV probability density function with $\mu = 0, \sigma = 1.2$ and $\xi = -0.2$ (Weibull), $\xi = 0.2$ (Fréchet), $\xi = 0$ (Gumbel).

Throughout this approach, the data themselves determine the most appropriate type of tail behaviour. But this model has disadvantages because it does not make use of all of the information available on the upper tail of the distribution. For instance, if the highest and second highest summer daily Tx over the historical record occurs during the same year, the second highest value would be ignored in the Block Maxima approach.

To deal with this problem, alternative approaches were developed. In the next sections, *the r-Largest Order Statistic model*, *the Peaks Over Threshold model*, *the Poisson-GPD model for excesses* and *the PP approach* are described with a brief mathematical foundation.

2.2 The r Largest Order Statistic model

In order to solve the problem of data scarcity for the model estimation, there are other modelling methodologies better than Block Maxima.

In this section one model based on the behaviour of the r largest order statistics within a block for small values of r is described. r stands for the number of peaks maximums at each block. The model formulation is as follows:

Suppose $X_1,...,X_n$ is a sequence of independent and identically distributed random variables, the limiting distribution as $n \to \infty$ of M_n, appropriately rescaled is GEV, as

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

was exposed in the previous section (section 2.1). This can be extended to other extreme order statistics, by defining

$$M_n^{(k)} = k^{\text{th}} \text{ largest of } \{X_1,...,X_n\},$$

and identifying the limiting behaviour of this variable, for fixed k, as $n \to \infty$. Then if the k^{th} order statistics in a block is normalized in exactly the same way as the maximum, the limiting distribution is of the form:

$$G_k(z) = \exp\{-\tau(z)\}\sum_{s=0}^{k-1}\frac{\tau(z)^s}{s!}$$

with

$$\tau(z) = \left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}$$

where (μ,σ,ξ) are the parameters of the limiting GEV distribution of the block maximum. Then the approximate distribution of $M_n^{(k)}$ is within the family $G_k(z)$.

Usually, each of the largest r order statistics is within each of several blocks, for some r, then $M_n^{(r)} = (M_n^{(1)},...,M_n^{(r)})$ the joint density function of the limit distribution is:

$$f(z^{(1)},...,z^{(r)}) = \exp\left\{-\left[1+\xi\left(\frac{z^{(r)}-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}\times\prod_{k=1}^{r}\sigma^{-1}\exp\left[-\frac{z^{(k)}-\mu}{\sigma}\right]^{-\frac{1}{\xi}-1}$$

where:

$-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$;

$z^{(r)} \le z^{(r-1)} \le ... \le z^{(1)}$ and $z^{(k)} : 1+\xi\left(z^{(k)}-\mu\right)/\sigma > 0$ for $k = 1,...,r$.

Like in Block Maxima approach there is the problem of block size amounts to trade-off between bias and variance, but this it is usually resolved by making a pragmatic choice, such as a block length of one year. Also, the number of order statistics used in each block comprises a bias-variance trade-off: small values of r generate few data leading to high variance and large values of r are likely to violate the asymptotic support for the model, leading to model bias. It is common to select the r as large as possible, subject to adequate model diagnostics.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

2.3 Peaks Over Threshold (POT) approach

Peaks Over Threshold model (POT) make use of more available information on the upper tail of the distribution than GEV models based on Block Maxima approach. In addition, this method is better than the r largest order statistics model, because it models more extreme events above a selected threshold than the r largest order statistics within a block.

The events exceeding some high threshold are considerate as extreme events, and the excesses over the threshold have an approximate Generalized Pareto Distribution (GPD) that governs the intensity of the events.

Figure 2.3 shows a time series of summer daily Tx at Valencia station, recorded over the period 1940-2010 with a threshold of 33ºC added.

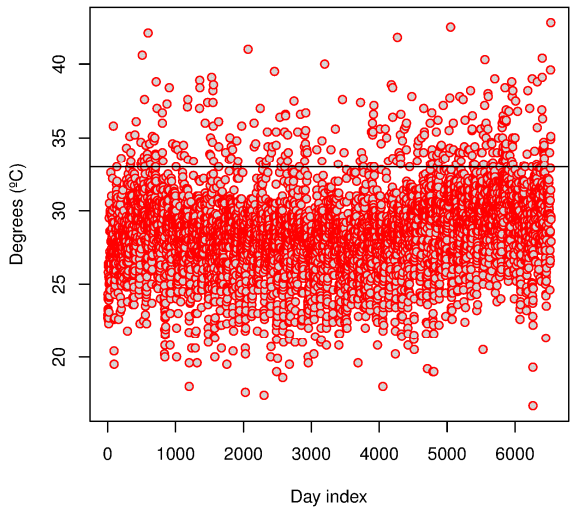


Figure 2.3. Summer daily Tx at Valencia station (red dots) from 1940 to 2010, with a selected threshold (horizontal black line).

The conditional probability can explain the stochastic behaviour of extreme events defining the distribution of threshold excesses u.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, having marginal distribution function F, denoting an arbitrary sequence by X then:

$$\Pr\{X > u + y \mid X > u\} = \frac{1 - F(u+y)}{1 - F(u)}, \qquad y > 0$$

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

The distribution of threshold excesses represents the probability that the data exceed a threshold. But in real applications, the distribution of threshold excesses is unknown, since the parent distribution F is also unknown as discussed above.

Asymptotic Model Characterization of GPD is basically described in the following theorem:

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with common distribution function F, and let $M_n = \max(X_1, \ldots, X_n)$.

Denote an arbitrary term in the X_i sequence by X, and for large n,

$$\Pr\{M_n \leq z\} \approx G(z)$$

where

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for some $\mu, \sigma > 0$ and ξ. Then for large enough u, the distribution function of $(X-u)$, conditional on $X > u$, is approximately

$$H(y) = 1-\left(1+\xi\frac{y}{\tilde{\sigma}}\right)^{-1/\xi}$$

defined on $\{y : y > 0$ and $\left(1+\xi\frac{y}{\tilde{\sigma}}\right) > 0$, where $\tilde{\sigma} = \sigma + \xi(u-\mu)$.

The interpretation of the theorem is if block maxima has approximating distribution G (i.e. GEV distribution), then threshold excesses have a corresponding approximate distribution within the GPD family with the parameters uniquely determined by those of the associated GEV distribution of block maxima. In particular, the parameter ξ in $H(y)$ is equal to that of the corresponding GEV distribution and also is dominant in determining the qualitative behaviour of the generalized Pareto distribution.

Therefore there are three possible types depending on the value of ξ:

i. $\xi < 0$, a bounded (or beta) distribution;

ii. $\xi > 0$, a heavy tailed (or Pareto) distribution;

iii. $\xi = 0$, a light-tailed (or exponential) distribution.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

The type (i) distribution of excesses has an upper bound of $u-\tilde{\sigma}/\xi$, the type (ii) distribution has no upper limit and the type (iii) distribution is also unbounded which should again be interpreted by taking the limit $\xi \to 0$ in $H(y)$, leading to

$$H(y)=1-\exp\left(-\frac{y}{\tilde{\sigma}}\right), \quad y>0$$

Corresponding to an exponential distribution with parameter $1/\tilde{\sigma}$.

Figure 2.4 presents the three types of GPD family distribution for different shape parameter.

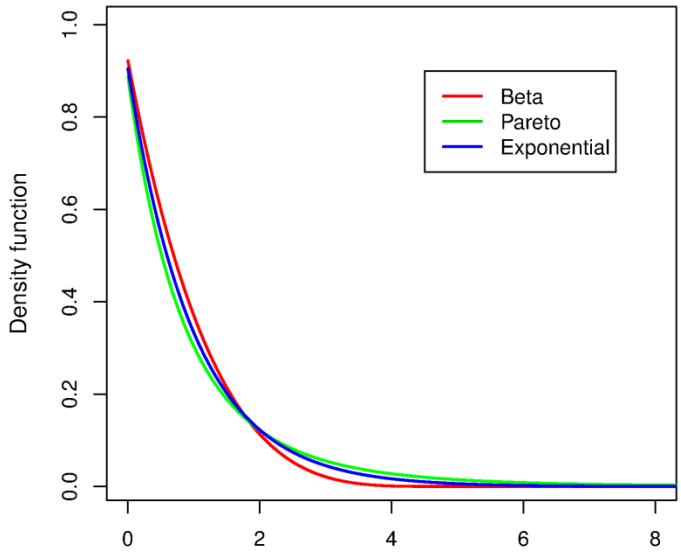


Figure 2.4. GPD density function with $\tilde{\sigma}=1$, $\xi=-0.2$ (Beta), $\xi=0.2$ (Pareto) and $\xi=0$ (exponential).

2.4 Poisson-GPD model for excesses

The Poisson-GPD model for excesses is closely related to the Peaks Over Threshold (POT) model originated in hydrology for carrying out a statistical modelling of floods (Todorovic and Zelenhasic, 1970)

This model is a joint distribution, the GPD, for the excesses values y, and a Poisson distribution for the number of excesses over a level u in any given year. Therefore, it can be estimated not only the intensity of the excesses, but also the frequency of these events.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

The model consists on:

1. The number, N, of excesses of the level u in any individual year has a Poisson distribution with mean λ and governs the occurrence of an extreme event in the form of exceeding a high threshold.
2. Conditionally on $N \geq 1$, the excesses values $Y_1,...,Y_N$ are independent and identically distributed from the GPD.

The probability that the annual maximum of the Poisson-GPD process is lower than a value x, with $x > u$, is:

$$F(x) = \exp\left\{-\lambda\left[1+\xi\left(\frac{x-u}{\tilde{\sigma}}\right)\right]^{-1/\xi}\right\}$$

There is a relationship between the GEV and GPD parameters by:

$$\tilde{\sigma} = \sigma + \xi(u-\mu), \qquad \lambda = \left(1+\xi\frac{u-\mu}{\sigma}\right)^{-1/\xi}$$

If these parameters are substituted in F(x), then the distribution function is reduced to the GEV form. Thus, the GEV and GPD models are entirely consistent with each other above the threshold u.

2.5 Point Process approach

In this thesis, the focus is placed on the PP approach, which will be used in this analysis. Hence this section describes not only the Asymptotic Model, but also the statistical modelling.

The PP approach provides and interpretation of extreme value behaviour that unifies all the asymptotic models introduced so far (Coles, 2001), namely, the Block Maxima model, the *r* Largest Order Statistic model and the POT model, all of them being special cases of the PP approach and likewise the POT models, this has several advantages over the Block Maxima and *r* Largest Order Statistic model, because it uses considerably more data on extremes, returning more reliable results. All inferences made using the PP methodology could equally be gained using the Poisson-GPD model; however, there are good reasons to adopt this approach:

- the model can be formulated in terms of the GEV parameters, which are invariant to the choice of the threshold and leads directly to a likelihood that

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

enables a more natural formulation of non-stationarity than the given by the GPD model

- it includes the threshold excess rate in the inference to account for the frequency of occurrence, which is modelled separately in a Poisson-GPD model
- avoid the need to combine uncertainty from two components as in Poisson-GPD model

The PP model combines the two components of the Poisson-GPD model: the modelling of the occurrence of excesses of a high threshold and their corresponding excesses into a single two-dimensional process. The asymptotic theory of threshold exceedances shows that under suitable normalization, this process behaves like a nonhomogeneous Poisson process with non-constant (or non-homogenous) rate parameter (Smith, 2003).

A point process on a set *A* is a stochastic rule for the occurrence and position of point events. *A* representing, for example, a period of time to modelling the occurrence of extreme temperature events.

The asymptotic model characterization is summarized as follows:

Let $X_1, X_2, \ldots$ be a series of independent and identically distributed random variables, with common distribution function *F* and well behaved in an extreme value sense. That is, with $M_n = \max(X_1, \ldots, X_n)$ that there are sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\{(M_n - b_n)/a_n \leq z\} \to G(z),$$

$$G(z) = \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

for some parameters μ, σ and ξ.

The sequence of point processes N_n are defined by:

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \ldots, n\}$$

The scaling in the first ordinate ensures that the time axis is always mapped to $(0,1)$; the scaling in the second ordinate stabilizes the behaviour of extremes as $n \to \infty$

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

On regions of the form $(0,1)\times[u,\infty)$, $N_n \xrightarrow{d} N$ as $n \to \infty$, where N is a non-homogeneous Poisson process, with intensity measure on $A=[t_1,t_2]\times(u,\infty)$, with $[t_1,t_2]\subset[0,1]$, given by:

$$\Lambda(A)=(t_2-t_1)\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

and for n_y years of observation

$$\Lambda(A)=n_y(t_2-t_1)\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}$$

and extended to non-stationary processes where the parameters μ, σ and ξ are time dependent as $\mu(t)$, $\sigma(t)$ and $\xi(t)$:

$$\Lambda(A)=n_y(t_2-t_1)\left[1+\xi(t)\left(\frac{u-\mu(t)}{\sigma(t)}\right)\right]^{-1/\xi(t)}$$

The parameters (μ,σ,ξ) associated with the PP model are the parameters of the corresponding annual maximum GEV distribution. Therefore, the applied model is reduced to the estimation of the three unknown parameters (μ,σ,ξ).

Likewise in the Poisson- GPD model, the relationship between the GEV and GPD parameters is of this form:

$$\tilde{\sigma}=\sigma+\xi(u-\mu), \qquad \lambda=\left(1+\xi\frac{u-\mu}{\sigma}\right)^{-1/\xi}$$

This approach can be taken as an indirect method of fitting data to the GEV distribution because it uses more information about the upper tail of the distribution than does Block Maxima approach.

2.5.1 Parameter estimation

To estimate the unknown parameters of the distribution, there are several methods. The most used are the Probability Weighted Moments (PWM) and the Maximum Likelihood Estimation (MLE). From both, the latter is the one that allows for a straightforward inclusion of covariates. Therefore, the statistical modelling approach adopted in this thesis is based on the MLE.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

The principle of the MLE is to adopt the model with greatest likelihood because this is the one that assigns highest probability to the observed data. The probability of the observed data as a function of θ is called the likelihood function, where θ is the vector containing all parameters which characterize the distribution. Values of θ that have high likelihood correspond to models that give high probability to the observed data. Therefore, the maximum likelihood estimator is the value of θ that maximizes the appropriate likelihood function. The model works as follows:

given an observed series of independent realizations of a random variable $x_1,...,x_n$ having probability density function $f(x;\theta)$, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

To facilitate calculations, it is convenient to take into account logarithms and work with the log-likelihood function

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$$

the log-likelihood takes its maximum at the same point as the likelihood function, because the logarithm function is monotonic, so the maximum likelihood estimator $\hat{\theta}$ also maximizes the corresponding log-likelihood function. The maximum likelihood estimate is found by maximizing this expression with respect θ. And it can be done by solving the equation:

$$\frac{\partial \ell(\theta)}{\partial \theta_i} = 0 \quad \text{for } i \text{ parameters}$$

Standard errors and confidence intervals of the model parameters can be assessed based on asymptotic properties of MLE.

It is worth noting that MLE has many optimal properties, among others:

- consistency (true parameter value that generated the data recovered asymptotically)
- sufficiency (complete information about the parameter of interest contained in its MLE estimator)
- efficiency (lowest-possible variance of parameter estimates achieved asymptotically)

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

- parameterization invariance (same MLE solution is achieved independently of the parameterization used).

One of the main advantages of the MLE method is the procedure is commonly and broadly applicable to many types of distributions, particularly for PP approach. In this case $\theta = (\mu, \sigma, \xi)$ and an approximate likelihood can be derived assuming the limiting Poisson process is an acceptable approximation to the process N_n on A and maximizing this likelihood leads to estimates of the parameters (μ, σ, ξ) of the limiting intensity function.

The likelihood function is:

$$L_A(\mu, \sigma, \xi; x_1, \ldots, x_n) = \exp\{-\Lambda(A)\}\prod_{i=1}^{N(A)} \lambda(t_i, x_i)$$

$$\propto \exp\left\{-n_y\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}\prod_{i=1}^{N(A)} \sigma^{-1}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}$$

where $N(A)$ is the number of observed points in the region A, $\{(t_1, x_1), \ldots, (t_{N(A)}, x_{N(A)})\}$.

The estimates derived from the PP likelihood are based on all those data greater than a specified threshold; therefore, the estimates will likely be more accurate than the estimates based on a direct fit of the GEV.

2.5.2 Incorporating non-stationarity into the model

Most of the climatological series present non-stationarity; hence, the model needs to be adapted to enable for non-stationary effects. This is direct and straightforward with the MLE, since it allows a simple incorporation of non-stationarity by modifying the likelihood function to include temporal or covariate effects in the parameters μ, σ or ξ. Time dependence in the parameters can be incorporated by allowing $\mu(t), \sigma(t)$ and $\xi(t)$ and dependence on a covariate z by $\mu(z), \sigma(z)$ and $\xi(z)$.

2.5.3 Uncertainty and confidence intervals

In any extreme model, it must be always taken into account the estimations uncertainty budget. Especially for analysis of extreme values, since it is likely to have more sources of uncertainty than most of other statistical analyses.

The main problem in the analysis of extremes is, by definition, the scarcity of data, along with the lack of continuity of good resolution and long enough time series. Then this limited sample of few and unusual extreme values should be modelled. This causes to the extreme models described to have a remaining uncertainty. Although each of the results is an asymptotic limit law achieved as the sample size increases to infinity, under regularity conditions, the results are approximations whose accuracy improves as n increases but they are not exact.

Modelling the extreme values means to describe a statistical model that better fits a set of observations. The goodness of fit is a measure that typically summarizes the discrepancy between observed values and the values expected under a specific model. The maximum-likelihood estimation is a method of estimating the parameters of a statistical model. When applied to a data set and given a statistical model, maximum-likelihood estimation provides estimates for the model's parameters.

Next, two methodologies to analyse the goodness-of-fit for each estimator based on the standard errors associated with the maximum likelihood estimator are described.

- Confidence intervals based on Fisher method

For $x=(x_1,...,x_n)$ independent realizations from a distribution within a parametric family, with the maximum value of the log-likelihood function, $\ell(x;\theta)$, and specific parameter estimates, $\hat{\theta}$, of the d-dimensional model parameter θ, such that the probability of the observed sample of extreme values which follow the theoretical model is maximum.

The parameter θ can be a scalar or can be a vector of parameters; for example, $\theta=(\mu,\sigma,\xi)$ in GEV family.

The sensitivity of the d-dimensional model with the estimated parameters $\hat{\theta}$ against the d-dimensional theoretical model with parameters θ can be quantified as follows:

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

$$I_{\mathrm{O}} = \begin{pmatrix} -\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_1^2} & \cdots & \cdots & -\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_1 \partial \theta_d} \\ \vdots & \ddots & -\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_i \partial \theta_j} & \vdots \\ \vdots & -\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_j \partial \theta_i} & \ddots & \vdots \\ -\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_d \partial \theta_1} & \cdots & \cdots & -\frac{\partial^2 \ell(\hat{\theta})}{\partial \theta_d^2} \end{pmatrix}$$

The I_{O} matrix is so-called observed information matrix, I_{O} and measures the observed curvature of the log-likelihood surface. An approximate $(1-\alpha)$ confidence interval for a single parameter θ_i is:

$$ci(\theta_i) = \hat{\theta}_i \pm z_{\alpha/2}\sqrt{\tilde{\psi}_{i,i}}$$

where $\tilde{\psi}_{i,i}$ are the terms of the diagonal of the inverse of I_{O} and $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution.

In addition to the confidence intervals of the estimated parameters, to calculate the confidence intervals of the probability of occurrence of such extreme event (high quantiles) is also required.

These probabilities are functions of the parameters of the distribution $g(\theta)$. Note that the maximum likelihood estimates are invariants, i.e., once the maximum likelihood estimate of θ has been calculated, the maximum likelihood estimate of any function of θ, $g(\theta)$, is calculated by simple substitution, then $g(\hat{\theta})$ is achieved.

The estimated variance of the function $g(\hat{\theta})$ can be calculated using the Delta method. Then, considering the $g(\hat{\theta})$ function and $\hat{\theta}$ as a vector composed by d parameters, the variance function is:

$$Var\left[g(\hat{\theta})\right] = \left(\frac{\partial g(\hat{\theta})}{\partial \theta_i}\right)^T \cdot \tilde{\psi}_{ij}(\hat{\theta}) \cdot \left(\frac{\partial g(\hat{\theta})}{\partial \theta_i}\right)$$

where $\tilde{\psi}_{ij}(\hat{\theta})$ is the $d \times d$ variance-covariance matrix. And the standard errors of the $g(\theta)$ function can be returned by:

$$se\left[g(\theta)\right] = \left(\sum_{i=1}^{d}\sum_{i=1}^{d} \frac{\partial g(\hat{\theta})}{\partial \theta_i}\frac{\partial g(\hat{\theta})}{\partial \theta_j}\tilde{\psi}_{i,j}^{-1}\right)$$

The Delta method use the approximate normality of the $g(\hat{\theta})$ to get the confidence intervals, like in the case of individual components of $\hat{\theta}$.

- Confidence intervals based on Profile Likelihood method

An alternative, and usually the best for finding accurate confidence intervals, is the method based on profile likelihood.

The log-likelihood for θ is $\ell(\theta_i;\theta_{-i})$, where θ_{-i} are all components of a parameter vector θ excluding θ_i. The profile likelihood for a particular component θ_i is gained by maximizing the likelihood with respect to all other parameters of the model and is defined by:

$$\ell_p(\theta_i) = \max_{\theta_{-i}} \ell(\theta_i, \theta_{-i})$$

An approximate $(1-\alpha)$ confidence region for θ_i can be obtained using de deviance function:

$$D_p(\theta_i) = 2\left\{\ell(\hat{\theta}) - \ell_p(\theta_i)\right\}$$

where $\ell(\hat{\theta})$ is the likelihood of the original model evaluated in their estimates and $\ell_p(\theta_i)$ is the likelihood of the parameter of interest (maximized with regard to the remaining parameters).

For a large n, under suitable regularity conditions, the deviance function satisfies:

$$D_p(\theta_i) \approx \chi^2_k$$

where χ^2_k is a chi-square distribution function with k degrees of freedom equal to the number of parameters in the model less the one of interest.

Finally, for a single component θ_i, the $(1-\alpha)$ confidence interval is:

$$C_\alpha = \left\{\theta_i : D_p(\theta_i) \le c_\alpha\right\}$$

where c_α is the $(1-\alpha)$ quantile of the χ^2_1 distribution.

The profile log-likelihood for ξ and for 20-year return level in PP model for daily maximum temperature data are plotted as examples in Figure 2.5 and Figure 2.6, respectively.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

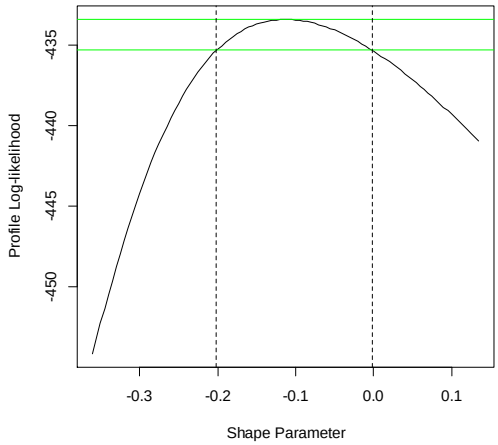


Figure 2.5. Profile likelihood for ξ parameter in PP model of daily maximum temperature data.

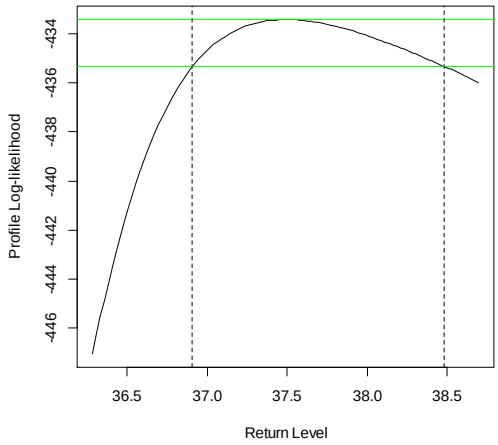


Figure 2.6. Profile likelihood for 20-year return level in PP model of daily maximum temperature data.

For combinations of parameters, such as return levels, the same technique is applied by transforming the parameters in the likelihood to reflect the desired combination.

2.5.4 Model diagnostics

To check if the resulting model, based on the MLE, is a good fit to the data, a graphical technique is commonly used (Coles, 2001:36). The procedure is as follows: suppose data $x_1,...,x_n$ are independent random variables from a common unknown distribution function F, which is estimated ($\hat{F}$), for instance, by MLE, and in order to assess the

plausibility that x_i are a random sample from $\hat{F}$ a model-free estimate of F is calculated empirically from the data. To get the empirical distribution function with uniform distribution of a sample data, the following approximation is taken:

$$\tilde{F}(x) = \frac{i}{n+1} \qquad \text{for } x_{(i)} \le x \le x_{(x+1)}$$

Then, the analysis of the goodness-of-fit is carried out by checking the empirical distribution $\tilde{F}(x)$ with the estimated $\hat{F}(x)$.

The development of this method is based on two graphical techniques for comparing two probability distributions in order to diagnose how well a specified theoretical distribution fits a set of measurements. One is the Probability-Probability plot and the other the Quantile-Quantile plot, also known as P-P plot and Q-Q plot.

Given an ordered sample of independent observations $x_{(1)} \le ... \le x_{(n)}$ from a population with estimated distribution function $\hat{F}(x)$:

- the P-P plot consists in the representation of the points:

$$\left\{\left(\hat{F}(x_{(i)}), \frac{i}{n+1}\right) : i = 1,...,n\right\}$$

  if $\hat{F}$ is a reasonable model for the population distribution, the points of the probability plot should lie close to the unit diagonal

- the Q-Q plot consists in the representation of the points:

$$\left\{\left(\hat{F}^{-1}\left(\frac{i}{n+1}\right), x_{(i)}\right) : i = 1,...,n\right\}$$

  if $\hat{F}$ is a reasonable estimate of F, then the points of the Q-Q plot should be also close to the unit diagonal.

The P-P plot and the Q-Q plot have the same information expressed on a different scale. On example of their representation can be seen in the Figure 2.7 which presents both diagnostic plots for PP model fitted to Valencia Tx data from 1940 to 2010.

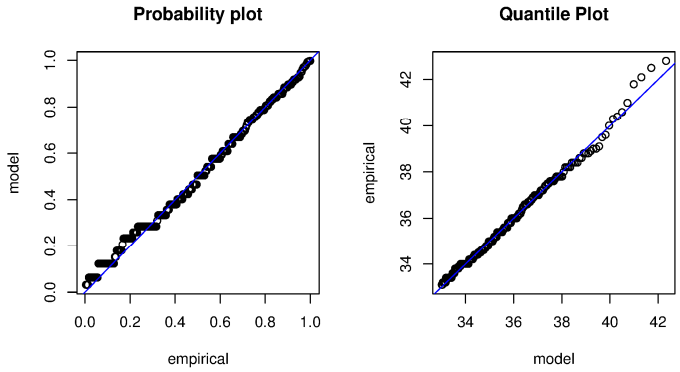


Figure 2.7. Diagnostic plots of PP model fitted to Valencia Tx series (1940-2010).

2.5.5 Model selection

The improvements to the model gained by introducing covariates can be examined using the likelihood ratio test, as this is most appropriate for comparing nested models fitted with fixed MLEs given by (Coles, 2001:35) and (Reiss and Thomas, 2007:118); whereby the difference in negative log-likelihood values between two models is tested for significance using a Chi-squared distribution.

$$D = 2\{\ell_1(M_1) - \ell_0(M_0)\} > c_\alpha$$

Where c_α is the (1-α) quantile of the χ^2_k, M_0 and M_1 are the two models, M_0 is nested in M_1, and the difference in dimensionality of the two models is *k*. This means, in fact, that M_0 is derived from M_1 by imposing *k* constraints on the parameters of M_1. If M_0 is true, then, approximately,

$$D \approx \chi^2_k$$

the chi-squared distribution with *k* degrees of freedom. Thus, hypothesis M_0 at significance level α is rejected if D is larger than the upper-α point of the χ^2_k distributions.

2.5.6 Effective return levels

The effective return levels are the return levels derived from fitting a non-stationary PP model. The idea of effective return levels relies on the fact that for each value of a covariate, you get what the return level would be if that were the fixed value.

The m-year return level z_m for a non-stationary point process model is obtained as:

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

$$1-\frac{1}{m}=\Pr\{\max(X_1,\ldots,X_n)\le z_m\}\approx\prod_{i=1}^{n}p_i$$

where

$$p_i=\begin{Bmatrix}1-n^{-1}\left[1+\xi_i(z_m-\mu_i)/\sigma_i\right]^{-1/\xi_i} & ,if\left[1+\xi_i(z_m-\mu_i)/\sigma_i\right]>0\\ 1 & otherwise\end{Bmatrix},$$

n is the number of observations in a year and (μ_i,σ_i,ξ_i) are the parameters of the point process model for observation i. Taking logarithms,

$$\sum_{i=1}^{n}\log p_i=\log(1-1/m)$$

which can be solved using standard numerical methods for non-linear equations.

A difficulty arises in the estimation of the standard errors or confidence intervals; actually this is still an active area of research in the field of extreme values theory (Gilleland and Katz, 2011).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

3 Data and methodology

3.1 The area in study

The region being analysed is mainland Spain, which lies in the south-west of Europe between 10º W and 5º E of longitude and 35º N to 45ºN of latitude (Figure 3.1). It runs from the Pyrenees in the north to the Gibraltar strait in the south and inland it has in its heart the Central Plateau, which is surrounded by mountains. Note that Spain is the highest European country after Switzerland and it is extremely diverse, ranging from the near-deserts of Almeria to the green countryside of the north and the beaches of the Mediterranean coast showing a remarkable amount of climate types and sub-types.

The analysed locations in this study are shown in Figure 3.1. In Table 3.1 is given their geographical details (station name, coordinates and elevation). They are scattered throughout the study region and covers reasonably well mainland Spain (Figure 3.1).

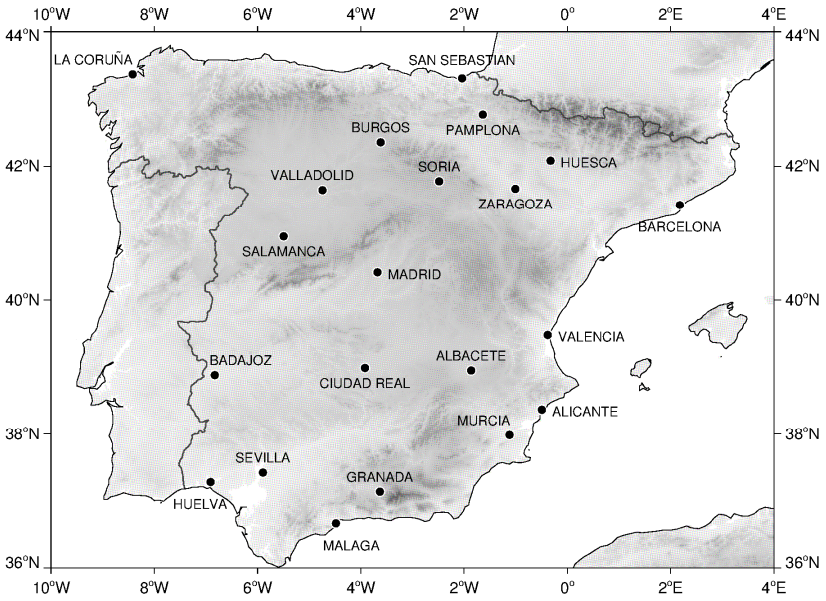


Figure 3.1. Location map of the 21 stations over mainland Spain with long daily records of temperature.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 3.1. Name, longitude, latitude and altitude of each station.

Station	Lon º	Lat º	altitude (m)
Albacete	-1.863	38.952	699
Alacant	-0.494	38.367	81
Badajoz	-6.829	38.883	185
Barcelona	2.177	41.418	420
Burgos	-3.616	42.356	881
Ciudad Real	-3.920	38.989	627
Granada	-3.631	37.136	685
Huelva	-6.910	37.280	19
Huesca	-0.326	42.083	541
La Coruña	-8.419	43.367	67
Madrid	-3.678	40.411	679
Malaga	-4.483	36.666	6
Murcia	-1.121	37.983	57
Pamplona	-1.639	42.768	452
Salamanca	-5.495	40.947	789
San Sebastian	-2.039	43.307	251
Sevilla	-5.896	37.421	31
Soria	-2.484	41.775	1083
Valencia	-0.381	39.480	11
Valladolid	-4.743	41.644	691
Zaragoza	-1.008	41.662	245

3.2 Data description

3.2.1 The Spanish Daily Adjusted Temperature Series (SDATS)

To model and analyse climatic events accurately and, especially, extreme events, high quality, reliable and homogeneous daily time series are required. This thesis is relying on the daily maximum (Tx) and minimum (Tn) temperature series taken from the Spanish Daily Adjusted Temperature Series (SDATS) developed by Brunet et al. (2006, 2008), which cover the 1850-2005 period.

The SDATS dataset is composed of the 22 longest, adjusted, most continuous and reliable series of daily temperatures recorded in Spain since the mid-19th century onwards, which were subjected to quality control (QC) procedures to identify non-systematic biases and to homogenization to minimize and adjust systematic biases existing in the raw series.

The procedures followed for developing the SDATS dataset are summarised as follows:

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

- An assessment of the reliability of the data and metadata sources used for generating the SDATS was undertaking by the authors first.
- Time-series QC procedures were applied to the raw daily Tx and Tn series, following Brunet et al. (2008) approach described in their WMO's guidance.
- An empirical minimisation of the *screen bias*, related to the changeover from old open stands to new Stevenson screens to protect thermometers from radiation and wetting, was past to the quality controlled series before the authors undertaken the homogenisation exercise.
- The application of the Standard Normal Homogeneity Test (SNHT) to the quality controlled and pre-adjusted time-series to account for the *screen bias* followed next. Annually and monthly averaged series were used to detect and adjust the validated breakpoints.
- Finally, the authors interpolated the estimated monthly adjustments into the daily scale following Vincent et al. (2002) approach.

To apply the methodology of PP approach it is necessary to count on continuous daily data. For this reason, the period 1940-2005 with less missing data has been chosen from the SDATS. This period has been extended to 2010 thanks to the data provided by the Spanish Meteorological Agency (AEMET). One out of the 22 stations comprised in the SDATS, Cadiz, was discarded because it was not made available in the updating provided by AEMET.

The percentages of missing data for both variables (Tx and Tn) for the period (1940-2010) are shown in Figure 3.2. The missing daily values of SDATS dataset have been filled in. To deal with this problem, the R package *mice* (Multivariate Imputation by Chained Equation) developed by van Buuren and Groothuis-Oudshoorn, (2011) was applied. The R package *mice* imputes incomplete multivariate data by chained equations.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

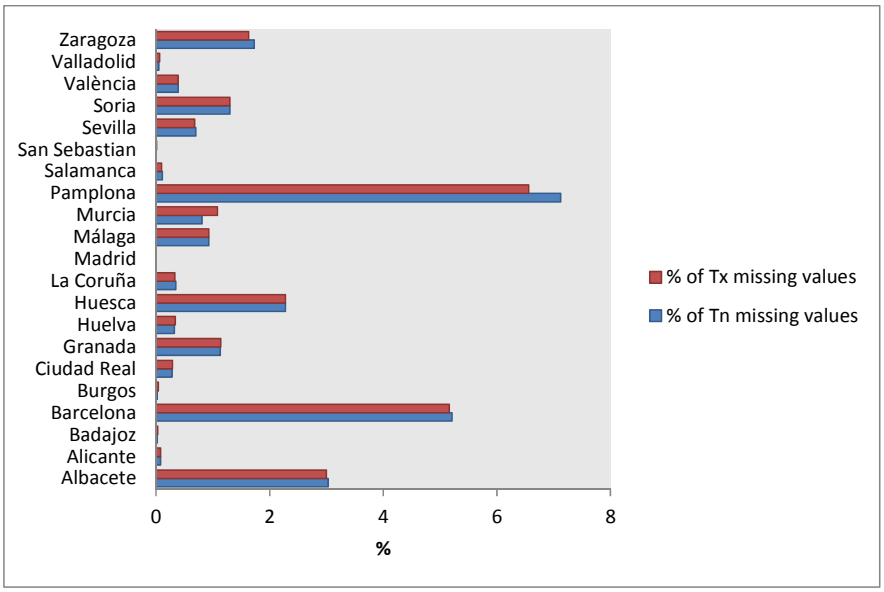


Figure 3.2. Percentage of missing daily values in Tx and Tn series for each station during the period 1940-2010.

mice generates multiple imputations for incomplete multivariate data by Gibbs sampling. Multiple imputation involves filling in the missing values multiple times, creating multiple "complete" datasets. Therefore, the missing values of daily Tx and Tn from 1940 to 2010 were filled in by Gibbs sampling. Gibbs sampling is a Markov chain Monte Carlo algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution, when direct sampling is difficult. Each series containing missing values was predicted from the four stations with better correlation. The correlations have been estimated using a Spearman pairwise correlation test and the lower correlation accepted was 0.84 for Tx and 0.86 for Tn. The prediction equations are used to impute plausible values for the missing data. The process iterates until convergence over the missing values is achieved. Bayesian linear regression was chosen to replace the missing data, hence, the imputation was made according to a linear imputation model. This method is fast and efficient if the model residuals are close to normal (van Buuren and Groothuis-Oudshoorn, 2011).

In order to assess whether the imputations created by *mice* algorithm are plausible, the probability densities of both the observed and in filled data were superimposed. For all stations and for both variables, insignificant differences in the densities between observed and completed values were detected (not shown), which means that the imputations are reasonable. Figure 3.3 show one example of the diagnostic plot for Tx in Barcelona from 1940-2010.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

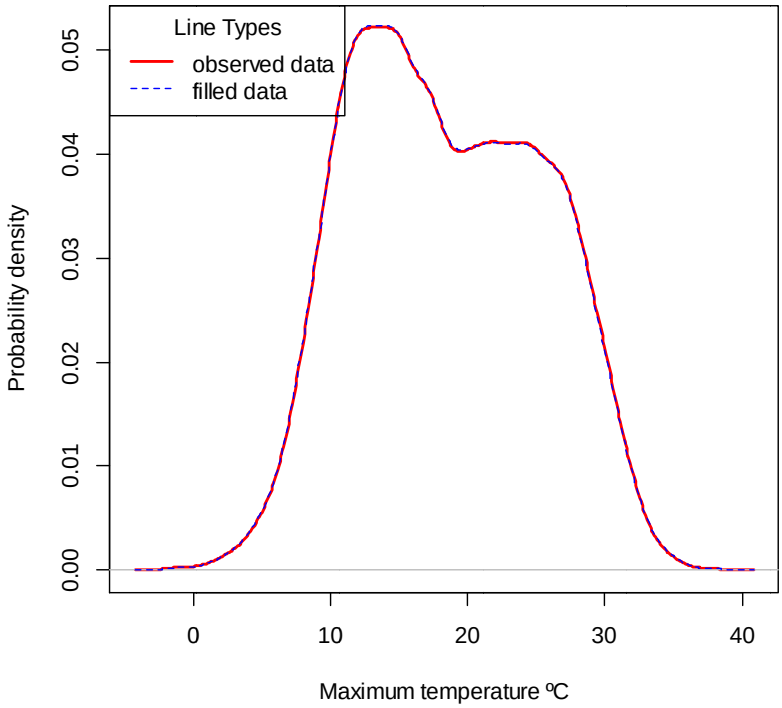


Figure 3.3. Probability density function of the observed data (red line) and probability density function of the completed data with *mice* (dashed blue line).

3.2.2 Large-scale datasets

The possible relationships between the dynamic and thermodynamic processes favouring the intensity of summer very hot days (VHD) and very hot nights (VHN) and these extremes over mainland Spain have been examined using three large-scale datasets, namely, daily sea surface temperature (SST), Sea Level Pressure (SLP) and Soil Moisture (SM) datasets, which are described next.

3.2.2.1 Sea Surface Temperature data

It is relatively easy to find and access to long-term gridded monthly SST datasets, but data on the daily scale is much more limited, since they are only available from 1982 onwards. However, the length of the period 1982-2010 used in this thesis is closer to the normal period (30-years long) recommended by WMO for providing climatological validity to the results.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Daily anomalies of SST were taken from the National Oceanic and Atmospheric Administration (NOAA). This is the version 2.0 produced using satellite data and in situ data from ships and buoys. The spatial grid resolution is 0.25 degrees, the coverage of the dataset is global. A detailed description of the complete analysis procedure can be found in Reynolds et al. (2007). The data is available at:
ftp://eclipse.ncdc.noaa.gov/pub/OI-daily-v2/NetCDF/1981/AVHRR/

For including de SST anomaly as a covariate into the model, all the Spanish Tx and Tn time series have been limited to the period 1982-2010 because the daily SST anomalies are only available from 1982.

In order to be able to make the process operative and minimize the computing time required for this analysis, the SST data have been re-sampled. The technique employed relies on the estimation of simple averages of the original grid-cell (0.25º x 0.25º) with SST anomaly data to 1º x 1º grid box for the windows 15º W - 10º E and 31º N - 48ºN. Figure 3.4 shows both maps, the original and the resampling map, where it can be seen the differences of the grid sizes.

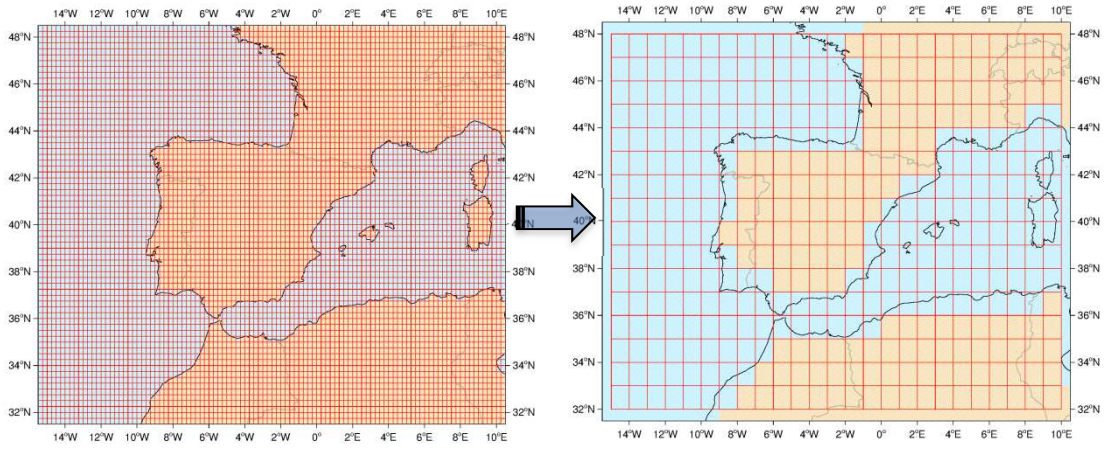


Figure 3.4. Resampling from 0.25º x 0.25º to 1º x 1º grid box resolution of SST anomalies data.

To relate the VHD and VHN to SST anomalies, it has been calculated a moving average of the previous 15 days to the targeted extreme temperature day to take into account the thermal inertial of the sea, which is introduced into the model. Different windows of moving average (10, 15 and 20 days) were tested and 15 days average showed better results.

3.2.2.2 Sea Level Pressure (SLP) dataset

To represent anomalous atmospheric conditions, Daily Averages of SLP form the National Centers for Environmental Prediction (NCEP) Reanalysis have been used (Kalnay et al., 1996). The NCEP Reanalysis data was provided by the NOAA[1]/OAR[2]/ESRL[3] PSD[4], Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. The dataset has a global coverage and provides daily data for the period 1948 to present. Nevertheless for this study a smaller geographical window (25º W to 20º E - 24º N to 60ºN 2.5º x 2.5º) has been chosen for the period 1948-2010. The spatial resolution of the SLP grid data is 2.5º x 2.5º of regular grid.

For each grid, daily anomalies have been calculated as the difference between each daily SLP data and the normal value for every day of the year. The normal value has been estimated for each day using the 1961-1990 reference period. Finally, only the summer months (JJA) of the period 1948-2010 has been selected.

3.2.2.3 Soil Moisture (SM) dataset

As the SM variable could play an important role intensifying extreme temperature events over mainland Spain, this variable has been also related to the extreme temperature days under scrutiny. To do so, the mean Daily Volumetric Soil Moisture between 0-10 cm Below Ground Level from NCEP Reanalysis data provided by the NOAA/OAR/ESRL PSD has been used. The dataset is available in the following webpage: http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html for the period 1948-onwards. Kalnay et al. (1996) provide complete details on the estimation fields and the calculations of the dataset source. The dataset is presented in a Gaussian grid but, for the convenience of data processing it has been regridded to a regular grid of 1.904º lat x 1.875º lon.

The selected domain for this study was 15º W to 10º E and 31º N to 48ºN for the period 1948-2010. The availability of the SM data only since 1948 has limited the investigation of the role of this covariate into the PP model to this period.

[1] NOAA: National Oceanic and Atmospheric Administration
[2] OAR: Oceanic and Atmospheric Research
[3] ESRL: Earth System Research Laboratory
[4] PSD: Physical Science Division

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Likewise in the SLP anomalies calculation, the SM daily anomalies have been estimated for each grid of the studied window. The normal value for each day of SM variable has been estimated using the 1961-1990 reference period. The model was tested with different time-windows of moving average (10, 20 and 30 days) and finally the 30-day moving average was selected because in this case more grids showed model improvement for this covariate.

3.3 The Methodology applied

Assessments of extreme temperature events using models based on extreme value theory (EVT) are increasing in recent years. Many publications have used the statistical modelling of extremes to analyse temperature extremes, most of them using non-stationary GEV distribution (Furió and Meneu, 2011; Kharin and Zwiers, 2005; Sillmann et al., 2011), although more straightforward to apply, GEV models use few information because annual maxima not catch all extremes. Fewer studies have investigated trends in the characteristics of the extreme observations using threshold excesses models (Abaurrea et al., 2007; Brown et al., 2008; Dalelane and Deutschländer, 2013; Furrer, 2010; Katsoulis and Hatzianastassiou, 2005), which are better alternative because they make use of more available information about the upper tail of the distribution than GEV models. In this thesis, a statistical model for extreme value analysis, so-called Point Process (PP) approach, is used. The model described in section 2.5 is based on extreme value distributions of threshold excess and can be extended to non-stationary processes by including covariates (Coles, 2001).

The inclusion of covariates in the statistical modelling of climate extremes enables the study of the relationship between a large-scale atmospheric pattern and the climate extreme (Sillmann et al., 2011). For instance, the influence of North Atlantic Atmospheric Blocking on extreme cold winter temperatures in Europe was analysed fitting the Generalized Extreme Value (GEV) distribution to monthly minimum temperatures of Europe with and indicator for atmospheric blocking condition being used as covariate (Sillmann et al., 2011). Also Brown et al. (2008) studied the global effect of the North Atlantic Oscillation (NAO) on extreme winter temperatures

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

introducing the NAO index[5] into a marked PP model of extreme values of daily temperatures.

SLP, SST and SM anomalies have been related to European extreme temperature events (Carril et al., 2008; Della-Marta et al., 2007; Jaeger and Seneviratne, 2011). Following the methodology of Sillmann et al. (2011) and Brown et al. (2008), three models were considered in this thesis to study the influence of dynamical and physical processes on summer extreme temperatures in mainland Spain including anomalies of SLP, SST and SM as covariates into the PP approach.

Besides, changes in temperature extremes could be linked with changes in the location, the scale and the shape of the extreme distribution (Brown et al., 2008; Coles, 2001; Kharin and Zwiers, 2005). Thus, in this thesis, it is also assessed changes in the characteristics of the Tx and Tn extreme distributions.

For the extreme value analysis, the package of *extRemes* (Gilleland and Katz, 2011) has been used as a main statistical tool, but also the *evd* package (Stephenson, 2004) has been applied. The *extRemes* package is a suite of functions for carrying out analyses on the extreme values of a process of interest and is specially indicated to weather and climate applications of Extreme Value Analysis (EVA). The *evd* package extends simulation, distribution, quantile and density functions to univariate and multivariate parametric extreme value distributions, and provides fitting functions which calculate maximum likelihood estimates for univariate and bivariate maxima models, and for univariate and bivariate threshold models. These packages are in the open source statistical programming language of R and are available at http://www.r-project.org/

The methodology applied to statistically modelling and analyse summer very hot events is described next by providing the general procedure followed to identify summer temperature extremes, such as VHD and VHN (sub-section 3.3.1), the approach for the threshold selection (sub-section 3.3.2), the declustering procedure (sub-section 3.3.3), the modelling using the PP approach (sub-section 3.3.4) and the suitability of the covariates (sub-section 3.3.5). In addition, the effective return levels calculation and the composite mapping technique are described in the sub-sections 3.3.6 and 3.3.7, respectively.

[5] The NAO index is defined as the standardized seasonal mean pressure difference between the Azores and Iceland (Walker and Bliss, 1932).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

3.3.1 Definition of VHD and VHN

An extreme weather and climate event does not only mean that an extreme occurs at an individual point, but more generally it has a certain impacted area and duration, which means that it is a regional extreme event (Ren et al., 2012). At the same time the extremity of a weather or climate event of a given magnitude depends on a geographic context. For instance, a temperature corresponding to the expected climatological daily maximum in one site could be an extreme event in another site. Therefore extreme events have not a standardized and universal definition. Globally there are large geographical variations in daily temperature extreme (Donat et al., 2013). Also over the area in study in this thesis, mainland Spain, has been found different spatial patterns for the summer counts of days exceeding Tx>90th and also of days exceeding Tn >90th (Brunet et al., 2007a).

Several definitions of extreme temperature events have been considered in climate studies in order to assess their frequency, intensity and persistence (Abaurrea et al., 2007). The definitions can be based on absolute thresholds (Klein Tank and Können, 2003), or on relative thresholds like percentiles (e.g. Abaurrea et al., 2007; Alexander et al., 2006; Brown et al., 2008) for a weather variable (e.g., daily maximum temperature or daily minimum temperature).

Zwiers et al. (2013) defined extreme events as weather or climate events which can be outlined in terms of measurable physical quantities, such as temperature, precipitation, wind speed, runoff levels or similar, which are rare within the current climate, i.e, located in the tails of the probability distribution. They are not expected to occur each year and correspond to very high quantiles, such as 95th, 99th or 99.9th percentiles.

The IPCC Special Report on extreme events (IPCC, 2012) defined an extreme event as the occurrence of a weather or climate variable value above (or below) a threshold near the upper (or lower) ends of the range of observed values of the variable.

This thesis is focused on the upper tail of the Spanish Tx and Tn series daily distributions and it is applied a statistical criterion based on percentile thresholds to identify the extreme events. Taking into account both definitions of extreme events (Zwiers et al., 2013 and IPCC, 2012), here very hot days (VHD) are defined as days exceeding a threshold over the 95th percentile of daily Tx for the summer season and

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

very hot nights (VHN) as the days in which the daily Tn values exceed the high threshold. The method of a threshold selection is described in the next section (3.3.2).

3.3.2 Threshold selection

When modelling threshold excesses with the PP approach, it is important to choose correctly the best threshold value u because the selected threshold has a significant impact on the fit of extreme value model. Too low threshold will incorporate non extreme events and will violate the asymptotic basis of the model, causing bias, while too high threshold will unnecessarily reduce the available data and will produce few excesses over the threshold leading to a high variance in the estimated values.

The methodology applied to identify the best threshold in each temperature series requires fitting the data to the GPD distribution at a range of thresholds and to look for stability of parameters estimates. The technique consist in plotting both estimated parameters, the generalized Pareto scale parameter reparameterized $\hat{\sigma}^*$ and the generalized Pareto shape parameter $\hat{\xi}$, against the possible thresholds, together with the confidence intervals for each of these quantities, and select the threshold as the lowest value of possible thresholds for which the estimates remain near-constant (Coles, 2001: 83). Figure 3.5 shows one example of the plots of parameters estimates against threshold for daily Tx in Barcelona during the period 1948-2010.

For most of the analysed series, the diagnostic plots (not shown) indicated that a threshold of around the 95th percentile was adequate.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

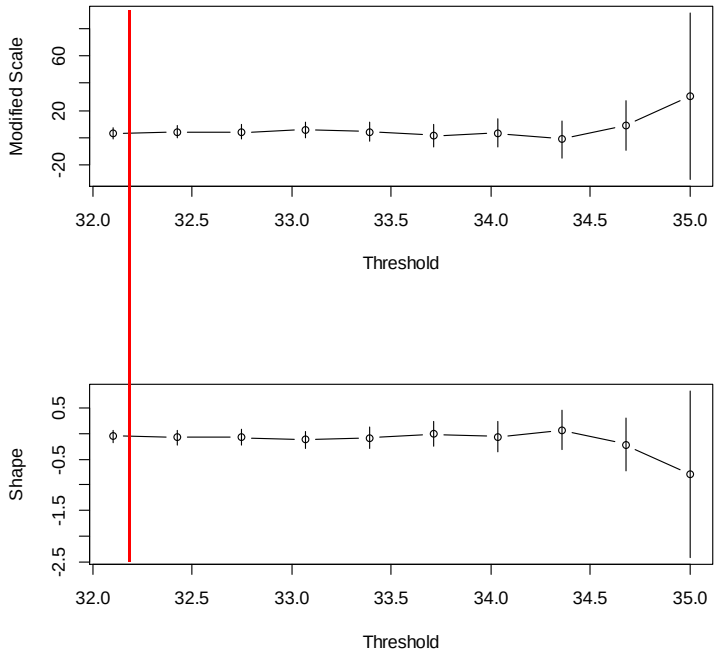


Figure 3.5. Parameter estimates from GPD fit for a range of 50 thresholds from 95th to 99.5th percentile of daily maximum temperature data from 1948-2010 for Barcelona. The red line indicates the chosen threshold (32.1ºC).

3.3.3 Declustering

The temporal dependence is a common issue in univariate extreme studies; particularly, the events exceeding high threshold have a tendency to occur in clusters. The occurrence of an extreme event one day may influence the probability of the occurrence of the following extreme event the next day. According to the asymptotic approximation the distribution of any one of the threshold excesses can be modelled using a GPD distribution, but the dependence in the observations make invalid the MLE method because the observations must be independent as indicated in section 2.5.1. There is no alternative likelihood function that incorporates the dependence between observations. The solution for dealing with the problem of dependent excesses in the model was addressed by the application of a declustering method (Coles, 2001: 99). The method consists in selecting the maximum intensity within clusters (measured by the highest temperature observed in each cluster).

After selecting the best threshold of Tx and Tn series following the procedure described before, the steps to declust the data are:

- Extremes separated by r=1 non-extremes belong to the same cluster.

were r is the run length.

It is assumed run length = 1 as a convention. For instance, clusters of maximum temperature above a threshold with one day with maximum temperature below are considered to belong to the same cluster, because the events are probably dependents and caused by the same triggers. And if the number of non-extreme days is bigger than one, then the clusters are different.

For r=1, it has been estimated the extremal index, which was close to one (if not equal to 1) in all series, suggesting no dependence in the extreme levels.

The extremal index is a parameter measuring the degree of clustering of extremes in a stationary process. It is between 0 and 1 and is the reciprocal of the mean cluster size (Coles, 2001; Ferro and Segers, 2003; Smith and Weissman, 1994) defined by:

$$\theta = (\text{limiting mean cluster size})^{-1}$$

where limiting is in the sense of cluster of excesses of increasingly high thresholds (Coles, 2001: 97).

If $\theta = 1$ then excesses of an increasing threshold occur isolated and if $\theta < 1$ the excesses tend to cluster.

- Identify the maximum excess within each cluster
- Assuming that cluster maxima (series of maximum of each cluster) are independent with conditional excess distribution given by the GPD.
- Fitting the GPD to the cluster maxima

It has to be ensured that clusters do not cross the summer seasonal boundaries, since the data covers several years but only for summer season. Then there are blocks, implying a natural clustering that has been preserved.

Hence, it has to take into account that the PP approach is vulnerable to determinations from the independent and identically distributed (i.i.d.) assumption and therefore the discrimination of independent cluster maxima is crucial. The separation of extreme events into cluster maxima is likely to be sensitive to the threshold. Here it has been considered a cluster to be active until an arbitrary choice of one value fallen below the threshold. The methodology used is simple but has its limitations because the results can be sensitive to the arbitrary choices made in cluster determination and there is wastage of information in discarding all data except the cluster maxima.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

3.3.4 Model: Non-stationary point process

PP approaches are used to test the influence of the anomalies of SLP, SST and SM on VHD and VHN at the 21 Spanish stations.

It has been assumed a linear dependence of the covariates (SLP, SST and SM anomalies) in the location parameter of the GEV distribution of Tx and Tn.

The effect of the covariate on the location parameter of the GEV corresponds to both an effect on the relative frequency of exceeding a high threshold (in terms of rate parameter for Poisson distribution) and an effect on the excess temperature over the high threshold (in terms of scale parameter of the GP distribution) how the next well-known equations shows (Katz et al., 2002):

$$\lambda = \left(1 + \xi \frac{u-\mu}{\sigma}\right)^{-1/\xi} \qquad \tilde{\sigma} = \sigma + \xi(u-\mu)$$

Next, the general procedure is described, which is the same for the three large-scale covariates.

After selecting the best threshold of the daily Tx and Tn series for the chosen period according to the methodology described in section 3.3.2:

- First, a stationary PP is fitted to estimate the initial parameters of the approach by MLE method.
- Second, the declustering in the stationary PP is done with r=1.
- Third, the non-stationarity is introduced allowing the location parameter to depend linearly on the covariate. For a gridded data it has a $(n \times m)$ matrix of covariates such as:

$$Z = \begin{pmatrix} z_{11} & \cdots & \cdots & z_{1n} \\ \vdots & \ddots & z_{ij} & \vdots \\ \vdots & z_{ji} & \ddots & \vdots \\ z_{m1} & \cdots & \cdots & z_{mn} \end{pmatrix}$$

where n and m are the number of grids in latitude and longitude respectively. Then by the MLE method is estimated a new location parameter for each covariate (grid) as follows:

$$\mu_{ij}(z) = \mu_0 + \mu_1 \cdot z_{ij}$$

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

where z_{ij} is the associated covariate of SST, SLP and SM anomalies.

It has been investigated the linear dependence because the linear form is considered an adequate approximation to any type of dependence (Brown et al., 2008). The existence of a linear dependence is enough to demonstrate the relationships with the covariate. The scale and shape parameter σ and ξ are assumed constants.

Furthermore, a non-stationary PP approach is employed to assess the time trend in the extreme distributions of Tx and Tn.

In theory, trends can be assumed in the three parameters μ, σ and ξ. However, if some of them have insignificant change in the considered series, it may be advantageous to keep these parameters constant (Kharin and Zwiers, 2005). Usually, the shape parameter is assumed constant in the statistics of extreme value because there are few observations in the tail and it is difficult to estimate it (Maraun et al., 2011).

In the present thesis, first, the scale and shape parameters are assumed as constants and, second, the shape parameter is only assumed to be constant. Hence, in the first model a trend is assumed for the location parameter; namely, the expected value μ is assumed to be linearly dependent on time, as follows:

$$\mu_i(t) = \mu_0 + \mu_1 \cdot t_i$$

Time is scaled to range from 0 to 1 and placed into the model as a vector:

$$t = (t_0, ..., t_i, ..., t)$$

In the second model, the location and scale parameters are hypothesized to be linear throughout time:

$$\begin{cases} \mu_i(t) = \mu_0 + \mu_1 \cdot t_i \\ \ln \sigma_i(t) = \sigma_0 + \sigma_1 \cdot t_i \end{cases}$$

To ensure the scale parameter is positive, a log-linear trend is assumed.

The intercept coefficients μ_0 and σ_0 are the parameters values at time t_0. The slope coefficients μ_1 and σ_1 characterize the rate of change in the GEV parameters estimates.

The shape parameter is assumed to be independent of time in this study. A shape parameter that changes over time implies that the underlying extreme distribution changes and this analysis is not on the focus of this thesis.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

3.3.5 Suitability of the covariates into the model

The improvements to the model reached by introducing covariates is ensured by using the likelihood ratio test or also called the deviance test (Coles, 2001:109): the difference in negative log-likelihood values between the two nested models ($M_0 \subset M_1$) is defined as:

$$D = 2\{\ell_1(M_1) - \ell_0(M_0)\}$$

where $\ell_0(M_0)$ and $\ell_1(M_1)$ are the maximized log-likelihoods under models M_0 and M_1 respectively.
The hypothesis M_0 is rejected at the significance level α if:

$$D = 2\{\ell_1(M_1) - \ell_0(M_0)\} > c_\alpha$$

where c_α is the (1-α) quantile of the χ^2_k distribution and k is the difference in the dimensionality of M_0 and M_1.

For both variables Tx and Tn, in this thesis, it has been calculated the likelihood-ratio test between the next models:

Model

1. Stationary model
2. Non-stationary model (μ linear dependence with SLP anomalies)
3. Non-stationary model (μ linear dependence with SST anomalies)
4. Non-stationary model (μ linear dependence with SM anomalies)
5. Non-stationary model (μ linear dependence with time)
6. Non-stationary model (μ and σ both linear dependence with time)

Model comparison with likelihood-ratio test

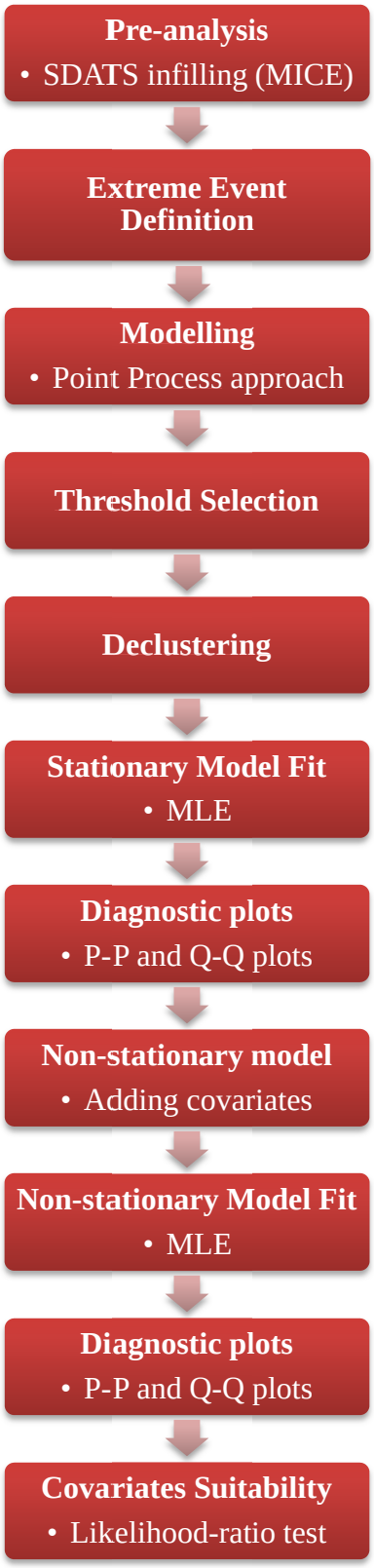
Model 1 vs Model 2
Model 1 vs Model 3
Model 1 vs Model 4
Model 1 vs Model 5
Model 5 vs Model 6

Finally, for each case the p-value associated to the χ^2_k test statistic of D with k degrees of freedom has been calculated, where the p-value is the probability of getting the results (or more extreme results) assuming that the null hypothesis is true. For p-values less than 0.05 significance level the null hypothesis was rejected.

The scheme of the general methodological procedure applied in this thesis to model summer VHD and VHN is presented next:

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

SCHEME PROCEDURE



UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

3.3.6 Effective return levels

The effective return levels are those derived from fitting a non-stationary PP model. The idea of effective return levels relies on the fact that for each value of a covariate, you get what the return level would be if that were the fixed value.

A difficulty arises in the estimation of the standard errors or confidence intervals; actually this is still an active area of research in the field of extreme values theory (Gilleland and Katz, 2011).

The effective return levels have been calculated with the new version of package extRemes, extRemes 2.0, provided by Eric Gilleland. extRemes v2.0 does contain functions for performing extreme value analysis and has a much better way of calculating the effective return levels (Gilleland and Katz, 2011). However confidence intervals (CI) of the return values estimated taking into account the tendency in the location parameter could not be provided because they are not yet available. Calculus of CI for effective return levels is being tested by Eric Gilleland and they will be in the next public version of extRemes 2.0. This aspect should have taken into consideration because estimates of events with a long return period may be subjected to large uncertainties.

3.3.7 Composite maps

To identify consistent spatial patterns of SLP, SST and SM associated to VHD and VHN the composite mapping technique was used. The procedure consists in selecting the dates corresponding to VHD and VHN, respectively, recorded at each of the twenty one stations in mainland Spain and averaging the corresponding maps of the large-scale field anomalies over these selected dates. The resulting map of SST, SLP and SM anomalies is the corresponding composite associated to VHD and VHN, respectively, at each station.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

4 Results and discussion

As discussed in chapter 1, anomalous SLP, anomalously warm SST and SM deficit have been linked with extreme temperature events. To better understand dynamical and physical processes favouring the occurrence and intensity of warm extreme temperature events over mainland Spain, a PP approach has been applied to investigate the influence of SLP, SST and SM anomalies on VHD and VHN and the results of this assessment are provided in section 4.1 and 4.2, respectively. The effects of each large-scale variable are included, individually, into the statistical model in terms of a linear dependence in the location parameter of the GEV distribution and they are analysed and discussed in different sub-sections. This dependence implies an effect on the relative frequency of exceeding a threshold and on the temperature excess over it (section 3.3.4). The excess over the high threshold would be a measure of the intensity of very hot events (VHD and VHN).

In addition, observed temporal changes and trends in Tx and Tn extreme temperatures are also explored in this Chapter. To fitting data to the GEV distribution, the PP approach is applied, since changes in extremes are related to changes in the characteristics of the extreme distribution. To deal with this study, two different periods (1940-1972 and 1973-2010) are analysed. First, changes and trends in the location parameter of the GEV distribution are assessed and discussed in section 4.3 and, second, changes in the scale parameter of GEV distribution during both periods are investigated and discussed in section 4.4. Finally, an assessment of the different effective return levels is provided in section 4.5.

4.1 Very Hot Days (VHD)

Results of the influence of the three large-scale variables, SLP, SST and SM anomalies, on VHD of the 21 stations analysed are presented in this section, which at the same time is divided into three sub-sections, one for the assessment of each variable.

4.1.1 Influence of SLP anomalies on summer VHD

Anomalous atmospheric conditions have been associated with the occurrence of extreme temperature events in Europe (Andrade et al., 2012; Carril et al., 2008; Della-

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Marta et al., 2007), in the Mediterranean Sea during warm summers (Xoplaki et al., 2003), in North-eastern Spain for very warm days (El Kenawy et al., 2012a) and at the local scale to assess extreme summer temperatures in Madrid (García-Herrera et al., 2005).

In this sub-section relationships between SLP anomalies and VHD over mainland Spain is presented and discussed. The assessment was focused on summer months (JJA) for the period 1948-2010 and for which daily SLP data are available. The geographical domain considered as representative for synoptic-scale influence on the IP is 25º W-20º E, 24º N-60ºN.

The composite mapping technique was used for the identification of spatial patterns of SLP associated to VHD. A composite map consists in an average of SLP anomalies over the selected dates of VHD recorded at each of the 21 stations in mainland Spain.

Distinct spatial patterns emerge from the examination of these composites. There are, basically, three distinguishable patterns associated with VHD over mainland Spain: Southerly, Weak south-westerly and North-westerly airflow patterns, although some slight differences inside each group have been observed. Next, each pattern is presented and discussed and two representative examples for each case are illustrated. The rest of the figures can be found in the Appendix A.

1. Southerly Airflow Pattern

This pattern shows a large area of strong positive SLP anomalies over central Europe expanded southward across Mediterranean Sea and North Africa, associated with anomalously low SLP over the Central and North Atlantic. Such a configuration enables a southerly component inflow of warm and dry air masses from northern Africa invading the IP and corresponds to three different Spanish areas: inland, north-eastern and north, although some slight differences among them can be highlighted.

- Inland Spain

The stations belonging to this group are Madrid, Ciudad Real, Badajoz, Salamanca, Sevilla and Huelva. For these stations the Southerly Airflow Pattern allows warm and dry air from northern Africa to be forced over inland Spain on a southern or south-eastern component which is in agreement with García-Herrera et al. (2005) who found the same pattern associated with daily extreme summer temperatures in Madrid. Figure

4.1 shows the Southerly Airflow Pattern for two representative stations from this group which are Madrid and Salamanca.

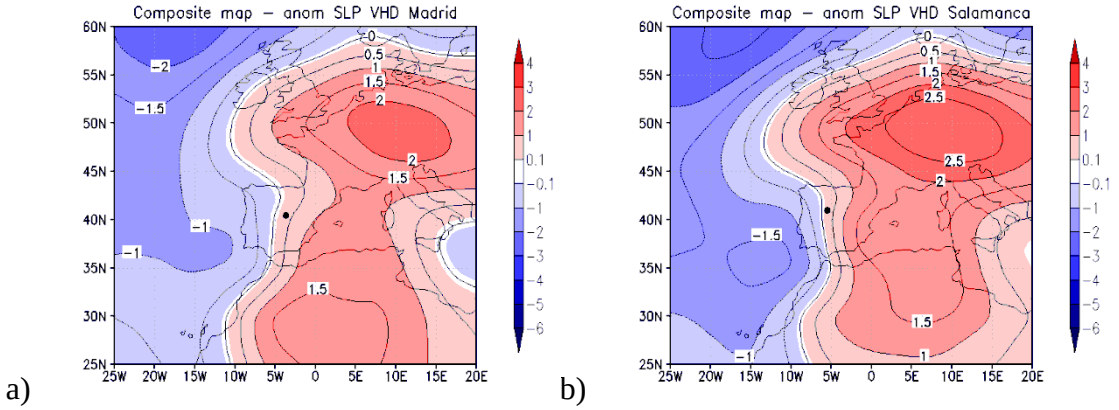


Figure 4.1 (a-b). Pattern 1. SLP anomalies (hPa) of summer VHD for Madrid (a) and Salamanca (b).

- North-eastern Spain

In the case of the stations in the north-eastern Spain (Zaragoza, Soria and Huesca and Barcelona), the extension of negative SLP anomalies from the North Atlantic over the IP modulates the flow on a south-westerly component as shown in Figure 4.2a for Barcelona. For Soria (Figure 4.2b), the positive SLP anomalies over the continent are stronger than in the case of the other stations from this group.

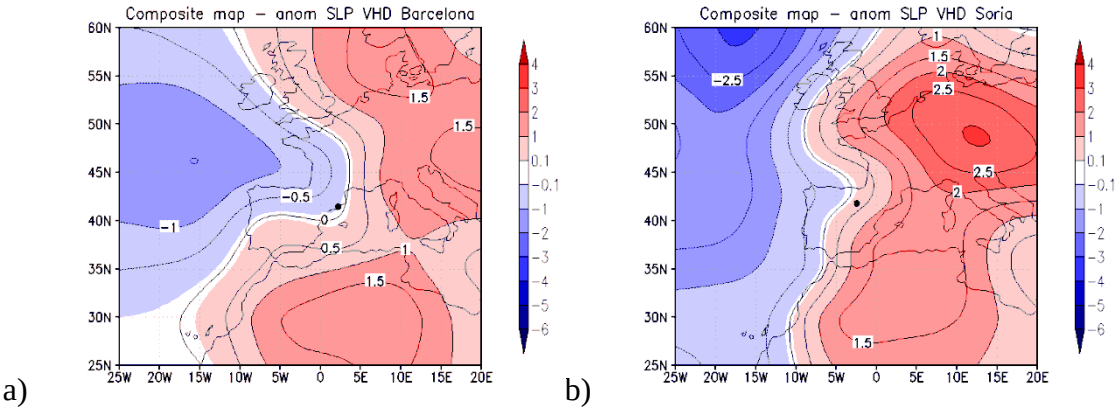


Figure 4.2 (a-b). Same as Figure 4.1(a-b) but for Barcelona (a) and Soria (b).

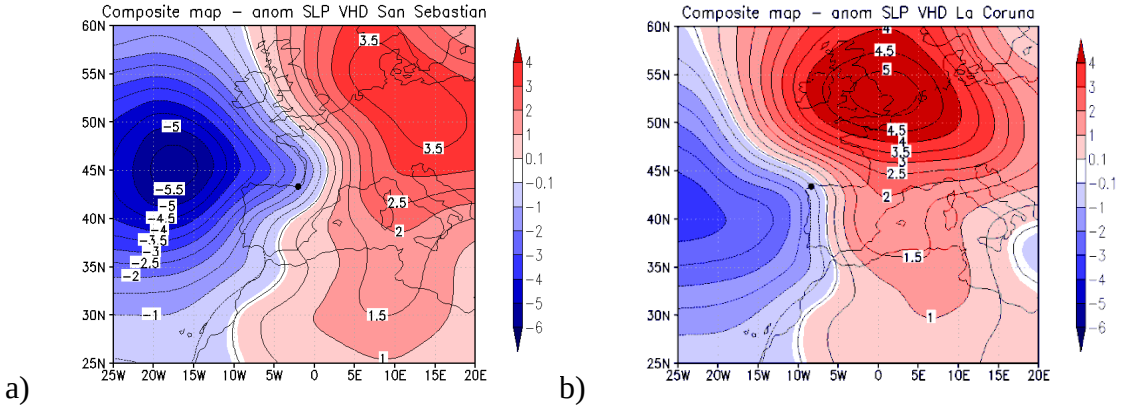
Similar results for SLP anomalies pattern have been obtained by El Kenawy et al. (2012a) who have investigated the most favourable synoptic conditions at different levels (SLP, 200 hPa, 500 hPa) for very warm days (VWD) and very cold days (VCD) over north-eastern Spain. They have identified strong positive anomalies of SLP and

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

geopotential height at different levels over central Europe, associated with negative anomalies over the Atlantic Ocean near the IP.

- North Spain

The Southerly Airflow Pattern for these stations (Burgos, Valladolid, Pamplona, La Coruña and San Sebastian) shows stronger amplitudes in the SLP anomalies than that corresponding to Inland and North-eastern stations. As, for example, for San Sebastian, the positive anomalies located over north-central Europe are about 3.5 hPa and the negative anomalies over the central North Atlantic Ocean are about -5.5 hPa (Figure 4.3a). For La Coruña, the positive SLP anomalies are even stronger (~ 5 hPa) and rather displaced over the British Isles (Figure 4.3b).



a) b)

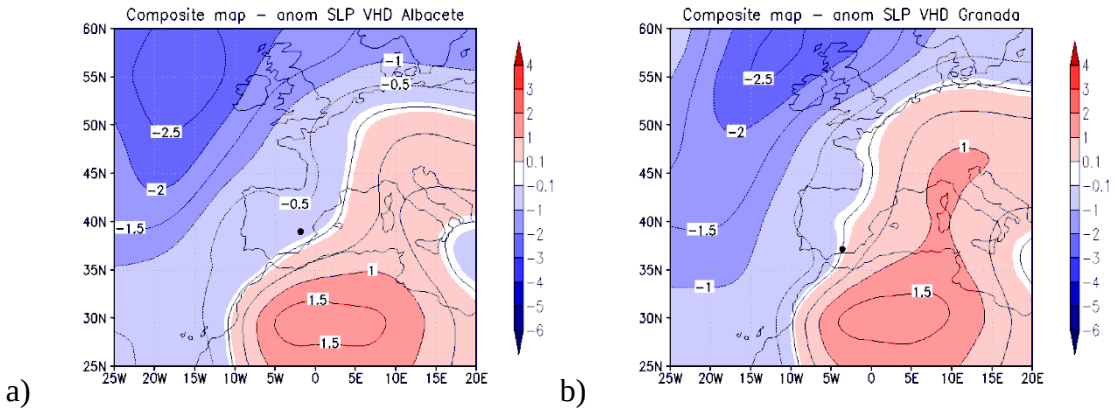
Figure 4.3 (a-b). Same as Figure 4.1 (a-b) but for San Sebastian (a) and La Coruña (b).

The Southerly Airflow Pattern presented above is in agreement with the second canonical correlation SLP mode of variability identified by Della-Marta et al. (2007). This mode was associated with heat waves in Western Europe and particularly over the IP. Also, similar SLP anomaly pattern associated with exceptionally warm summer days over Western Europe have been reported by Andrade et al. (2012).

2. Weak South-westerly Airflow Pattern

Moderate positive anomalies (~ 1.5hPa) are centred over north-western Africa, extending over the Western Mediterranean Basin and central and south-eastern Europe, meanwhile a centre of negative anomalies (~ 2.5hPa) is located over the North Atlantic in front of the British Isles. This configuration favours the advection of a weak warm

south-westerly airflow over the IP and corresponds to the pattern for the south-eastern part of the IP (Albacete and Granada). In Figure 4.4 is shown this pattern for both locations: Albacete (Figure 4.4a) and Granada (Figure 4.4b). It is apparent how the configuration enforces inflow of African air masses towards these locations.



a) b)

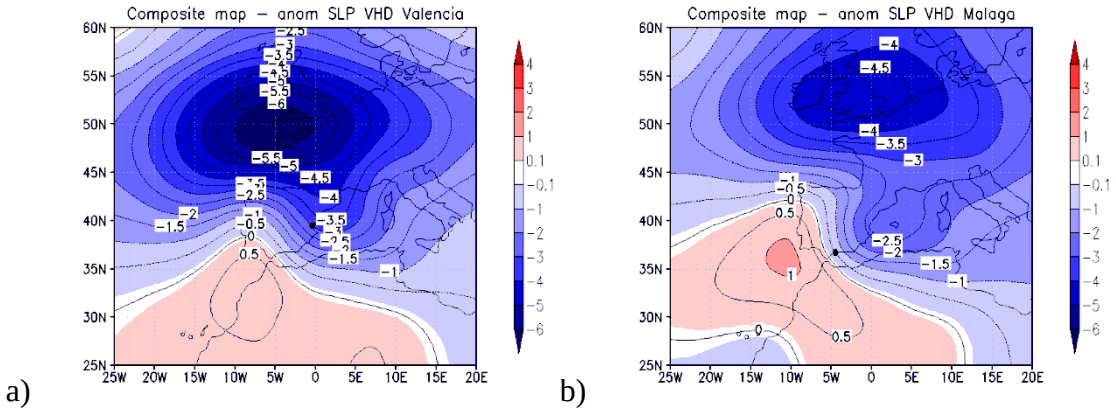
Figure 4.4 (a-b). Pattern 2. SLP anomalies (hPa) of summer VHD in Albacete (a) and Granada (b).

The patterns identified so far, namely Southerly Airflow and Weak South-westerly Airflow present a large blocking high pressure system centred either over the western part of Europe or over northern Africa, respectively. Such a configuration blocks the westerly flow and induces warm air advections over Europe.

3. North-westerly Airflow Pattern

A strong centre of negative SLP anomalies is located over the United Kingdom (UK), covering the northern half part of the IP and a weak positive centre over north-western Africa and southern Portugal, which returns the warmed north-westerly airflow when it crosses the IP. This pattern has been found for VHD at stations located in centre and southern Mediterranean coast of the IP (Valencia, Alicante, Murcia and Malaga). To illustrate this pattern two examples are displayed in Figure 4.5(a,b) for Valencia (Figure 4.5a) and Malaga (Figure 4.5b). The example pattern for Valencia shows the strongest representation of negative anomalies, being as higher as ~ -6hPa over the UK.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



a) b)

Figure 4.5 (a-b). Pattern 3. SLP anomalies (hPa) of summer VHD for Valencia (a) and Malaga (b).

Once identified the configuration of SLP anomalies associated to VHD, their influence on the intensity and frequency of daily Tx is investigated. It has been fitted the PP model in which the location parameter of the extreme distribution is linearly related to SLP anomalies. The analysis consists in comparing the stationary and the non-stationary models by using the likelihood-ratio test. The test rejects the stationary model in favour of the non-stationary when it has larger values than the 0.95 quantile of the χ_1^2 distribution. Thus the inclusion of SLP anomalies as covariates in the PP model of extreme Tx enables to find out whether the synoptic situation statistically explains extreme warm conditions recorded in Spain.

To model threshold excesses with the PP approach, it is important to choose an appropriate threshold value as discussed in section 3.3.2. The methodology applied to choose the threshold consist in fitting the temperature data to GPD distribution at a range of thresholds and select the lowest value of possible threshold for which the parameters estimates remain near-constant (see section 3.3.2). Table 4.1 shows the best thresholds selected.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.1. Best thresholds selected for daily Tx and their percentile for the period 1948-2010.

Station	Threshold (ºC)	Percentile (%)
Albacete	37.3	96.1
Alicante	33.3	95.0
Badajoz	39.8	95.2
Barcelona	32.1	95.0
Burgos	33.4	95.0
Ciudad Real	38.4	95.0
Granada	38.2	95.0
Huelva	37.5	95.0
Huesca	35.8	95.0
La Coruña	26.4	95.1
Madrid	36.2	96.3
Malaga	36.6	95.0
Murcia	38.2	95.6
Pamplona	35.4	95.0
Salamanca	35.2	95.4
San Sebastian	28.9	95.0
Sevilla	40.5	95.0
Soria	33.6	95.0
Valencia	33.0	95.0
Valladolid	35.6	95.0
Zaragoza	37.3	95.0

The stationary model has been modelled using declustered daily Tx data for the period 1948-2010. Details on the declustered methodology have been provided in section 3.3.3. The goodness-of-fit was assessed by diagnostic plots, probability plot (P-P plot) and quantile plot (Q-Q plot), based on the comparison between the empirical distribution function from observations and the estimated distribution function by the model. For all analysed series the estimated distribution function is a reasonable model of the population distribution because most of the points of the P-P plot and Q-Q plot lie close to the unit diagonal. Therefore, for the selected thresholds, the diagnostic plots show the suitability of the MLE of GPD fitted from PP approach. As an example is shown the diagnostic plots for Barcelona (Figure 4.6), where it can be seen the linear distribution of most of the points except few of them for very high quantiles.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

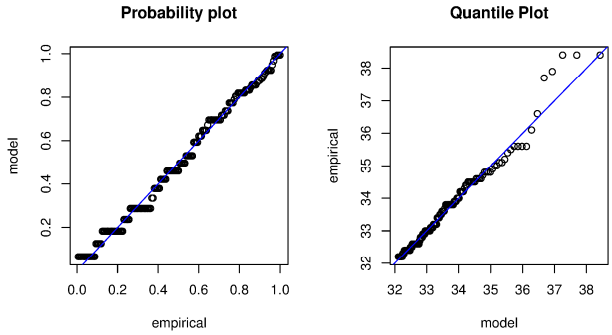


Figure 4.6. Diagnostic plots of stationary PP model fitted to Barcelona daily maximum temperatures series from 1948 to 2010.

The non-stationary model has been performed assuming the SLP anomaly of every grid box (285 in total) as linear covariate in the location parameter of the extreme distribution derived by PP approach for declustered Tx series. For all the analysed series, the diagnostic plots are approximately linear in all grid boxes, which imply a good fit of the GPD distribution, but not perfect, especially at the upper end of the distribution, where there are some points away from the diagonal. The plots are shown as an example for Barcelona in Figure 4.7.

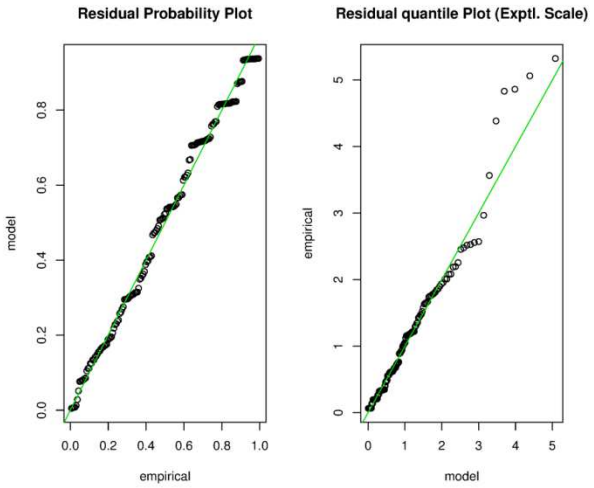


Figure 4.7. Same as Figure 4.6 but for the non-stationary model (SLP anomalies of the grid 15ºE 35ºN in the location parameter of the GEV distribution).

The suitability of a covariate into the model has been assessed using the likelihood-ratio test. This test is based on approximately χ_1^2 distribution (1 degree of freedom under the null hypothesis) which provides the probability of getting the results (or more extreme results) assuming that the null hypothesis is true (p-value). The null hypothesis

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

has been rejected when the p-value turned out to be less than the 0.05 significance level. In this case, the null hypothesis is that the model is stationary with no trend in any parameter of the GEV distribution.

Grid maps of p-values have been plotted to visualize the spatial distribution of grid points with significant improvement of the model when SLP anomalies are taken into account. The same stations presented as examples for composite maps of SLP anomalies associated to VHD have been selected to display examples of grid maps. Grid maps of the Southerly Airflow Pattern are presented in Figure 4.8(a-f) for Madrid, Salamanca, Barcelona, Soria, San Sebastian and La Coruña, respectively. Grid maps exemplifying the Weak South-westerly Airflow Pattern are depicted in Figure 4.9(a-b) for Albacete and Granada, respectively. In Figure 4.10(a-b) are shown grid maps of the North-westerly Airflow Pattern for Valencia and Malaga. The other grid maps for the rest of the stations are gathered in the Appendix B. The results show a strong relationship between SLP anomalies and VHD in all the examined locations, without any exception. The grid points with high SLP anomalies both positives and negatives, included as covariate into the model, improved it at least at the 0.001 significance level.

Analysing the areas with p-values lower than 0.001, which means that the observed deviation from the stationary model is statistically significant, it can be seen they are in good agreement with the highest anomalies of large-scale patterns of SLP anomalies. Thus, the statistical model improved when the location parameter is assumed linearly dependent on SLP anomalies and therefore they are related to intensity and frequency of VHD.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

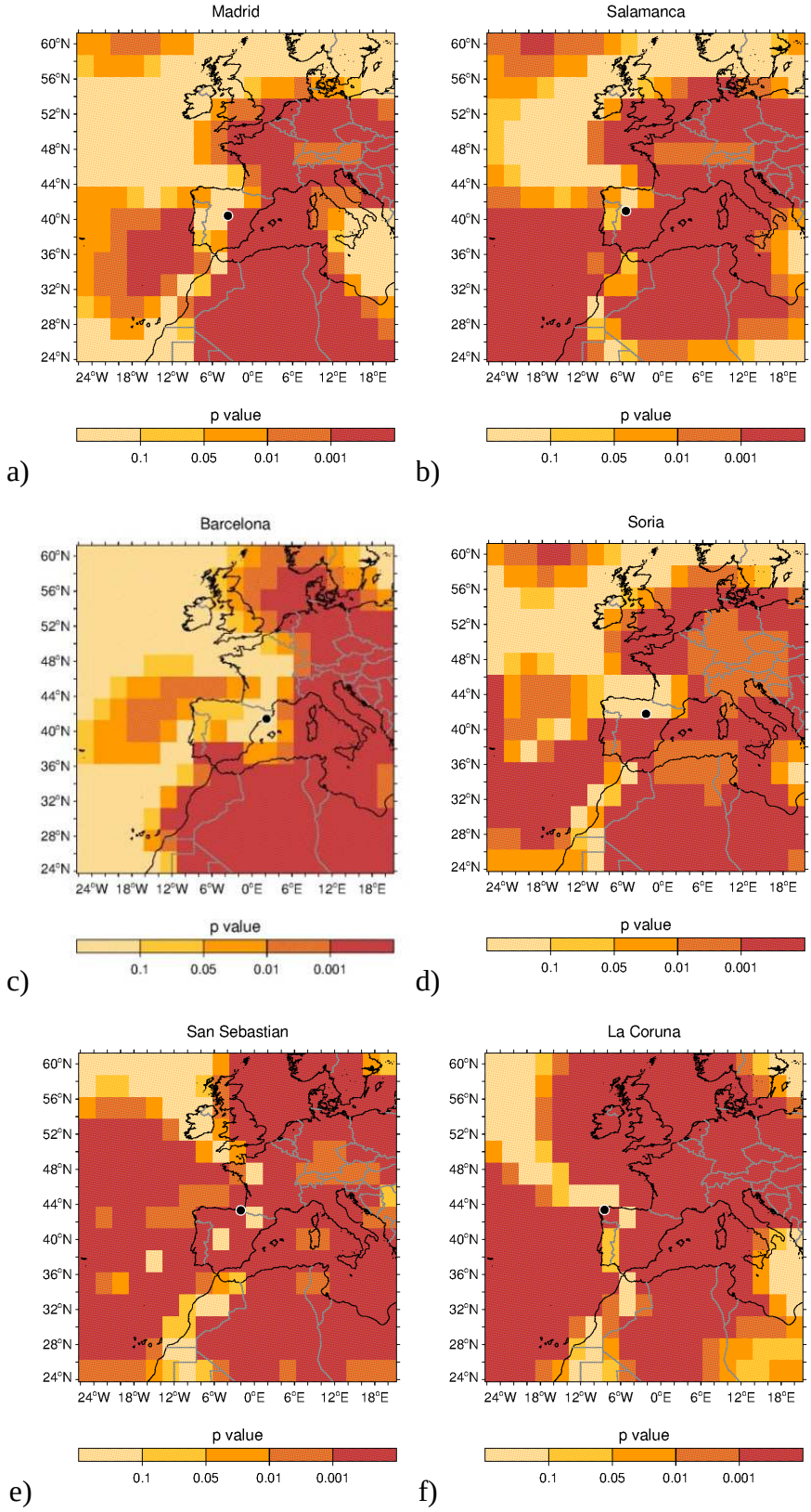


Figure 4.8 (a-f). Gridded maps of p-values between the base model and the model with SLP anomalies as covariate in the location parameter of the GEV distribution for Madrid (a), Salamanca (b), Barcelona (c), Soria (d), San Sebastian (e) and La Coruña (f).

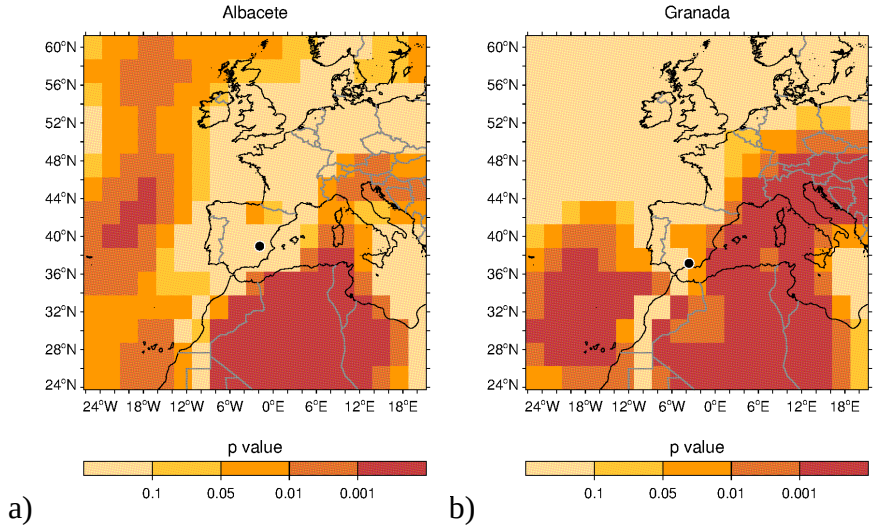


Figure 4.9 (a-b). Same as Figure 4.8 (a-f) but for Albacete (a) and Granada (b).

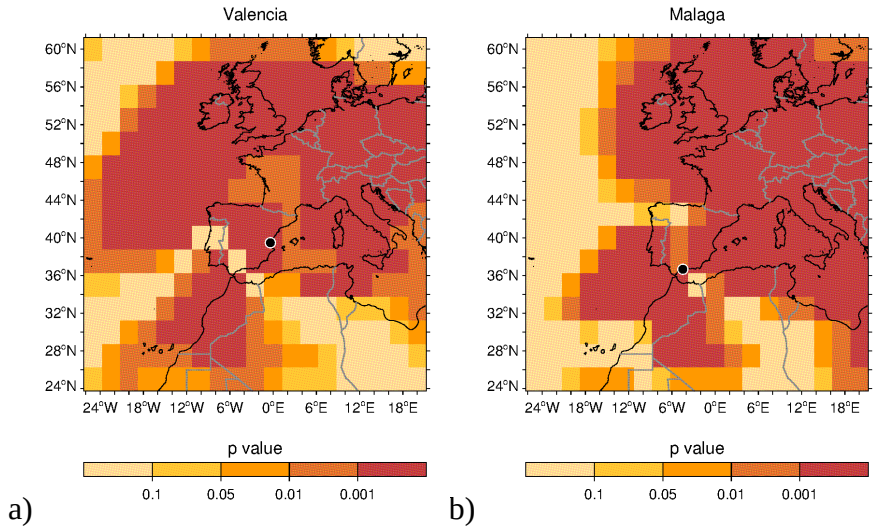


Figure 4.10 (a-b). Same as Figure 4.8 (a-f) but for Valencia (a) and Malaga (b).

4.1.2 Influence of SST anomalies on summer VHD

Warm summers in Europe have been also associated with the movement of anomalously warm SST across the North Atlantic during spring months (Colman and Davey, 1999). Carril et al. (2008) related frequent extreme temperature events in the Euro-Mediterranean region during early summer to anomalous warming of the Mediterranean SST and Della-Marta el al. (2007) demonstrated the North Atlantic SST anomalies represent an important factor in the decadal modulation of heat waves. Other studies focused on the European summer 2003 heat wave related this extreme event to exceptional SST anomalies in the Mediterranean Sea (Carril et al., 2008; Feudale and

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Shukla, 2007 and 2011 and Garcia-Herrera et al., 2010). In particular, during the European summer 2003 heat wave the Mediterranean SST was exceptionally warm reaching anomalies from 2ºC to 4ºC with the warmest values observed over the north-western part of the Mediterranean Sea (Feudale and Shukla, 2007 and 2011).

Therefore, after analysing relationships between VHD and SLP anomalies, this section is intended to identify relationships between intensity and frequency of VHD over mainland Spain and SST anomalies from seas and ocean surrounding the IP. Temporal focus of this assessment is put on summer months (JJA) for the period 1982-2010, for which daily SST data are available. The domain has been limited to 25ºW-10ºE - 32ºN-52ºN because the analysis carried out returned this spatial window as the most influential for the Spanish VHD, since SST impact was degraded over farther marine regions.

First of all, to see the mean state of SST anomalies during the days under scrutiny and to better understand the results returned by the statistical modelling, composites maps for mean summer VHD have been elaborated. Figure 4.11(a-b) shows two illustrative examples of the mean state of SST anomalies during VHD for Barcelona and Madrid. The remaining figures are provided in the Appendix C.

Composites maps are similar for all locations, drawing same pattern, although intensity is appreciably higher in coastal locations, such as Barcelona, Valencia and Malaga, along with locations in the Ebro Basin: Zaragoza and Huesca. SST anomalies along the IP seashores are positive almost everywhere. The highest values are observed over western Mediterranean, but also there are high anomalies along the Iberian coastal zones and in the central North Atlantic Ocean (southwest of the IP). A similar pattern was identified by Della-Marta et al. (2007) using canonical correlation analysis between heat wave index and SST anomalies during summer.

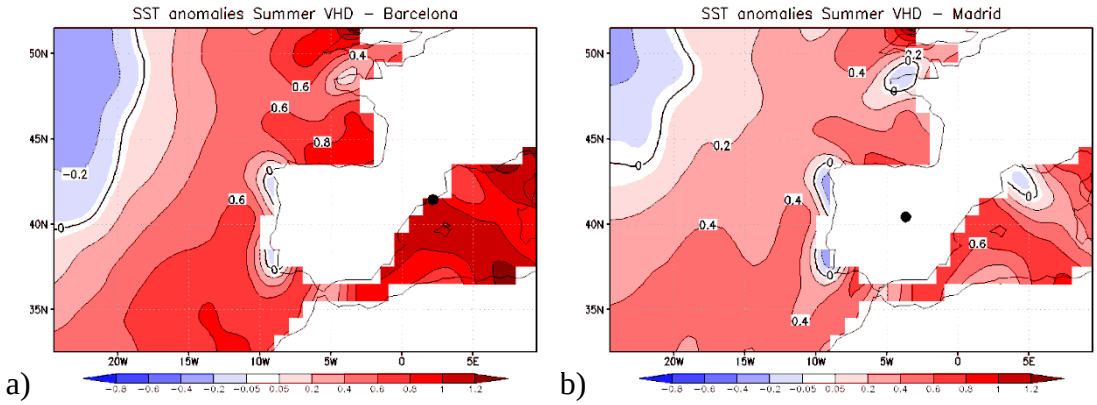


Figure 4.11 (a-b). Composite maps SST anomalies for summer VHD in Barcelona (a) and Madrid (b).

Next, the statistical modelling of VHD has been carried out following the methodology introduced in section 3.2.

While the composite maps explain only the mean state of SST anomalies for VHD during summer months, the inclusion of these anomalies into the statistical model could improve the fit of their extreme distribution, explaining relationships between SST anomalies and these daily temperature extreme events. To deal with statistical analysis, first, it has been selected the thresholds (see Table 4.2) following the methodology described in section 3.2.2. Second, the stationary model, also named base model in this study, has been modelled using declustered daily Tx data. For all stations (not shown), the expected values are close to the observed ones and are approximately linear in the diagnostic plots, which indicates the suitability of the MLE of GPD fitted from PP approach for the selected thresholds.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.2. Best thresholds selected for daily Tx and their percentiles for the period 1982-2010.

Station	Threshold (ºC)	Percentile (%)
Albacete	37.2	95.0
Alicante	34.0	95.9
Badajoz	40.0	95.0
Barcelona	32.6	95.0
Burgos	34.2	95.0
Ciudad Real	38.8	95.0
Granada	38.5	95.0
Huelva	37.9	95.0
Huesca	36.4	95.0
La Coruña	26.6	95.0
Madrid	36.5	95.0
Malaga	37.0	95.2
Murcia	38.3	95.1
Pamplona	36.1	95.1
Salamanca	35.2	95.0
San Sebastian	29.4	95.0
Sevilla	41.3	96.0
Soria	34.0	95.0
Valencia	34.0	95.0
Valladolid	36.2	95.0
Zaragoza	37.8	95.0

Figure 4.12 shows diagnostic plots exemplified for Barcelona. All the points are approximately linearly distributed along the diagonal, although some of the highest quantiles show the largest departures from the diagonal.

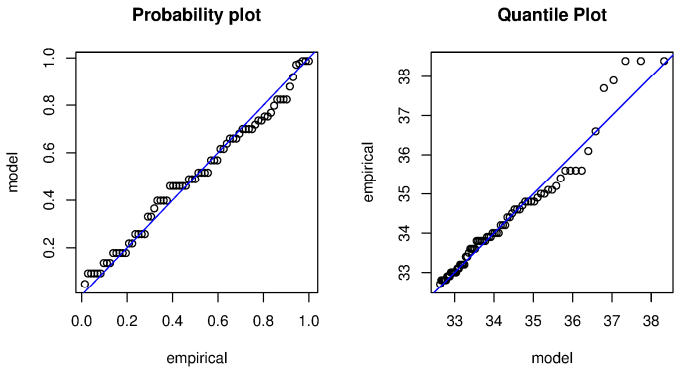


Figure 4.12. Diagnostic plots of stationary PP model fitted to Barcelona daily Tx series from 1982 to 2010.

To know whether SSTs have a statistically significant role in the intensity and frequency of VHD in Spain, as well as identifying the spatial distribution of SST anomalies related to these extreme events, SST anomalies were included into the statistical model.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

At each station, the SST anomaly for every grid box (507 in total) has been introduced as a covariate in the location parameter of the corresponding annual maximum extreme distribution derived by the PP approach in Tx series.

The goodness-of-fit was checked by diagnostic plots, P-P plot and Q-Q plot. For both, the stationary and non-stationary PP model, and for all the grid boxes analysed, the points approximately lie on a line showing a good fit of GPD distribution for the threshold and for the covariate selected. There are not shown for all grid points, only one example in Figure 4.13 is provided, which shows diagnostic plots for the non-stationary model for Barcelona and for a particular grid point of SST anomaly. Comparing these diagnostics with those of the stationary model in Figure 4.12, the goodness-of-fit of the non-stationary model is improved. In these diagnostic plots, it can be seen how the highest quantiles are closer to the diagonal indicating a better fit.

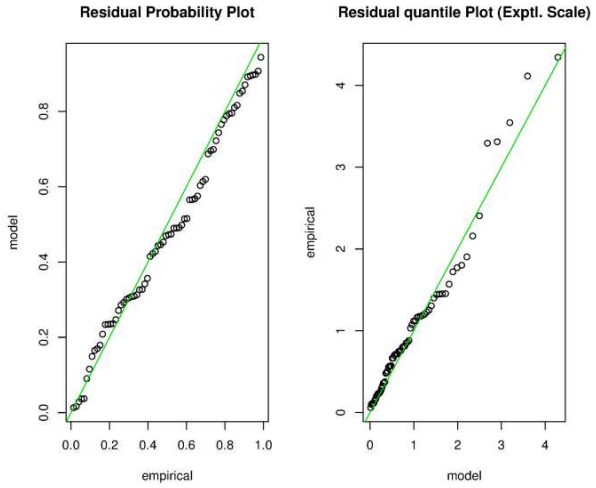


Figure 4.13. Same as Figure 4.12 but for the non-stationary model (SST anomalies of the grid 4.5ºE 41.5N in the location parameter).

To quantify the improvement of the model, plots of the p-values of the likelihood-ratio test have been displayed. The whole p-values plotted give spatial patterns of relationships between SST anomalies and VHD. Generally speaking, when SST anomalies are taking into account, the VHD statistic model improves, especially for coastal locations. However, this effect gradually vanishes towards inland locations of the IP. Hence, the most influenced VHD by SST anomalies are coastal locations of Barcelona, La Coruña and Malaga, while locations with the lowest influence are

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

situated in the southern Spanish Plateau (Madrid and Ciudad Real, along with the northernmost station of Pamplona). According to the spatial distribution of p-values, the locations have been classified into three main groups, depending on how the SST anomalies influence the VHD at their specific location. According to this criterion the stations may have:

A. both Mediterranean and North Atlantic SST anomaly influence
B. mainly North Atlantic SST anomaly influence
C. little SST anomaly influence

The spatial distribution of p-values is presented in Figure 4.14(a-f) for two stations belonging to each group defined above: Barcelona and Malaga, Huelva and Badajoz, Madrid and Pamplona, respectively. The remaining figures for the other locations are presented in the Appendix D. A short description of the characteristics of each group is presented next.

A. **Both Mediterranean and North Atlantic SST anomaly influence**

This group is composed of nine stations for which the spatial distribution of p-values with high probability indicates both Mediterranean and North Atlantic SST anomaly influence on VHD. These stations are located either in the north and the north-east of Spain (San Sebastian, Burgos, Huesca, Zaragoza and Barcelona) and along the central and south of the Mediterranean coast of IP (Valencia, Alicante, Murcia and Malaga). VHD in Barcelona are the most influenced by the Mediterranean and North Atlantic SST anomalies as Figure 4.14a shows. The Mediterranean SST also seems to play an important role on VHD at the stations of Malaga (Figure 4.14b), Valencia and Murcia all of them located along the Mediterranean coast of IP, according to the significant p-values reached (<0.001). The rest of the stations have smaller area of SST anomalies improving the model at the 0.001 significance level. A link between anomalous warming of the Mediterranean SST and the frequency of early summer extreme temperature events in the Euro-Mediterranean was also reported by Carril et al. (2008).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

B. **Mainly North Atlantic SST anomaly influence**

The stations belonging to this group (Huelva, Badajoz, La Coruña, Salamanca and Burgos) are located in the western half of IP, fact that may explain the prevalent influence of North Atlantic SST anomalies on the VHD at these locations. Figure 4.14 (c,d) shows the grid map of p-value spatial distribution for Huelva and Badajoz where it can be seen the clear influence of the North Atlantic SST anomalies as a representative example of this group. However, for the case of La Coruña quite significant influence of western Mediterranean SST anomalies has been found.

C. **Little SST anomaly influence**

According to the results emerging from the p-value analysis, little influence of SST anomalies on VHD has been identified for a group of seven stations. They are located inland in the Spanish southern (northern) plateau as Madrid, Ciudad Real and Albacete (Soria) or in erosion basins (Pamplona and Granada) in northern and southern Spain, respectively. All these stations have the altitude higher than 450 m, some are topographically sunken and surrounded by mountains, as in the case of Pamplona and Granada. The inland situation, the altitude of the station and their topographically isolation might be an explanation for little or negligible influence of SST anomalies on VHD at these locations, based on no significant p-values estimated for them. Figure 4.14 (e,f) shows the grid map of p-values spatial distribution for two representative examples of this group, for Madrid and Pamplona respectively.

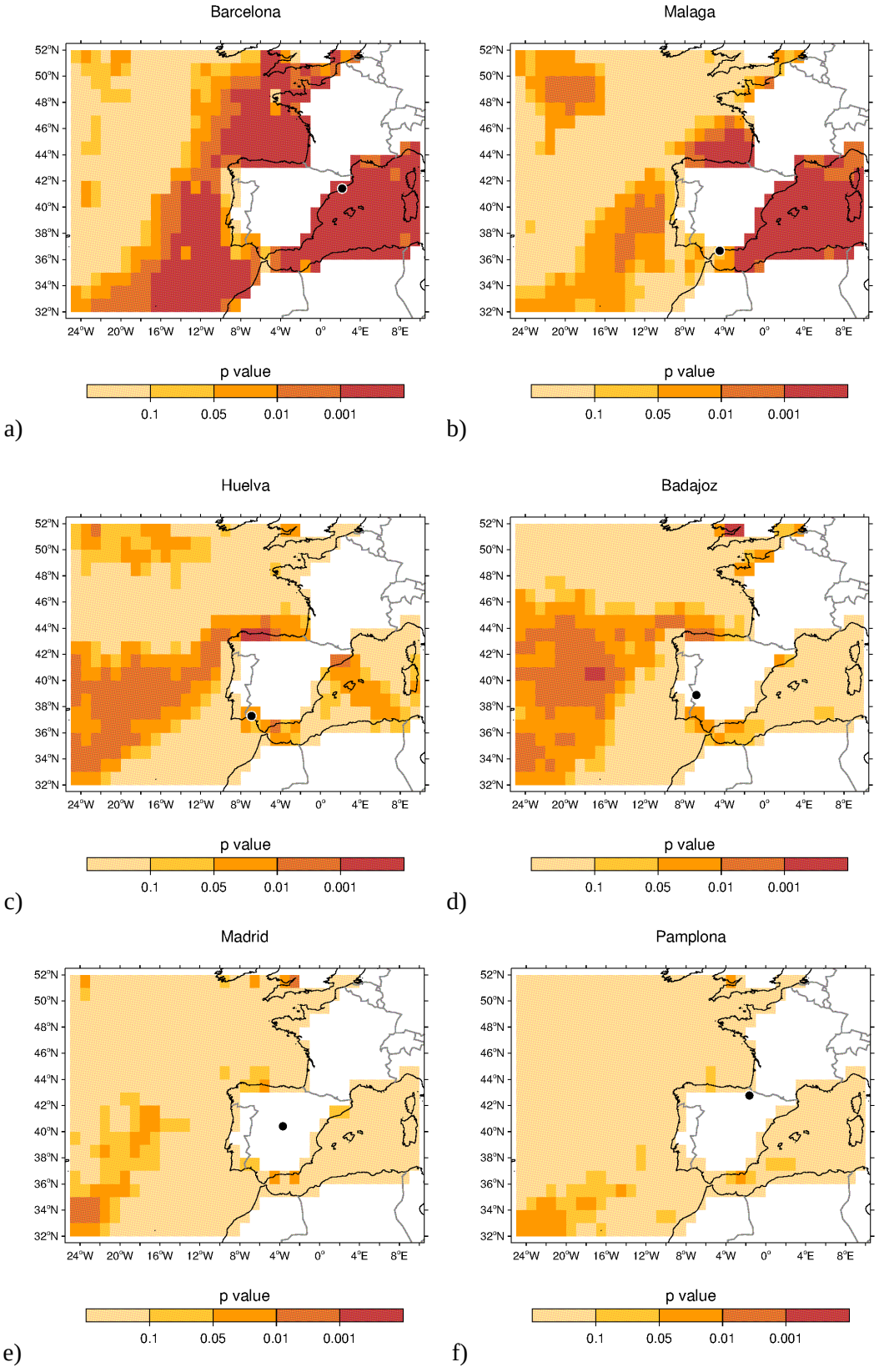


Figure 4.14 (a-f). Gridded maps of p-values between the base model and the model with SST anomalies as covariate in the location parameter of GEV distribution for Barcelona (a), Malaga (b), Huelva (c), Badajoz (d), Madrid (e) and Pamplona (f).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Additional insights on the influence of SST anomalies on VHD at stations in the IP is obtained when analysing the grid maps of p-values in association with the corresponding composite map for each station. The comparative analysis may clarify whether the SST anomalies have a significant influence on VHD at each location.

Composite patterns of positive SST anomalies resembles the grid map patterns of the locations with significant influence (p-values<0.01) of both the North Atlantic and the Mediterranean on VHD, increasing the significance with the amplitude of the anomalies, as it can be seen in the example of Barcelona in Figure 4.14a compared with the composite map Figure 4.11a. Therefore, it is apparent the role of positive SST anomalies in VHD in the most influenced locations.

In contrast, for the locations with little influence of SST anomalies on VHD, as shown in the example provided in Figure 4.14(e,f), the grid maps show few grids with significance, although the composite maps show patterns of positive SST anomalies (Figure 4.11(a-b)). Therefore, although the mean state of SST anomalies for VHD shows positive anomalies, these do not have a significant impact on the frequency and intensity of VHD in these stations may be due to their localization, far from the coast and/or to their altitude and isolation, as it has been mentioned before.

High values of SST anomalies have been reported to be associated with extreme temperature events as for example during the European 2003 heat wave when the Mediterranean SST anomalies were exceptionally strong. The highest anomalies (2ºC to 4ºC) were observed over the north-western Mediterranean (Feudale and Shukla, 2007; 2011).

4.1.3 Influence of SM anomalies on summer VHD

In addition to the effects of anomalous SLP and SST, surface temperatures are also influenced by SM due to the effects in changes of the local surface energy balance, determining the partition of the surface heat flux into sensible and latent components. For example, the lack of SM causes a strong reduction of the evapotranspiration with a subsequent latent flux cooling, which is compensated by the enhanced sensible heat and, hence, forcing surface temperature rise.

SM anomalies were associated to the development of drought episodes (Beguería et al., 2010; Sheffield and Wood, 2008 and Zampieri et al., 2009) over different regions that point out to droughts as the result of pre-existing SM deficits and accumulation of

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

precipitation deficits and/or evapotranspiration excesses (IPCC, 2012). This physical variable has been also related to temperature and precipitation extremes (Jaeger and Seneviratne, 2011) and to intensity and persistence of heat waves (Fischer et al., 2007; Lorenz et al., 2010). Della-Marta et al. (2007) related mean temperatures and heat waves over western Europe to precipitation, which was used as a proxy of SM due to its importance in strengthening land-atmosphere feedback processes. In Spain, a number of studies have used indirect measures of SM to analyse drought conditions using a climatic drought index (Lorenzo-Lacruz et al., 2010; Vicente-Serrano et al., 2010), however, to my knowledge the relationship between SM and the occurrence and intensity of extreme temperatures over Spain is investigated for the first time in this thesis. To complement the investigation on the driving factors of the development of summer VHD over mainland Spain, the possible impact of SM deficit is investigated in this section. The period explored is 1948-2010 because data availability for both variables (daily SM and Tx data) and the spatial window have been bounded to 15º W to 10º E - 31º N to 48ºN.

Analysing VHD composite maps for SM anomalies, the same pattern emerges for all locations. The patterns show a clear deficit in soil water content over IP, south France and North Africa, although there are two parts considerably more arid, with maximum negative anomalies of SM. The first one is centred in the middle of the Ebro Basin where it is located the desert of Bardenas Reales (Navarra) and the Monegros (Aragon) and the other is located in North Africa, over the Algerian Saharan Desert. Figure 4.15. shows one representative example of mean SM anomalies over mainland Spain related to VHD in Barcelona. The other figures, for the rest of the locations, are not shown since the pattern is very similar.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

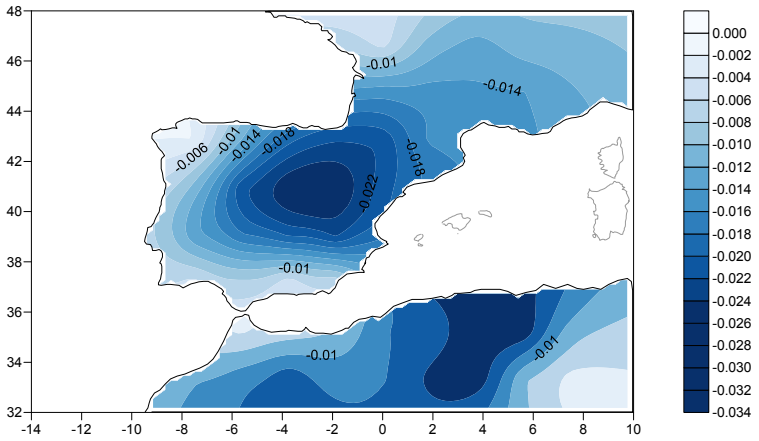


Figure 4.15. Composite map of soil moisture anomalies for summer VHD for Barcelona (1948-2010).

To ascertain whether SM anomalies have influence on the intensity and frequency of summer VHD in the analysed locations, the statistical modelling has been applied. Because the analysed period is the same as the used for SLP anomalies analysis, the thresholds selected correspond to those presented in Table 4.1. The stationary model is the same, as well, and an example of diagnostic plots for Barcelona is shown in (Figure 4.6).

To consider SM anomalies into the model, for each station the anomalies of every grid box have been introduced as a linear covariate in the location parameter of the extreme distribution.

With regard to P-P and Q-Q plots of the fitted non-stationary PP model, they suggest a good fit of the GPD distribution for all the stations. However, difficulty arises when fitting the highest quantiles. As it is evident from the example shown in Figure 4.16 for Barcelona, there are some points away from the diagonal. However, present study is focused on finding out improvement of the model when the explored covariate is taken into account, and this goal has been achieved.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

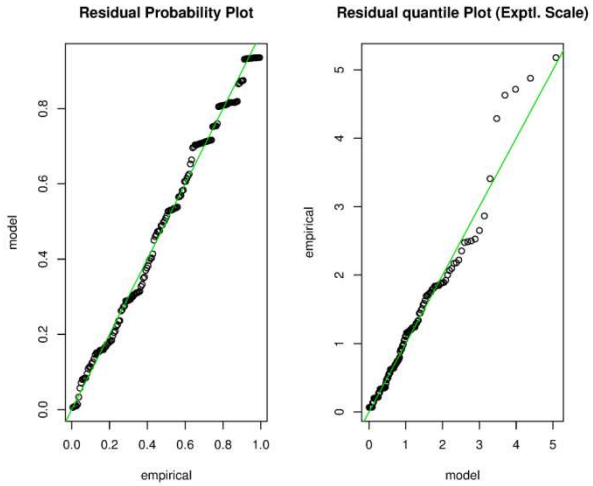


Figure 4.16. Diagnostic plots of PP model fitted to daily maximum temperatures in Barcelona (1948-2010) for the non-stationary model (soil moisture anomalies of the grid 1.9ºW 41ºN in the location parameter of GEV distribution).

The results show that SM anomalies play an important role in the intensity and frequency of VHD in all the analysed locations according to the p-values calculated between the stationary model and the model with the SM anomalies introduced as covariates in the location parameter. Thus, the model improves significantly when SM anomalies are introduced into it for most of the grid points; however, the enhancement is not uniform for all locations, since different significant areas emerge from their analysis. The significance is determined by the likelihood-ratio test and is presented in Figure 4.17(a-i) on a grid point basis at different significance levels for both the most and less influenced locations, while the remaining figures are shown in the Appendix E.

For Zaragoza, Huesca, Pamplona, Burgos, Valladolid and Madrid many grids of SM anomalies improve the model at the significance level of at least 0.001. Therefore, SM anomalies of the study area seem to be related with VHD at these stations.

For other stations along the coastal areas, like Malaga, La Coruña and San Sebastian lower improvement of the model is reached.

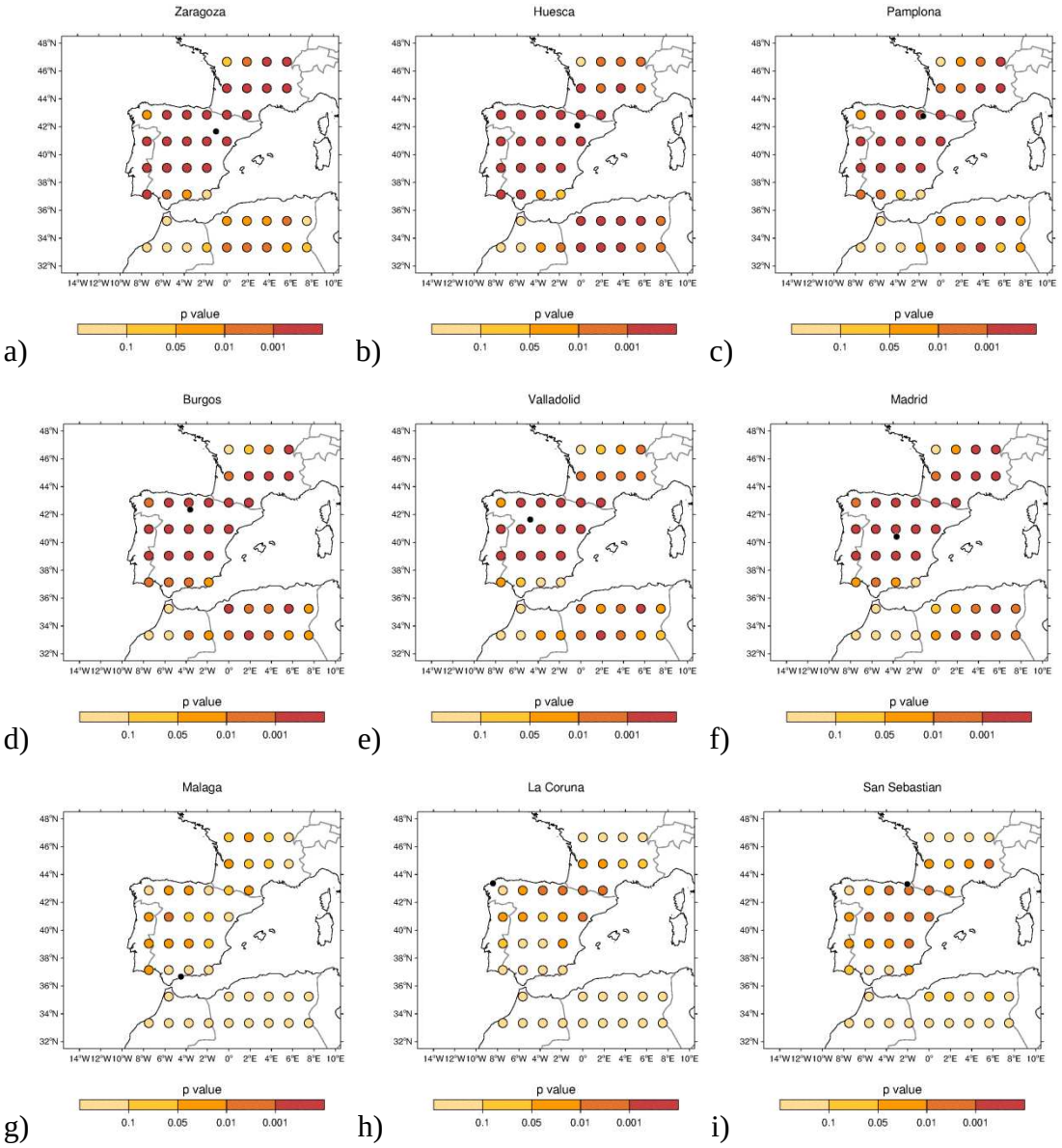


Figure 4.17 (a-i). Maps of p-values between the base model and the model with soil moisture anomalies as covariate in the location parameter of GEV distribution for Zaragoza (a), Huesca (b), Pamplona (c), Burgos (d), Valladolid (e), Madrid (f), Malaga (g), La Coruña (h) and San Sebastian (i).

Figure 4.18 provides an overall view of the relationships between SM anomalies and VHD by showing for each station the lowest p-value obtained among all the grid points analysed.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

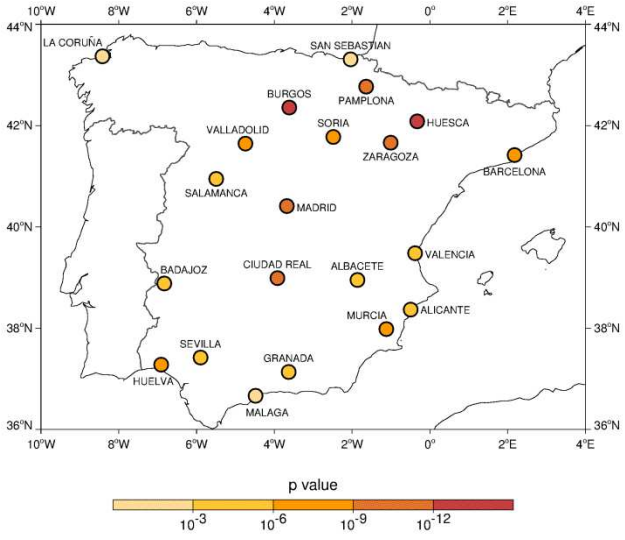


Figure 4.18. Map of the lowest p-values from the likelihood-ratio test for each station analysed (VHD).

VHD in Ebro Basin and central Spain exhibit the most significant relationship with SM anomalies according to the likelihood-ratio test. Besides, the North coast stations and Malaga exhibit the weakest association (though statistically significant). In any case the best relationships are always found with the SM anomalies for grid points within the IP. The importance of soil moisture-climate interactions on European temperature extremes was shown by Jaeger and Seneviratne (2011). They proved the role of SM both in the persistence and in the absolute values of extreme temperatures. Fisher et al. (2007) indicated that spring SM anomalies play an important role in the evolution of European summer heat waves and suggested that SM may strongly amplify temperature anomalies in an extreme summer such as in 2003. Brabson et al. (2005) and Lorentz et al. (2010) also related heat waves to SM deficit. In particular Brabson et al. (2005) examined the relationships between temperature extremes and SM in eastern England and pointed out that all the occurrences in the upper tail of temperature distribution took place in periods characterized by SM deficit over eastern England. In addition, the author found longer spells of temperature extremes will arise from both the statistical increase in extremes frequency and the extended periods of low SM.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

4.2 Very Hot Nights

This section is divided in three subsections and presents the results of the influence of SLP, SST and SM anomalies on VHN in the 21 stations analysed. The same procedure as that carried out for VHD have been used.

4.2.1 Influence of SLP anomalies on summer VHN

In this subsection the relationship between SLP anomalies and VHN for the period (1948-2010) has been analysed and the findings have been described and discussed.

The SLP patterns identified for VHN resemble those identified for VHD, namely: Southerly, Weak South-westerly and North-westerly Airflow Patterns. Although they show similar configurations there are slight differences with respect to the spatial distribution and amplitude of SLP anomalies. Next, the description of these patterns is presented and two representative examples for each one are shown in Figure 4.19-Figure 4.23. The composites for the remaining stations are presented in the Appendix F.

1. The Southerly Airflow Pattern for VHN

This pattern shows a large area of SLP positive anomalies located over central Europe with a strong southward extension along the Mediterranean Sea towards North Africa, which in some cases is divided into two centres. Associated with this large area of SLP positive anomalies, SLP negative anomalies are located over most of the North Atlantic Ocean and the western half of the IP. Comparing with the configurations for VHD, these ones are similar but negative anomalies are in general more extended from the Atlantic Ocean toward the IP.

Likewise for VHD, three different parts of Spain (Inland, North-western and North) are affected for this configuration.

- Inland Spain

The configuration returns dominant warm southerly component airflow that affects the stations in this area: Madrid, Ciudad Real, Badajoz, Salamanca, Sevilla and Huelva. This pattern is illustrated in Figure 4.19(a,b) for Madrid and Badajoz. In these figures it is evident the dry and warm southerly airflow towards Madrid, and south-easterly airflow towards Badajoz, respectively.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

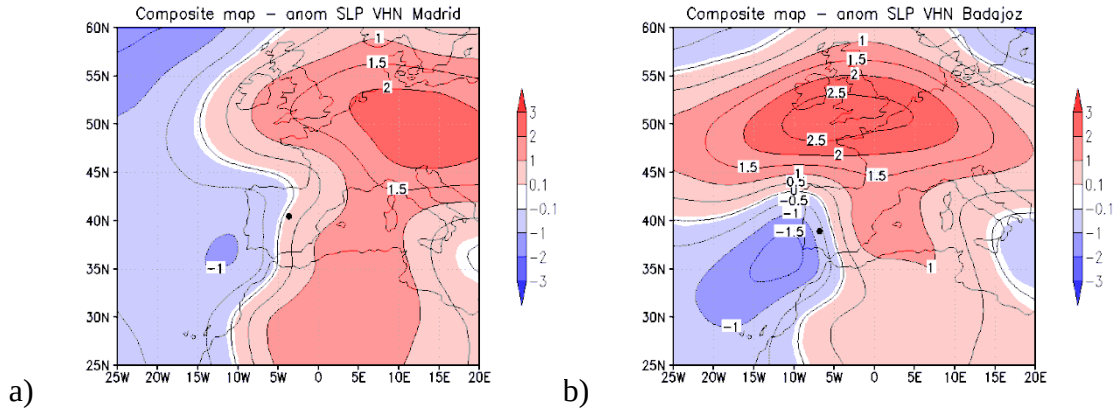


Figure 4.19 (a-b). Pattern 1. SLP anomalies (hPa) of summer VHN for Madrid (a) and Badajoz (b).

This patterns of influence corresponds well with the patterns found by (García-Herrera et al., 2005) for Tn>95th percentile during extreme hot events in Madrid.

- North-eastern Spain

Likewise VHD configuration, for the north-eastern locations of Zaragoza, Soria, Huesca and Barcelona the negative SLP anomalies from the North Atlantic are extended over the IP modulating the flow on a south-westerly component as shown in Figure 4.20a for Barcelona and Figure 4.20b for Soria.

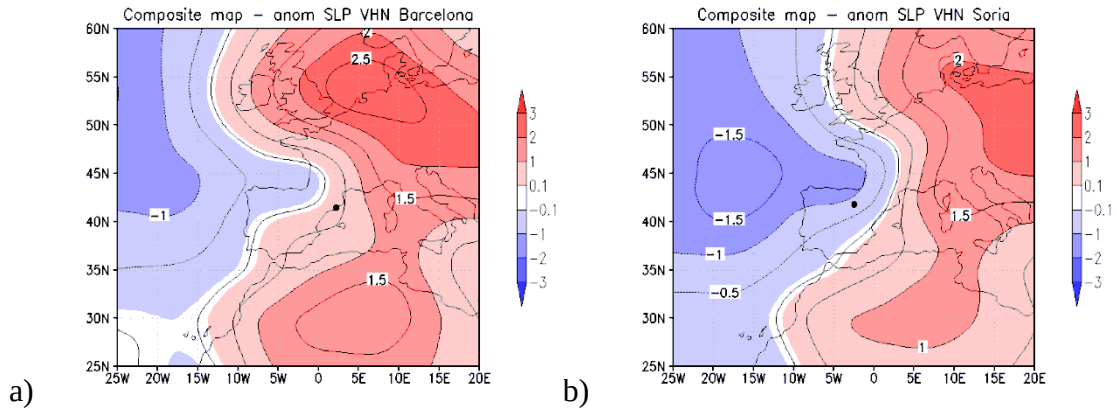
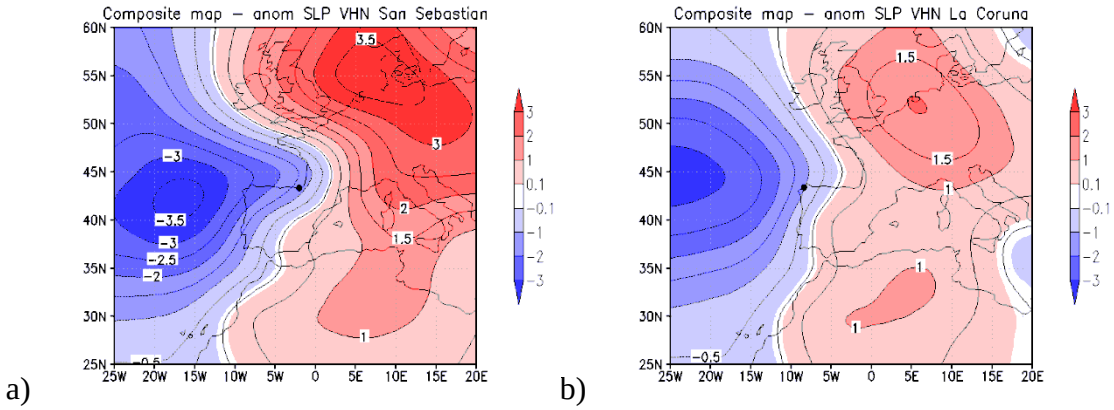


Figure 4.20 (a-b). Same as Figure 4.19 (a-b) but for Barcelona (a) and Soria (b).

- North Spain

The pattern of SLP anomalies affecting Burgos, Valladolid, Pamplona, San Sebastian and La Coruña presents a centre of negative SLP anomalies in the North Atlantic Ocean and a large area of positive SLP anomalies centred over north-western Europe which are stronger than in the case of the stations situated Inland and North-eastern Spain. This is particularly true for Valladolid and San Sebastian where positive

anomalies of ~3.5hPa are observed (Figure 4.21a), although less intense anomalies are found for La Coruña (Figure 4.21b). The dipole-like pattern of negative SLP anomalies along with positive SLP anomalies favours the intensification of a south-westerly component flow over the IP.

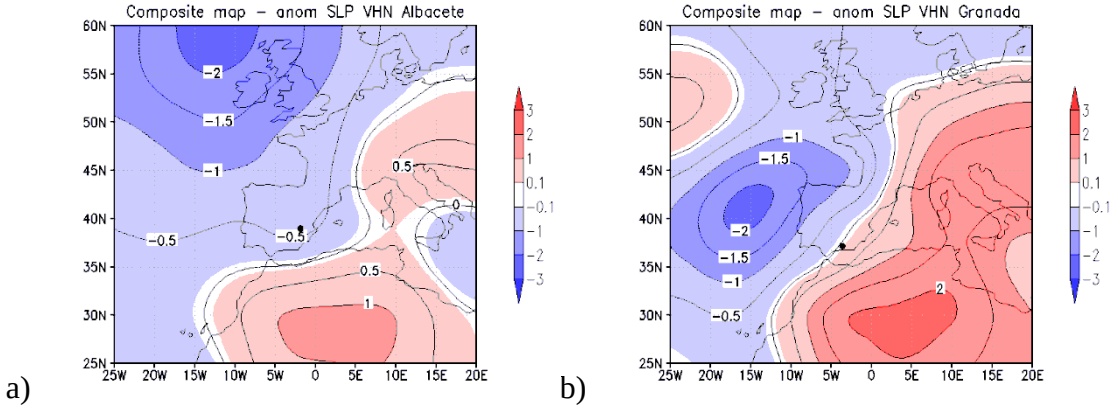


a) b)

Figure 4.21 (a-b). Same as Figure 4.19 (a-b) but for San Sebastian (a) and La Coruña (b).

2. Weak South-westerly Airflow Pattern for VHN.

Moderate positive anomalies centred over north Africa, and expanded across the Mediterranean Sea to the Eastern Europe, and a large area of negative anomalies centred over northern coast of the British Isles (Figure 4.22a) or over Azores islands (Figure 4.22b) characterize this pattern. It allows a weak westerly or south-westerly flow to influence the stations of Albacete and Granada. The patterns of SLP anomalies associated with VHN are very similar to those found for VHD for these locations.



a) b)

Figure 4.22 (a-b). Pattern 2. SLP anomalies (hPa) of summer VHN for Albacete (a) and Granada (b).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

3. North-westerly Airflow Pattern for VHN

An area of weak positive SLP anomalies centred over North Africa associated with a large area of moderate negative anomalies centred over the IP or over the British Isles characterizes this pattern of SLP anomalies associated with VHN. This pattern enables rather calm conditions on the locations over mid and south Mediterranean coast of the IP (Valencia, Alicante, Murcia and Malaga).

This pattern is similar in terms of the spatial distribution of anomalies found for VHD for these locations but with weaker negative anomalies, which goes along with weaker airflow. Next, it is shown two examples for Valencia (Figure 4.23a) and Malaga (Figure 4.23b).

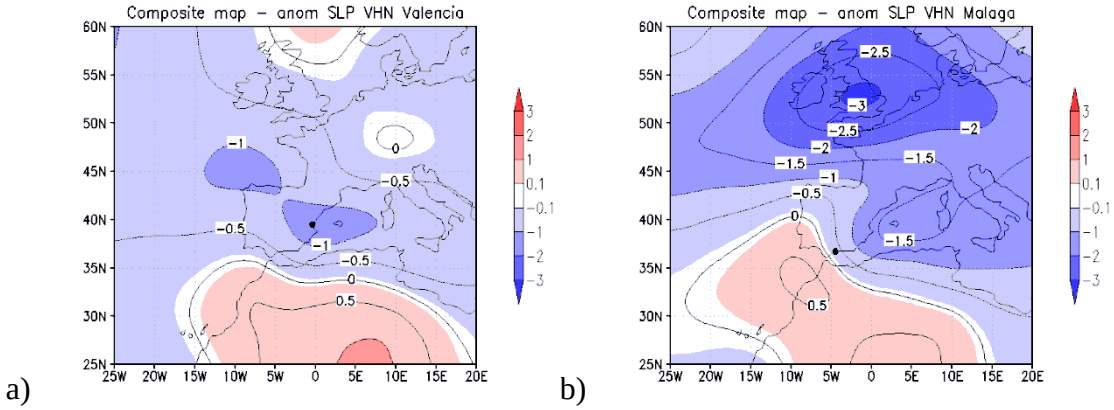


Figure 4.23 (a-b). Pattern 3. SLP anomalies (hPa) of summer VHN for Valencia (a) and Malaga (b).

Once the mean SLP anomalies associated with summer VHN have been identified, the statistical modelling has been performed to verify relationships between SLP anomalies and Tn extreme events. Table 4.3 shows the thresholds selected according to the techniques presented in section 3.3.2. The stationary model has been set up using daily Tn data after declustering. Figure 4.24 shows diagnostic plots namely P-P plot and Q-Q plot for Barcelona, as an example. As the plots show, points are close to the unit diagonal, indicating the suitability of the MLE of GPD fitted for the selected threshold. Then, the non-stationary model has been performed assuming a linear dependence of SLP anomalies for each grid box in the location parameter estimated for the extreme distribution. The diagnostic plots of the non-stationary PP model fitted to Tn (not show) present the distribution of the points near the unit diagonal in all grids, standing for a good fit of the GPD, as well. Figure 4.25 shows an example of diagnostic plots for the

non-stationary model fitted for Barcelona for a grid point (2.5ºE 30ºN) with high significant p-value.

Table 4.3. Best thresholds selected for daily Tn and their percentile for the period 1948-2010.

Station	Threshold (ºC)	Percentile (%)
Albacete	19.4	95.0
Alicante	23.1	95.0
Badajoz	20.8	95.0
Barcelona	22.4	95.0
Burgos	15.3	96.1
Ciudad Real	22.3	95.8
Granada	20.4	95.0
Huelva	21.4	95.0
Huesca	20.5	95.7
La Coruña	18.2	95.8
Madrid	22.2	95.0
Malaga	23.4	95.0
Murcia	23.0	95.0
Pamplona	18.0	95.3
Salamanca	16.4	95.0
San Sebastian	18.8	95.0
Sevilla	23.0	95.0
Soria	15.8	95.0
Valencia	23.6	95.0
Valladolid	17.6	95.3
Zaragoza	21.1	95.0

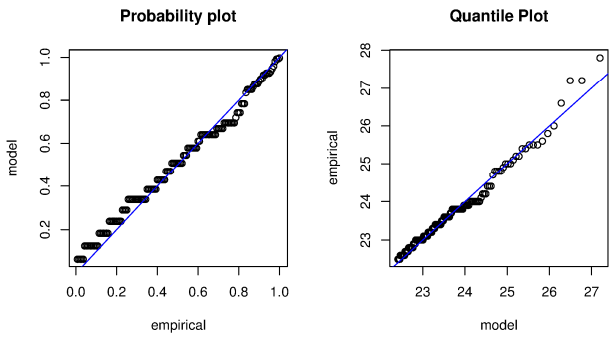


Figure 4.24. Diagnostic plots of stationary point process model fitted to Barcelona daily minimum temperature (Tn) series from 1948 to 2010.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

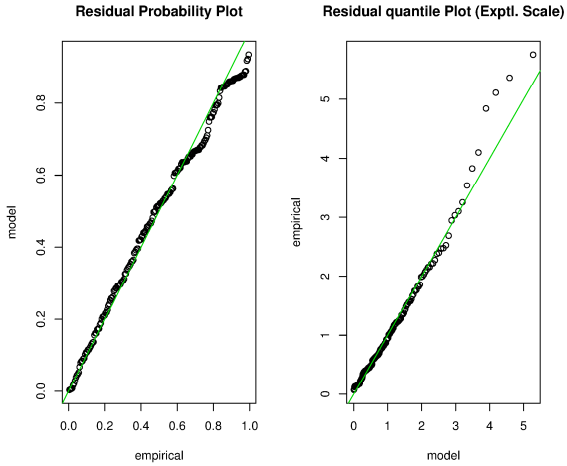


Figure 4.25. Same as Figure 4.24 but for the non-stationary model (SLP anomalies of the grid 2.5ºE 30ºN as covariate in the location parameter of the GEV distribution).

To assess the improvement of the model when SLP anomalies are taken into account, the likelihood-ratio test has been calculated between the stationary and the non-stationary model. Grid plots of p-values have been displayed to visualize regions with significant improvement. In Figure 4.26, Figure 4.27 and Figure 4.28 the same examples chosen to illustrate each pattern and sub-pattern with the composite maps have been presented to show the grid maps of p-values distribution. The rest of the figures are presented in the Appendix G.

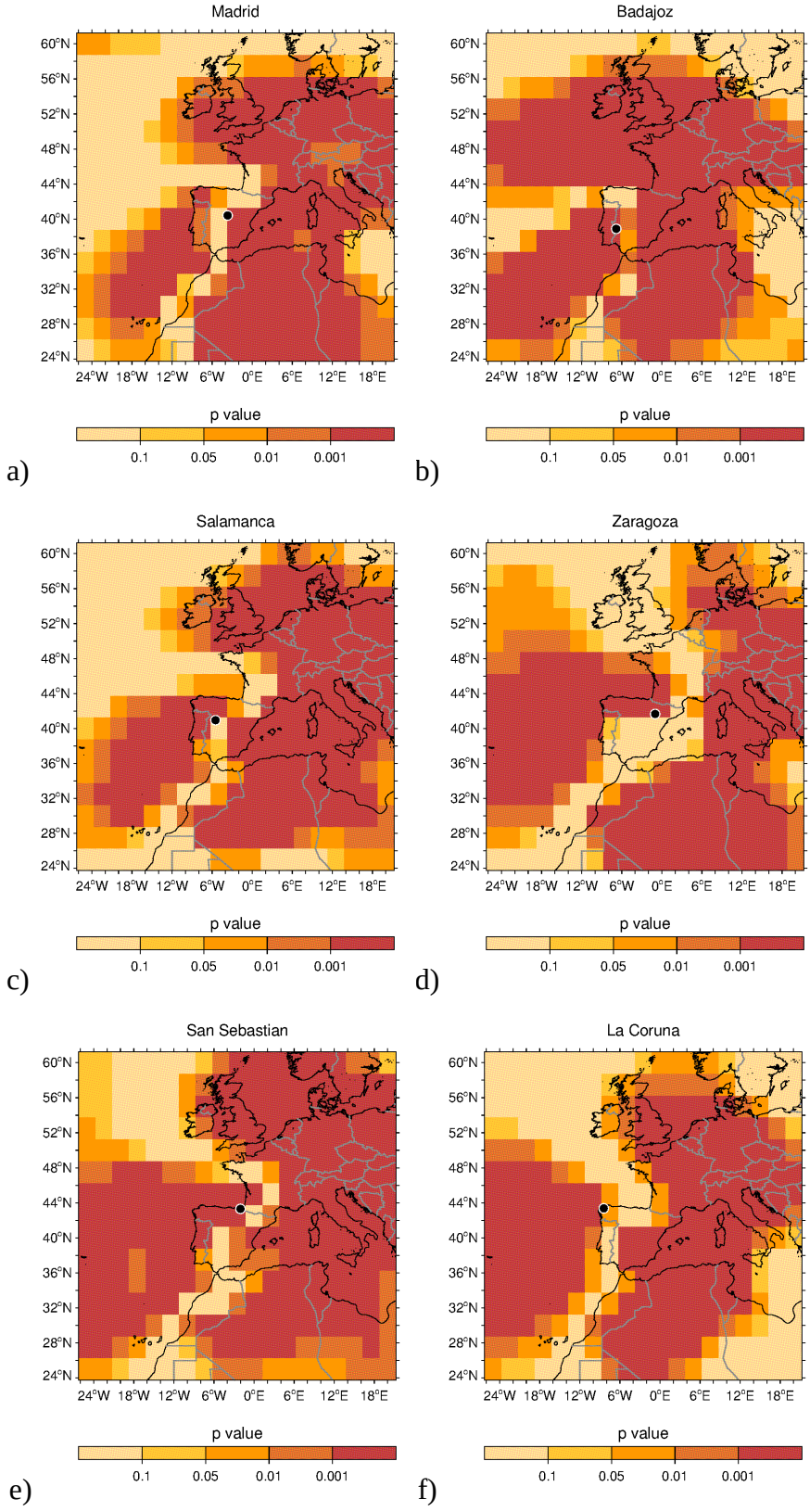
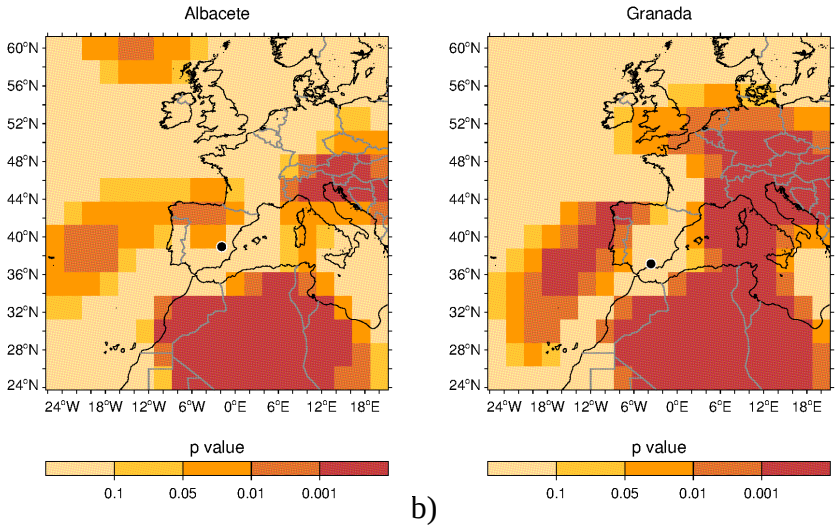


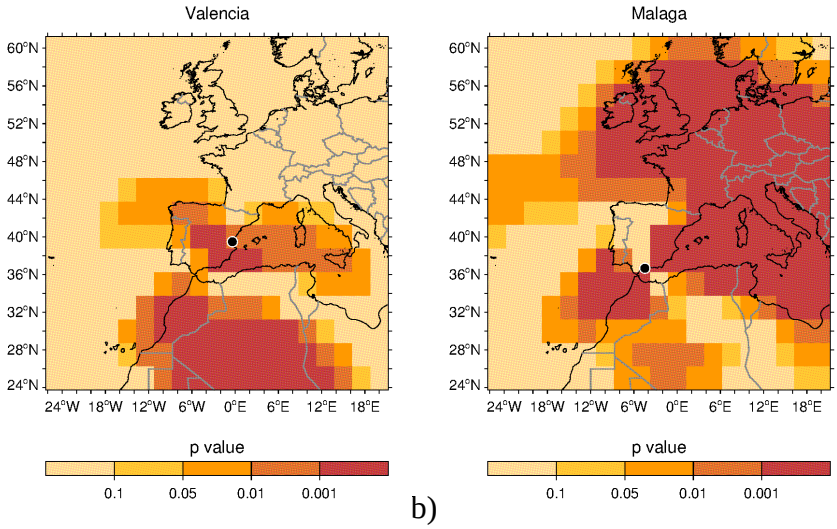
Figure 4.26 (a-f). Gridded maps of p-values between the base model and the model with SLP anomalies as a covariate in the location parameter of the GEV distribution for Madrid (a), Badajoz (b), Salamanca (c), Zaragoza (d), San Sebastian (e) and La Coruña (f).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



a) b)

Figure 4.27 (a-b). Same as Figure 4.26 (a-f) but for Albacete (a) and Granada (b).



c) b)

Figure 4.28 (a-b). Same as Figure 4.26 (a-f) but for Valencia (a) and Salamanca (b).

Likewise the results presented for VHD, relationships between VHN in the 21 explored locations and large-scale SLP anomalies have been analysed. The stationary model improves significantly at the 0.001 level for grids where higher SLP anomalies, both positive and negative, have been identified. Analysing the areas with small p-values in association with the corresponding composite maps, it can be seen they are in good agreement with the highest anomalies of the large-scale patterns of SLP anomalies.

4.2.2 Influence of SST anomalies on summer VHN

The same methodology used in section 4.1 has been employed to investigate possible influences of SST anomalies on VHN. Equally, the assessment is done for summer months (JJA) and for the period 1982-2010 due to the availability of the daily SST data.

For the domain 25ºW-10ºE, 32ºN-52ºN, composite maps of mean summer VHN have been computed. They show, in general, the same pattern for all the locations, but exhibits stronger anomalies over the Western Mediterranean Sea for the locations of Valencia, Murcia and Zaragoza, where it has been estimated around two degrees of positive anomalies. In general, this pattern shows the highest positive anomalies in the Western Mediterranean Sea, strong positive anomalies in the North Atlantic Ocean and coastal areas of the Cantabrian Sea and British Isles. The lowest negative anomalies have been identified over the Portuguese coast of the Central Atlantic Ocean.

Comparing with the corresponding composite maps for VHD (e.g. Figure 4.11b), in the case of VHN, the Atlantic Ocean shows a more extended area of anomalously high SST values and, in addition, the Mediterranean Sea is hotter during VHN events over mainland Spain except for Salamanca, Huesca and Barcelona.

Two composite maps, for Valencia and Madrid, are shown in the Figure 4.29 (a-b) as representative examples of mean state of SST anomalies associated to summer VHN. Figures for other locations are given in the Appendix H.

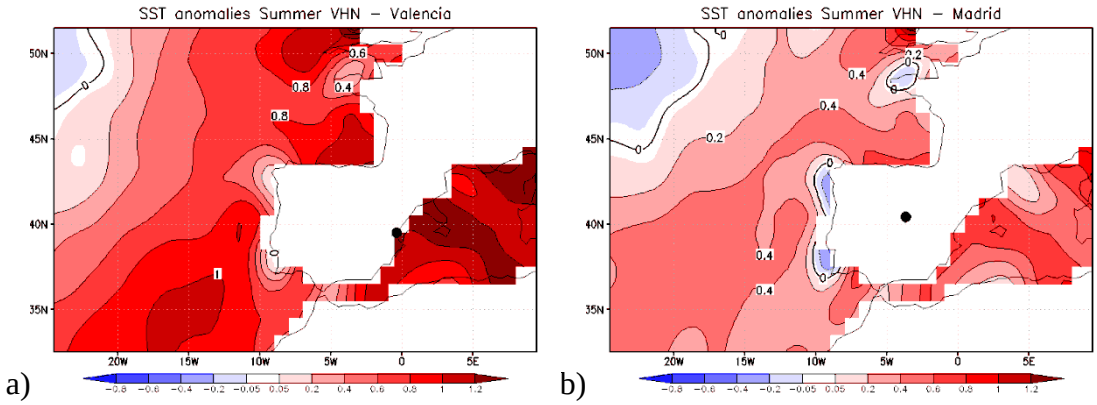


Figure 4.29 (a-b). Composite maps of SST anomalies of SST for summer VHN in Valencia (a) and Madrid (b).

Composite maps show the mean state of SST anomalies of VHN during summer months and the statistical model, considering these anomalies can explain relationships

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

between SST anomalies and Tn extremes. Therefore, next the statistical modelling of VHN has been carried out.

To deal with statistical analysis, first, it is has been selected the thresholds (Table 4.4) for each location following the methodology presented in section 3.2.2. Then, the stationary model has been set up using the declustered daily Tn (see section 3.2.3).

Table 4.4. Best thresholds selected for daily Tn and their percentile for the period 1982-2010.

Station	Threshold (ºC)	Percentile (%)
Albacete	20.0	95.0
Alicante	23.6	95.0
Badajoz	21.4	95.0
Barcelona	23.0	95.0
Burgos	15.4	95.0
Ciudad Real	22.6	95.0
Granada	21.0	95.0
Huelva	22.0	95.0
Huesca	21.2	96.0
La Coruña	18.3	95.2
Madrid	22.8	95.0
Malaga	24.0	95.0
Murcia	23.6	95.1
Pamplona	18.4	95.0
Salamanca	16.8	95.4
San Sebastian	19.4	95.8
Sevilla	23.8	95.0
Soria	16.0	95.0
Valencia	24.1	95.4
Valladolid	18.2	95.0
Zaragoza	21.6	95.0

Figure 4.30 shows one example of diagnostic plots, probability plot (P-P plot) and quantile plot (Q-Q plot) for Barcelona. For all the stations (not shown, but as shown in the example Figure 4.30), the expected values are close to the observed ones and are approximately linear, indicating suitability of the MLE of GPD fitted from a PP approach for the selected threshold. Next, the non-stationary model has been computed assuming the SST anomaly of each grid box behaves linearly in the location parameter estimated from the extreme distribution. P-P plots and Q-Q plots of the non-stationary PP model fitted to declustered Tn data have the slope near the unit diagonal in all grid points. Therefore, the expected values are close to the observed ones, meaning a good fit of the GPD, as well. Figure 4.21 shows one example of diagnostic plots for the non-stationary model for Barcelona and for a particular grid point of SST anomaly.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

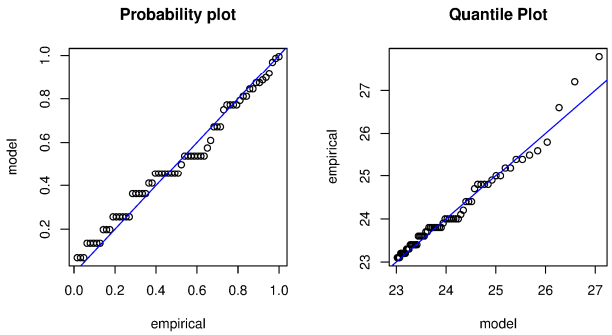


Figure 4.30. Diagnostic plots of stationary point process model fitted to Barcelona daily minimum temperature (Tn) series from 1982 to 2010.

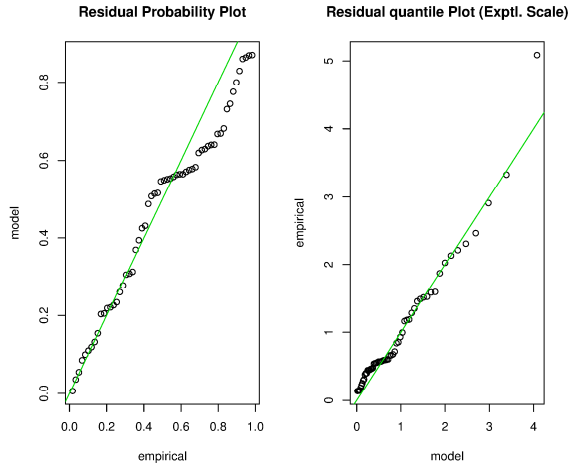


Figure 4.31. Same as Figure 4.30 but for the non-stationary model (SST anomalies of the grid 7.5ºW 36.5ºN in the location parameter).

To quantify the improvement of the model when SST anomalies are taken into account, it has been applied the likelihood ratio test (section 3.2.5) and results are displayed in grid plots to identify areas of SST anomalies related to VHN. According to the p-values VHN are influenced by SST anomalies in all the analysed locations, since many grids have significantly improved the stationary model. With the exception of Barcelona, the whole area of SST anomalies related to VHN is larger than for VHD, for both coastal and inland sites.

The locations have been classified into three groups, as follows:

D. both Mediterranean and North Atlantic SST anomaly influence
E. mainly North Atlantic SST anomaly influence
F. mainly Mediterranean and Cantabrian SST anomalies influence

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

In Figure 4.32 (a-f) are depicted two examples for each group, namely Valencia and La Coruña for group D, Badajoz and Huelva for group E and Zaragoza and Huesca for group F. The remaining figures for other locations are presented in the Appendix I.

D. Both Mediterranean and North Atlantic SST anomaly influence

This group is composed of the locations situated on south-eastern Spain, namely Ciudad Real, Albacete, Valencia, Alicante, Murcia, Malaga, Granada and Sevilla, along with locations situated in north Spain, as La Coruña, San Sebastian and Burgos. SSTs of the Mediterranean Sea and the Atlantic Ocean are related to VHN of these locations, in agreement with the results returned by testing one model against the other. The relationships are not equally for all the locations. The most influenced is Valencia followed by Burgos and La Coruña, and the less influenced is Granada. Two examples are shown in Figure 4.32(a,b) for Valencia and La Coruña.

E. Mainly North Atlantic SST anomaly influence

This group is composed of locations over western Spain: Badajoz, Huelva, Valladolid and Salamanca, along with the central location of Madrid. In these places there is clear the influence of the Atlantic Ocean and there is no influence or very little from the Mediterranean Sea. Figure 4.32(c,d) show two representative grid map of p-values for Badajoz and Huelva.

F. Mainly Mediterranean and Cantabrian SST anomalies influence

This group is composed of locations situated in north-eastern Spain: Barcelona, Zaragoza, Huesca, Soria and Pamplona. VHN in these places are mainly influenced by the Mediterranean and Cantabrian Seas, although Pamplona and Zaragoza are also affected by the Atlantic Ocean. Barcelona is more influenced by the Mediterranean Sea than by the Cantabrian Sea. In Figure 4.32(e,f) it can be seen two representative examples for Zaragoza and Huesca.

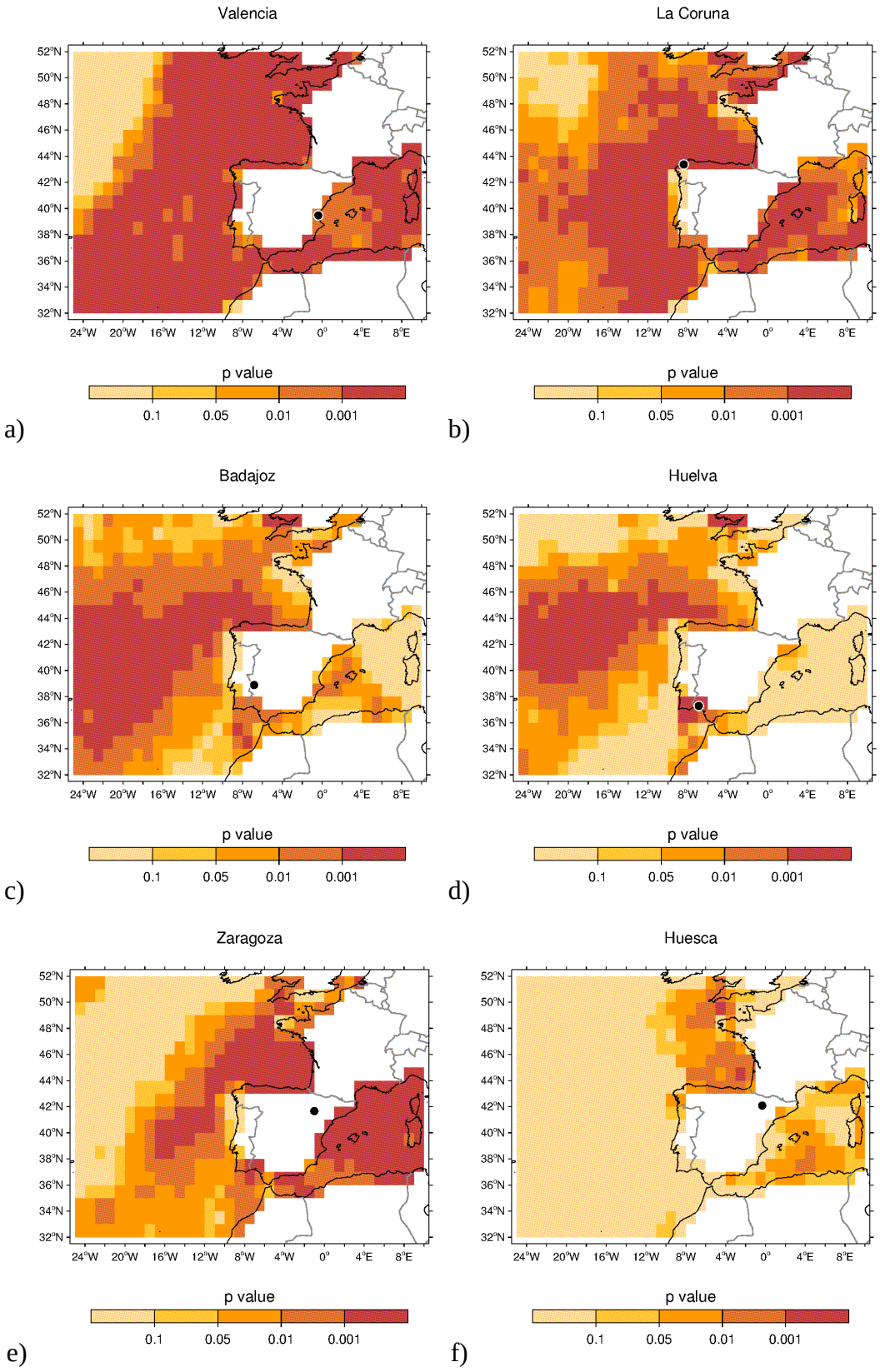


Figure 4.32(a-f). Gridded maps of p-values between the base model and the model with anomalies of SST as covariate in the location parameter for Valencia (a), La Coruña (b), Badajoz (c), Huelva (d), Zaragoza (e) and Huesca (f).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

4.2.3 Influence of SM anomalies on summer VHN

In this sub-section potential influence of SM anomalies on summer VHN is investigated, as in the sub-section 4.1.3 for VHD. First, it has been described the climatological behaviour of SM anomalies during the most extreme summer Tn events and, second, it has been investigated the possible links between SM anomalies and summer VHN by applying the PP approach.

The results from the analysis of the SM anomalies composite maps associated with summer VHN provide a general pattern with negative anomalies over the IP, south France and North Africa for all examined locations. This pattern is very much alike as the corresponding to VHD, highlighting two areas of strong negative SM anomalies, one in the desert of Bardenas Reales (Navarra) and the Monegros Desert (Aragon), and the other over the Algerian Sahara Desert.

In Figure 4.33, a representative composite map for SM anomalies of VHN in Barcelona is introduced. SM composites for the rest of the locations are the same (not shown).

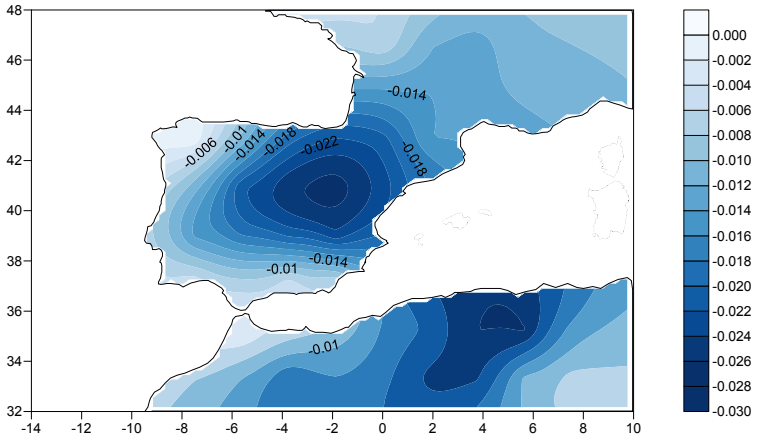


Figure 4.33. Composite map of soil moisture anomalies of summer VHN for Barcelona (1948-2010).

The statistical modelling has been carried out to look for relationships between SM deficit during VHN and the extreme Tn values recorded.

First, the best thresholds, ranging between 95th and 99.5th percentiles, have been selected (Table 4.3). They are the same as those used for analysing SLP anomalies, and the stationary model fitted to the declustered data is therefore the same too (Figure 4.24). Second, non-stationarity is inserted into the model by assuming the location

parameter of the extreme distribution has a linear dependence with SM anomalies. The model has been performed for each grid of SM anomaly and for each station.

Figure 4.34 presents one example of diagnostic plots of the non-stationary model for Barcelona.

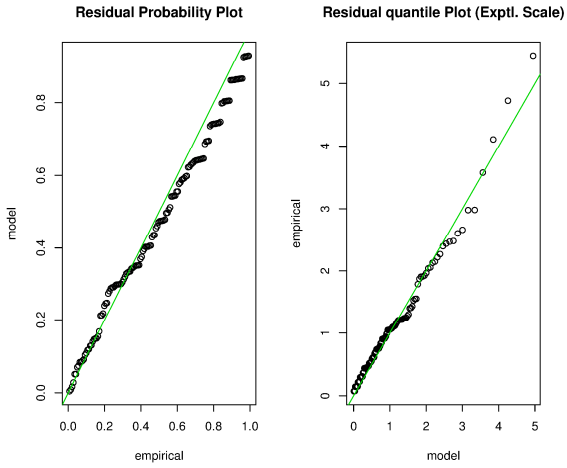


Figure 4.34. Diagnostic plots of PP model fitted to daily minimum temperatures in Barcelona (1982-2010) for the non-stationary model (Soil moisture anomalies of the grid 1.88ºE 35.2ºN in the location parameter).

In both cases, the points are close to the unit diagonal, which stands for a good fit of the GPD fitted for the selected threshold and, also, for the covariate explored in the non-stationary case.

When SM anomalies are taken into account, results suggest a statistically significant relationship between SM anomalies in the studied domain and VHN for the 21 selected locations. Figure 4.35(a-i) displays grid maps of the p-values derived from the test applied between the stationary and the non-stationary models for the most and for the less influenced locations by SM anomalies. The grid maps for the rest of the locations are compiled and shown in the Appendix J.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

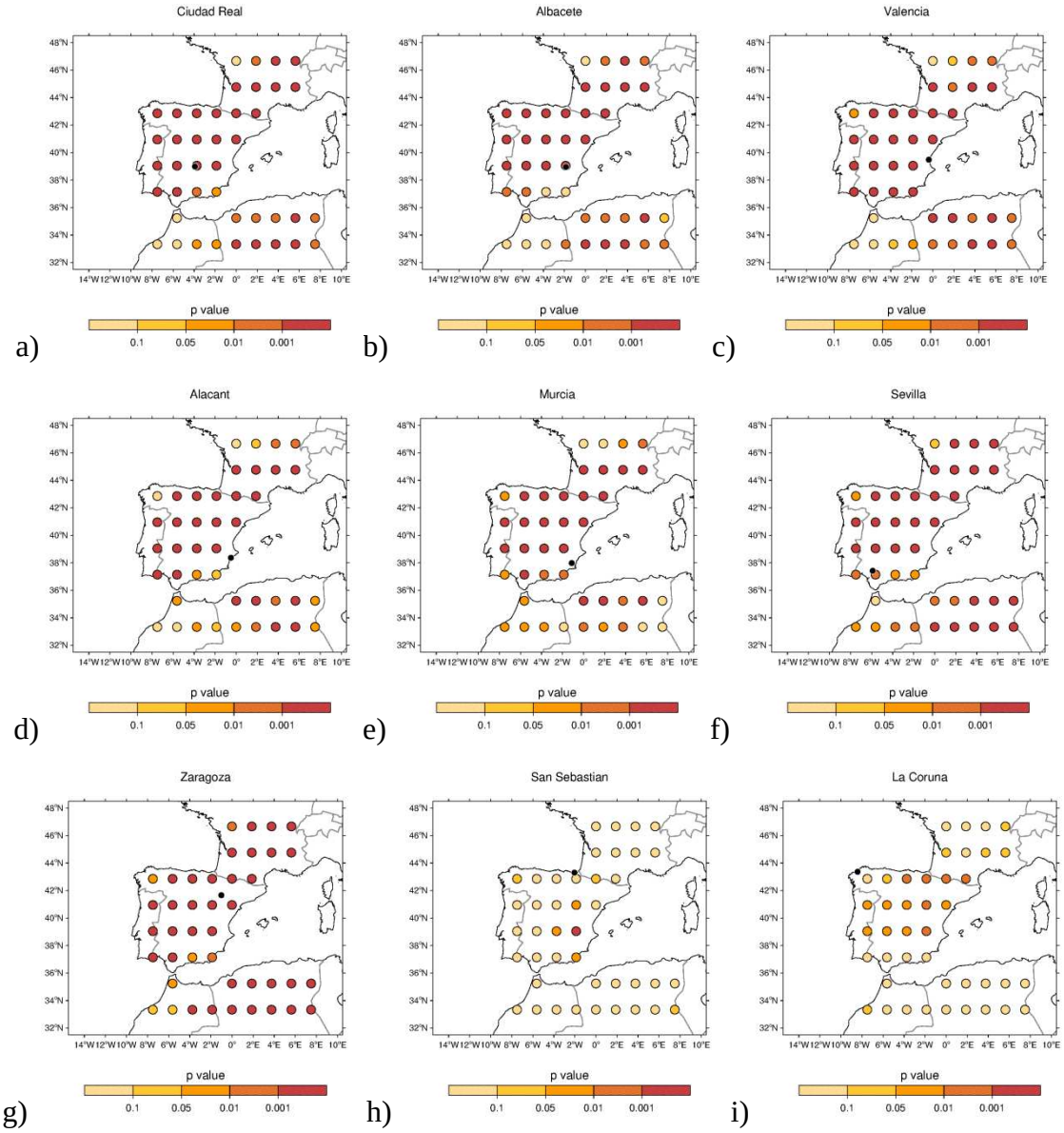


Figure 4.35 (a-i). Maps of p-values between the base model and the model with soil moisture anomaly as a lineal covariate in the location parameter in Ciudad Real (a), Albacete (b), Valencia (c), Alicante (d), Murcia (e), Sevilla (f), Zaragoza (g), San Sebastian (h) and La Coruña (i).

According to the p-values given by the likelihood-ratio test, the most influenced locations are situated over south-eastern Spain (Ciudad Real, Albacete, Valencia, Alicante, Murcia, Malaga and Sevilla), along with Zaragoza located in north-eastern Spain. In these locations, the SM deficit over the IP, south France and North Africa (mainly over Algeria) improve the stationary model at least at the 0.001 significance level. As for VHD, for the stations in Spain´s northern coast, i.e. La Coruña and San Sebastian, VHN are less affected by SM anomalies.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

It should be noted that the NCEP/NCAR-reanalysis field for SM have not been estimated by assimilating in situ SM data (Kalnay et al., 1996). Therefore the data analysed should rather represent an estimate of the larger-scale component of SM field. The connections of eastern Spain station extremes with Algerian SM should be interpreted in the context of the large-scale spatial correlation of SM field. A similar interpretation could be given for the apparent relations with SM in Southern France.

As for the study of VHD, Figure 4.36 shows the lowest p-value obtained among all the SM grid points analysed for the VHN of each station. These results indicate that VHN in south-eastern Spain present the most significant relationship with SM anomalies. Again, the best relationships are found with the SM anomalies for grid points within the IP.

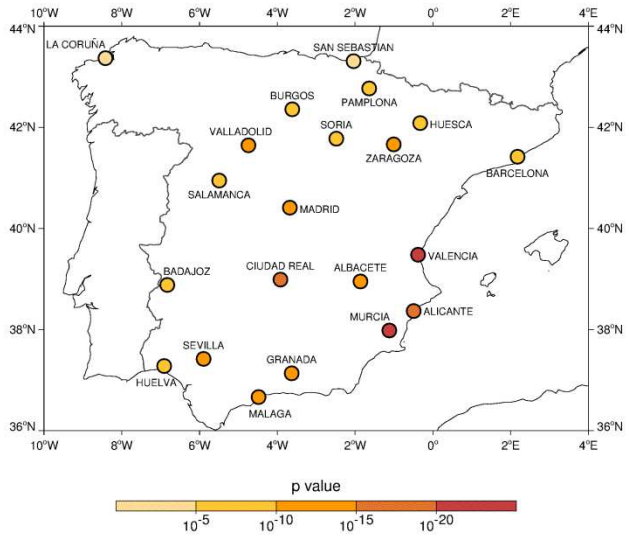


Figure 4.36. Map of the lowest p-values from the likelihood-ratio test at each station analysed (VHN).

For both VHD and VHN, the overall results from the analysis suggest that such a relationship between the extreme temperatures and SM anomalies do exist, in general, for all locations, although it appears to be weak for San Sebastian and La Coruña (both located within the most humid zone of Spain). Nevertheless, the geography of the best relationships differs between VHN and VHD. Future research using additional SM datasets (with in-situ measurements) could shed further light onto the temperature extremes–soil moisture relationship and its robustness.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

4.3 Observed time trends in extreme temperatures

This section is devoted to study the non-stationarity of the Tx and Tn extreme distributions. 1) A non-stationary model assuming a linear trend in the location parameter μ and both the scale σ and shape ξ parameters maintained constant has been tested against the stationary model. 2) A non-stationary model assuming a linear trend in both the location μ and scale σ parameters and maintaining constant the shape ξ parameter has been tested against the non-stationary model with a linear trend only in the location parameter μ. The hypothesis of time linear dependence of these parameters is important because any decrease or increase in μ shifts the temperature extreme distribution toward lower or higher values, which implies reduction or increasing in the occurrence and intensity of all extremes, equally. Additionally, decreases or increases in σ imply a reduction or an increasing of the temporal variability of extremes.

This analysis has been carried out using daily Tx and Tn data for two different periods: 1940-1972 and 1973-2010, respectively. The selection of these periods is based on the findings of Brunet et al. (2006, 2007a,b), who identified for mainland Spain the year 1973 as a breakpoint in temperature series, when an unprecedented and strong rise in temperatures took place. Therefore, this year has been considered in this study as a change point for temperature evolution.

The thresholds of temperature excesses for daily Tx and Tn have been selected according to the methodology described in section 3.3.2 around the 95th percentile, for both periods. It is worth noting that the range of the Tx threshold varies between 25.9ºC-40.4ºC and 26.7ºC-41.0ºC for the periods 1940-1972 and 1973-2010, respectively. As for Tn, the thresholds vary between 14.7ºC-22.6ºC and 15.4ºC-24.1ºC, respectively for the same periods. It is apparent the increase of the thresholds in the second period, both for Tx and Tn.

4.3.1 Observed trend in the location parameter (1940-1972)

For the earlier period 1940-1972, it has been tested the non-stationary model assuming a linear trend in the location parameter μ against the stationary model. The results show significant negative trend in the location parameter of the Tx extreme distribution, rejecting the hypothesis of no trend at the significance level of 0.05 or better in nine stations, as shown in Table 4.5. The location parameter of the Tn extreme distribution shows significant negative trend at the same level only for five stations (see Table 4.5).

The p-values testing the suitability of the linear time dependence in the location parameter are assessed using the likelihood-ratio test. They are shown in Figure 4.37 (a-b) for different significance level at each station. In Table 4.5 the statistical parameters of the models for both (Tx and Tn) extreme distributions are provided. These are the intercept coefficients μ_0, the slop coefficients μ_1, the scale σ and the shape ξ parameters of the extreme distribution estimated by MLE of PP approach for the non-stationary model. Also, the significance of the improvement as a result of a likelihood-ratio test considering a linear dependence in the μ parameter is indicated in Table 4.5.

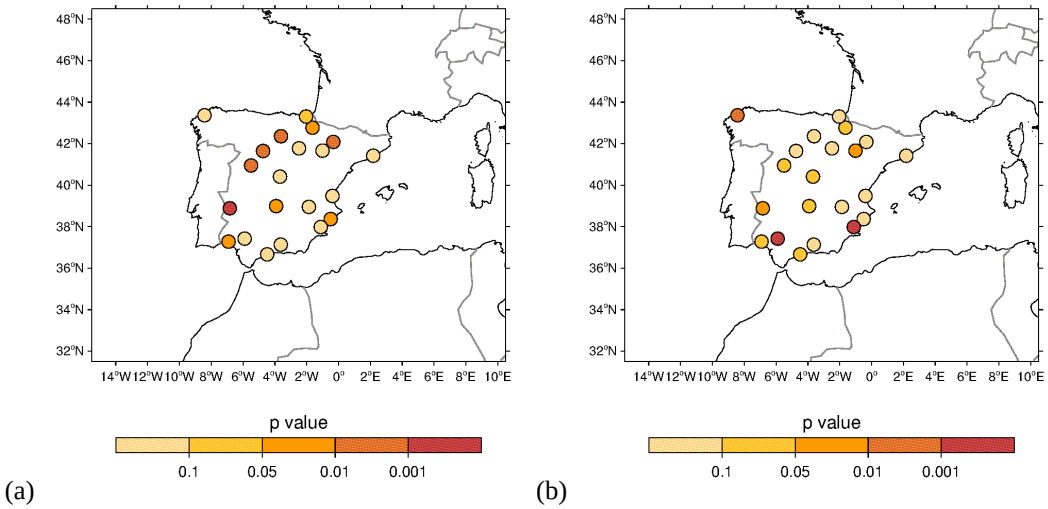


Figure 4.37 (a-b). Maps of p-values between the base model and the model with time dependence as a covariate in the location parameter of GEV distribution during the period 1940-1972 for Tx (a) and for Tn (b).

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.5. Parameters (MLE) of PP approach with time trend into the location parameter for the period 1940-1972 and statistical significance (*p-value<0.05, **p-value < 0.01 and ***p-value <0.001).

1940-1972	Tx					Tn				
Station	Location		Scale	Shape	Significance	Location		Scale	Shape	Significance
	μ_0	μ_1	σ	ξ		μ_0	μ_1	σ	ξ	
Albacete	38.1	-0.6	1.4	-0.3		20.1	-0.1	0.9	-0.3	
Alicante	35.5	-1.0	1.4	0.0	*	23.2	0.0	0.9	-0.2	
Badajoz	42.1	-2.7	1.5	-0.3	***	22.6	-1.4	1.3	-0.3	*
Barcelona	32.7	-0.5	1.3	-0.3		23.2	-0.4	1.3	-0.2	
Burgos	35.4	-2.4	1.5	-0.4	**	16.9	-0.5	1.4	-0.2	
Ciudad Real	39.7	-1.4	1.2	-0.2	*	22.2	1.0	1.1	-0.4	
Granada	39.4	-0.5	1.2	-0.5		21.2	-0.3	1.1	-0.3	
Huelva	39.6	-1.7	1.5	-0.2	*	22.1	-0.7	1.0	-0.1	
Huesca	37.3	-2.0	1.4	-0.3	**	21.2	-0.4	1.2	0.1	
La Coruña	28.8	-1.0	2.1	-0.3		19.7	-1.2	1.0	-0.1	**
Madrid	35.7	-0.7	1.4	-0.2		22.7	-0.9	1.0	-0.3	
Malaga	39.0	0.4	2.1	-0.3		24.0	0.9	1.3	-0.1	
Murcia	39.5	-0.3	1.5	-0.2		23.7	-1.6	1.0	-0.2	***
Pamplona	37.3	-1.6	1.6	-0.5	*	19.3	-1.0	1.4	-0.1	
Salamanca	37.8	-2.5	1.5	-0.4	**	19.2	-1.5	1.4	-0.3	
San Sebastian	34.0	-1.7	2.4	-0.5		20.3	-0.6	1.5	0.0	
Sevilla	41.5	-0.5	1.3	-0.1		24.0	-2.4	1.0	-0.4	***
Soria	35.3	-1.1	1.4	-0.3		16.8	0.2	1.1	-0.2	
Valencia	36.0	-0.5	2.2	-0.2		23.4	-0.1	0.6	-0.3	
Valladolid	37.7	-2.1	1.3	-0.3	**	19.3	-0.7	1.4	-0.3	
Zaragoza	38.6	-0.8	1.4	-0.2		22.0	-0.9	1.0	-0.2	*

During the first observational period, with the exception of Malaga, the slop coefficients μ_1 of Tx extreme distribution are negative for all locations, indicating a negative trend of μ. The improvement in allowing the temporal effect is significant at least at the 0.05 significance level only at 9 out of 21 stations (Table 4.5). The highest negative trends were estimated for the stations of Salamanca, Valladolid and Badajoz situated in the western part of Spain along with Burgos and Huesca located in the north and north-eastern part respectively, all of them exceeding -2ºC (Table 4.5).

Regarding the analysis of Tn extreme distribution, the estimated slop coefficients μ_1 are all negative except for Alicante, Ciudad Real, Malaga and Soria, but only for a few locations they are statistically significant at least at the adopted 0.05 significance level. The highest negative trends were identified for Sevilla and Murcia located in the

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

southern, Badajoz located in the western and La Coruña located in the north-western coast of Spain and all of them exceeding -1ºC (-2ºC in the case of Sevilla) (Table 4.5). Therefore, for both variables, Tx and Tn, a decrease in μ determines a shift of the distribution to lower values and, consequently, the corresponding reduction of all extreme values have been observed in agreement to the findings previously reported by Brunet et al. (2007a) over mainland Spain. They found reductions of extreme warm days and nights with Tx>90th and Tn>90th percentile, respectively, for summer season, although the trend for warm nights was nonsignificant. Also in the same way, Brunet et al. (2007b) associated the period 1950- 1972 with reduction in the annual number of nights with Tn above the 90th, 95th and 98th percentile.

In order to exemplify the amplitude of change during the first period, the estimated density function of extreme Tx in the first and in the last days of the analysed period, June, the 1st 1940 (solid black line) and August, the 31st 1972 (dashed blue line), is shown in Figure 4.38 (a-e) for the locations with the highest negative trends in μ. Figure 4.39 (a-d) shows same information as in Figure 4.38 (a-e), but for the estimated density function of extreme Tn for locations with the highest negative trends in μ.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

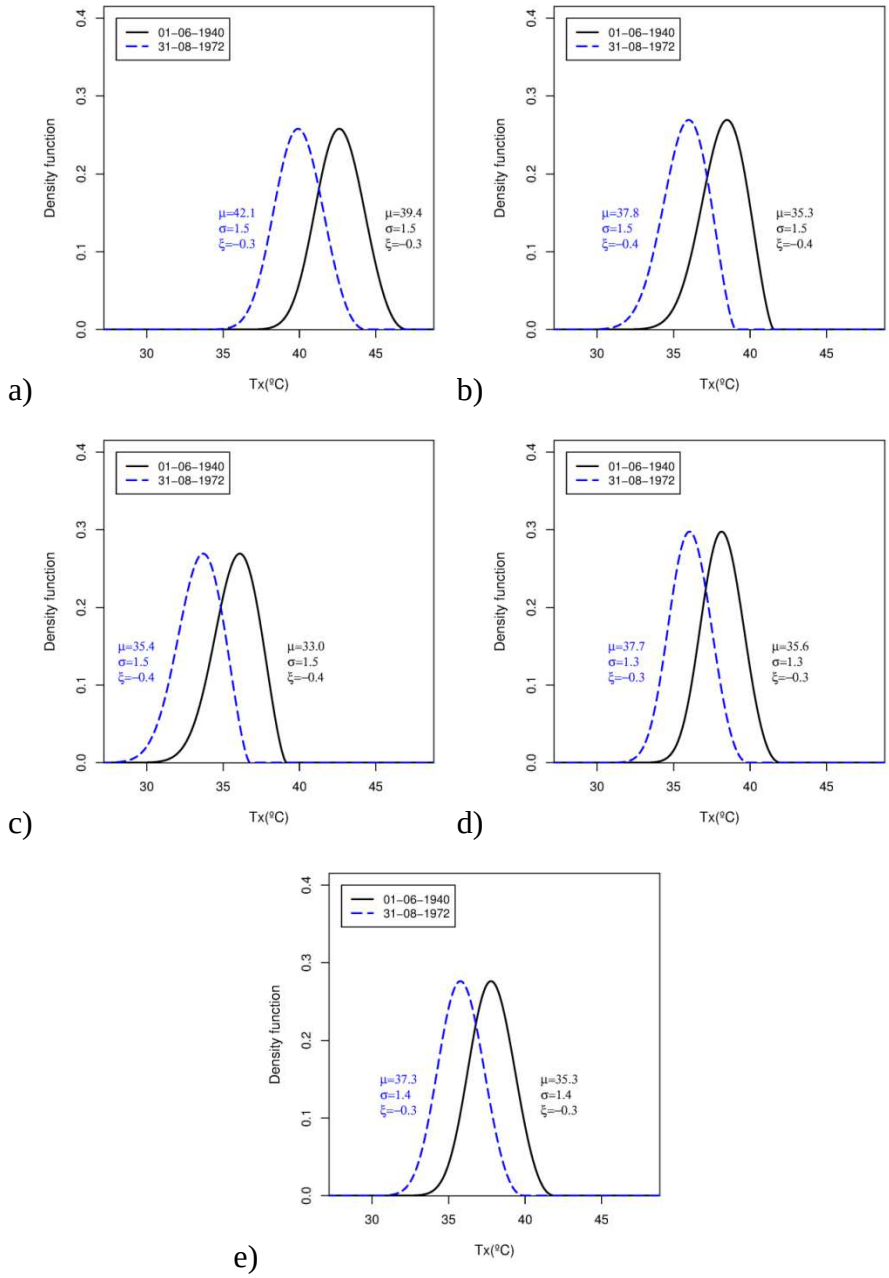


Figure 4.38 (a-e). Maximum temperature (Tx) extreme probability density functions. GEV parameters estimated. Solid black (dashed blue) curves display the estimated density function on June, the 1st 1940 (August, the 31st 1972). For Badajoz (a), Salamanca (b), Burgos (c), Valladolid (d) and Huesca (e).

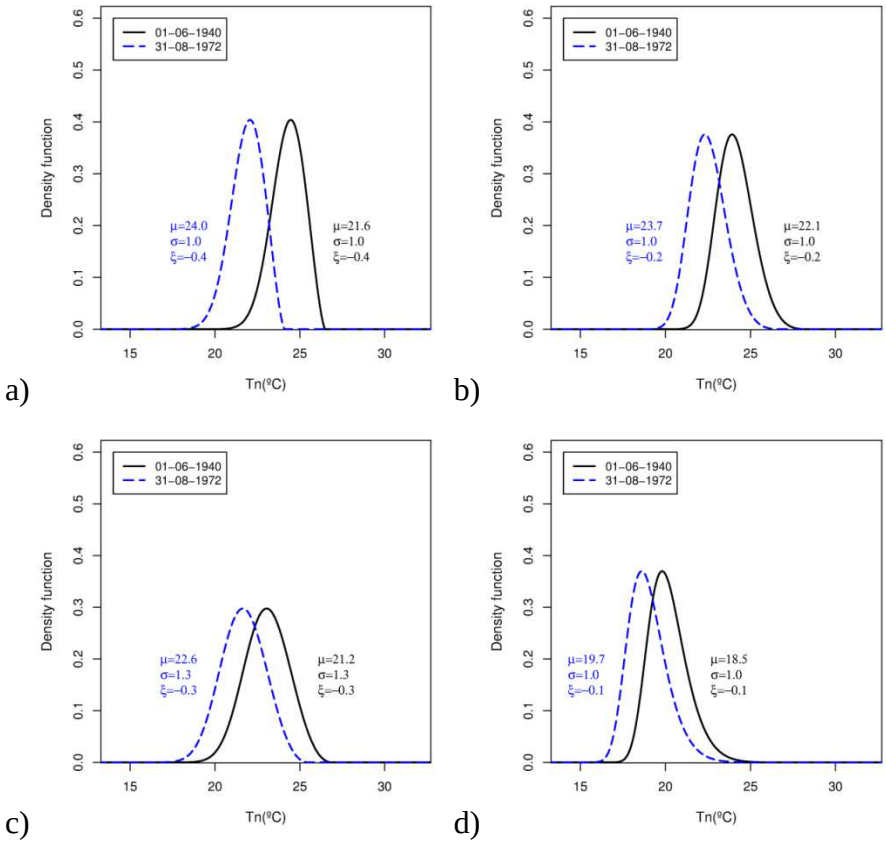


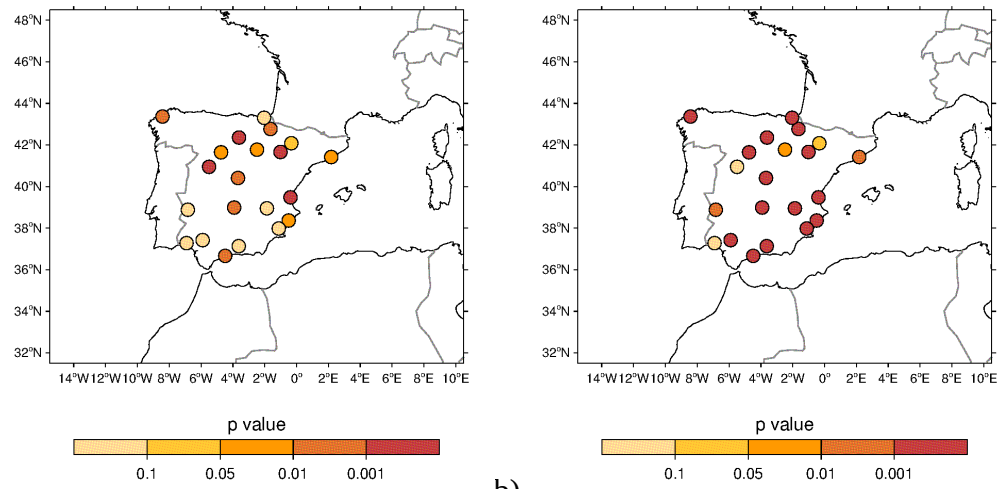
Figure 4.39 (a-d). Same as Figure 4.38 (a-e) but for the minimum temperature (Tn) extreme probability density functions and for Sevilla (a), Murcia (b), Badajoz (c) and La Coruña (d).

4.3.2 Observed trends in the location parameter (1973-2010)

When repeating the analysis for the recent period 1973-2010 comparing the non-stationary model assuming a linear trend in the location parameter μ with the stationary model, the results are completely different to those found for the former period. Not only the coefficient trends are positive, but also more locations reached statistical significance, particularly over northern Spain for both daily Tx and especially for Tn series. This implies positive trends in the upper tails of Tx and Tn distributions.

Figure 4.40(a-b) shows the p-values provided by the likelihood-ratio test, which indicate the significance of the improvement of the model when a linear trend in the location parameter of Tx and Tn extreme distributions is introduced. Likewise the Table 4.6 shows for both variables the intercept μ_0 and slop μ_1 coefficients and the scale σ and shape ξ parameters of the extreme distribution estimated by MLE with the location parameter varying in time, along with the statistical significance by the likelihood-ratio test.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



a) b)

Figure 4.40 (a-b). Same as Figure 4.37 (a-b) but during the period 1973-2010 for Tx (a) and for Tn (b).

For Tx extreme values, according to the likelihood ratio test, there is a significant time dependence in μ in 13 locations. Trend is positive everywhere, and the highest values are estimated in different Spanish places, such as the coastal locations of Valencia, Malaga and La Coruña and the north-western locations of Salamanca and Burgos, all of them exceeding 2.5ºC, while for Valencia the highest increasing trend (3.6ºC) was estimated (Table 4.6).

For Tn extreme values, the results show a highly significant improvement of the model at the 0.001 significance level with positive trends in most of the locations. The evidence of a temporal effect on μ is overwhelming, particularly for Sevilla, Ciudad Real, Murcia and Valladolid, where the coefficient trends exceeded 3ºC (4ºC in the case of Sevilla), while for Albacete, Madrid, Malaga and Granada trends exceeded 2ºC. Exceptions have been found for Huesca, Salamanca and Huelva, since time effect has not brought improvement to the model (Table 4.6).

Table 4.6. Same as Table 4.5 but for the period 1973-2010.

1973-2010	Tx					Tn				
Station	Location		Scale	Shape	Significance	Location		Scale	Shape	Significance
	μ_0	μ_1	σ	ξ		μ_0	μ_1	σ	ξ	
Albacete	38.0	0.8	1.1	-0.1		19.6	2.5	0.8	-0.3	***
Alicante	34.5	1.4	1.6	-0.1	*	23.3	1.8	0.7	-0.2	***
Badajoz	40.5	0.8	1.2	-0.2		21.6	1.8	1.3	-0.2	**
Barcelona	33.3	1.5	1.5	-0.2	*	23.1	1.3	1.2	-0.2	**
Burgos	33.9	2.6	1.3	-0.3	***	15.8	1.5	1.1	-0.1	***
Ciudad Real	39.1	1.2	1.0	-0.2	**	21.7	3.9	1.0	-0.4	***
Granada	39.3	0.7	1.3	-0.5		21.0	2.1	1.4	-0.1	***
Huelva	39.3	0.7	1.7	-0.3		22.7	0.5	1.1	-0.1	
Huesca	36.8	1.1	1.3	-0.1		21.4	0.7	1.2	0.0	
La Coruña	27.6	2.5	2.1	0.0	**	18.3	1.3	0.8	-0.1	***
Madrid	36.2	1.9	1.1	-0.3	**	22.4	2.1	1.0	-0.4	***
Malaga	37.7	2.6	2.1	-0.2	**	23.6	2.5	1.0	-0.2	***
Murcia	40.1	0.6	1.6	-0.1		22.0	3.4	0.7	-0.2	***
Pamplona	36.3	1.8	1.5	-0.2	**	18.8	1.4	1.2	0.0	***
Salamanca	34.7	2.8	1.2	-0.3	***	17.6	0.5	1.2	-0.2	
San Sebastian	32.5	1.3	2.6	-0.4		18.9	1.6	1.0	-0.1	***
Sevilla	41.6	1.0	1.4	-0.2		21.8	4.2	1.2	-0.2	***
Soria	34.6	0.9	1.0	-0.2	*	16.9	0.9	1.1	-0.2	***
Valencia	34.6	3.6	2.3	-0.3	***	24.0	1.1	0.6	-0.2	***
Valladolid	36.7	1.1	1.1	-0.2	*	17.7	3.2	1.4	-0.2	***
Zaragoza	37.5	2.1	1.2	0.0	***	21.5	1.7	0.8	-0.3	***

These results are in qualitative agreement with findings of Brown et al. (2008) for Europe at the global scale, who using similar methodology, non-stationary marked PP model, reported significant positive trends in the location parameter of both Tx and Tn extreme distributions, although they found lower coefficient trends of 1.1ºC and 1.4ºC, respectively, for the period 1950-2004 in Europe. Main difference lies on the amplitude of the trends estimated in both studies, which could be mainly explained by the different length of the periods analysed. Whilst in Brown et al. (2008) the trend analysis is carried out for a longer period and include 18 years pre-1973 characterised by stagnant temperatures in Spain, in this thesis the trend analysis is made from 1973 onwards, when a strong rise in Spanish temperatures has been identified (Brunet et al. 2007a). Also, at regional scale the results agree well with the findings of El Kenawy et al. (2011) for north-eastern Spain. They investigated trends in daily Tx and Tn extremes

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

during the period 1960-2006 and reported a significant increase in the frequency and intensity of warm nights and days.

Generally, for both analysed variables, the extreme distributions have shifted towards warmer values during the period 1973-2010. To illustrate this, in Figure 4.41(a-f) the estimated density function of Tx extreme in the first and in the last day of the second period; namely, June the 1st 1973 (solid black line) and August the 31st 2010 (dashed red line) are provided for locations with the highest positive trends. The corresponding 20-year return values (dashed grey line) are indicated, as well. Figure 4.42(a-h) shows same information as in Figue 4.41(a-f), but for the density function of extreme Tn of the locations with the highest positive trends.

With regard to the shape and scale parameters, they are assumed to be stationary in time at each location, although it is worth to mention they have spatial variations among the different geographical locations. The shape parameter mostly takes negative values ranging from -0.5 to 0 for Tx extreme distribution and from -0.4 to 0 for Tn extreme distribution, indicating a bounded (or Weibull) distribution in most of the locations. Good matching, in the sign and in the values of the shape parameter for all the analysed stations, is found when compared with the corresponding values for Europe and particularly for Spain reported by Brown et al. (2008).

The scale parameter shows an appreciable variability among locations with a range of 1.6 (0.8) σ for Tx (Tn) extreme distributions (Table 4.6). Meanwhile the results of the time dependence analysis during the period 1940-1972 suggest fewer VHD and VHN, for the period 1973-2010 the estimated trends point out towards more VHD and even more VHN. These results are indeed in agreement with the study of Brunet et al. (2007a) for mainland Spain who found significant trends for summer extreme warm days and nights with Tx>90th and Tn>90th, respectively during the period 1973-2005. Also are consistent with the findings of Brunet et al. (2007b) at the annual scale. They reported significant positive trend for the number of days with Tx and Tn above the 95th percentile for the period 1973-2005. Additionally, larger trends for Tn extremes compared with Tx extremes was reported, indicating that hot and extremely hot nights become slightly more frequent compared with hot and extremely hot days long-term trends according to the results found in this thesis. This generally agrees with the earlier study of Donat et al. (2013) who present a collation and analysis of gridded land-based dataset of indices derived from daily Tx/Tn and precipitation observations. They

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

showed widespread significant changes in temperature extremes consistent with warming, especially for those indices derived from daily Tn over the whole 110 years of record (1901-2010) but with stronger trends in more recent decades and significant warming in summer season.

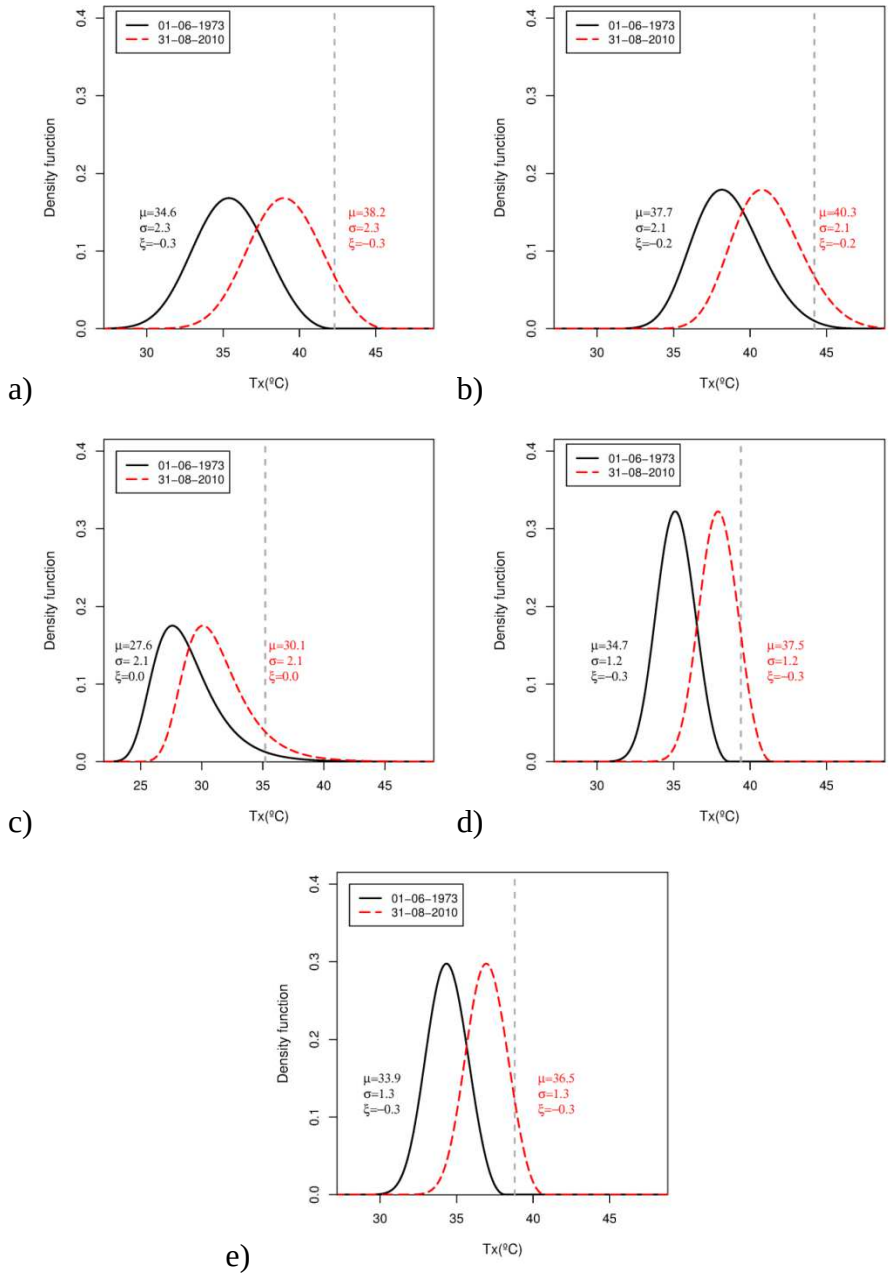


Figure 4.41 (a-f). Tx extreme probability density functions. GEV parameters estimated. Solid black (dashed red) curves display the estimated density function in June the 1st 1973 (August the 31st 2010) for Valencia (a), Malaga (b), La Coruña (c), Salamanca (d) and Burgos (e). The corresponding 20-year return values are indicated by grey dashed line.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

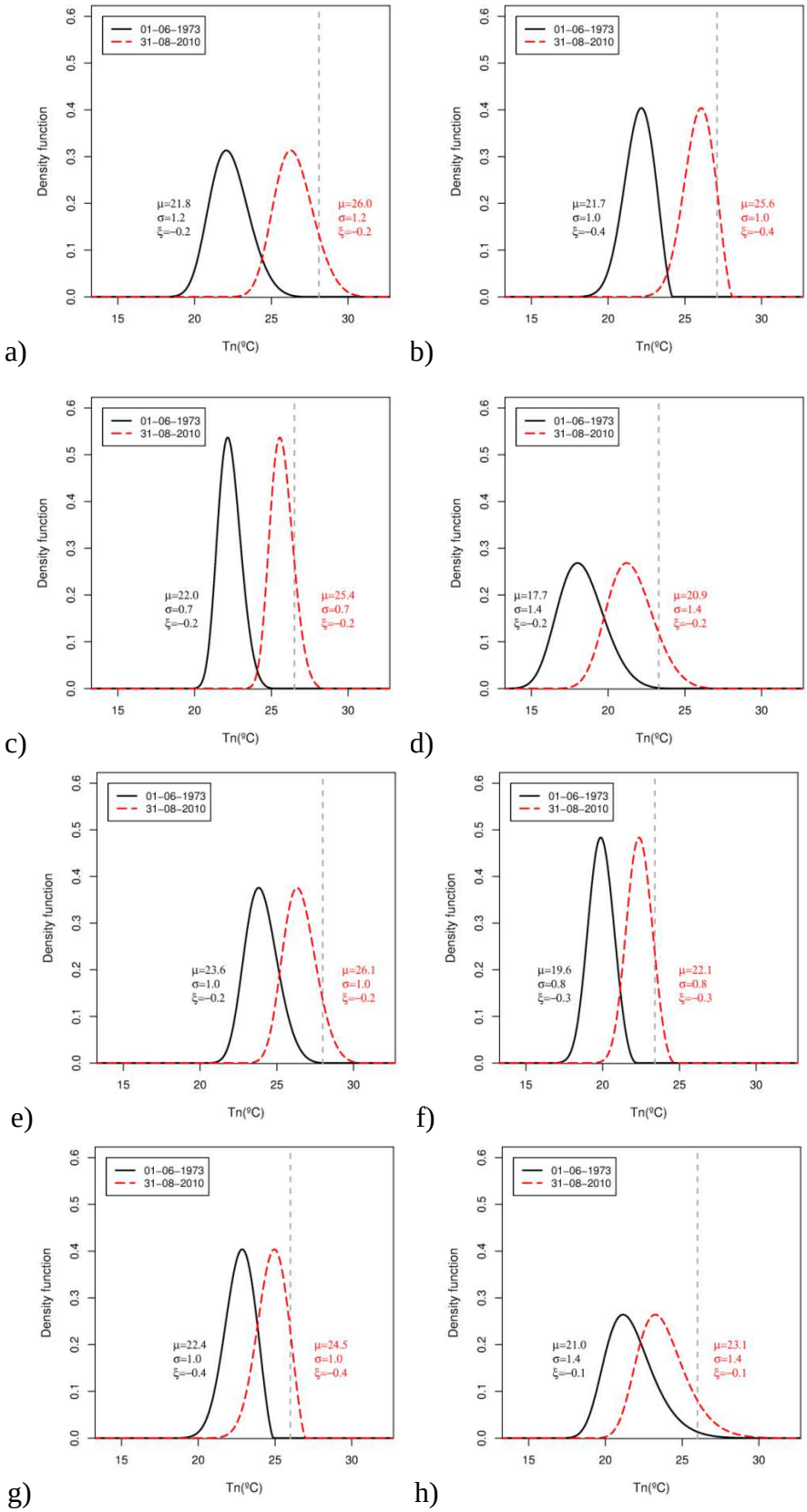


Figure 4.42 (a-h). Same as Figure 4.41 (a-f) but for Tn extreme probability density functions and for Sevilla (a), Ciudad Real (b), Murcia (c), Valladolid (d), Malaga (e), Albacete (f), Madrid (g) and Granada (h). The corresponding 20-year return values are indicated by grey dashed line.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Using a non-homogeneous Poisson Process to assess daily Tx, Abaurrea et al. (2007) explored the intensity of extreme hot events in the mid Ebro Basin during the period 1951-2004 and concluded the intensity was stationary, unlike the results reported in this thesis for stations located in the same region. The discrepancy is likely due to the different analysed periods; since in the former authors analysed altogether the two sub-periods that in this thesis are assessed separately.

The results can be compared with findings of Furió and Meneu (2011), who investigated the statistical behavior of extreme temperatures in various Spanish stations such as Bilbao (1947-2009), Madrid (1900-2009 and 1960-2009), Valencia (1938-2009) and Sevilla (1922-2009) using block maxima approach allowing for a linear trend in the location parameter of GEV distributions. The authors found evidence of significant trends in temperature extremes for most of the considered series. In particular, summer Tn increased over time from 0.3ºC to 0.4ºC per decade in Bilbao, Sevilla and Valencia, but not in Madrid, while for Tx they found evidence of an increasing linear trend only for Madrid 0.06ºC per decade. In this thesis and for summer Tn, it was estimated for the period 1973-2010 positive trends in Valencia (0.3ºC/decade), Sevilla (1.1ºC/decade) and in Madrid (0.6ºC/decade), while for summer Tx both Madrid and Sevilla showed significant trends with 0.5ºC and 1ºC per decade, respectively.

4.4 Observed trends in the scale parameter

After studying the time dependence in the location parameter, it has been examined whether the scale parameter has also a temporal trend and, if so, how is its behaviour in each station for both Tx and Tn variables during the two periods of analysis. Again, the shape parameter was modelled as invariant. In this case, the non-stationarity has been modelled allowing the location and the log-transformed scale parameter to depend linearly on time. Table 4.7 and Table 4.8 show the intercept coefficients μ_0 and σ_0, slop coefficients μ_1 and σ_1, and the shape parameters ξ estimated by MLE of PP approach for the new non-stationary model for Tx and Tn extreme during the period 1940-1973. In addition, it is indicated the significance calculated by the likelihood-ratio test between both non-stationary models (one with μ time varying and the other with μ and σ time varying). Table 4.9 and Table 4.10 show similar information than that shown in Table 4.7 and Table 4.8, but for the period 1973-2010.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

When comparing the two non-stationary models to see whether the model with time variation in the location parameter improves significantly when the log-transformed scale parameter is also assumed as time dependent, the results show that there is no evidence to support the model with μ and σ time varying in most of the locations and, hence, there is no reason to adopt a temporal effect in σ. The model has been enhanced for few locations during both periods and for both variables (Table 4.7 – Table 4.10) and the locations are different for each case. However, it is interesting to note for both variables and for both periods, as an exception of Tx extreme in Huesca for the period 1940-1972 (Table 4.7 – Table 4.10), the significant trends in the scale parameter are negative, which implies extreme distributions are narrowing at the end of the period. This also implies a reduction in variability of extreme temperatures on the daily scale.

Consequently, the results support the conclusions that in Spain the general changes in extreme distributions have been mainly attributed to changes in the location parameter rather than in the scale parameter, comprising temporal effects, either increasing or decreasing the extremes, in agreement with Kharin and Zwiers (2005) for globally averaged changes. And different behaviour for the changes estimated in the location of the extreme distributions for each period has been identified.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.7. Parameters (MLE) of PP approach with time trend into the location and scale parameters of Tx GEV distribution for the period 1940-1972 and the significance (*p-value < 0.05 and **p-value < 0.01).

1940 – 1972	Tx					
Station	Location		Scale		Shape	Significance
	μ_0	μ_1	σ_0	σ_1	ξ	
Albacete	38.1	-0.7	0.3	-0.1	-0.3	
Alicante	35.3	-0.8	0.3	0.2	0.0	
Badajoz	42.1	-2.7	0.4	0.0	-0.3	
Barcelona	32.9	-0.8	0.5	-0.6	-0.2	
Burgos	35.4	-2.3	0.6	-0.4	-0.4	
Ciudad Real	39.6	-1.4	0.1	0.2	-0.2	
Granada	39.5	-0.6	0.1	0.2	-0.5	
Huelva	39.6	-1.8	0.4	-0.1	-0.2	
Huesca	37.4	-1.7	-0.1	1.3	-0.7	*
La Coruña	28.9	-1.1	0.8	0.0	0.0	
Madrid	35.7	-0.7	0.3	0.0	-0.2	
Malaga	39.5	-0.6	1.0	-0.5	-0.3	*
Murcia	39.6	-0.6	0.5	-0.2	-0.2	
Pamplona	37.3	-1.5	0.5	0.0	-0.5	
Salamanca	37.9	-2.6	0.5	-0.3	-0.3	
San Sebastian	34.1	-1.8	0.8	0.1	-0.5	
Sevilla	41.5	-0.4	0.2	0.0	-0.1	
Soria	35.4	-1.4	0.6	-0.5	-0.4	
Valencia	36.1	-0.6	0.8	0.0	-0.2	
Valladolid	37.6	-2.1	0.5	-0.6	-0.2	
Zaragoza	38.6	-0.9	0.3	0.0	-0.2	

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.8. Same as Table 4.7 but for Tn GEV distribution.

1940 – 1972	Tn					
Station	Location		Scale		Shape	Significance
	μ_0	μ_1	σ_0	σ_1	ξ	
Albacete	20.0	-0.1	-0.1	0.1	-0.3	
Alicante	23.5	-0.4	0.1	-0.5	-0.2	
Badajoz	22.6	-1.5	0.3	-0.2	-0.2	
Barcelona	23.2	-0.4	0.3	0.0	-0.2	
Burgos	17.3	-1.3	0.6	-0.5	-0.2	
Ciudad Real	22.2	0.9	0.3	-0.3	-0.4	
Granada	21.0	0.2	-0.2	0.4	-0.3	
Huelva	22.8	-2.0	0.5	-1.2	-0.2	**
Huesca	21.2	-0.5	0.2	0.0	0.1	
La Coruña	19.6	-1.0	-0.1	0.2	-0.2	
Madrid	22.7	-0.9	0.0	0.0	-0.3	
Malaga	23.5	1.8	0.0	0.5	-0.1	
Murcia	23.7	-1.6	0.0	-0.1	-0.2	
Pamplona	19.9	-2.3	0.7	-0.9	-0.1	**
Salamanca	19.7	-2.5	0.7	-0.9	-0.4	**
San Sebastian	20.3	-0.7	0.4	0.0	0.0	
Sevilla	24.0	-2.4	0.1	-0.4	-0.4	
Soria	16.8	0.3	0.0	0.1	-0.2	
València	23.5	-0.2	-0.4	-0.1	-0.3	
Valladolid	19.3	-0.8	0.4	0.0	-0.2	
Zaragoza	22.4	-1.7	0.3	-0.8	-0.2	**

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.9. Same as Table 4.7 but for the period 1973-2010.

1973 – 2010	**Tx**					
Station	Location		Scale		Shape	Significance
	μ_0	μ_1	σ_0	σ_1	ξ	
Albacete	38.1	0.6	0.6	-0.9	-0.3	*
Alicante	34.4	1.8	0.3	0.2	-0.1	
Badajoz	40.5	0.9	0.1	0.1	-0.2	
Barcelona	33.2	1.5	0.4	0.0	-0.2	
Burgos	33.9	2.6	0.7	-0.6	-0.4	
Ciudad Real	39.2	1.0	0.2	-0.3	-0.2	
Granada	39.3	0.8	0.3	-0.1	-0.5	
Huelva	39.3	0.6	0.6	-0.1	-0.3	
Huesca	37.0	0.7	0.5	-0.5	-0.1	
La Coruña	27.7	2.3	0.8	-0.1	0.0	
Madrid	36.4	1.3	0.9	-1.3	-0.3	
Malaga	38.1	1.9	1.2	-0.9	-0.4	**
Murcia	40.1	0.5	0.5	0.0	-0.1	
Pamplona	36.6	1.3	0.7	-0.5	-0.3	
Salamanca	34.6	2.9	0.0	0.3	-0.3	
San Sebastian	32.5	1.4	1.0	-0.2	-0.4	
Sevilla	41.8	0.5	0.6	-0.5	-0.2	
Soria	34.7	0.7	0.2	-0.3	-0.3	
Valencia	34.4	3.8	0.7	0.1	-0.2	
Valladolid	36.9	0.6	0.3	-0.5	-0.2	
Zaragoza	37.5	2.2	0.1	0.1	0.0	

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Table 4.10. Same as Table 4.8 but for the period 1973-2010.

1973 – 2010	Tn					
Station	Location		Scale		Shape	Significance
	μ_0	μ_1	σ_0	σ_1	ξ	
Albacete	19.6	2.4	-0.1	-0.2	-0.3	
Alicante	23.3	1.7	-0.2	-0.3	-0.3	
Badajoz	21.6	1.7	0.3	-0.1	-0.2	
Barcelona	23.1	1.4	0.1	0.1	-0.2	
Burgos	15.7	1.6	0.1	0.1	-0.1	
Ciudad Real	21.7	3.9	-0.2	0.4	-0.3	
Granada	21.6	1.2	0.8	-0.8	-0.2	*
Huelva	22.5	0.9	-0.1	0.4	-0.1	
Huesca	21.5	0.5	0.3	-0.3	-0.1	
La Coruña	18.2	1.5	-0.6	0.5	-0.1	
Madrid	22.5	2.0	0.0	-0.1	-0.5	
Malaga	23.7	2.4	0.2	-0.3	-0.2	
Murcia	21.9	3.6	0.0	-0.4	-0.3	
Pamplona	18.7	1.6	0.1	0.2	0.0	
Salamanca	18.1	-0.6	0.5	-0.6	-0.2	
San Sebastian	18.8	1.7	-0.2	0.4	-0.1	
Sevilla	21.8	4.3	0.3	-0.1	-0.2	
Soria	17.3	0.2	0.5	-0.7	-0.3	**
Valencia	24.0	1.1	-0.6	0.0	-0.2	
Valladolid	17.5	3.5	0.1	0.4	-0.2	
Zaragoza	21.4	1.9	-0.5	0.6	-0.1	

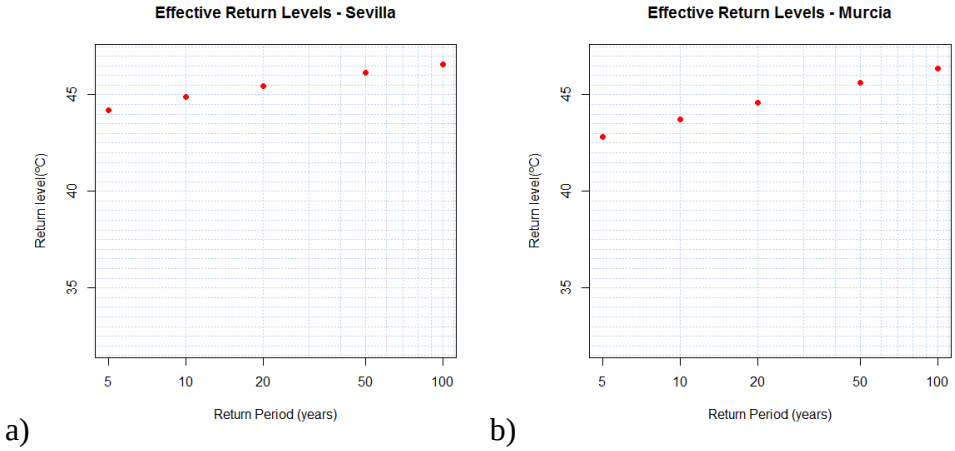
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

4.5 Effective return levels of temperatures extremes (1973-2010)

In order to better understand the behaviour of Tx and Tn extremes and to know how such exceptionally events will be in the future, the 5, 10, 20, 50 and 100-year return levels, derived from fitting a non-stationary PP model with location parameter linearly depending on time have been calculated for the warmer period (1973-2010). The results introduced in this section assume a linear trend in μ and no trend in σ and ξ parameters. Figure 4.43 (a-e) provides the maximum effective return levels of extreme Tx for the locations with the highest values of Tx estimated and Figure 4.44 (a-d) gives the maximum effective return levels of extreme Tx for locations with the lowest values of Tx estimated. For the extreme Tx, the results show the estimated return levels for each return period pointing out the expected increase of Tx at all stations, with southern parts of Spain expecting to experience higher values. It seems reasonable, to find the highest values over the southernmost part of Spain, namely at the stations of Sevilla, Murcia, Badajoz, Malaga and Huelva, where for example the estimated return levels for the 20-year return period are 46.6ºC, 46.3ºC, 44.9ºC, 45.8ºC and 44.3ºC, respectively. The lowest values are expected in the northernmost part of Spain in the locations of La Coruña, Soria, Barcelona and San Sebastian and for the 20-year return period the return values are estimated as 38.3 ºC, 39.2 ºC, 39.4 ºC and 38.2 ºC, respectively. However, the north coast of Spain, La Coruña and San Sebastian, is expected to suffer the greatest increase in the 20-yeatr return values of extreme Tx exceeding until 4ºC. These results are in the line of the findings reported by Abaurrea et al. (2007) for NE of the Iberian Peninsula. They projected extreme hot events up to 2050 and found an increase of monthly signal between 1980 and 2040 nearly 4ºC in August.

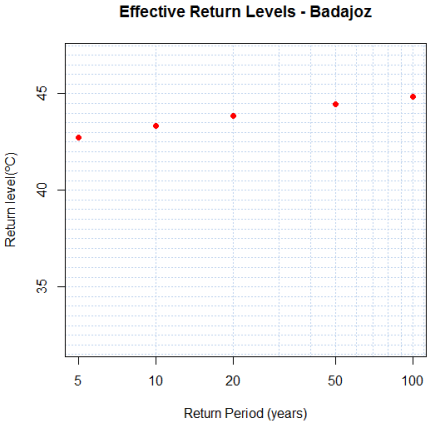
Furió and Meneu (2010) by extrapolating the observed trends into the future found summer maximum temperatures in Madrid is expected to reach 37.3ºC, 39.1ºC and 40.6ºC in 2020, 2050 and 2075 respectively, with a probability of 90%, meanwhile findings of this thesis show the increases will be larger for Madrid, reaching 40.0 ºC for the 20-years return period.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

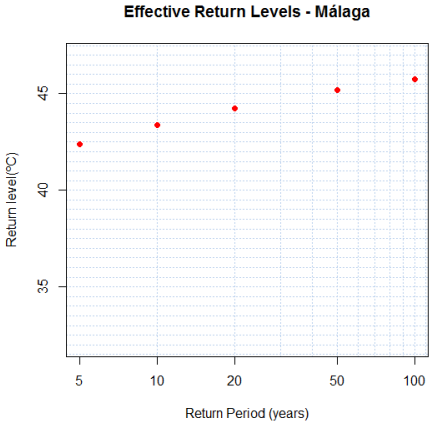


a)

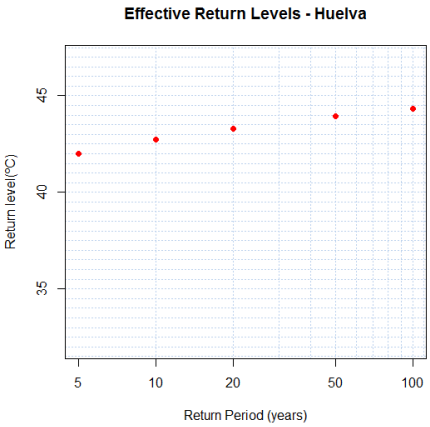
b)



c)



d)



e)

Figure 4.43(a-e). The highest effective return levels for 5, 10, 20, 50 and 100 years for Tx extreme.

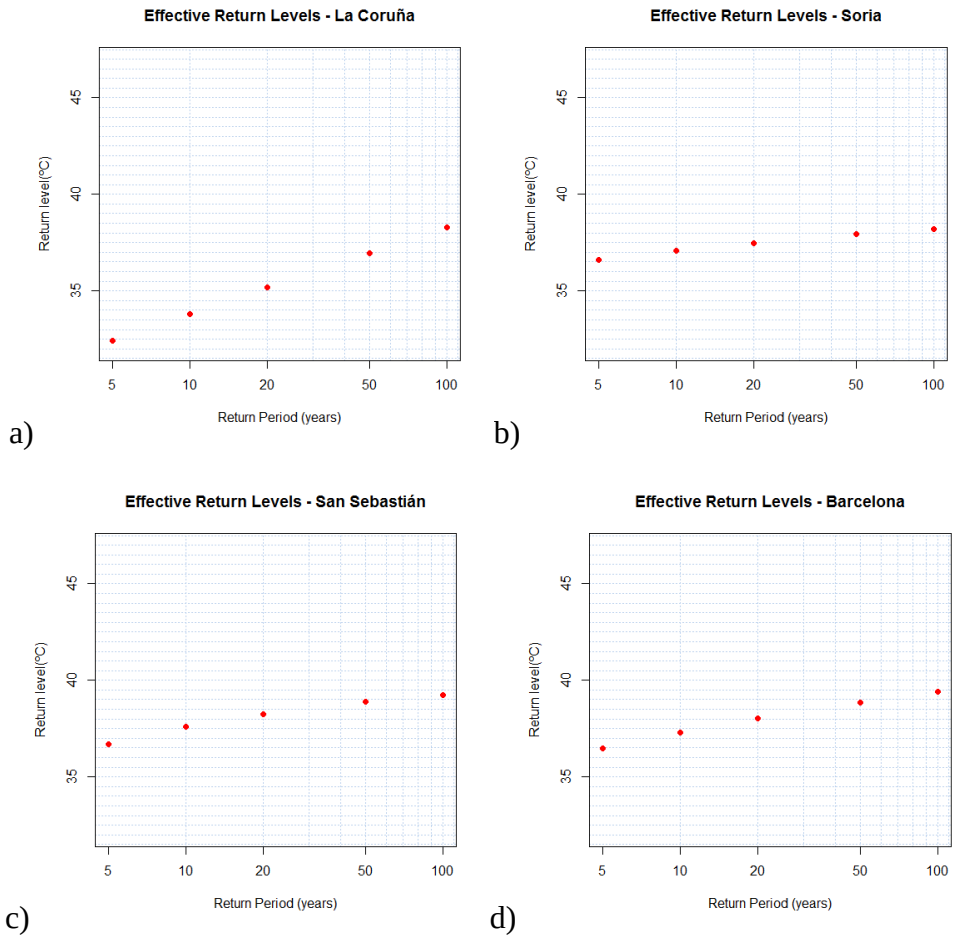


Figure 4.44 (a-d). The lowest effective return levels for 5, 10, 20, 50 and 100 years for Tx extreme.

Figure 4.45 (a-e) provides the maximum effective return levels of extreme Tn for the locations with the highest values of Tn estimated and Figure 4.46 (a-d) presents the maximum effective return levels of extreme Tn for the locations with the lowest values of Tn estimated. For the extreme Tn, the results show the expected increase of the estimated return levels at all stations, with higher values not only over southern Spain, but also over inland and coastal areas. The highest estimated return values are expected in Sevilla, Malaga, Ciudad Real and Barcelona with 28.1 ºC, 28.0ºC, 27.1ºC and 26.8ºC for 20-year level respectively. The lowest return values of extreme Tn are estimated for the northernmost part of Spain, likewise for the return values of extreme Tx, but in this case the lowest 20-year return values estimated are expected for Burgos, Soria, Salamanca and La Coruña with 19.9 ºC, 20.0 ºC, 20.7 ºC and 21.2 ºC respectively. However, it is interesting to highlight that the highest increases are expected in Pamplona and Huesca, both situated at North-eastern Spain, with a rate of 3ºC.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

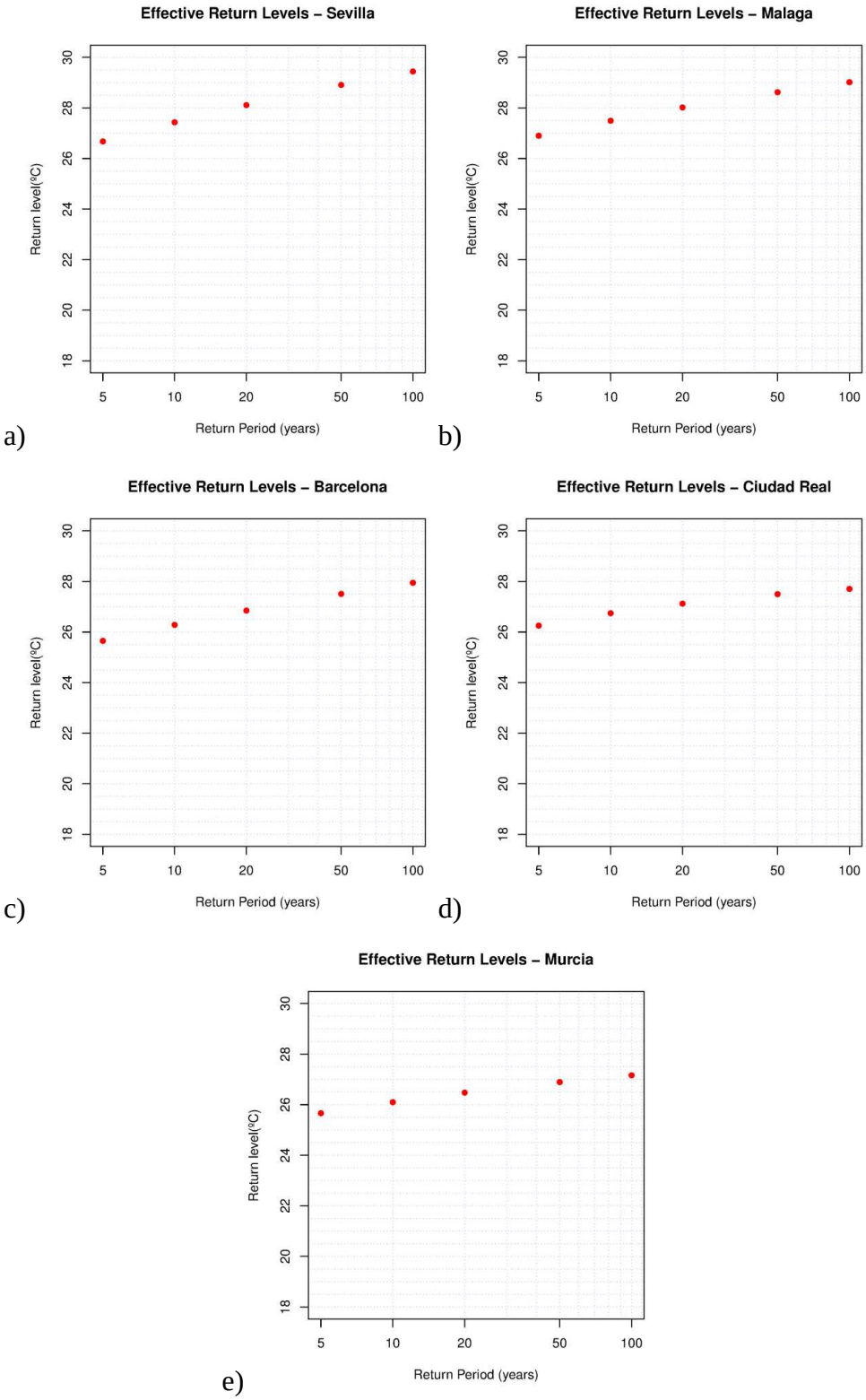


Figure 4.45 (a-e). Same as Figure 4.43 (a-e) but for Tn extreme.

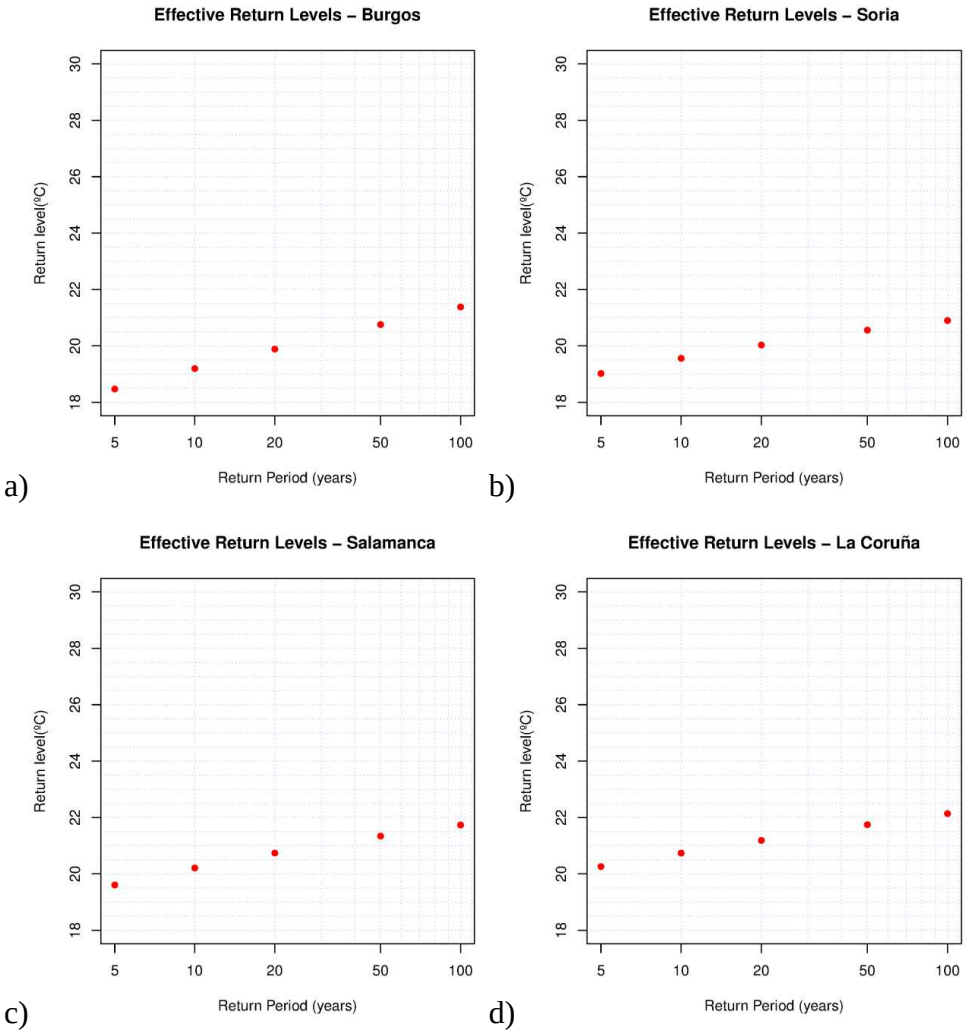


Figure 4.46 (a-d). Same as Figure 4.44 (a-d) but for Tn extreme.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

5 Summary, conclusions and outlook

In this chapter, the main results and findings of this thesis dealing with statistical modelling and analysis of summer very hot events in mainland Spain are summarized. In section 5.1 the main results of the influence of the large-scale variables: SLP, SST and SM anomalies on VHD and VHN are presented. The observed changes and trends in both Tx and Tn extreme distributions and the expected extreme temperature values to be likely achieved in the future are reviewed in section 5.2. Finally, in section 5.3, the final conclusions and outlook are provided.

All the scientific questions addressed in the introductory chapter have been answered through the results of the analysis.

5.1 Influence of large-scale variables on VHD and VHN

The relationships between the large-scale variables of Sea Level Pressure (SLP), Sea Surface Temperature (SST) and Soil Moisture (SM) and the occurrence and intensity of very hot days (VHD) and very hot nights (VHN) in mainland Spain have been investigated. The methodology used relies on the application of the Point Process (PP) approach which is based on Extreme Value Theory (EVT) to daily data from the Spanish Daily Adjusted Series (SDATS). The PP approach has been chosen because it has several advantages over other asymptotic models like the Block Maxima, *r* Largest Order Statistic model or Poisson-GPD model (Coles, 2001) to model extreme events. This methodology has been used, for example, to model hot spells and heat waves (Furrer, 2010) and extreme waves (Galiatsatou and Prinos, 2011). An appropriate technique to identify thresholds, which define the extreme events of VHD and VHN, has been used. It consists in fitting the data to the GPD distribution in a range of thresholds between the 95^{th} and 99.5^{th} percentiles and to look for the stability of the estimated parameters. Composite maps of SLP, SST and SM anomalies for VHD (VHN) corresponding to the days when the maximum (minimum) temperatures were above the threshold have been drawn.

The results of the statistical modelling show that the behaviour of the VHD and VHN is well modelled by the PP approach. Meaningful results have been gained by the

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

inclusion of large-scale covariates (SLP, SST and SM anomalies) into the model. Next, there are summarized the main findings.

For mainland Spain the results from the analysis of the influence of SLP anomalies show that there are three particular patterns related to VHD and VHN. Although atmospheric configurations are rather similar, SLP anomalies are, in general, stronger during the formation of VHD conditions than for VHN. The patterns identified are:

1. Southerly Flow Pattern

The first pattern namely the Southerly Flow Pattern presents high positive SLP anomalies over central Europe, which expands southward across the Western Mediterranean until North Africa, while anomalously low SLP above central and northern Atlantic basin, with stronger negative anomalies affecting most parts of the IP in the case of VHN. This configuration enforces a southerly component inflow of warm and dry air masses from African Sahara to the IP affecting inland, north-eastern and north Spain locations, with slight differences between these three cases.

2. Weak South-westerly Airflow Pattern

The second pattern identified is the Weak South-westerly Airflow Pattern, which presents moderately positive anomalies centred over north-western Africa that extend over the Western Mediterranean Basin and south-eastern Europe, along with an area of negative anomalies centred over the North Atlantic favouring the advection of a weak warm south-westerly airflow over the IP. This pattern corresponds to VHD and VHN in the south-eastern of the IP.

This and the first pattern summarized above correspond to blocking-like patterns over Western Europe and northern Africa, respectively, advecting warm air over Europe.

3. North-westerly Airflow Pattern

The third pattern identified in this thesis is the North-westerly Airflow Pattern. It shows a weak positive centre over north-western Africa and southern Portugal and a large area of negative SLP anomalies with its centre located over or near the UK, which is much stronger during the VHD. This configuration for VHD enforces the warmed north-westerly airflow when it crosses the IP, while for VHN gives rather weak north-westerly flow. This pattern has been found for central and southern Mediterranean coastal locations of the IP.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

By the statistical modelling through PP approach it was found correspondence of the grid maps of p-values with the large-scale SLP anomaly patterns identified; therefore, it was statistically confirmed the role of SLP anomalies when forcing these extreme temperature events. In particular, according to the methodology applied, the effect of the SLP anomalies corresponds to both an effect on the occurrence and the intensity of VHD and VHN.

For mainland Spain, main results about influences of SST anomalies on summer VHD and VHN reveal positive anomalies in the North Atlantic Ocean, Biscayne Bay and with the highest anomalies over the Western Mediterranean coast both during the summer VHD and VHN recorded, although the findings show warmer Atlantic Ocean during the extreme nights.

Throughout the statistical modelling it was found that the SST anomalies near the IP are related to the occurrence and intensity of extreme temperatures events on various locations in mainland Spain. Nevertheless, relationships are heterogeneous in space, since SST anomalies do not affect all locations equally. For VHD, it has been observed the continentality effect, since the SST impact is larger in coastal locations and gradually diminishes towards inland locations, especially over those surrounded and isolated by mountainous systems. Three groups of stations according to the different marine areas of influence have been identified:

A. both Mediterranean and North Atlantic SST anomaly influence
B. mainly North Atlantic SST anomaly influence
C. little SST anomaly influence

The results suggest that some locations in north Spain and in the Spanish Mediterranean coast are influenced by the Western Mediterranean and the Atlantic Ocean, while western Spain is mostly influenced by the North Atlantic Ocean and the inland locations are little influenced by SST anomalies.

For VHN, the findings are somewhat different, although also three groups are identified:

D. both Mediterranean and North Atlantic SST anomaly influence
E. mainly North Atlantic SST anomaly influence
F. mainly Mediterranean and Cantabrian SST anomalies influence

VHN of south-eastern and north Spain are influenced by the Mediterranean Sea and the Atlantic Ocean, western locations along with Madrid are influenced by the Atlantic

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Ocean and north-eastern Spain places are influenced by the Mediterranean and Cantabrian Sea. Inland locations are more influenced by SST anomalies during VHN than for VHD. Based on this analysis, it was found SST anomalies over the Western Mediterranean play an important role in the occurrence of summer VHD and VHN in locations situated over the Mediterranean coast of the IP.

Next, main results highlighting relationships between SM anomalies and summer VHD and VHN over mainland Spain are summarized.

The composite maps showed during the previous days to any extreme temperature event, for both VHD and VHN, there is a SM deficit over the IP, south France and North-west Africa. Two regions were found remarkably drier: one located in north-eastern Spain including mid Ebro Basin and Bardenas Reales (Navarra) and Monegros (Aragon) deserts, while the other one sited in northern Africa over the Algerian Sahara Desert.

By the statistical modelling, it was found statistically significant relationships between SM anomalies over the IP, south France and North Africa and the occurrence and intensity of both VHD and VHN for virtually all the locations except for La Coruña and San Sebastian where lower improvement of the model was reached.

In particular, VHD in Ebro Basin and central Spain exhibit the most significant relationship with SM anomalies according to the likelihood-ratio test. As for VHN the most influenced locations for which SM anomalies improve the model are situated over south-eastern and north-eastern Spain.

In both cases, for VHD and VHN, the best relationships with the SM anomalies are always found for grid points located in the IP. The influence of SM anomalies on extreme temperatures in Spain has been confirmed in this study.

5.2 Observed changes in extreme distributions and return levels

Main findings of the analysis of observed changes and trends in the location and scale parameters for both Tx and Tn GEV distributions are presented in this section. The analysis has been performed for two different periods, which are 1940-1972 and 1973-2010. These two periods have been selected based on the results of Brunet et al. (2007 a,b), who identified the year 1973 as the start of a strong temperature increase in mainland Spain.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

For both variables, Tx and Tn, the observed changes have been mainly associated to changes in the location parameter rather than in the scale parameter of the GEV distribution. Furthermore, different behaviour of the changes estimated in the location parameter has been observed for each period.

During the period 1940-1972, a negative trend in the location parameter for both extreme distributions (Tx and Tn) was detected in almost all locations, although statistical significance was reached only in a few of them (9 and 5 respectively) at least at the 0.05 significance level. This change determines a shift of the extreme distribution towards lower values and, consequently, diminishes all the temperature extremes in terms of their occurrence and intensity. The highest negative trends in the location parameter of Tx extreme distribution exceed -2ºC at the stations of Salamanca, Valladolid and Badajoz situated in the western of Spain along with Burgos and Huesca located in the north and north-eastern Spain, respectively. The highest negative trend in the location parameter of Tn extreme distribution exceeded -1ºC in Murcia, Badajoz and La Coruña, which are located in the south-eastern, in the western and in the extreme north-west of the IP respectively, along with Sevilla situated in the south and exceeding 2ºC.

For the period 1973-2010, the trend in the location parameter was positive and more stations presented statistical significance, especially for the Tn extreme distribution, which showed an evident time tendency in most of the locations at the 0.001 significance level.

These results suggest that extreme distributions have shifted towards warmer values that lead to an increase in the occurrence and magnitude of hot extreme events. Tx extreme distribution showed high positive trends in the location parameter in several Spanish locations. The coastal locations of Valencia, Malaga and La Coruña along with the north-western locations of Salamanca and Burgos exhibit the highest values, above 3.5ºC for Valencia and above 2.5ºC for the rest. Regarding the Tn extreme distribution, the results returned the highest positive trends at stations over southern part of IP and Valladolid, where trends in the location parameter exceeded 3ºC (4ºC in the case of Sevilla), whilst Albacete, Madrid and Granada trends exceeded 2ºC.

In summary, decrease in VHD and VHN have been detected for mainland Spain over the period 1940-1972, while for the period 1973-2010 a meaningful increase of VHD and specially of VHN have been observed.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

To know how these extreme events could be in the future, the 5, 10, 20, 50 and 100-year return levels of maximum and minimum extreme temperatures have been estimated, taking into account time tendency in the location parameter. The results suggest increases of extreme temperatures in all the analysed series. Particularly for the 20-years return level analysis, the highest values were estimated over southern Spain for Tx extremes and over southern and also inland and coastal locations for Tn extremes. However, the largest increases in daily maximum extreme temperatures have been found in the north coast of Spain, in La Coruña and San Sebastian exceeding 4.4ºC and 5.1ºC respectively, whilst the largest increases in daily minimum extreme temperatures have been found in Huesca and Pamplona, both located over north-eastern Spain, with 3.0 ºC and 3.1 ºC respectively.

5.3 Conclusions and outlook

This thesis provides a statistical modelling study on extreme temperature events over mainland Spain. In the analysis, relationships between VHD and VHN and large-scale field anomalies of SST, SLP and SM have been investigated. Observed change in the statistical characteristics of these extreme events during two different periods (1940-1972 and 1973-2010) was analysed. Finally, some return values of extreme temperatures both related to warm Tx and Tn were estimated.

The methodology used in this study enabled not only knowing the mean state of the dynamical and physical forcing factors during the VHD and VHN, but also understanding specific large-scale conditions affecting the intensity and frequency of the extreme temperatures events in each location.

The application of the same approach for Tx and Tn at the 21 stations enabled the identification and assignment of the role played by SST, SLP and SM anomalies on the local extreme temperature events, along with their associated anomaly patterns.

From the analysis the following conclusions have been drawn. There are three large-scale atmospheric circulation patterns associated to extreme temperature events in mainland Spain. The three patterns are basically the same for both VHD and VHN, although stronger SLP anomalies during VHD are evident from the analysis. The first is the Southerly Flow Pattern, which enables a southerly component inflow of warm and dry air masses from Africa to affect the IP; the second is the Weak South-westerly

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Airflow Pattern, which forces a weak warm westerly or south-westerly airflow over the IP, and the third is the North-westerly Airflow Pattern, which returns a warmed north-westerly airflow when it crosses the IP but much weaker for VHN episodes. SST anomalies preceding an extreme temperature event have, in general, an important role in the intensity and frequency of VHD and VHN in mainland Spain, although the effect is not homogenous in space. SM deficit during the previous days of an extreme event has an important contribution to its occurrence and intensity in all the locations analysed except in the northern coast locations. Changes in extreme temperature are generalized, but not homogeneous in space and time with different behaviour at each analysed period over the whole domain. The highest rate increases of the 20-year return level for Tx and Tn extreme are expected in the northernmost part of Spain.

The main and original contribution of this work consists in the novel application of a PP approach to model extreme temperatures in mainland Spain. The PP approach provides more reliable results with less uncertainty in the estimations than other statistical models extensively used in many studies on extreme values, such as the Block Maxima approach.

According to the results, this PhD thesis represents a significant contribution to better understanding the dynamical and physical factors related to the occurrence and intensity of VHD and VHN events in mainland Spain. Also the analysis of the statistical characteristics and trends of the observed extreme temperatures gives a better perspective on their expected changes.

Related to the low density of the stations used in this thesis, a detailed regional analysis of extreme temperature events based on stations network of high spatial density should be envisaged as future work, as well as using long-term gridded climate data and Reanalysis products at daily scale, such as the new 20th Century Reanalysis (20CR) V2 (from NOAA/OAR/ESRL) or the ECMWF[6] Reanalysis product (ERA-20C). This study will benefit our knowledge on the factors forcing multi-decadal variability of extreme events and provide more complete and accurate, both spatially and temporally, results.

Also, using additional in-situ SM measurements would be essential in future research to capture local effects of this variable and provide more accurate results.

[6] ECMWF: European Centre for Medium-Range Weather Forecasts

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Given that the results showed stronger and significant role of the SLP, SST and SM anomalies in the intensity and frequency of extreme temperatures in Spain, a complementary research not undertaken in this thesis, but worth trying in the future, would be to take into consideration the connections of former large-scale physical variables on the duration of VHD and VHN, i.e, the lengths of spells of extreme temperatures.

The fixed threshold could be replaced in future work by a threshold that would change over time according to the long-term trend of summer temperature extremes, such as Brown et al. (2008) applied in their study on global changes in extreme daily temperatures.

Finally, another interesting topic is projections of these extreme events (VHD and VHN) under different climate change scenarios.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

6 References

Abaurrea, J., J. Asín, A. C. Cebrián, and A. Centelles (2007), Modeling and forecasting extreme hot events in the central Ebro valley, a continental-Mediterranean area, *Global Planet. Change, 57*, 43-58.

Abaurrea, J. and A. C. Cebrián (2002), Drought analysis based on a cluster Poisson model: Distribution of the most severe drought, *Climate Research, 22*, 227-235.

AghaKouchak, A., D. Easterling, and K. Hsu (2013), *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*, vol. 65, Springer.

Alexander, L., X. Zhang, T. Peterson, J. Caesar, B. Gleason, A. Klein Tank, M. Haylock, D. Collins, B. Trewin, and F. Rahimzadeh (2006), Global observed changes in daily climate extremes of temperature and precipitation, *Journal of Geophysical Research: Atmospheres (1984–2012), 111*.

Andrade, C., S. Leite, and J. Santos (2012), Temperature extremes in Europe: overview of their driving atmospheric patterns, *Natural Hazards and Earth System Science, 12*, 1671-1691.

Beguería, S., S. M. Vicente-Serrano, and M. Angulo-Martínez (2010), A Multiscalar Global Drought Dataset: The SPEIbase: A New Gridded Product for the Analysis of Drought Variability and Impacts, *Bull. Am. Meteorol. Soc., 91*, 1351-1356.

Beirlant, J. (2004), *Statistics of Extremes: Theory and Applications*, 490 pp., John Wiley & Sons, Chichester.

Beniston, M. and H. F. Diaz (2004), The 2003 heat wave as an example of summers in a greenhouse climate? Observations and climate model simulations for Basel, Switzerland, *Global Planet. Change, 44*, 73-81.

Brabson, B., D. Lister, P. Jones, and J. Palutikof (2005), Soil moisture and predicted spells of extreme temperatures in Britain, *Journal of Geophysical Research, 110*.

Brown, S., J. Caesar, and C. Ferro (2008), Global changes in extreme daily temperature since 1950, *Journal of Geophysical Research: Atmospheres (1984–2012), 113*.

Brunet, M., P. D. Jones, J. Sigró, O. Saladié, E. Aguilar, A. Moberg, P. M. Della□ Marta, D. Lister, A. Walther, and D. López (2007a), Temporal and spatial temperature variability and change over Spain during 1850–2005, *Journal of Geophysical Research: Atmospheres (1984–2012), 112*.

Brunet, M., O. Saladié, P. Jones, J. Sigró, E. Aguilar, A. Moberg, D. Lister, A. Walther, and C. Almarza (2008), A case-study/guidance on the development of long-term daily adjusted temperature datasets, *World Meteorological Organization WCDMP-66/WMO-TD-1425, Geneva*.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Brunet, M., J. Sigró, P. D. Jones, O. Saladié, E. Aguilar, A. Moberg, D. Lister, and A. Walther (2007b), Long-term changes in extreme temperatures and precipitation in Spain, *Contrib Sci*, *3*, 331-342.

Brunet, M., O. Saladié, P. Jones, J. Sigró, E. Aguilar, A. Moberg, D. Lister, A. Walther, D. Lopez, and C. Almarza (2006), The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850?2003), *Int. J. Climatol.*, *26*, 1777-1802.

Carril, A. F., S. Gualdi, A. Cherchi, and A. Navarra (2008), Heatwaves in Europe: areas of homogeneous variability and links with the regional to large-scale atmospheric and SSTs anomalies, *Clim. Dyn.*, *30*, 77-98.

Cassou, C., L. Terray, and A. S. Phillips (2005), Tropical Atlantic influence on European heat waves, *J. Clim.*, *18*, 2805-2811.

Clark, R. T., S. J. Brown, and J. M. Murphy (2006), Modeling Northern Hemisphere summer heat extreme changes and their uncertainties using a physics ensemble of climate sensitivity experiments, *J. Clim.*, *19*, 4418-4435.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, 208 pp., Springer, London.

Colman, A. and M. Davey (1999), Prediction of summer temperature, rainfall and pressure in Europe from preceding winter North Atlantic Ocean temperature, *Int. J. Climatol.*, *19*, 513-536.

Colombo, A. F., D. Etkin, and B. W. Karney (1999), Climate variability and the frequency of extreme temperature events for nine sites across Canada: implications for power usage, *J. Clim.*, *12*, 2490-2502.

Dalelane, C. and T. Deutschländer (2013), A robust estimator for the intensity of the Poisson Point Process of extreme weather events, *Weather and Climate Extremes*.

Davison, A. C. and R. L. Smith (1990), Models for exceedances over high thresholds, *Journal of the Royal Statistical Society.Series B (Methodological)*, 393-442.

Della-Marta, P., J. Luterbacher, H. von Weissenfluh, E. Xoplaki, M. Brunet, and H. Wanner (2007), Summer heat waves over western Europe 1880–2003, their relationship to large-scale forcings and predictability, *Clim. Dyn.*, *29*, 251-275.

Donat, M., L. Alexander, H. Yang, I. Durre, R. Vose, R. Dunn, K. Willett, E. Aguilar, M. Brunet, and J. Caesar (2013), Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset, *Journal of Geophysical Research: Atmospheres*.

El Kenawy, A., J. López-Moreno, and S. Vicente-Serrano (2011), Recent trends in daily temperature extremes over northeastern Spain (1960–2006), *Nat.Hazards Earth Syst.Sci*, *11*, 2583-2603.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

El Kenawy, A., J. I. López-Moreno, N. A. Brunsell, and S. M. Vicente-Serrano (2012a), Anomalously severe cold nights and warm days in northeastern Spain: their spatial variability, driving forces and future projections, *Global Planet. Change*.

El Kenawy, A., J. I. López-Moreno, and S. M. Vicente-Serrano (2012b), Summer temperature extremes in northeastern Spain: spatial regionalization and links to atmospheric circulation (1960-2006), *Theoretical and Applied Climatology*, *1*, 232.

Ferro, C. A. and J. Segers (2003), Inference for clusters of extreme values, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*, 545-556.

Feudale, L. and J. Shukla (2007), Role of Mediterranean SST in enhancing the European heat wave of summer 2003, *Geophys. Res. Lett.*, *34*, L03811.

Feudale, L. and J. Shukla (2011), Influence of sea surface temperature on the European heat wave of 2003 summer. Part I: an observational study, *Clim. Dyn.*, *36*, 1691-1703.

Fischer, E. M., S. Seneviratne, P. Vidale, D. Lüthi, and C. Schär (2007), Soil moisture-atmosphere interactions during the 2003 European summer heat wave, *J. Clim.*, *20*, 5081-5099.

Fisher, R. A. and L. H. C. Tippett (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, Mathematical Proceedings of the Cambridge Philosophical Society.

Furió, D. and V. Meneu (2011), Analysis of extreme temperatures for four sites across Peninsular Spain, *Theoretical and applied climatology*, *104*, 83-99.

Furrer, E. M. (2010), Statistical modeling of hot spells and heat waves, *Climate research*, *43*, 191-205.

Galiatsatou, P. and P. Prinos (2011), Modeling non-stationary extreme waves using a point process approach and wavelets, *Stochastic Environmental Research and Risk Assessment*, *25*, 165-183.

García-Herrera, R., J. Díaz, R. Trigo, and E. Hernández (2005), Extreme summer temperatures in Iberia: health impacts and associated synoptic conditions, *Ann. Geophys.*, *23*, 239-251.

Garcia-Herrera, R., J. Diaz, R. Trigo, J. Luterbacher, and E. Fischer (2010), A Review of the European Summer Heat Wave of 2003, *Crit. Rev. Environ. Sci. Technol.*, *40*, 267-306.

Gilleland, E. and R. W. Katz (2011), New software to analyze how extremes change over time, *Eos, Transactions American Geophysical Union*, *92*, 13.

Gumbel, E. J. (1958), *Statistics of Extremes*, DoverPublications. com.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Haan, L. d. L. and A. Ferreira (2006), *Extreme Value Theory*, Springer.

Hatfield, J. L., K. J. Boote, B. Kimball, L. Ziska, R. C. Izaurralde, D. Ort, A. M. Thomson, and D. Wolfe (2011), Climate impacts on agriculture: implications for crop production, *Agron. J., 103*, 351-370.

Hirschi, M., S. I. Seneviratne, V. Alexandrov, F. Boberg, C. Boroneant, O. B. Christensen, H. Formayer, B. Orlowsky, and P. Stepanek (2010), Observational evidence for soil-moisture impact on hot extremes in southeastern Europe, *Nature Geoscience, 4*, 17-21.

IPCC, 2012*: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.

Jaeger, E. and S. Seneviratne (2011), Impact of soil moisture–atmosphere coupling on European climate extremes and trends in a regional climate model, *Clim. Dyn., 36*, 1919-1939.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, and J. Woollen (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc., 77*, 437-471.

Katsoulis, B. and N. Hatzianastassiou (2005), Analysis of hot spell characteristics in the Greek region, *Clim. Res., 28*, 229-241.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour., 25*, 1287-1304.

Kharin, V. V. and F. W. Zwiers (2005), Estimating extremes in transient climate change simulations, *J. Clim., 18*, 1156-1173.

Klein Tank, A. and G. Können (2003), Trends in indices of daily temperature and precipitation extremes in Europe, 1946-99, *J. Clim., 16*, 3665-3680.

Klein Tank, A. M. G., F. W. Zwiers, and X. Zhang, 2009: Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. Climate Data and Monitoring Rep. WCDMP 72, WMO-TD 1500, 56 pp.

Lana, X., M. Martínez, A. Burgueño, and C. Serra (2009), Statistics of hot and cold events in Catalonia (NE Spain) for the recording period 1950–2004, *Theoretical and applied climatology, 97*, 135-150.

Lorenz, R., E. B. Jaeger, and S. I. Seneviratne (2010), Persistence of heat waves and its link to soil moisture memory, *Geophys. Res. Lett., 37*.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Lorenzo-Lacruz, J., S. M. Vicente-Serrano, J. I. López-Moreno, S. Beguería, J. M. García-Ruiz, and J. M. Cuadrat (2010), The impact of droughts and water management on various hydrological systems in the headwaters of the Tagus River (central Spain), *Journal of Hydrology, 386*, 13-26.

Maraun, D., T. J. Osborn, and H. W. Rust (2011), The influence of synoptic airflow on UK daily precipitation extremes. Part I: Observed spatio-temporal relationships, *Clim. Dyn., 36*, 261-275.

Meehl, G. and C. Tebaldi (2004), More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century, *Science, 305*, 994-997.

Moberg, A. and P. D. Jones (2005), Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99, *Int. J. Climatol., 25*, 1149-1171.

Moberg, A., P. D. Jones, D. Lister, A. Walther, M. Brunet, J. Jacobeit, L. V. Alexander, P. M. Della-Marta, J. Luterbacher, and P. Yiou (2006), Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000, *Journal of Geophysical Research: Atmospheres (1984–2012), 111*.

Prasad, P., K. Boote, L. Allen Jr, J. Sheehy, and J. Thomas (2006), Species, ecotype and cultivar differences in spikelet fertility and harvest index of rice in response to high temperature stress, *Field Crops Res., 95*, 398-411.

Reiss, R. and M. Thomas (2007), *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields*, Springer.

Ren, F., D. Cui, Z. Gong, Y. Wang, X. Zou, Y. Li, S. Wang, and X. Wang (2012), An objective identification technique for regional extreme events, *J. Clim., 25*, 7015-7027.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of Climate*, *20*(22).

Rodríguez-Puebla, C., S. Ayuso, M. Frias, and L. Garcia-Casado (2007), Effects of climate variation on winter cereal production in Spain, *Climate Research, 34*, 223.

Rodríguez-Puebla, C., A. H. Encinas, L. A. García-Casado, and S. Nieto (2010), Trends in warm days and cold nights over the Iberian Peninsula: relationships to large-scale variables, *Clim. Change, 100*, 667-684.

Serra, C., M. Martínez, X. Lana, and A. Burgueño (2010), Extreme normalised residuals of daily temperatures in Catalonia (NE Spain): sampling strategies, return periods and clustering process, *Theoretical and Applied Climatology, 101*, 1-17.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Sheffield, J. and E. F. Wood (2008), Global trends and variability in soil moisture and drought characteristics, 1950-2000, from observation-driven simulations of the terrestrial hydrologic cycle, *J. Clim., 21*, 432-458.

Sillmann, J., M. Croci-Maspoli, M. Kallache, and R. W. Katz (2011), Extreme Cold Winter Temperatures in Europe under the Influence of North Atlantic Atmospheric Blocking, *Journal of Climate, 24*, 5899.

Smith, R. L. (2003), Statistics of extremes, with applications in environment, insurance and finance, *Extreme values in finance, telecommunications and the environment*, 1-78.

Smith, R. L. and I. Weissman (1994), Estimating the extremal index, *Journal of the Royal Statistical Society.Series B (Methodological)*, 515-528.

Stephenson, A. (2004), A user's guide to the 'EVD'package (version 2.1), *Department of Statistics.Macquarie University.Australia*.

Tobías, A., B. Armstrong, I. Zuza, A. Gasparrini, C. Linares, and J. Diaz (2012), Mortality on extreme heat days using official thresholds in Spain: a multi-city time series analysis, *BMC Public Health, 12*, 133.

Tobías, A., P. de Olalla, C. Linares, M. Bleda, J. Caylà, and J. Díaz (2010), Short-term effects of extreme hot summer temperatures on total daily mortality in Barcelona, Spain, *International Journal of Biometeorology, 54*, 115-117.

Todorovic, P. and E. Zelenhasic (1970), A stochastic model for flood analysis, *Water Resour. Res., 6*, 1641-1648.

van Buuren, S. and K. Groothuis-Oudshoorn (2011), mice: Multivariate Imputation by Chained Equations in R, *Journal of Statistical Software, 45*.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno (2010), A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *J. Clim., 23*, 1696-1718.

Vincent, L. A., X. Zhang, B. Bonsal, and W. Hogg (2002), Homogenization of daily temperatures over Canada, *J. Clim., 15*, 1322-1334.

Walker, G. T. and E. Bliss (1932), World weather, *V.Mem.Roy.Meteor.Soc., 4*, 53-84.

Xoplaki, E., J. Gonzalez-Rouco, J. Luterbacher, and H. Wanner (2003), Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs, *Clim. Dyn., 20*, 723-739.

Zampieri, M., F. D'Andrea, R. Vautard, P. Ciais, N. de Noblet-Ducoudré, and P. Yiou (2009), Hot European summers and the role of soil moisture in the propagation of Mediterranean drought, *J. Clim., 22*, 4747-4758.

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

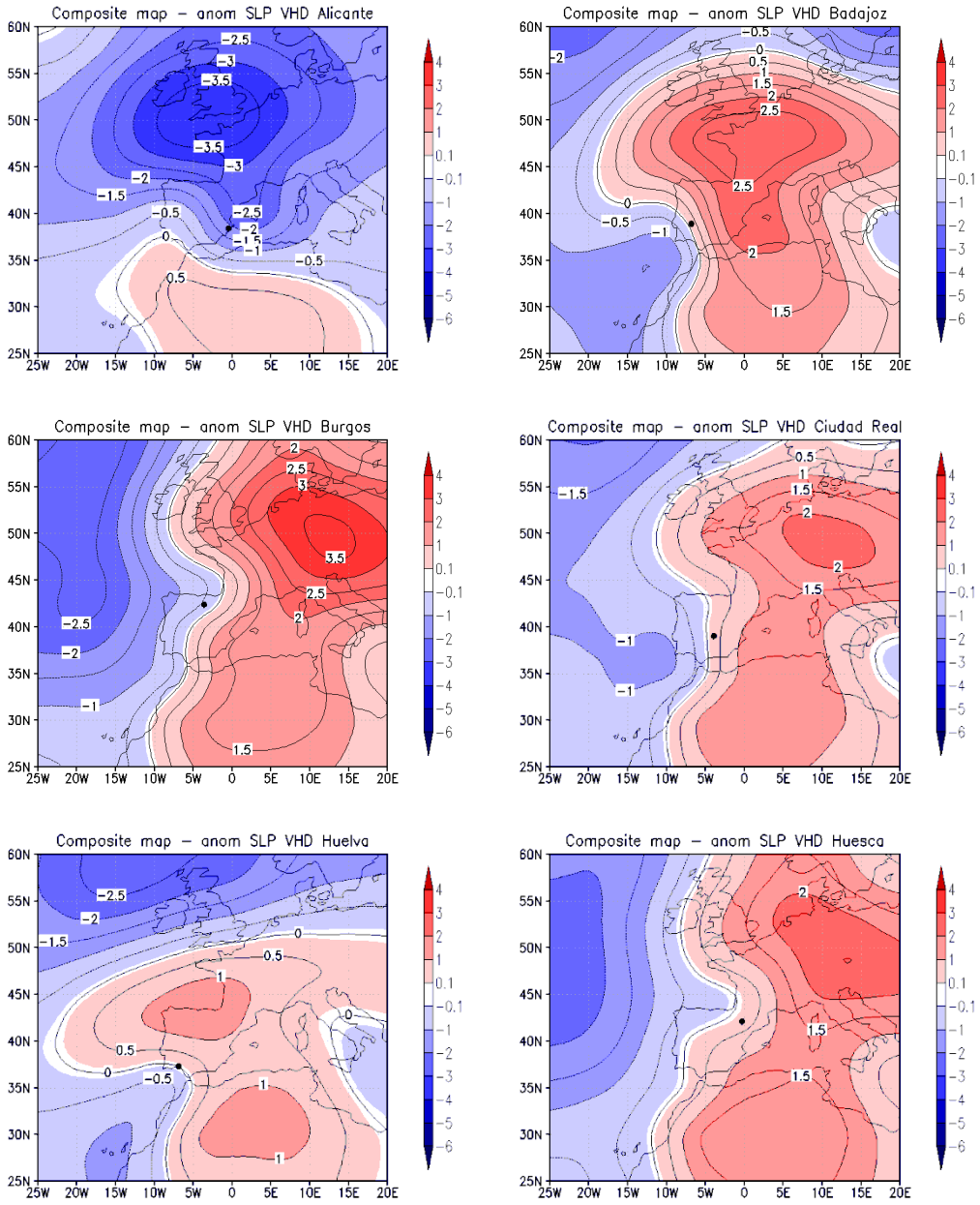
Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers (2011), Indices for monitoring changes in extremes based on daily temperature and precipitation data, *Wiley Interdisciplinary Reviews: Climate Change, 2*, 851-870.

Zwiers, F. W., L. V. Alexander, G. C. Hegerl, T. R. Knutson, J. P. Kossin, P. Naveau, N. Nicholls, C. Schär, S. I. Seneviratne, and X. Zhang (2013), Climate extremes: Challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events, in *Climate Science for Serving Society* Anonymous , pp. 339-389, Springer.

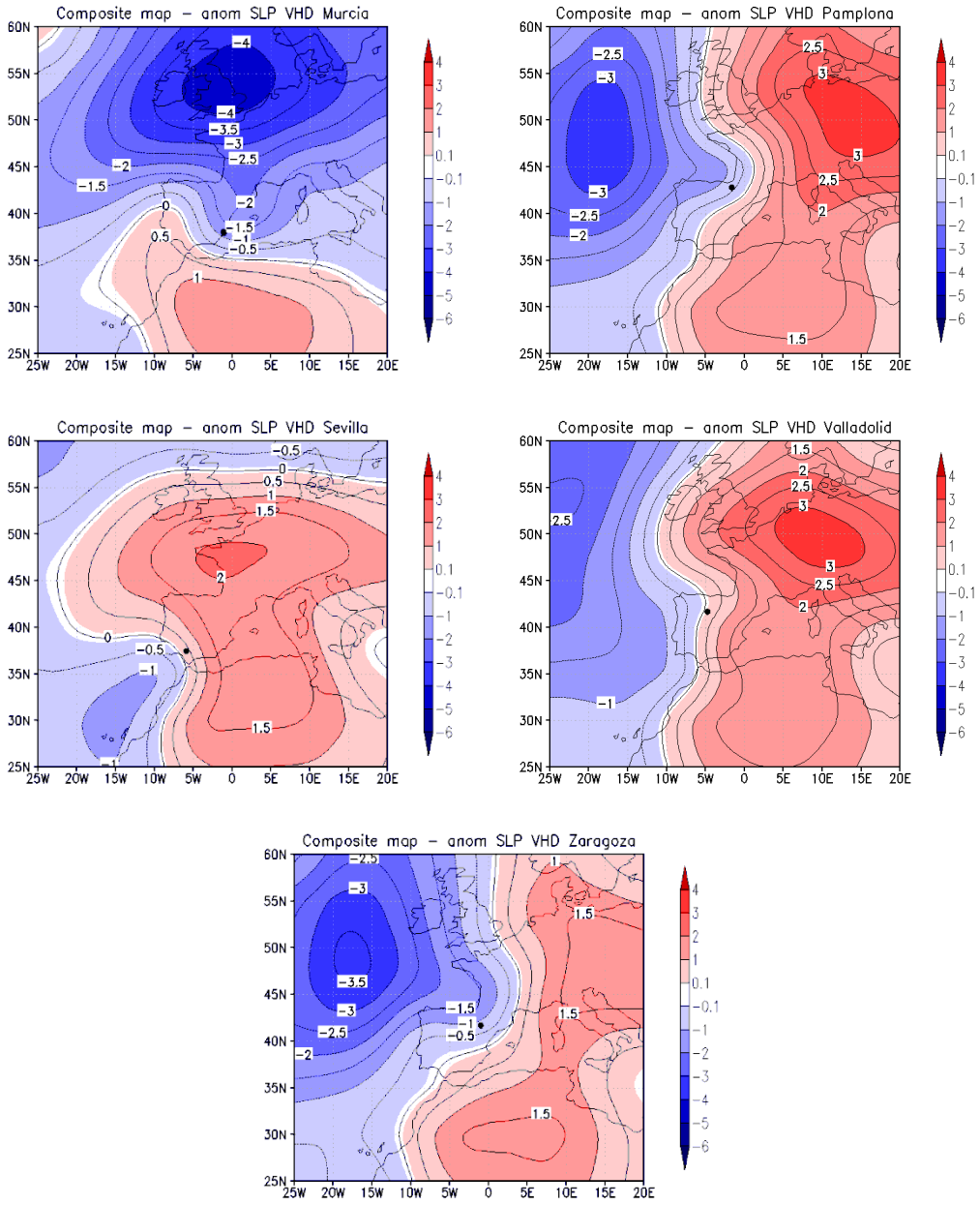
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Appendix

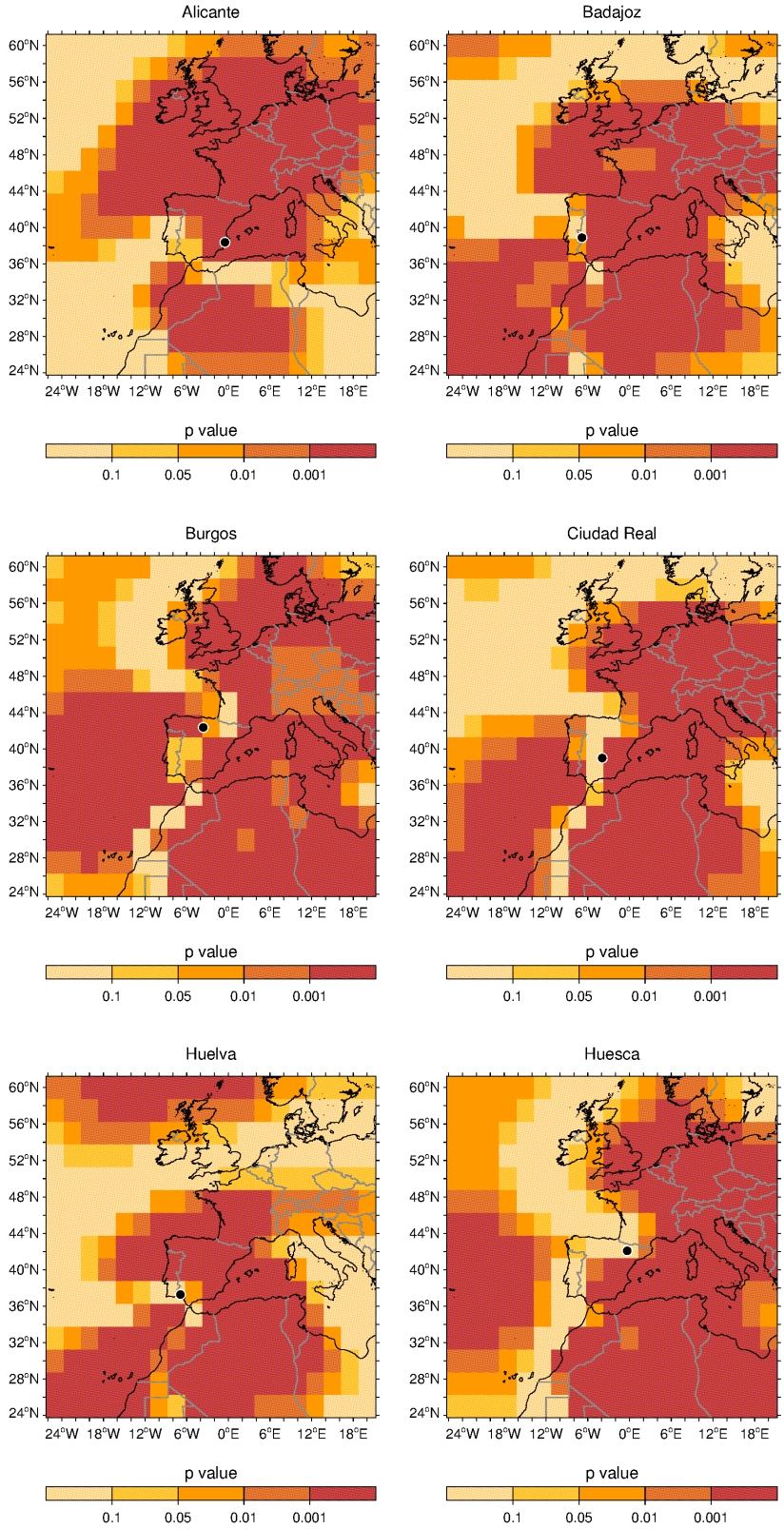
A. SLP anomaly patterns of summer VHD for the stations not shown in the text



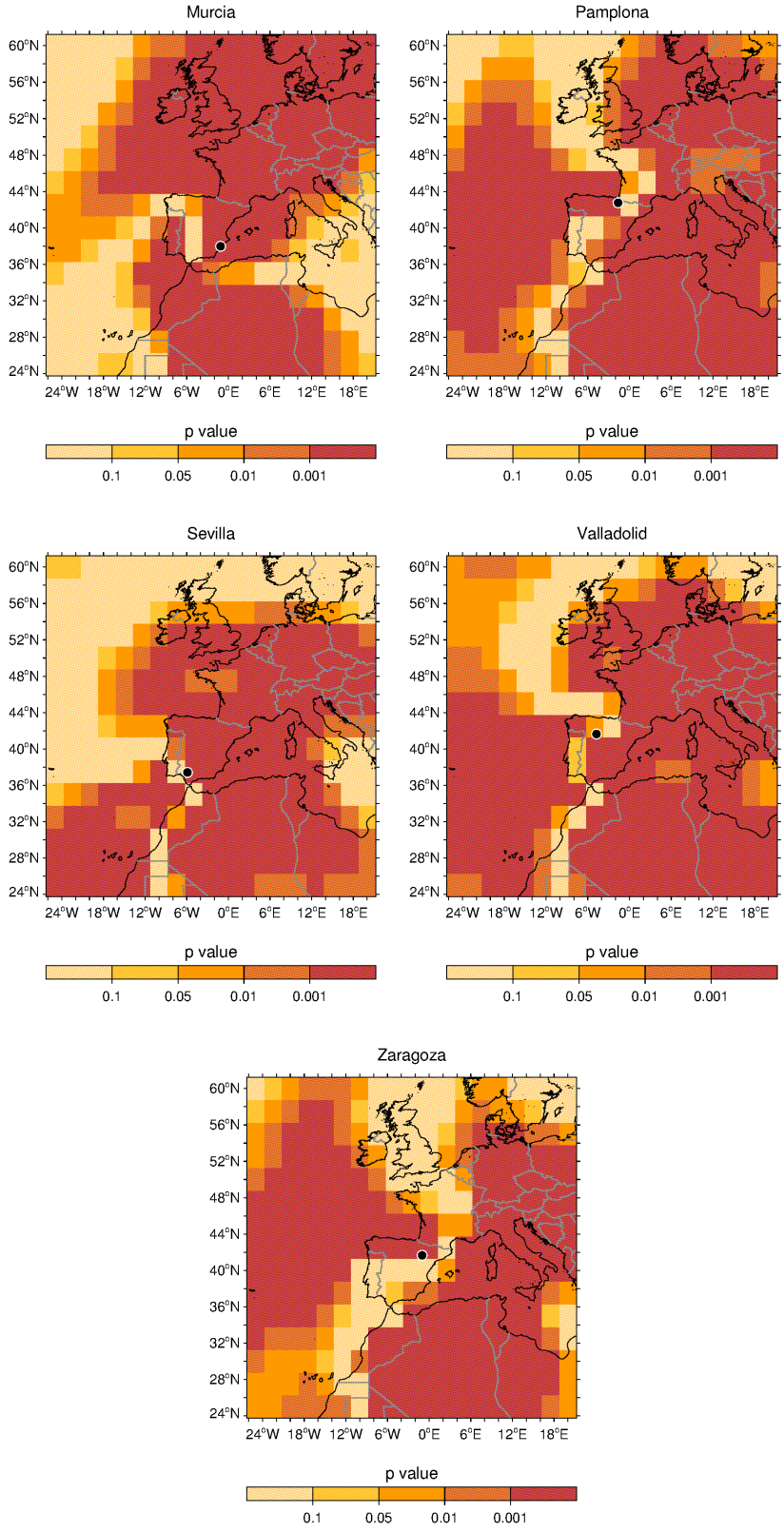
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



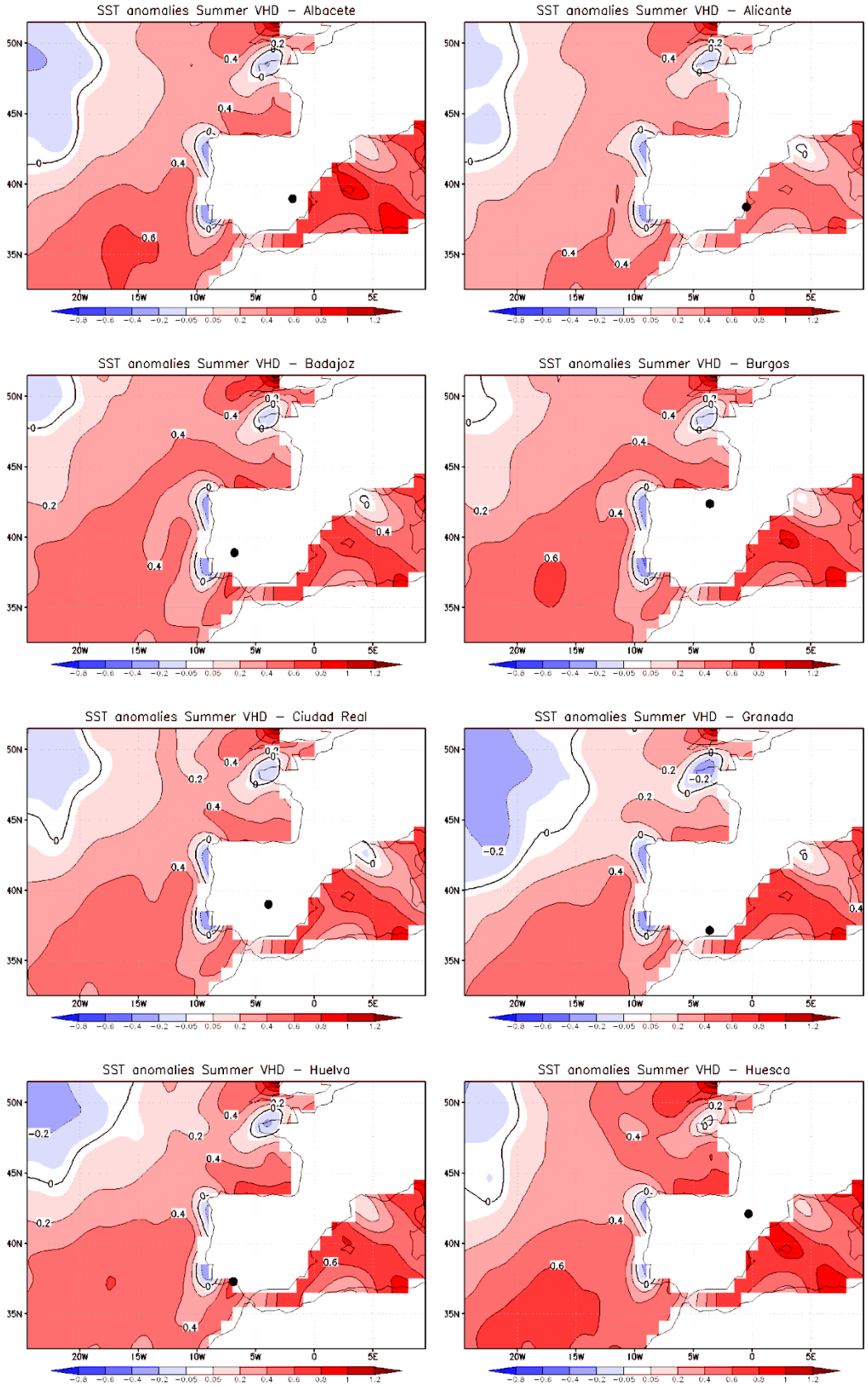
B. Gridded maps of p-values between the base model and the model with SLP anomalies as covariate in the location parameter of the Tx extreme distribution for the stations not shown in the text

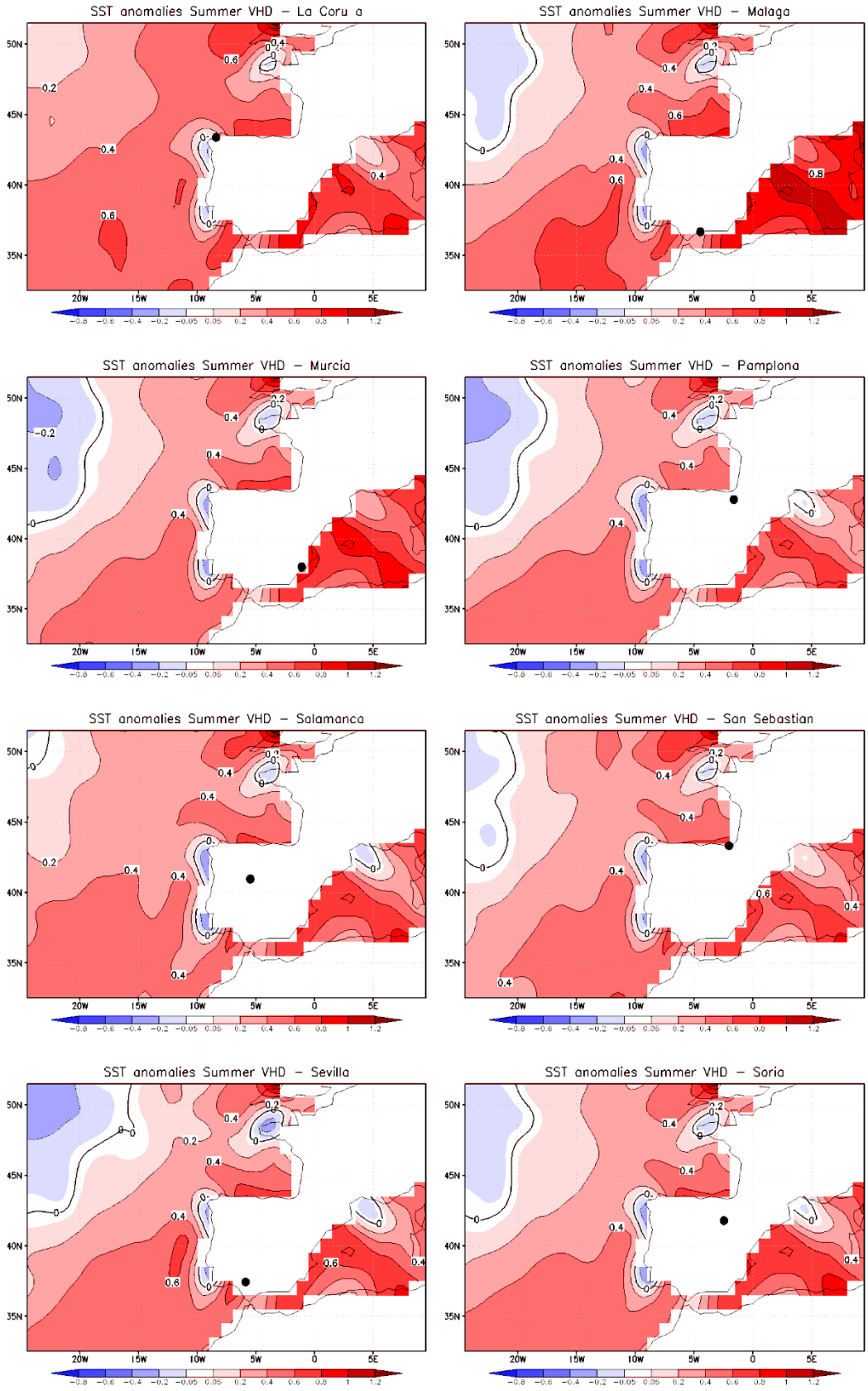


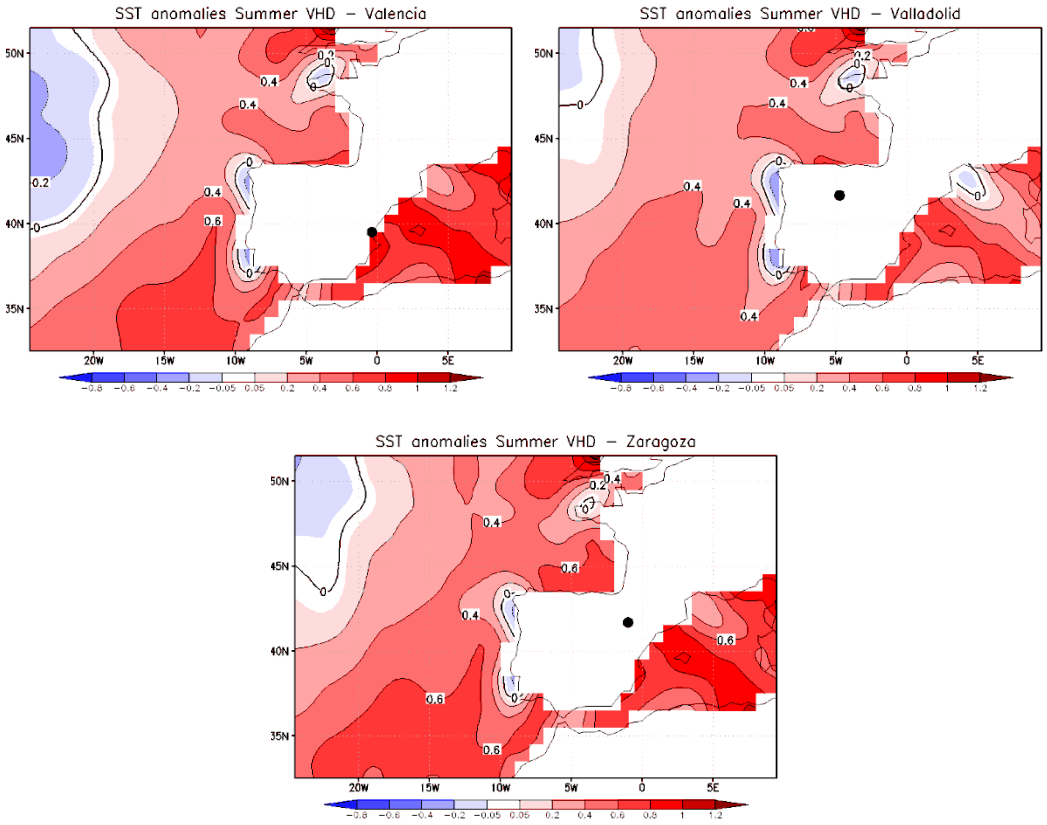
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



C. Composite maps of SST anomalies for summer VHD for the stations not shown in the text

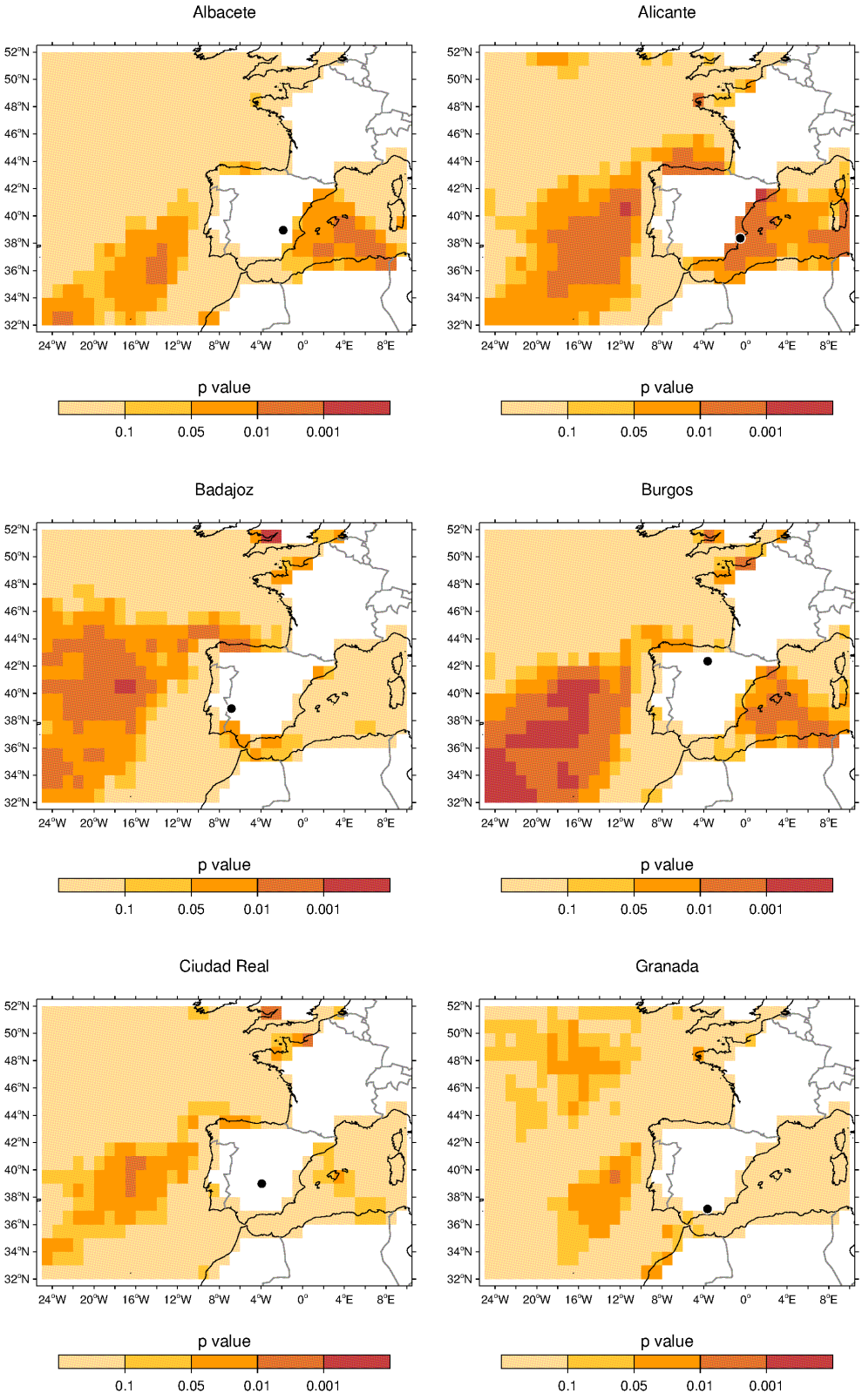



SST anomalies Summer VHD – La Coruña
SST anomalies Summer VHD – Malaga
SST anomalies Summer VHD – Murcia
SST anomalies Summer VHD – Pamplona
SST anomalies Summer VHD – Salamanca
SST anomalies Summer VHD – San Sebastian
SST anomalies Summer VHD – Sevilla
SST anomalies Summer VHD – Soria
50N
45N
40N
35N
20W
15W
10W
5W
0
5E

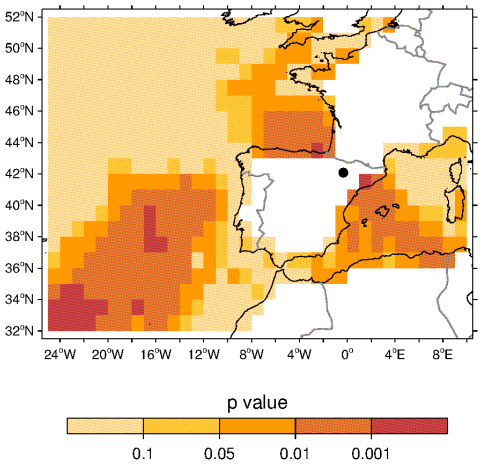

SST anomalies Summer VHD – Valencia
SST anomalies Summer VHD – Valladolid
SST anomalies Summer VHD – Zaragoza
50N
45N
40N
35N
20W
15W
10W
5W
0
5E

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

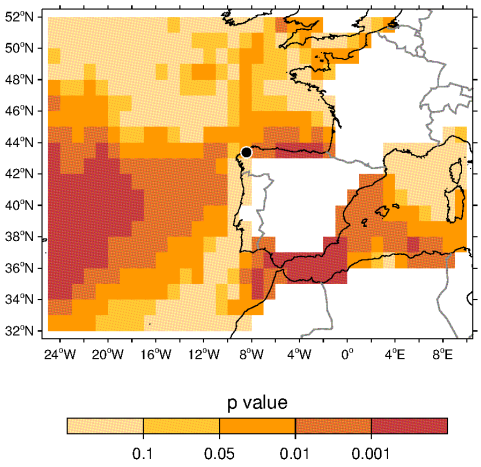
D. Gridded maps of p-values between the base model and the model with SST anomalies as covariate in the location parameter of the Tx extreme distribution for the stations not shown in the text



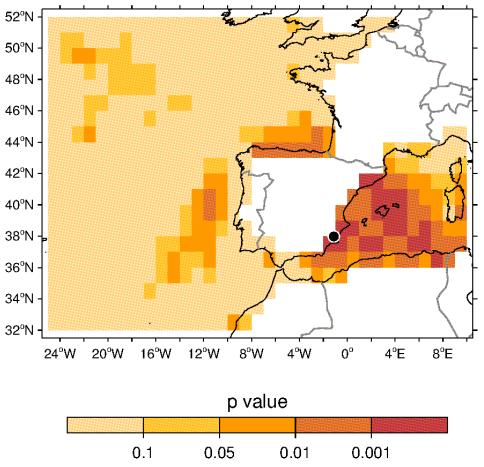
Huesca



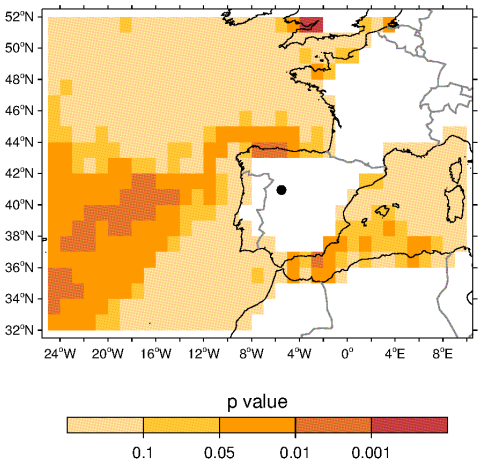
La Coruna



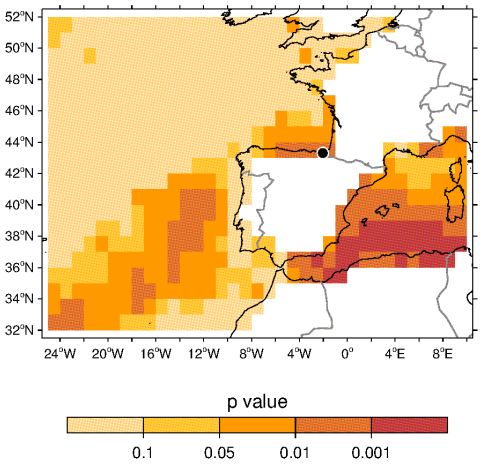
Murcia



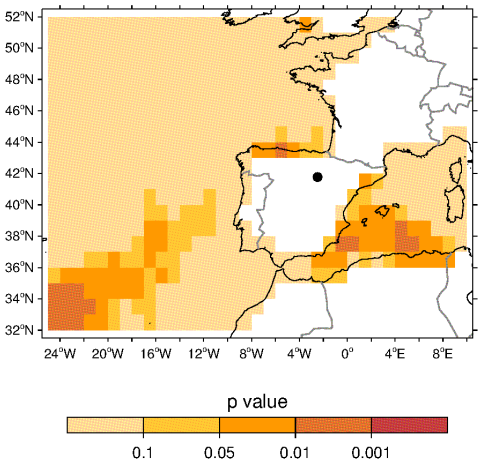
Salamanca



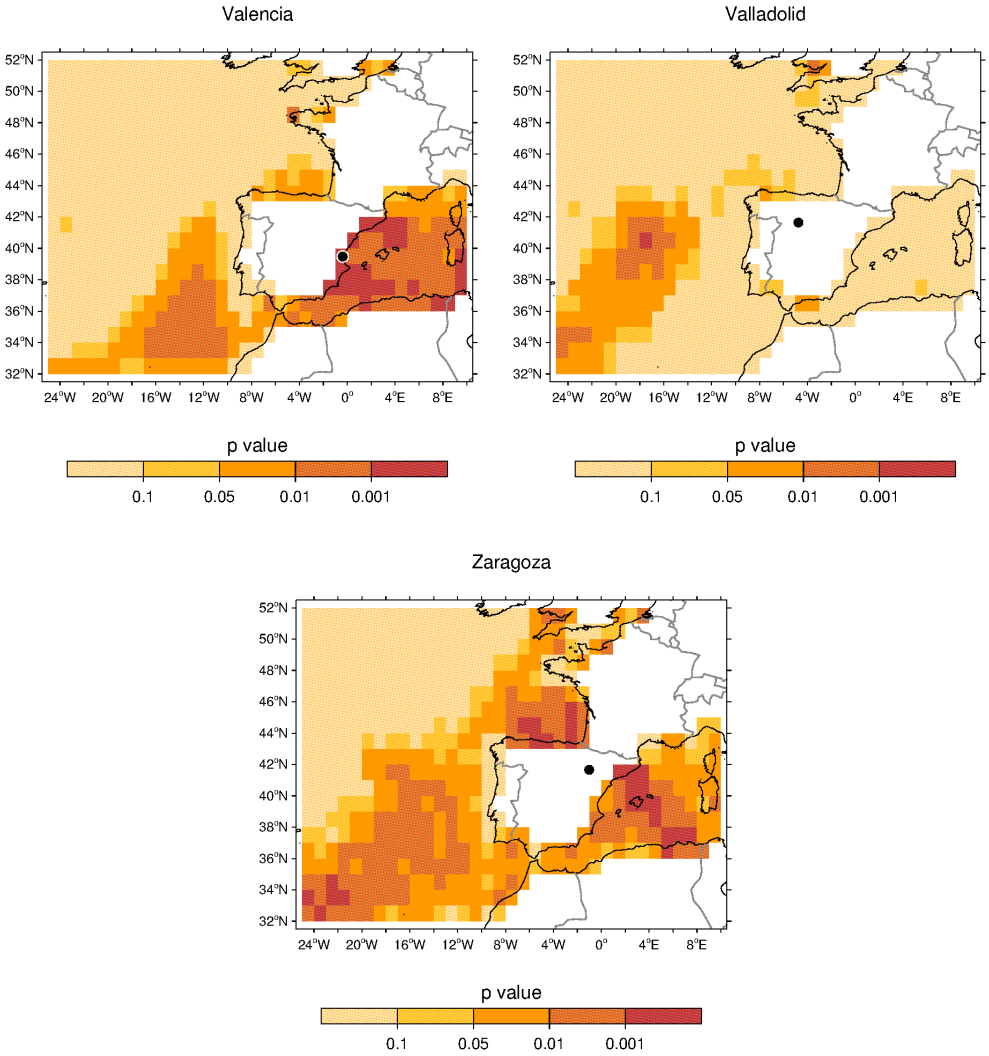
San Sebastian



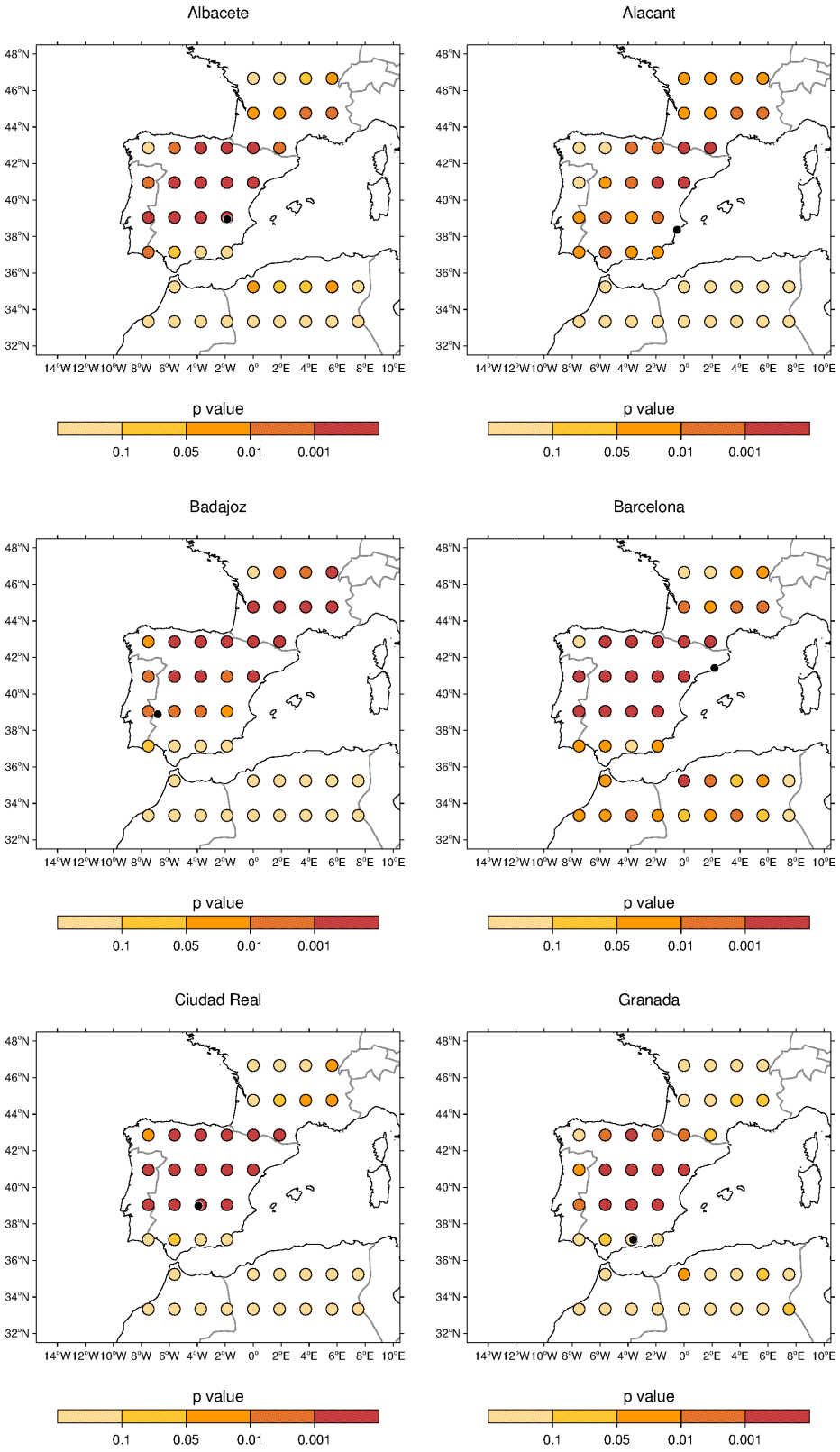
Soria



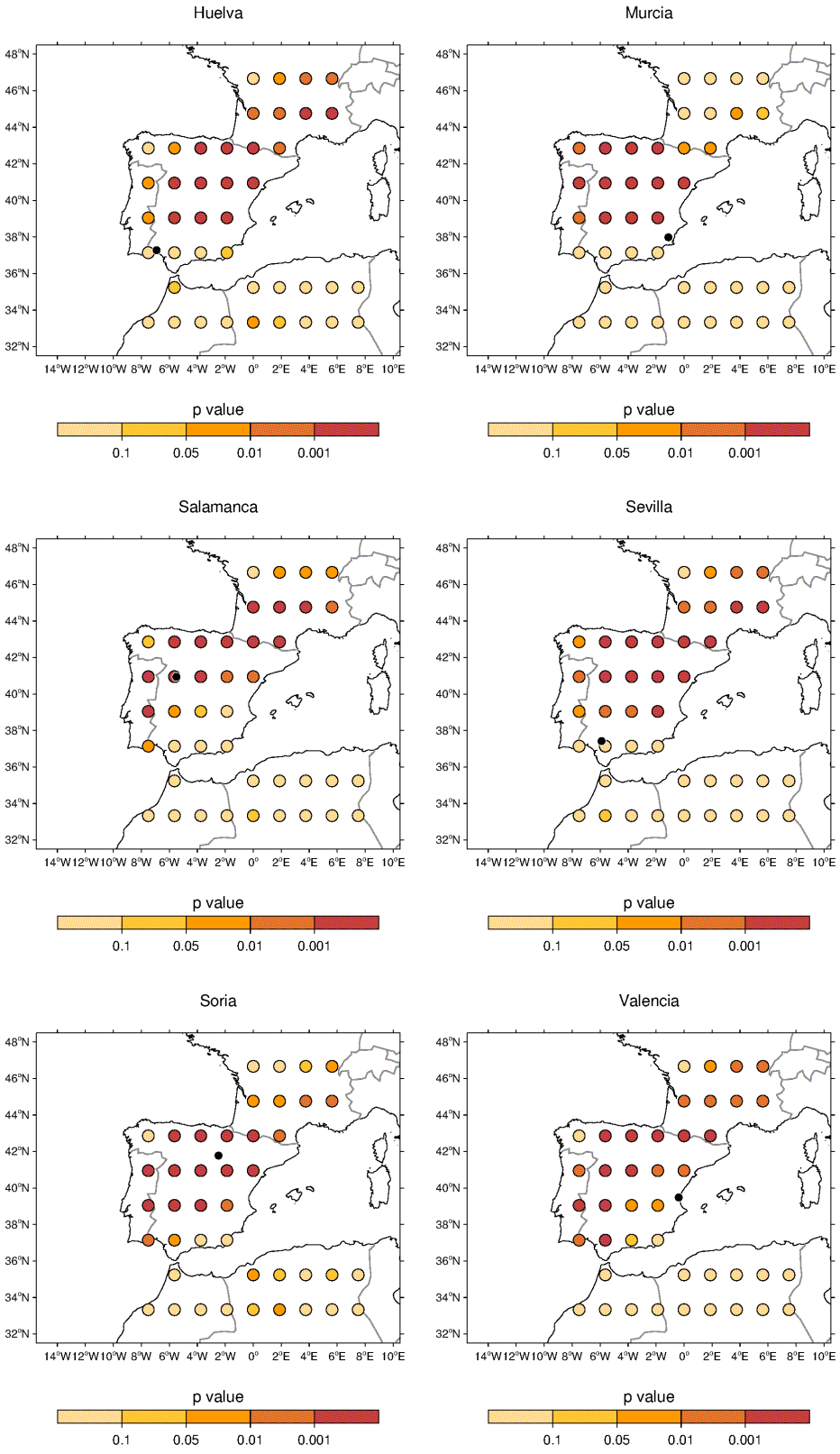
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



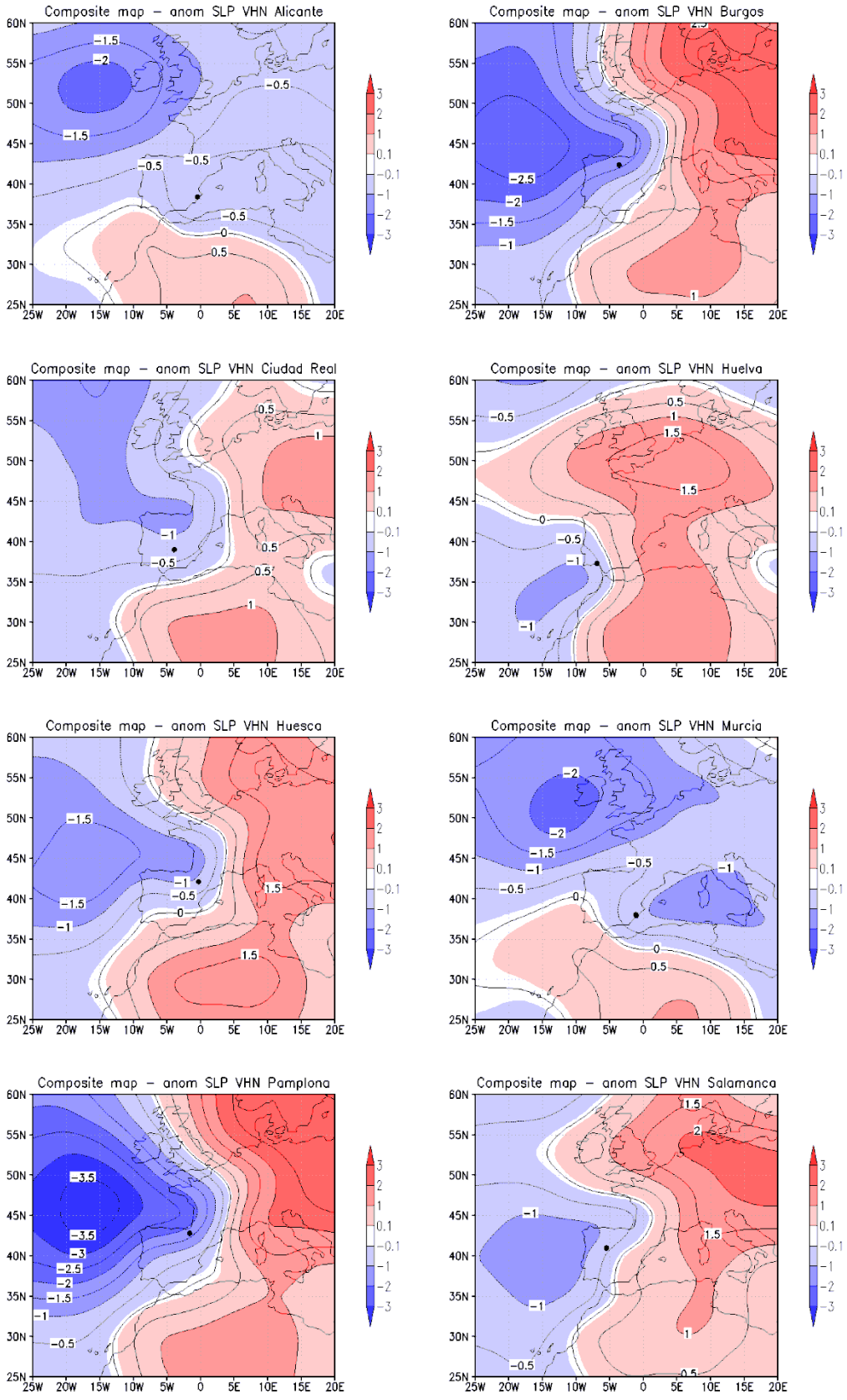
E. Maps of p-values between the base model and the model with SM anomalies as covariate in the location parameter of the Tx extreme distribution for the stations not shown in the text

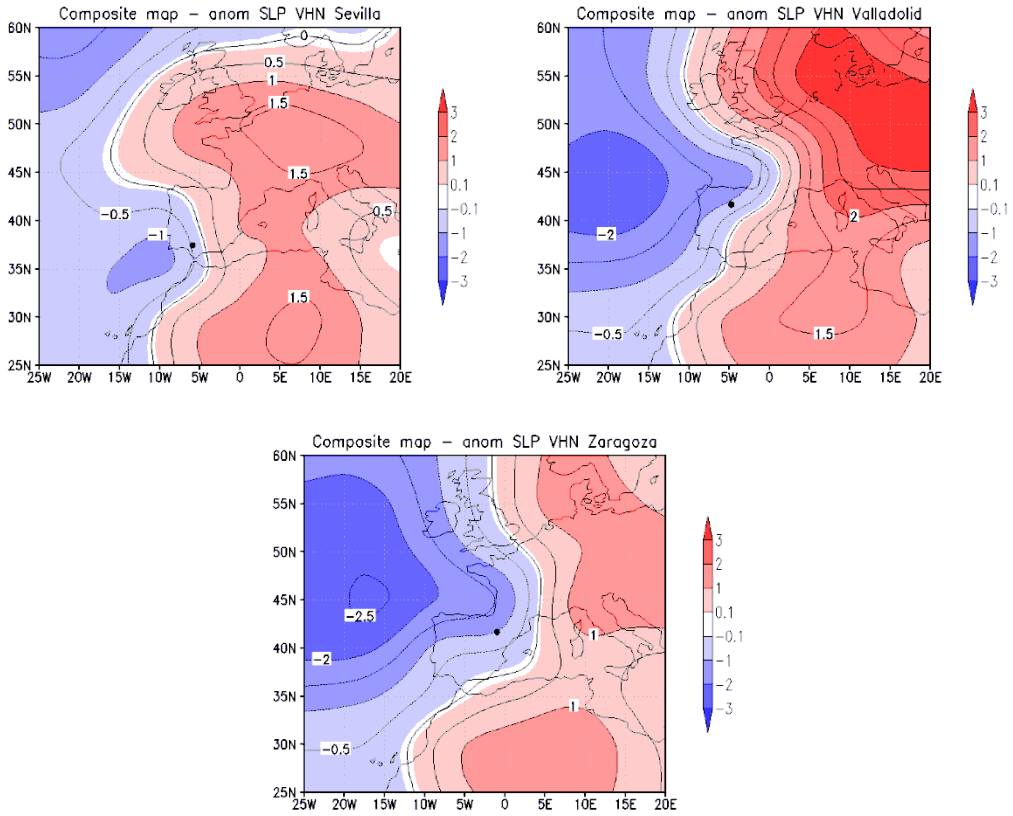


UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

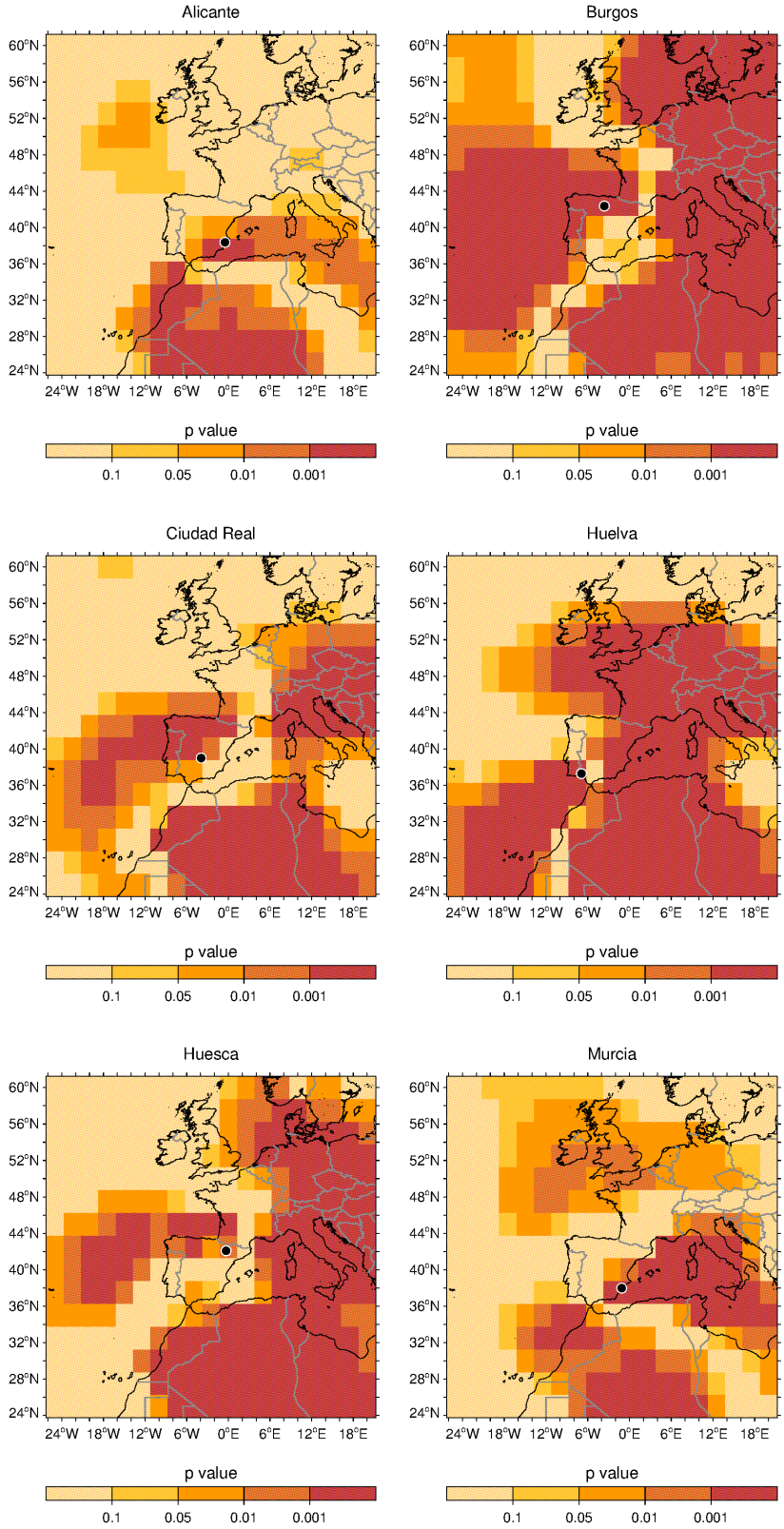


F. SLP anomaly patterns of summer VHN for the stations not shown in the text

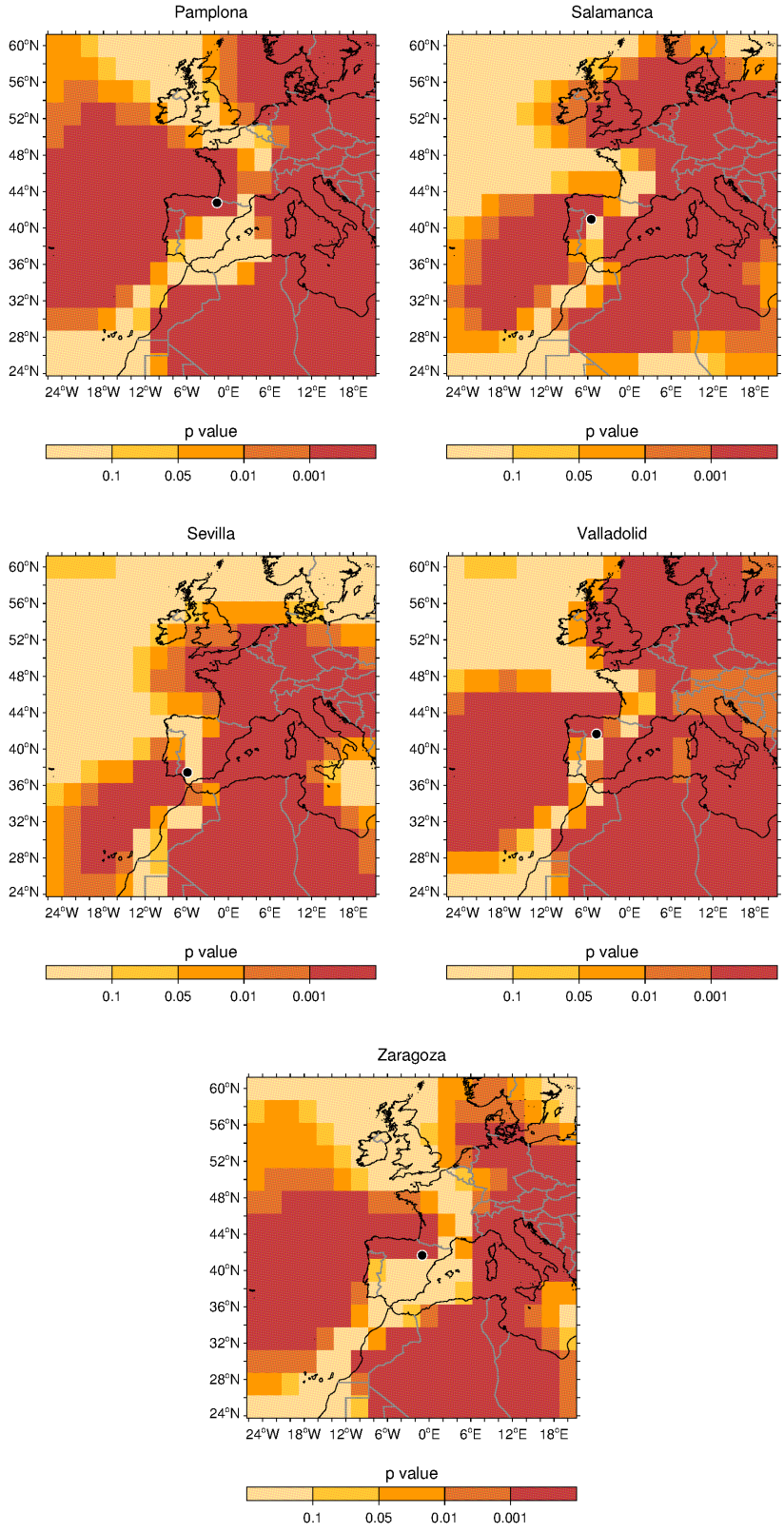



Composite map – anom SLP VHN Sevilla
Composite map – anom SLP VHN Valladolid
Composite map – anom SLP VHN Zaragoza
60N
55N
50N
45N
40N
35N
30N
25N
25W 20W 15W 10W 5W 0 5E 10E 15E 20E
3
2
1
0.1
−0.1
−1
−2
−3

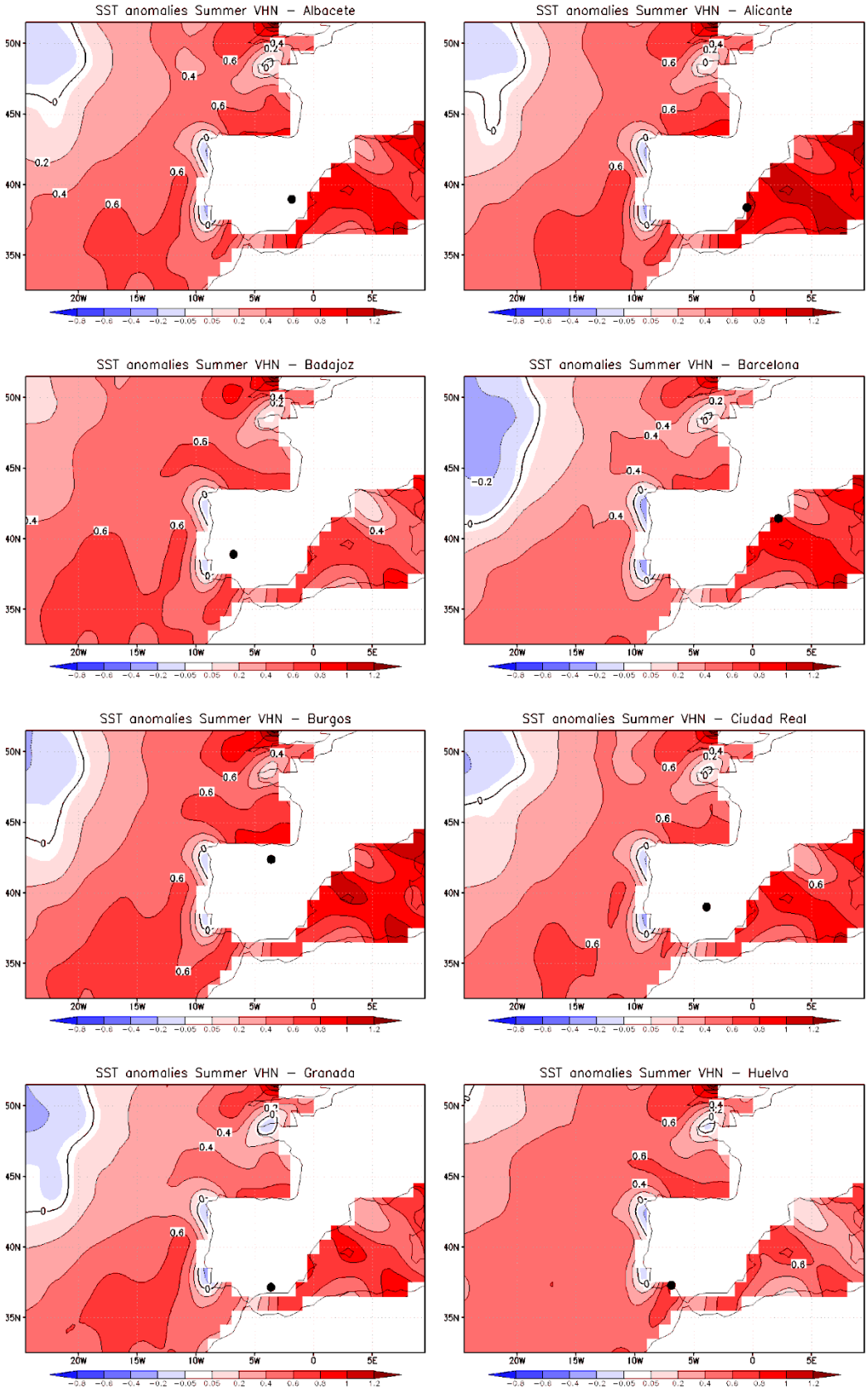
G. Gridded maps of p-values between the base model and the model with SLP anomalies as covariate in the location parameter of the Tn extreme distribution for the stations not shown in the text

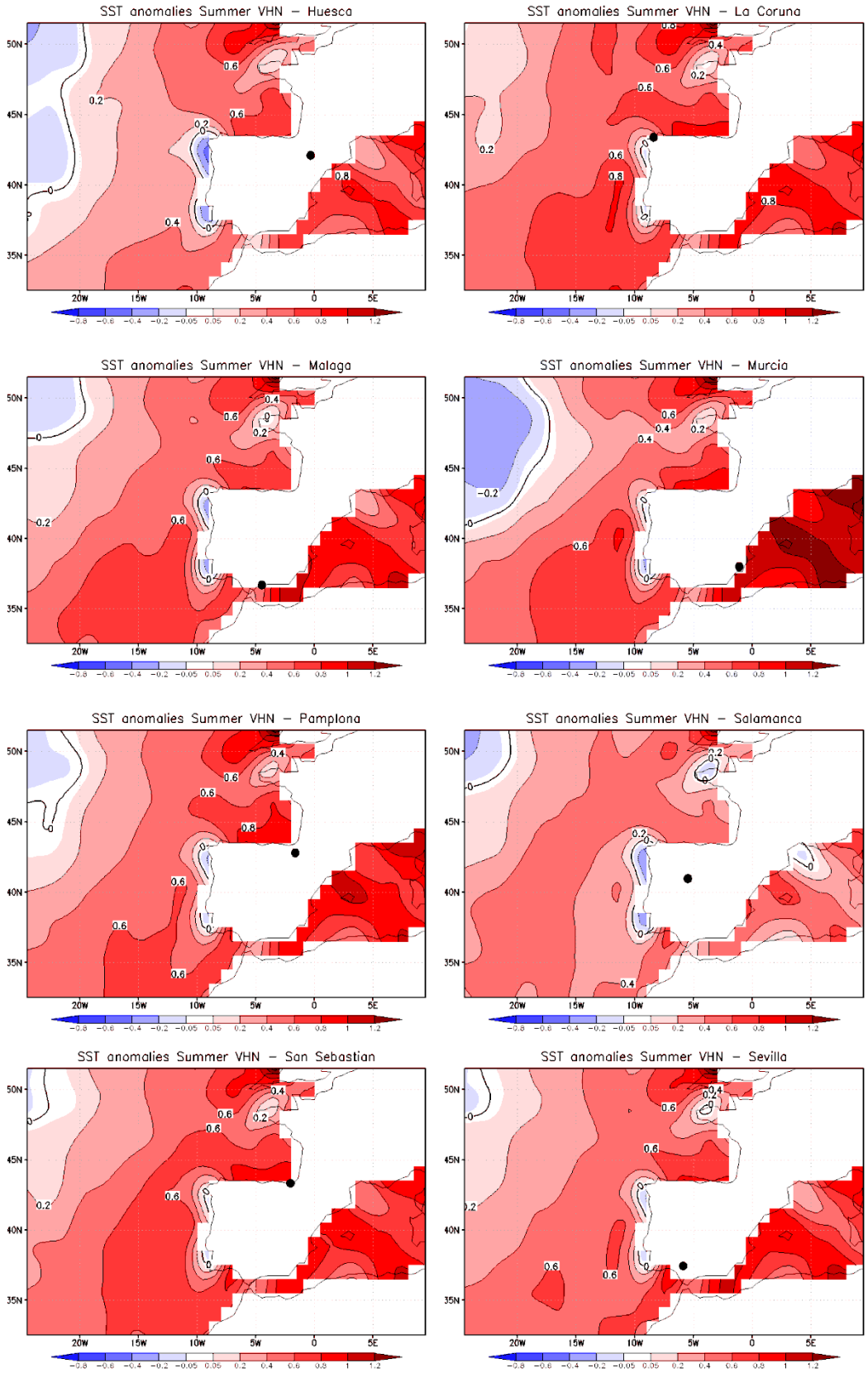


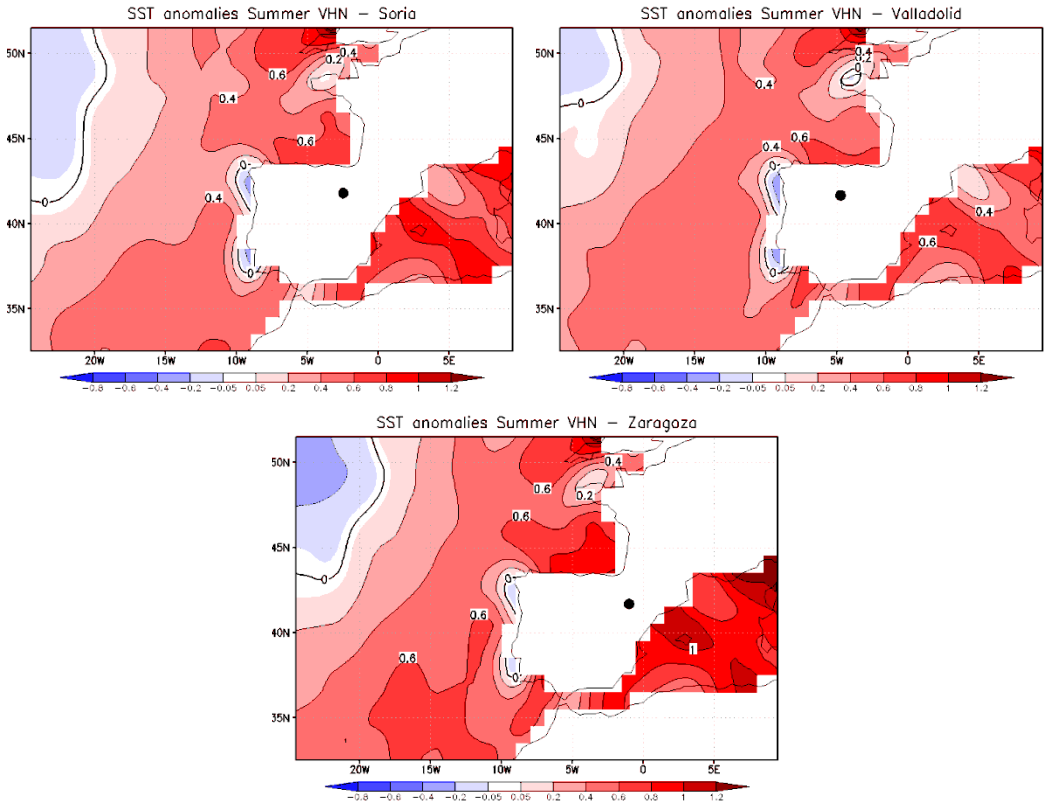
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014



H. Composite maps of SST anomalies for summer VHN for the stations not shown in the text



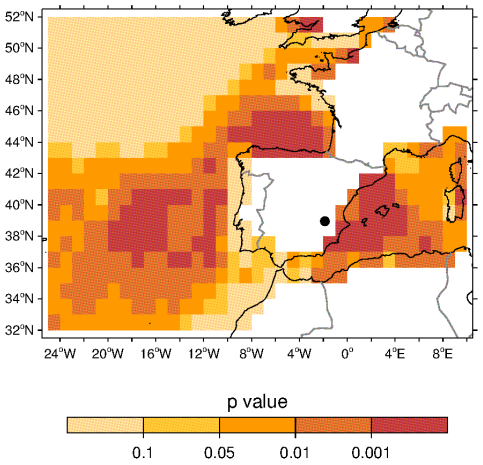

SST anomalies Summer VHN – Huesca
SST anomalies Summer VHN – La Coruna
SST anomalies Summer VHN – Malaga
SST anomalies Summer VHN – Murcia
SST anomalies Summer VHN – Pamplona
SST anomalies Summer VHN – Salamanca
SST anomalies Summer VHN – San Sebastian
SST anomalies Summer VHN – Sevilla
50N
45N
40N
35N
20W
15W
10W
5W
0
5E


SST anomalies Summer VHN – Soria
SST anomalies Summer VHN – Valladolid
SST anomalies Summer VHN – Zaragoza
50N
45N
40N
35N
20W
15W
10W
5W
0
5E

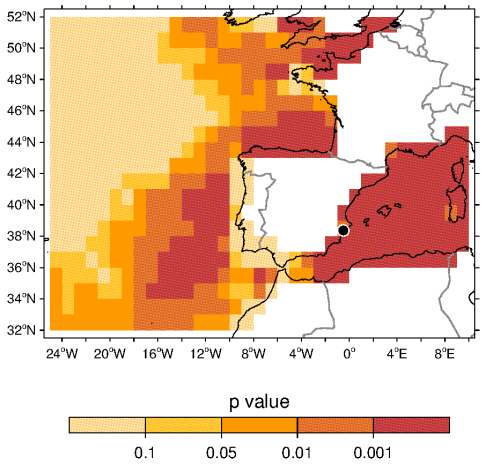
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

I. Gridded maps of p-values between the base model and the model with SST anomalies as covariate in the location parameter of the Tn extreme distribution for the stations not shown in the text

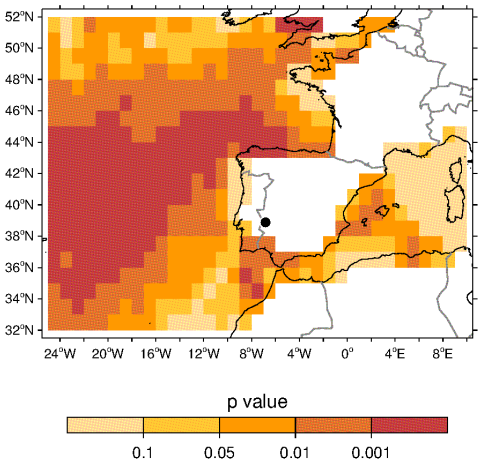
Albacete



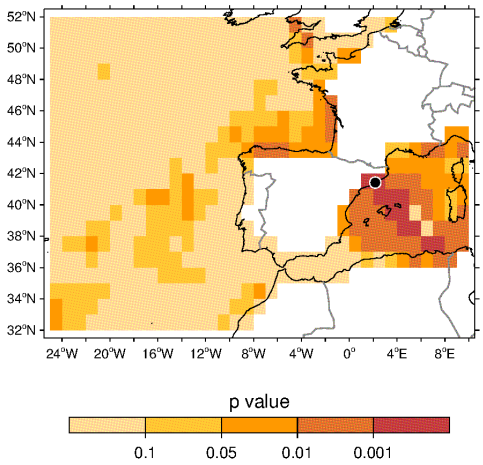
Alicante



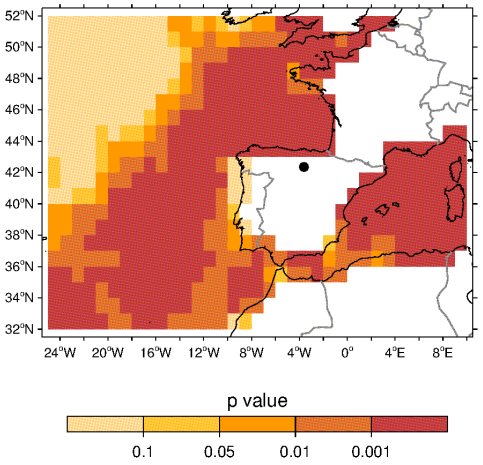
Badajoz



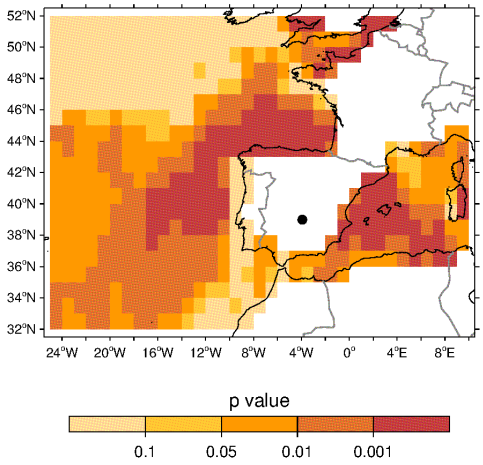
Barcelona



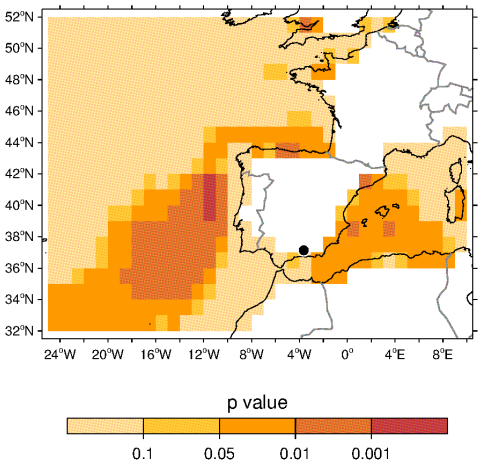
Burgos



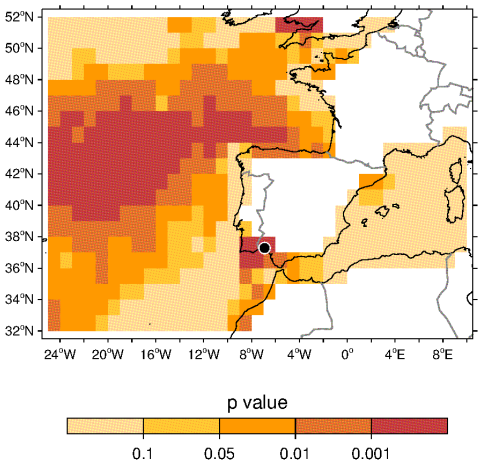
Ciudad Real



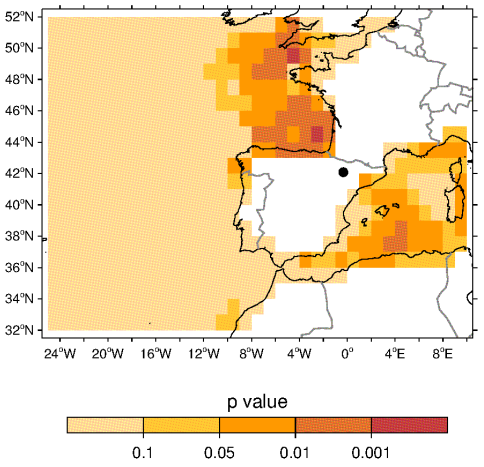
Granada



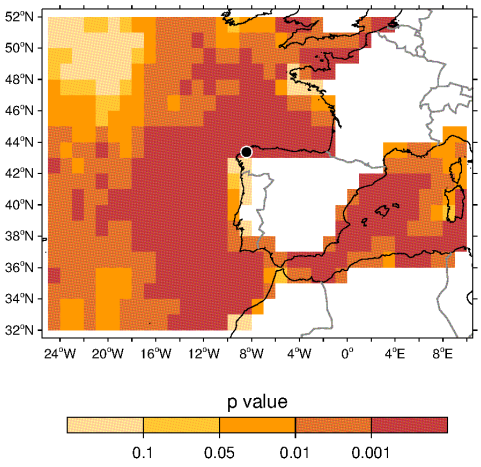
Huelva



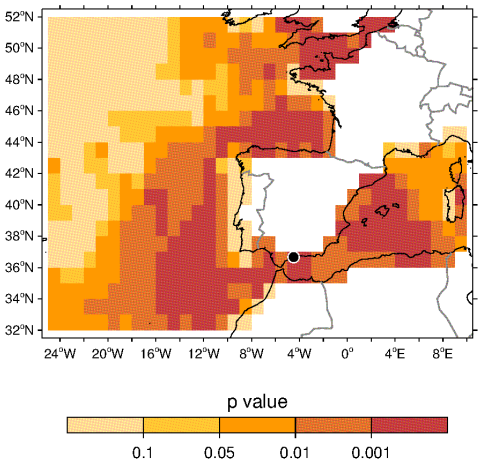
Huesca



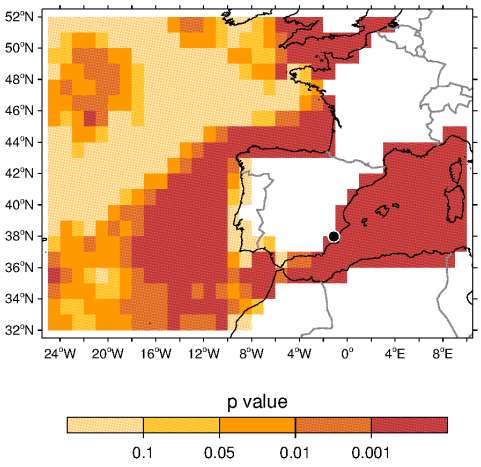
La Coruna



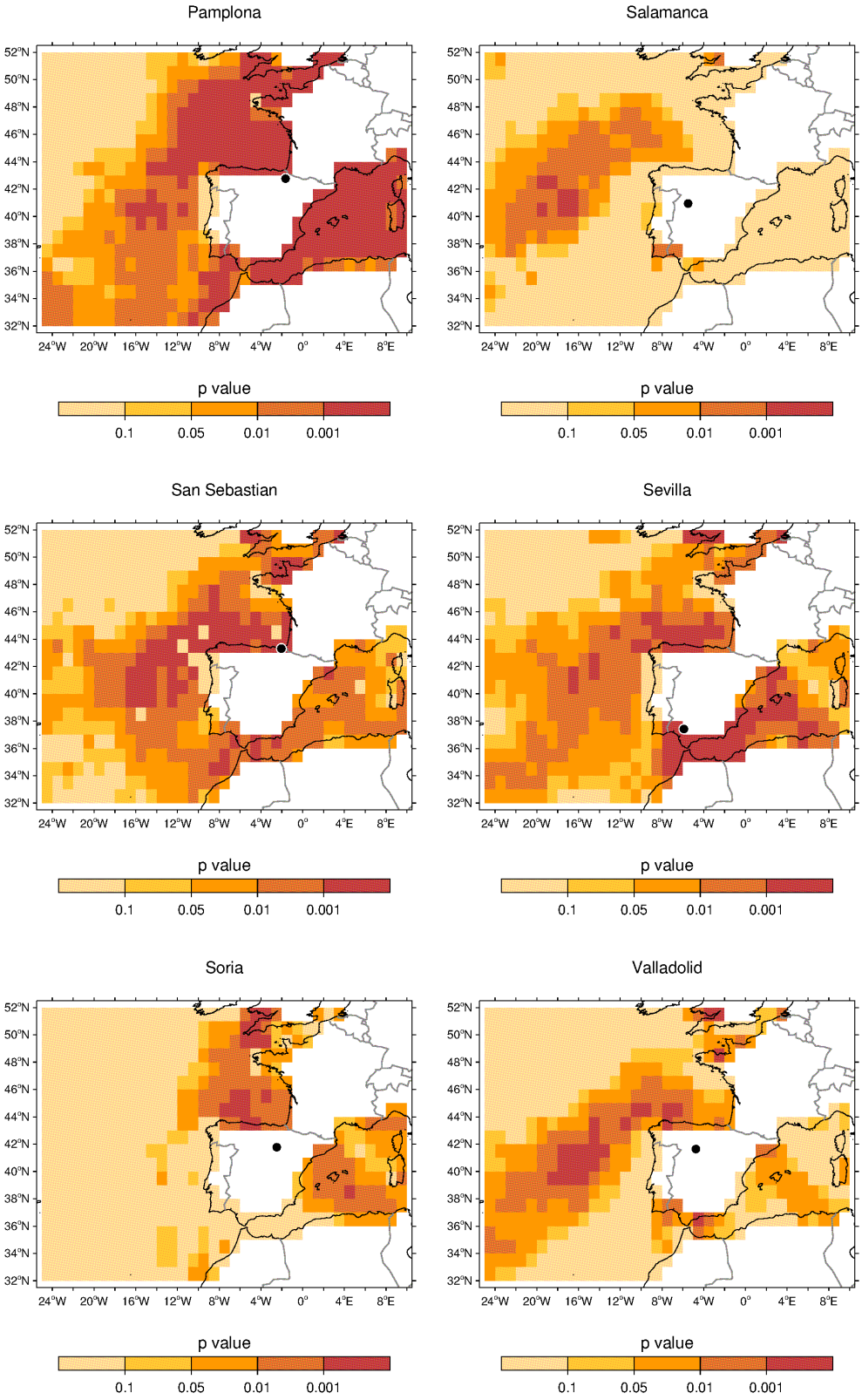
Malaga

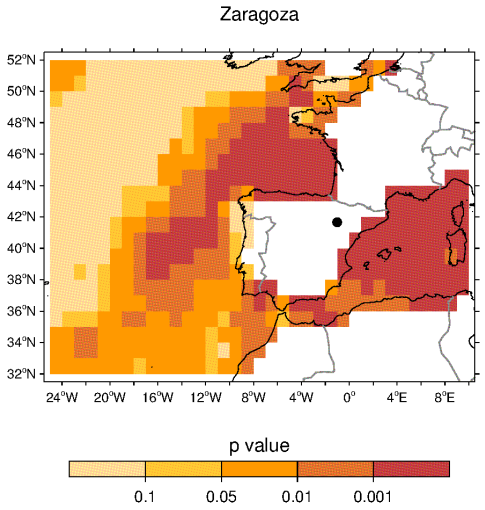


Murcia



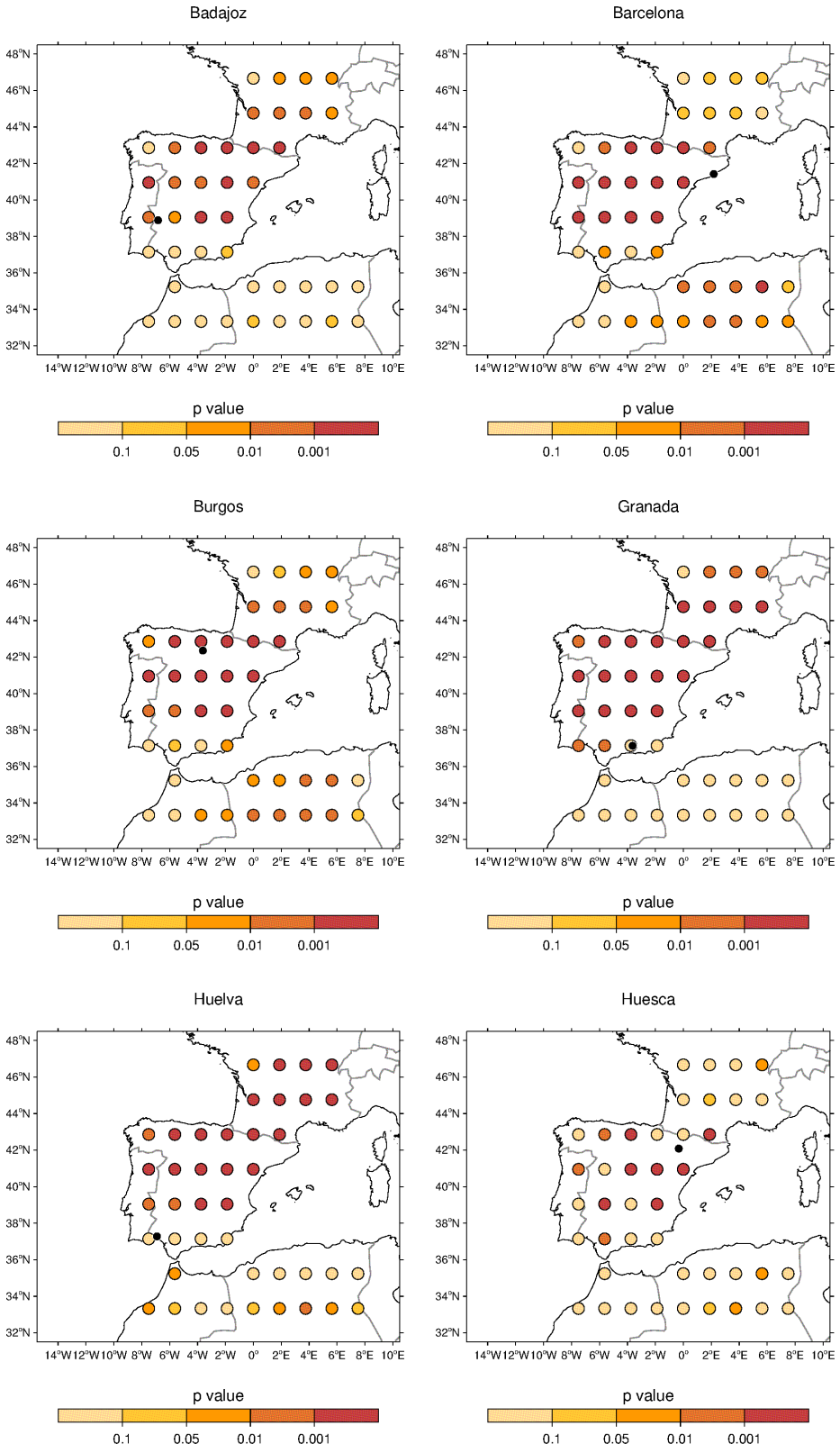
UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

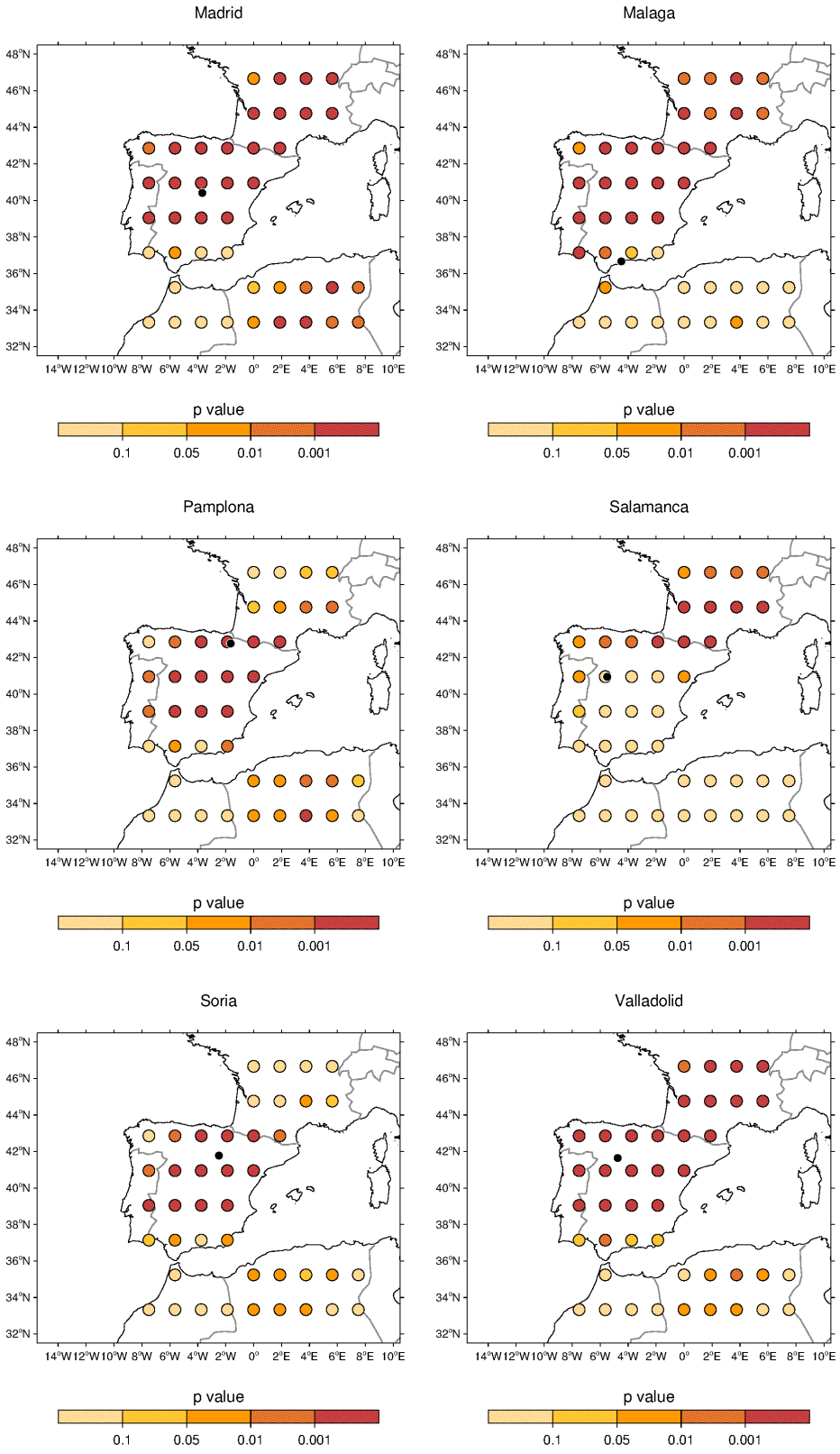



Zaragoza
52°N
50°N
48°N
46°N
44°N
42°N
40°N
38°N
36°N
34°N
32°N
24°W
20°W
16°W
12°W
8°W
4°W
0°
4°E
8°E
p value
0.1
0.05
0.01
0.001

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

J. Maps of p-values between the base model and the model with SM anomalies as covariate in the location parameter of the Tn extreme distribution for the stations not shown in the text




Madrid
Malaga
Pamplona
Salamanca
Soria
Valladolid
48°N
46°N
44°N
42°N
40°N
38°N
36°N
34°N
32°N
14°W 12°W 10°W 8°W 6°W 4°W 2°W 0° 2°E 4°E 6°E 8°E 10°E
p value
0.1 0.05 0.01 0.001

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

List of Figures

Figure 1.1. Schematic diagram showing a change in the location of the GEV distribution. 6

Figure 2.1. Annual maximum of daily maximum temperature (Tx) series at Albacete station for the 1940-2010 period. 12

Figure 2.2. Plot of the GEV probability density function with $\mu = 0$, $\sigma = 1.2$ and $\xi = -0.2$ (Weibull), $\xi = 0.2$ (Fréchet), $\xi = 0$ (Gumbel). 15

Figure 2.3. Summer daily Tx at Valencia station (red dots) from 1940 to 2010, with a selected threshold (horizontal black line). 17

Figure 2.4. GPD density function with $\tilde{\sigma} = 1$, $\xi = -0.2$ (Beta), $\xi = 0.2$ (Pareto) and $\xi = 0$ (exponential). 19

Figure 2.5. Profile likelihood for ξ parameter in PP model of daily maximum temperature data. 28

Figure 2.6. Profile likelihood for 20-year return level in PP model of daily maximum temperature data. 28

Figure 2.7. Diagnostic plots of PP model fitted to Valencia Tx series (1940-2010). 30

Figure 3.1. Location map of the 21 stations over mainland Spain with long daily records of temperature. 33

Figure 3.2. Percentage of missing daily values in Tx and Tn series for each station during the period 1940-2010. 36

Figure 3.3. Probability density function of the observed data (red line) and probability density function of the completed data with *mice* (dashed blue line). 37

Figure 3.4. Resampling from 0.25º x 0.25º to 1º x 1º grid box resolution of SST anomalies data. 38

Figure 3.5. Parameter estimates from GPD fit for a range of 50 thresholds from 95th to 99.5th percentile of daily maximum temperature data from 1948-2010 for Barcelona. The red line indicates the chosen threshold (32.1ºC). 44

Figure 4.1 (a-b). Pattern 1. SLP anomalies (hPa) of summer VHD for Madrid (a) and Salamanca (b). 53

Figure 4.2 (a-b). Same as Figure 4.1(a-b) but for Barcelona (a) and Soria (b). 53

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Figure 4.3 (a-b). Same as Figure 4.1 (a-b) but for San Sebastian (a) and La Coruña (b). 54
Figure 4.4 (a-b). Pattern 2. SLP anomalies (hPa) of summer VHD in Albacete (a) and Granada (b). 55
Figure 4.5 (a-b). Pattern 3. SLP anomalies (hPa) of summer VHD for Valencia (a) and Malaga (b). 56
Figure 4.6. Diagnostic plots of stationary PP model fitted to Barcelona daily maximum temperatures series from 1948 to 2010. 58
Figure 4.7. Same as Figure 4.6 but for the non-stationary model (SLP anomalies of the grid 15ºE 35ºN in the location parameter of the GEV distribution). 58
Figure 4.8 (a-f). Gridded maps of p-values between the base model and the model with SLP anomalies as covariate in the location parameter of the GEV distribution for Madrid (a), Salamanca (b), Barcelona (c), Soria (d), San Sebastian (e) and La Coruña (f). 60
Figure 4.9 (a-b). Same as Figure 4.8 (a-f) but for Albacete (a) and Granada (b). 61
Figure 4.10 (a-b). Same as Figure 4.8 (a-f) but for Valencia (a) and Malaga (b). 61
Figure 4.11 (a-b). Composite maps SST anomalies for summer VHD in Barcelona (a) and Madrid (b). 63
Figure 4.12. Diagnostic plots of stationary PP model fitted to Barcelona daily Tx series from 1982 to 2010. 64
Figure 4.13. Same as Figure 4.12 but for the non-stationary model (SST anomalies of the grid 4.5ºE 41.5N in the location parameter). 65
Figure 4.14 (a-f). Gridded maps of p-values between the base model and the model with SST anomalies as covariate in the location parameter of GEV distribution for Barcelona (a), Malaga (b), Huelva (c), Badajoz (d), Madrid (e) and Pamplona (f). 68
Figure 4.15. Composite map of soil moisture anomalies for summer VHD for Barcelona (1948-2010). 71
Figure 4.16. Diagnostic plots of PP model fitted to daily maximum temperatures in Barcelona (1948-2010) for the non-stationary model (soil moisture anomalies of the grid 1.9ºW 41ºN in the location parameter of GEV distribution). 72

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Figure 4.17 (a-i). Maps of p-values between the base model and the model with soil moisture anomalies as covariate in the location parameter of GEV distribution for Zaragoza (a), Huesca (b), Pamplona (c), Burgos (d), Valladolid (e), Madrid (f), Malaga (g), La Coruña (h) and San Sebastian (i). 73
Figure 4.18. Map of the lowest p-values from the likelihood-ratio test for each station analysed (VHD). 74
Figure 4.19 (a-b). Pattern 1. SLP anomalies (hPa) of summer VHN for Madrid (a) and Badajoz (b). 76
Figure 4.20 (a-b). Same as Figure 4.19 (a-b) but for Barcelona (a) and Soria (b). 76
Figure 4.21 (a-b). Same as Figure 4.19 (a-b) but for San Sebastian (a) and La Coruña (b). 77
Figure 4.22 (a-b). Pattern 2. SLP anomalies (hPa) of summer VHN for Albacete (a) and Granada (b). 77
Figure 4.23 (a-b). Pattern 3. SLP anomalies (hPa) of summer VHN for Valencia (a) and Malaga (b). 78
Figure 4.24. Diagnostic plots of stationary point process model fitted to Barcelona daily minimum temperature (Tn) series from 1948 to 2010. 79
Figure 4.25. Same as Figure 4.24 but for the non-stationary model (SLP anomalies of the grid 2.5ºE 30ºN as covariate in the location parameter of the GEV distribution). 80
Figure 4.26 (a-f). Gridded maps of p-values between the base model and the model with SLP anomalies as a covariate in the location parameter of the GEV distribution for Madrid (a), Badajoz (b), Salamanca (c), Zaragoza (d), San Sebastian (e) and La Coruña (f) 81
Figure 4.27 (a-b). Same as Figure 4.26 (a-f) but for Albacete (a) and Granada (b). 82
Figure 4.28 (a-b). Same as Figure 4.26 (a-f) but for Valencia (a) and Salamanca (b). 82
Figure 4.29 (a-b). Composite maps of SST anomalies of SST for summer VHN in Valencia (a) and Madrid (b). 83
Figure 4.30. Diagnostic plots of stationary point process model fitted to Barcelona daily minimum temperature (Tn) series from 1982 to 2010. 85

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Figure 4.31. Same as Figure 4.30 but for the non-stationary model (SST anomalies of the grid 7.5ºW 36.5ºN in the location parameter)..... 85
Figure 4.32(a-f). Gridded maps of p-values between the base model and the model with anomalies of SST as covariate in the location parameter for Valencia (a), La Coruña (b), Badajoz (c), Huelva (d), Zaragoza (e) and Huesca (f). 87
Figure 4.33. Composite map of soil moisture anomalies of summer VHN for Barcelona (1948-2010)..... 88
Figure 4.34. Diagnostic plots of PP model fitted to daily minimum temperatures in Barcelona (1982-2010) for the non-stationary model (Soil moisture anomalies of the grid 1.88ºE 35.2ºN in the location parameter). 89
Figure 4.35 (a-i). Maps of p-values between the base model and the model with soil moisture anomaly as a lineal covariate in the location parameter in Ciudad Real (a), Albacete (b), Valencia (c), Alicante (d), Murcia (e), Sevilla (f), Zaragoza (g), San Sebastian (h) and La Coruña (i)..... 90
Figure 4.36. Map of the lowest p-values from the likelihood-ratio test at each station analysed (VHN). 91
Figure 4.37 (a-b). Maps of p-values between the base model and the model with time dependence as a covariate in the location parameter of GEV distribution during the period 1940-1972 for Tx (a) and for Tn (b). 93
Figure 4.38 (a-e). Maximum temperature (Tx) extreme probability density functions. GEV parameters estimated. Solid black (dashed blue) curves display the estimated density function on June, the 1st 1940 (August, the 31st 1972). For Badajoz (a), Salamanca (b), Burgos (c), Valladolid (d) and Huesca (e)..... 96
Figure 4.39 (a-d). Same as Figure 4.38 (a-e) but for the minimum temperature (Tn) extreme probability density functions and for Sevilla (a), Murcia (b), Badajoz (c) and La Coruña (d). 97
Figure 4.40 (a-b). Same as Figure 4.37 (a-b) but during the period 1973-2010 for Tx (a) and for Tn (b). 98
Figure 4.41 (a-f). Tx extreme probability density functions. GEV parameters estimated. Solid black (dashed red) curves display the estimated density function in June the 1st 1973 (August the 31st 2010) for Valencia (a),

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

Malaga (b), La Coruña (c), Salamanca (d) and Burgos (e). The corresponding 20-year return values are indicated by grey dashed line. 101

Figure 4.42 (a-h). Same as Figure 4.41 (a-f) but for Tn extreme probability density functions and for Sevilla (a), Ciudad Real (b), Murcia (c), Valladolid (d), Malaga (e), Albacete (f), Madrid (g) and Granada (h). The corresponding 20-year return values are indicated by grey dashed line. 102

Figure 4.43(a-e). The highest effective return levels for 5, 10, 20, 50 and 100 years for Tx extreme. 110

Figure 4.44 (a-d). The lowest effective return levels for 5, 10, 20, 50 and 100 years for Tx extreme. 111

Figure 4.45 (a-e). Same as Figure 4.43 (a-e) but for Tn extreme. 112

Figure 4.46 (a-d). Same as Figure 4.44 (a-d) but for Tn extreme. 113

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

UNIVERSITAT ROVIRA I VIRGILI
STATISTICAL MODELLING AND ANALYSIS OF SUMMER VERY HOT EVENTS IN MAINLAND SPAIN
Mercè Castellà Sánchez
Dipòsit Legal: T 962-2014

List of Tables

Table 3.1. Name, longitude, latitude and altitude of each station. 34
Table 4.1. Best thresholds selected for daily Tx and their percentile for the period 1948-2010..... 57
Table 4.2. Best thresholds selected for daily Tx and their percentiles for the period 1982-2010..... 64
Table 4.3. Best thresholds selected for daily Tn and their percentile for the period 1948-2010..... 79
Table 4.4. Best thresholds selected for daily Tn and their percentile for the period 1982-2010..... 84
Table 4.5. Parameters (MLE) of PP approach with time trend into the location parameter for the period 1940-1972 and statistical significance (*p-value<0.05, **p-value < 0.01 and ***p-value <0.001). 94
Table 4.6. Same as Table 4.5 but for the period 1973-2010..... 99
Table 4.7. Parameters (MLE) of PP approach with time trend into the location and scale parameters of Tx GEV distribution for the period 1940-1972 and the significance (*p-value < 0.05 and **p-value < 0.01)..... 105
Table 4.8. Same as Table 4.7 but for Tn GEV distribution. 106
Table 4.9. Same as Table 4.7 but for the period 1973-2010..... 107
Table 4.10. Same as Table 4.8 but for the period 1973-2010..... 108