Universidade de Aveiro Departamento de Física,
2012

MARIA JOÃO MELO TAVARES DE CARVALHO

AVALIAÇÃO DA PERFORMANCE DE MODELOS CLIMÁTICOS NA PENÍNSULA IBÉRICA

PERFORMANCE EVALUATION OF CLIMATE MODELS FOR THE IBERIAN PENINSULA

MARIA JOÃO MELO TAVARES DE CARVALHO

AVALIAÇÃO DA PERFORMANCE DE MODELOS CLIMÁTICOS NA PENÍNSULA IBÉRICA

PERFORMANCE EVALUATION OF CLIMATE MODELS FOR THE IBERIAN PENINSULA

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requesitos necessários à obtenção do grau de Mestre em Meteorologia e Oceanografia Física, realizada sob a orientação científica de Alfredo Rocha, Professor do Departamento de Física da Universidade de Aveiro e Paulo de Melo-Gonçalves, Investigador Post-Doc do Departamento de Física da Universidade de Aveiro.

o júri / the jury

presidente / president

Professor Doutor José Manuel Henriques Castanheira
Professor Auxiliar da Universidade de Aveiro

vogais / examiners committee

Professor Doutor João Carlos Andrade dos Santos
Professor Auxiliar com Agregação da Universidade de Trás-os-Montes e Alto Douro

Professor Doutor Alfredo Moreira Caseiro Rocha
Professor Associado com Agregação da Universidade de Aveiro

Doutor Paulo de Melo-Gonçalves
Investigador no CESAM - Centro de Estudos do Ambiente e do Mar

agradecimentos / acknowledgements

Aproveito esta oportunidade para agradecer aos meus orientadores, o Professor Doutor Alfredo Rocha e o Doutor Paulo de Melo-Gonçalves pelo apoio que me prestaram durante o desenvolvimento desta tese. Aos colegas e amigos que mostraram sempre disponibilidade para ajudar no desenvolvimento de ideias e na procura de soluções para os problemas que surgiam. E especialmente à minha mãe que fez de mim quem sou e sempre apostou na minha eduação pessoal e académica, e ao João Teixeira que me tem acompanhado nesta incrível viagem desde o primeiro dia.

Palavras-chave modelos climáticos, avaliação de performance, temperatura, precipitação, Península Ibérica

Resumo

O objetivo deste trabalho é a análise da performance de Modelos Climáticos Regionais (RCMs) e Globais (GCMs) na Península Ibérica. Para tal, foram utilizados dados de precipitação, temperatura máxima e mínima (aos 2 metros) para o passado recente (1962-2000) disponibilizados pelo projeto ENSEMBLES, bem como dados observados da *European Climate Assessment & Dataset* (ECA&D). A análise foi separada em duas secções: a avaliação de performance dos modelos regionais (utilizando o ensemble das simulações forçadas pelas reanálises ERA40) e a avaliação dos modelos globais (utilizando o ensemble de simulações de modelos regionais forçados por modelos globais). Foram calculadas as climatologias sazonais das variáveis originais, bem como as tendências de índices, de forma a verificar a capacidade dos ensembles de simular as variáveis, os seus extremos e variabilidade temporal. Para além disso, calcularam-se parâmetros estatísticos, como o Erro Quadrático Médio (RMSE), o viés, o desvio padrão e o Coeficiente de Correlação de Pearson. Com o objetivo de verificar se as distribuições modeladas dos índices são estatisticamente semelhantes, utilizou-se o teste de Kolmogorov-Smirnov. Concluiu-se que o ensemble das simulações forçadas pelas reanálises ERA40 mostram melhor performance, enquanto que a incerteza associada ao ensemble das simulações forçadas por GCMs é superior. Os índices de precipitação que mostram a ocorrência de dias secos têm melhor performance que os que contabilizam os dias molhados. Ambas as temperaturas mostram diferenças entre observações e ensembles inferiores às da precipitação, bem como melhor representação dos padrões espaciais. No entanto, os dados modelados mostram pior performance em zonas de orografia complexa e ao longo de zonas costeiras, tanto para as variáveis originais como para os índices extremos. A análise dos dados observados mostra que há, em geral, aumento no número de dias em que a temperatura máxima esteve acima do percentil 90 e uma diminuição no número de dias em que a precipitação foi acima do percentil 90 e a temperatura mínima foi abaixo do percentil 10. Embora não apresentando sempre significância estatística, as tendências destes índices apontam para a existência de alterações climáticas que ocorreram na Península Ibérica durante o século XX.

Key Words climate models, performance evaluation, temperature, precipitation, Iberian Peninsula

Abstract The goal of this work is to analyse the performance of both Regional Climate Models and simulations where these are driven by Global Climate Models over the Iberian Peninsula. To do so, ENSEMBLES datasets of precipitation, maximum and minimum temperature (at 2 meters) for the recent-past (1961-2000) were used, as well as the European Climate Assessment $\&$ Dataset (ECA$\&$D) observed data. The analysis is performed in two parts: RCM performance evaluation (using ERA40-driven simulation ensemble) and GCM performance evaluation (using the ensemble of RCM-GCM simulations). Seasonal climatologies of the original variables, as well as trends of the indices were analysed in order to evaluate the ensembles' ability to simulate the variables, their extremes and their time variability. The skill of the ensembles was measured using statistics such as the root mean square error, BIAS, standard deviation error and Pearson correlation coefficient. Two sample Kolmogorov-Smirnov tests were performed to verify the similarity of the distributions of the yearly indices. It was found that ERA40-driven ensemble shows results more closely related to the observations than GCM-driven ensemble. Furthermore, the uncertainty associated with the GCM-driven ensemble is generally higher than that of ERA40-driven. Precipitation indices that evaluate the occurrence of dry days perform better than the wet days indices. Temperature shows smaller difference between observations and models and higher spatial pattern reproduction. However, models show lower skill and performance in areas of complex orography and along the coastline, for the original variables, as well as the extreme indices. The analysis of the observed data, showed an overall increase in number of days with maximum temperature over the $90th$ percentile and a decrease in number of days with precipitation over the $90th$ percentile and minimum temperature under the $10th$ percentile. Although not always statistically significant these trends point to changes in the Iberian climate that have already taken place during the $20th$ century.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

As defined by the World Meteorological Organization (WMO), climate is the statistical description of the central tendencies of variables such as temperature, precipitation, atmospheric pressure, humidity and winds or combination of elements as weather types and phenomena typical for a location or region, for a period of time ranging from months to thousands of years. As such, climate change (CC) is represented by an alteration in the statistical properties of one or more of these elements (or groups of elements). According to the Intergovernmental Panel for Climate Change (IPCC), CC is defined as a statistically significant change of the mean and/or variability of the of variables which define the climate during a long time scale, typically larger than a decade.

The climatic system is composed by several different elements and although there are many governing and forcing factors, the main one is radiation. The amount of radiation that reaches the surface of the Earth depends largely on the season and the location itself which influences not only the energy density but also the number of sun exposure hours. Of the radiation emitted by the sun that reaches the Earth's atmosphere, 99 % is of short wavelength ranging from 0.15 to 4 μm of which 9 % is ultraviolet (UV) radiation and 45 % is in the visible range. The shorter wavelength radiation is absorbed on the highest atmospheric levels by the oxygen (O_2) and the ozone (O_3) and in the lower levels by water vapour (H_2O) and carbon dioxide (CO_2), making it so that the radiation reaching the Earth's surface is reduced to wavelengths between the 0.3 and 2 μm. When it reaches the surface of the Earth the radiation is either absorbed or reflected, depending on surface composition as well as rugosity. The albedo (ratio between the reflected and absorbed energy amounts - %) varies very widely from lower values (water, for example) to high ones (snow) with an average for the Planet of $\sim 30\,\%$.

The atmosphere is heated from the surface up by the IR energy emitted. However it is not the main atmospheric gases, O_2 and nitrogen N_2 that absorb most of IR radiation. In fact, the radiative properties of the atmosphere are dominated by water vapour, CO_2 and O_3. Each of these gases has a different absorption spectrum inherent to their different molecular structure. It is then expected that alterations in the atmospheric composition, not only with the introduction of new elements but also by altering existing elements' concentrations, would have effects on the way the atmosphere behaves towards the radiation emitted by the Sun and the Earth itself.

Since pre-industrial times, human activities have increased gas emissions, such as CO_2, methane (CH_4) and nitrous oxide (N_2O) through deforestation and fossil fuel use - CO_2 - agriculture - CH_4 e N_2O. The IPCC claims that the emission of these greenhouse gases (GHGs) has been growing since the 1750's and grew approximately 70 % between 1974 and 2004. Furthermore, these GHGs have a long life period in the atmosphere and therefore its influence in the atmospheric radiation processes can be felt several years after its initial emission.

The changes that have occurred in the climate in the recent past can be easily determined by analysing observed data. However, the changes that might occur in the future can only be ascertained using modelled data. There are several types of Global and Regional Climate Models and therefore the results obtained using each one is different, creating a range of possible outcomes.

In the past two decades there have been changes in climate and environmental policies that change the goals of emission reduction which, together with the application of new and more efficient technologies, using cleaner sources of energy, can influence GHG emissions. All this makes it very difficult to predict how GHG emissions will progress in the next 100 years. In order to contemplate all spectrum of possibilities, future emission scenarios were developed by the IPCC and have been used by several

groups to produce climate simulations. These simulations then produce a set of data that, together with the different models used, can provide a spectrum of the possible future climate change outcomes.

Although Global Climate Models (GCMs) provide useful information, their resolution is low and therefore climate change assessment on a local scale becomes almost impossible. Recently, downscaling techniques have been developed and implemented, making it so that the resolution is increased and therefore local studies are made possible. These involve a Regional Climate Model (RCM) using higher resolution nested in GCMs and producing simulations for a smaller area. The RCM-GCM combination offers another source of variability in the outcomes, which in turn adds to the uncertainty of future climate projections.

This work aimed to assess the performance of, firstly a set of RCMs and then, several RCM-GCM combinations, using recent past (1961-2000) simulations. This was accomplished by using statistical tools, in order to compare modelled data with observations and therefore determine the accuracy of model results. This in itself provided an important resource in order to better interpret model results for future CC assessment, since it makes the understanding of model behaviour in different areas of the domain. This study was pursued for the Iberian Peninsula, which comprehends both continental Portugal and Spain.

Chapter 2

State of The Art

Climate Change (CC) plays an important role in several areas of everyday life (Field et al., 2012) and should therefore be studied in depth, not only to understand the changes that have already occurred but also and foremost, to foresee the ones in the future. The first are easily studied using observations, while the later can only be achieved though the use of Climate Models. Originally, Global Climate Models (GCMs) started being used, with a somewhat reduced resolution. More recently, the downscaling of those models has been performed by taking Regional Climate Models (RCMs), with higher resolution, and forcing them with GCMs. Prömmel et al. (2010) conducted a study where the 2-meter temperature results obtained by using ERA40 reanalysis and REMO5.0 driven by the ERA40 reanalysis were compared to observations, during the 1985 and 1998 period in the Great Alpine Region. They found that both showed good representation of the temporal variability although with high BIASes. The use of a Regional Model - REMO5.0 - had added value specially in areas of more complex orography and coastal region.

According to Déqué et al. (2007, 2012), there are four sources of errors is these types of simulations: sampling error, model uncertainty, radiative uncertainty and boundary uncertainty. Déqué et al. (2007) studied the relative weight of these uncertainty sources in Europe. They used surface temperature data from the PRUDENCE project and found that most dominant was the boundary forcing (GCM used), while the radiative forcing was only influential for summer surface temperature. Déqué et al. (2012) found the same results using the ENSEMBLES project datasets and added that the uncertainty associated to the forcing varies geographically, with its maximum influence in the southern western Europe, whereas the continental parts were more sensitive to the choice of RCM.

Gallardo et al. (2012) used the Köppen-Trewartha (K-T) climate classification in order to reproduce the current climate in Europe for an ensemble of 15 regional model simulations nested in six GCMs. They compared each of the individual simulations as well as the ensemble of those simulations with observations. They found that the ensemble performed better than any individual model, with 83.5 % of the grid points in agreement with observations. These results show that the uncertainty associated with the choice of a specific RCM/ GCM combination can be mitigated by using the ensemble of several distinct simulations. Furthermore, Maraun (2012) found that by performing a BIAS correction or even eliminating the BIAS outlier simulations, the performance of the ensemble of the simulations can be improved.

In addition to these sources of error and the difficulties already mentioned, there is also the possibility that the land surface scheme used by the model can influence the results obtained, as studied by Roy et al. (2011) using the Canadian Regional Climate Model (CRCM) driven by reanalysis (NCEP-NCAR) using two different land use datasets. When analysing the performance of the two simulations using different land use in reproducing several indices of both temperature and precipitation, they found that the latest version improved the results in general and in particular the intensity of extreme hot summer temperature, diurnal temperature range, wet days occurrence and seasonal dry spell.

The performance assessment of a model or simulation cannot be performed by using only one climate statistical characteristic, since some are easier to reproduce than other. Maxino et al. (2008) used precipitation, maximum and minimum temperatures in order to assess the performance of the models submitted for the Intergovernmental Panel for Climate Change's Fourth Assessment (AR4) over the Murray-Darling basin (Australia). They attempted to do so by using the mean and the Probability Distribution Functions (PDFs) of each of the variables and found that overall, models would accurately reproduce the observed means but showed worst results for the PDFs.

In a changing climate, there is the need for a deep understanding, not only of the climate system, but also of the model results that allow for a projection of future changes and therefore the development of adaptation and mitigation measures. Studies of the changes that have already occurred were performed by several authors and for different parts of the world (Alexander et al., 2006; Easterling et al., 2000; Allan and Soden, 2008), that in a warmer climate, extreme precipitation events will become more common and that recent measurements have showed that these are being underestimated by models.

The study of climate change has increasingly focused on the extreme events of temperature and precipitation since these are cause for destruction and loss of life (for example the 2003 European heat wave (Bucker, 2005) or the Madeira Island extreme precipitation event (Luna et al., 2011)) and the skill of a model simulating a given climate statistic is not the same as for the extreme events.

Tanking these facts into account, the performance evaluation of RCM-GCM simulations for the recent-past using observed data is an important tool to understand the results of the future climate simulations and perform the necessary corrections, in order to obtain the most accurate projections, with the lowest uncertainty possible.

Chapter 3

Method and Data

3.1 Data

The ENSEMBLES project, funded by the European Commission's 6th Framework Programme through contract GOCE-CT-2003-505539 (http://ensemblesrt3.dmi.dk/) created a database of recent past (1951-2000) and future (2001-2100) climates. GCMs produced global results that were then downscaled, producing higher resolution simulations. The integration area used can be seen in figure 3.1, with a 0.22° (25 km) horizontal resolution.

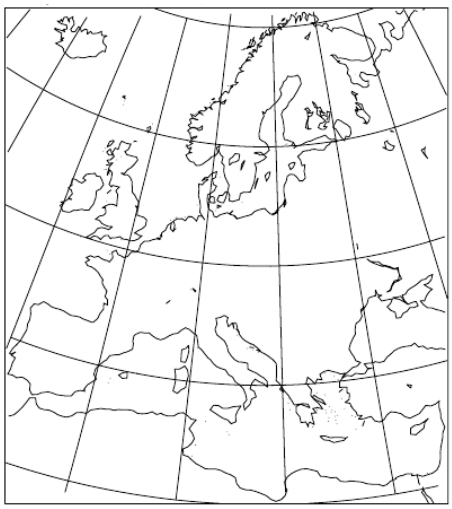


Figure 3.1: Integration area used by the ENSEMBLES project in order to produce the downscaled simulations.

The simulations were performed for the A1B scenario (Houghton et al., 2001) using a list of RCMs being driven by several GCMs. Also, observed datasets (E-OBS) from the EU-FP6 project ENSEMBLES provided by the European Climate & Dataset (ECA&D) (http://eca.knmi.nl) were used in order to evaluate model performance. The observed data has the same horizontal resolution as the modelled ones.

The observed and ERA40-driven datasets use a rotated grid so only the RCM-GCM simulations that use this rotated grid were used in order to avoid the need for interpolation, which could work as an error source. In addition, some of the series did not contain the full simulations with segments of years missing and were therefore also not considered. All RCMs were driven by ERA40 as well as GCMs. Consequently, the RCM-GCM simulations used were those shown in table 3.1. Hence, the simulations used contain eight RCM-GCM combinations, six ERA40-driven RCMs and one observed dataset. All these simulations produced a list of variables of which the maximum and minimum daily temperatures (tasmax and tasmin respectively) at 2-meters and daily precipitation (pr) values were used in this work.

The focus of this work was a domain containing only the Iberian Peninsula (IP) (topography of which can be seen in figure 3.2). The simulations contain values for all the domain while the observed data is only for land grid points. This domain is characterized by a large coastline to the West, North and South. In addition, its topography is rich in mountains and elevations, the main of which are the Pyrenees, located on the Spain-France border, the Meseta Central, the inner Iberian Plateau, the

Table 3.1: Regional Climate Model and driving Global Climate Model Combination used from the ENSEMBLES project. Three runs were performed for the HadRM and HadCM using different sensitivities: Low Sensitivity (3Q3), Normal Sensitivity (3Q0) and High Sensitivity (3Q16).

RCM		Driving Global Climate Model					
		BCM	ECHAM5	HadCM			ERA40
				Low S. 3Q3	Normal S. 3Q0	High S. 3Q16	
RACMO			X				
HadRM	Low S. (3Q3)			X			
	Normal S. (3Q0)				X		
	High S. (3Q16)					X	X
REMO			X				
RCA		X	X		X		

Cantabrian Mountains to the North of the Peninsula with an East to West setting and Sierra Nevada in the Southeaster part of Spain. These mountain ranges have a maximum hight of 1993 m in Portugal and 3500 m in Spanish territory.

3.2 Method

Initially, in order to compare the observations with both the ERA40-driven simulations and GCM-driven simulations, the seasonal climatologies were computed for all three variables. To reduce the amount of data, the ensemble mean of the two sets of simulations were determined, considering all models to have an equitable weight: ERA40-driven ensemble and GCM-driven ensemble. The ensemble mean should always be analysed together with a measure of its uncertainty. In this case, the uncertainty measure chosen was the Interquartile Range (IQR), which is the difference between the third and first quartiles of a distribution (Wilks, 1995). The IQR was determined for each grid point considering the models of each ensemble and therefore used eight data points for GCM-driven ensemble and six for the ERA40-driven ensemble.

The statistical significance of the difference between modelled and observed data was determined for both the seasonal climatologies of the original variables - precipitation, maximum and minimum temperatures - for both ERA40-driven ensemble and GCM-driven ensemble, as well as the GCM-driven ensemble of precipitation and temperature indices at the 5% confidence level, by applying a two sided Student's t-test (Wilks, 1995).

For the domain under study, several procedures were taken in order to evaluate model performance, which is divided into two main parts: RCM performance evaluation and RCM-GCM combination performance evaluation.

3.2.1 RCM Performance Evaluation (Part I)

For this first part, only the ERA40-driven simulations will be used, in order to evaluate how well the RCMs are reproducing the observations, without dealing with the forcing as well which, in this case is the same for all RCMs.

A basic measure of the performance of the models is the comparison between observed and modelled 40 year all-year-around distribution. In order to do so, the ensemble mean of all six ERA40-driven simulations was determined. Afterwards, the Kolmogorov-Smirnov (K-S) test was applied in order to test the hypothesis that observed and modelled distributions are equal, at the 5% significance level. For the grid points where the distributions are considered to be equal, the test statistic was plotted

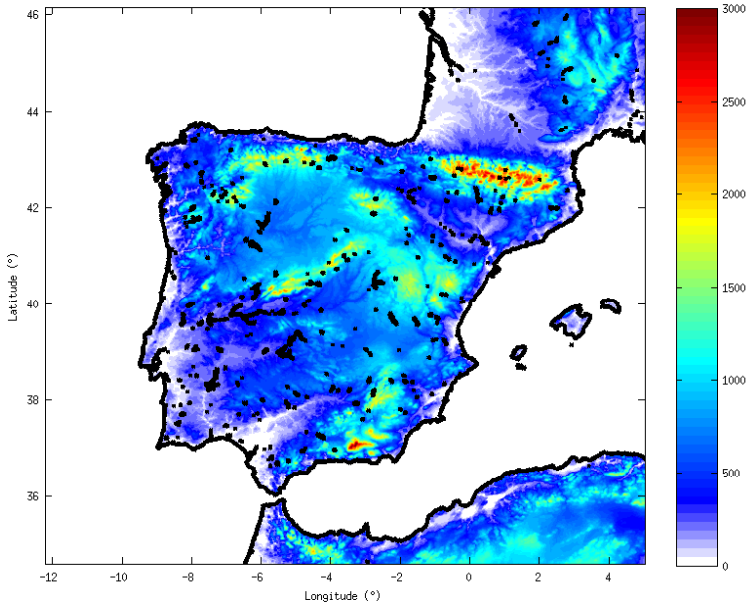


Figure 3.2: Topography of the study area containing the Iberian Peninsula using the USGS GTOPO30 database.

---

and is shown. The K-S test statistic - D_n - is the maximum difference between the two Cumulative Distribution Functions (CDFs) (observed and modelled, in this case) and was determined as can be seen in equation 3.1, adapted from Wilks (1995).

$$D_n = max \mid M_i - O_i \mid \tag{3.1}$$

where the M stands for modelled data and O for observed data.

When studying CC there is the need to assess how models behave when simulating the tails of distributions and their evolution in time, since these are the representation of the extremes of a variable and therefore of extreme events which are the subject of most concern when studying CC.

Taking the simulations only for a control period, 1961-1990, the seasonal daily accumulated precipitation and maximum daily temperature $90th$ percentile was determined for each season: winter (December, January and February - DJF), spring (March, April and May - MAM), summer (June, July and August - JJA) and autumn (September, October and November - SON). Also, minimum daily temperature $10th$ percentile was determined. This produces 4 values for each variable, for each grid point of the domain. Afterwards, the number of days per year, for each season, in which precipitation and maximum temperature were above the control $90th$ percentile (pr90p and tasmax90p respectively) and minimum temperature is below the $10th$ percentile (tasmin10p) was determined for all simulations, producing one yearly result per season, per grid point. This was done for the entire period of simulations between 1961 and 2000. Afterwards, the trend of this number of days, together with its statistical significance was computed. Furthermore, the ensemble mean, together with the corresponding IQR was determined for each of the fields.

The method presented so far is only focused on the ensemble mean of the pr90p, tasmax90p and tasmin10p trends. However, individual performance evaluation of ERA40-driven simulations can also be evaluate d. Therefore, some parameters of skill measure were determined, using the pr90p, tasmax90p and tasmin10p seasonal trend fields. Considering that M represents the modelled data while O stands for the observed data and the overbar represents the average. Skill parameters were deter-

mined as follows (adapted from Wilks (1995)):

- Root Mean Square Error

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2\right]^{1/2} \tag{3.2}$$

- Standard Deviation Error

$$STDE = \left[\frac{1}{N}\sum_{i=1}^{N}\left((M_i - O_i) - \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)\right)^2\right]^{1/2} \tag{3.3}$$

- BIAS

$$BIAS = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i) \tag{3.4}$$

- Pearson product-moment coefficient of linear correlation

$$CORR = \frac{1}{S_M S_O}\sum_{i=1}^{N}\left[\left(M_i - \overline{M}\right)\left(O_i - \overline{O}\right)\right] \tag{3.5}$$

  where

$$S_M = \left[\sum_{i=1}^{N}\left(M_i - \overline{M}\right)^2\right]^{1/2}$$

  and

$$S_O = \left[\sum_{i=1}^{N}\left(O_i - \overline{O}\right)^2\right]^{1/2}$$

  are the modelled (S_M) and observed (S_O) Standard Deviations.

3.2.2 RCM-GCM simulation performance evaluation (Part II)

The second part of model performance evaluation focuses on the GCM-driven simulations. In order to do so, together with the previously described indices - tasmax90p, pr90p and tasmin10p - the following ten CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) were computed.

Precipitation Indices:

- Heavy Precipitation Days (R10mm) - Number of days with precipitation over 10 mm;
- Very Heavy Precipitation Days (R20mm) - Number of days with precipitation over 20 mm;
- Extremely Wet Days (R99p) - Number of days with precipitation over $99th$ percentile calculated for wet days (on basis of 1961-1990 period);
- Consecutive Dry Days (CDD) - Greatest number of days with precipitation under 1 mm;
- Consecutive Wet Days (CWD) - Greatest number of days with precipitation over 1 mm.

Temperature Indices:

- Consecutive Summer Days (CSU) - Greatest number of consecutive days with maximum temperature over 25 °C;
- Consecutive Frost Days (CFD) - Greatest number of consecutive days with minimum temperature under 0 °C;

- Ice Days (ID) - Number of days with maximum temperature under 0 °C;
- Extreme Temperature Range (ETR) - Maximum difference between maximum and minimum daily temperatures.

Firstly, the seasonal index was determined for each year and grid point, for each of the eight simulations used. These were then used to determine the ensemble mean of the seasonal indices, together with the uncertainty (IQR). The climatology of these indices was determined, for both the ensemble and observations in order to compare the first with the later. Furthermore, using the yearly values of the indices, the K-S test was applied in order to ascertain if the modelled distributions are in accordance with the observed ones.

Using the yearly index value, the 40 year trend was determined for the GCM-driven ensemble. With the goal of evaluating the ability of the ensemble to reproduce the indices, Taylor Diagrams were computed. These diagrams take into account the individual standard deviation (S_M and S_O) as well as the Root Mean Square ($RMSE$) Error and Correlation Coefficient ($CORR$) of models relative to observations.

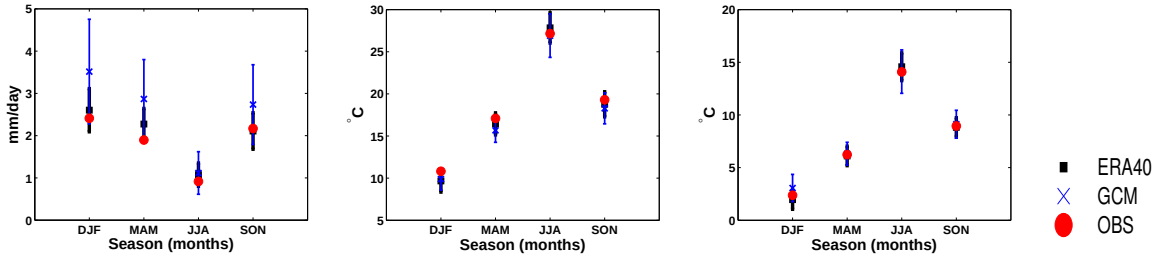
Due to space constrains, only results for CDD, CWD and ETR are shown in this thesis.

Chapter 4

Results and Discussion

A first evaluation of model performance was pursued by analysing the seasonal climatologies using the entire 1961-2000 simulations for the three variables under analysis. This evaluation is performed by comparing the modelled results with the observations. Also, uncertainty fields associated with the ensemble means is evaluated.

The spatial average of seasonal climatologies was calculated for observations, ERA40-driven ensemble mean and GCM-driven ensemble mean and can be seen in figure 4.1, where the bars are a representation of the IQR (and therefore uncertainty) of the ensembles..



(a) Precipitation (b) Max. Temperature (c) Min. Temperature

Figure 4.1: Spatial average of precipitation (mm/day), maximum and minimum temperatures (°C) seasonal climatologies over the IP for observations, ERA40-driven and GCM-driven ensemble mean (together with ensemble mean uncertainty bars).

During summer months, the spatial mean of the precipitation climatology (figure 4.1a), the three sets of data (observations, ERA40-driven ensemble mean and GCM-driven ensemble mean) to be very close to each other. For the remaining seasons, GCM-driven ensemble mean consistently overestimates climatologies (by ~ 1 mm/day for winter and spring and ~ 0.5 mm/day for autumn months). However far observations are from ERA40 ensemble mean, the first is always inside the uncertainty bar for the second. Furthermore, precipitation ensemble mean shows larger uncertainty, specially for GCM-driven simulations. For this variable, the lowest ensemble uncertainty is for summer months, both for ERA40 and GCM-driven simulations.

Temperatures, both maximum and minimum, show higher consistency among observations and ensemble means. In addition, ensemble mean uncertainty is also lower. Largest differences are for summer maximum and minimum temperature mean climatologies. GCM-driven ensemble mean shows its maximum uncertainty in this season (nearly 5 °C). As for precipitation, temperature observations are always inside the uncertainty range of both ERA40 and GCM-driven ensemble means.

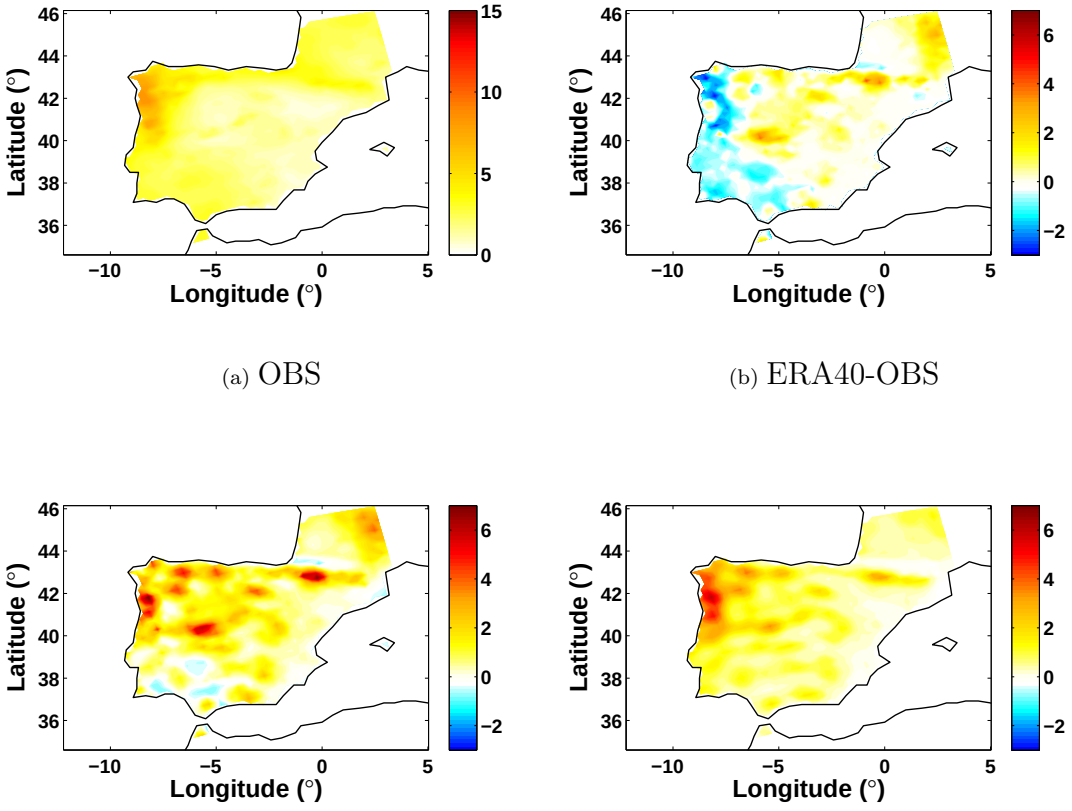
The difference between the ERA40-driven ensemble mean and observations is considered to be statistically significant at a 5% significance level for all seasons' maximum temperature climatologies, winter, spring and summer precipitation climatologies and winter and summer minimum temperature climatologies. On the other hand, differences between GCM-driven ensemble mean and observations are statistically significant at the same significance level for all seasons' precipitation and maximum temperature climatologies and winter minimum temperature.

4.1 Overview of climatologies

To better evaluate the performance of simulations, the spatial distributions of climatologies have been computed, namely:

- Observed climatologies - (a) OBS;
- ERA40-driven ensemble mean minus observed climatologies - (b) ERA40-OBS;
- GCM-driven ensemble mean minus observed climatologies - (c) GCM-OBS;
- GCM-driven minus ERA40-driven ensemble means climatologies - (d) GCM-ERA40.

4.1.1 Precipitation



(a) OBS

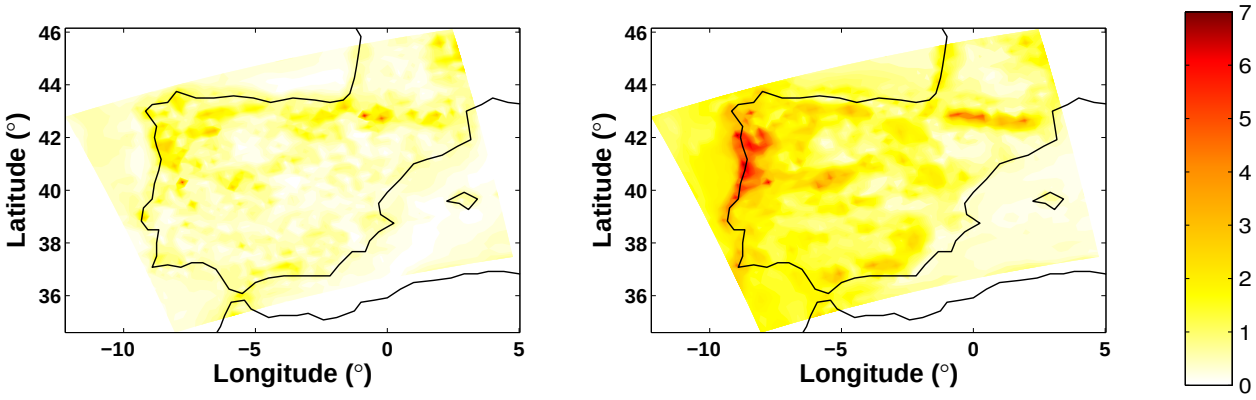
(b) ERA40-OBS

(c) GCM-OBS

(d) GCM-ERA40

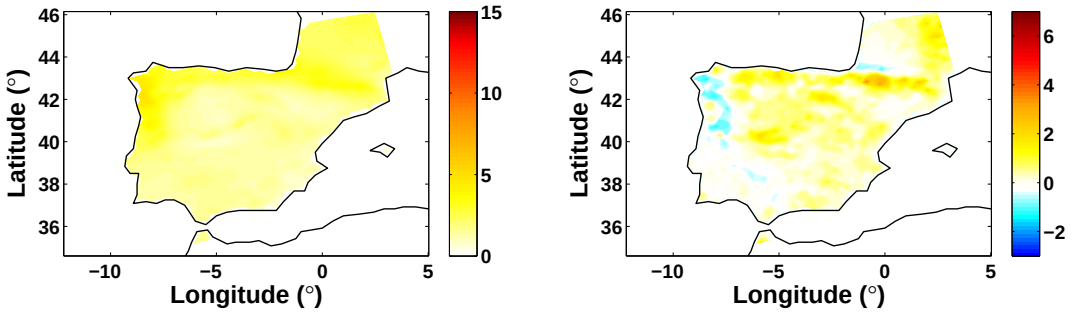
Figure 4.2: Winter precipitation climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d) for the 1961-2000 period (mm/day).

As can be seen in figure 4.2a, precipitation distribution in the IP during the winter shows a clear maximum ($\sim$ 7 mm/day) along the north-western coastal region, that extends from the north until the centre. Also, the central to eastern part of the IP shows a clear dryer pattern for this season.

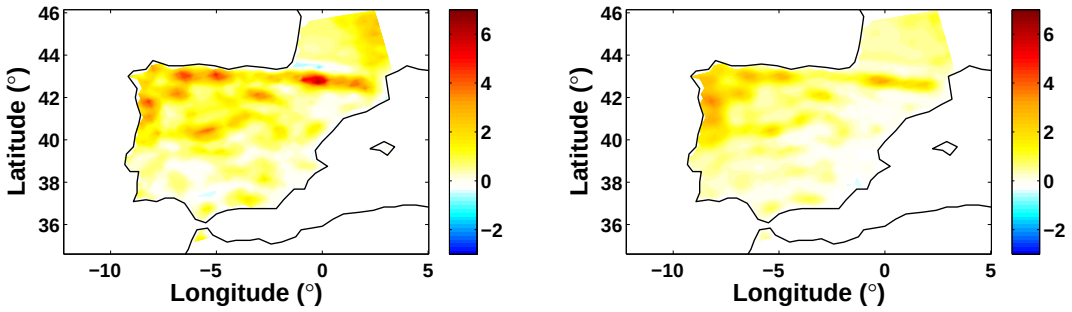


(a) ERA40-driven (b) GCM-driven

Figure 4.3: Interquartile Range associated with the ERA40-driven ensemble mean (a) and the GCM-driven ensemble mean (b) for winter precipitation climatology (mm/day).



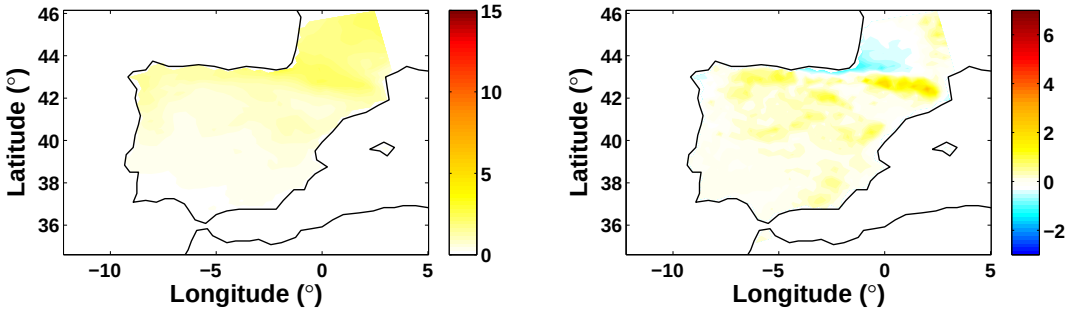
(a) OBS (b) ERA40-OBS



(c) GCM-OBS (d) GCM-ERA40

Figure 4.4: Spring precipitation climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d) for the 1961-2000 period (mm/day).

On the other hand, by looking at figure 4.2b one can see that ERA40 ensemble mean overestimates precipitation in some areas of the IP and underestimate it in others. In most of the Portuguese territory, as well as in the south-western part of Spain, precipitation is underestimated by ERA40 ensemble mean simulations meaning that, overall, models are simulating less rain than what is observed. This underestimation is higher in the north, reaching a maximum difference of about 2 mm/day. On the other hand, the rest of the domain (north, north-east of the IP) shows overestimation of precipitation by the ERA40 ensemble mean, of about 1 mm/day, reaching a maximum of approximately 3 mm/day in the centre of the IP and in the Pyrenees.



(a) OBS

(b) ERA40-OBS

(c) GCM-OBS

(d) GCM-ERA40

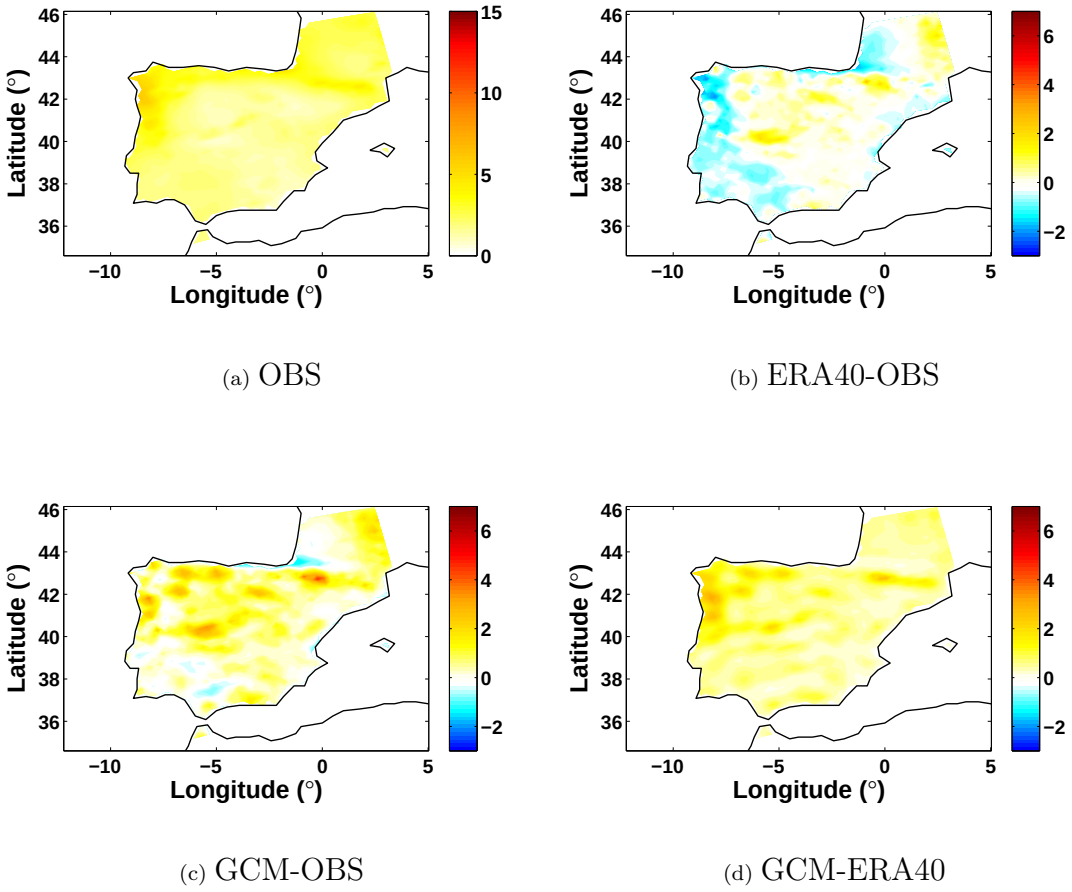
Figure 4.5: Summer precipitation climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d) for the 1961-2000 period (mm/day).

Contrary to what happens for the ERA40-OBS, GCM-OBS (figure 4.2c) shows mainly overestimation of winter precipitation, with only a couple of isolated areas where this difference is negative. Furthermore, the differences between GCM-driven winter climatology ensemble mean and observed climatology are higher than for ERA40, with the maximum reaching 6 mm/day in areas such as the northern west and north Iberian coast, the centre of the IP and along the Pyrenees. Similarly to ERA40-OBS, the area where the GCM ensemble mean shows lower deviation from observed climatology is along the eastern Iberian coast.

Since along the north-weastern coastline, ERA40-OBS field shows negative values, while GCM-OBS shows positive values, it would be expected that this is the area of largest difference between GCM and ERA40 climatology ensemble mean (figure 4.2d). Differences between these two ensembles have

an average value of 1.5 mm/day. Also as expected, the area where the difference between GCM and ERA40 ensemble mean is lower is along the eastern Iberian coast, where both ensembles present results close to observed as well.

As said before, associated to the ensemble mean is always uncertainty related to the spread of the models relative to the ensemble mean. In this case, the uncertainty measure chosen was the IQR, which is shown in figure 4.3 for both the ERA40-driven ensemble mean (4.3a) and GCM-driven ensemble mean (4.3b).



(a) OBS

(b) ERA40-OBS

(c) GCM-OBS

(d) GCM-ERA40

Figure 4.6: Autumn precipitation climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d) for the 1961-2000 period (mm/day).

As can be seen in the IQR representations (figure 4.3), the overall field presents higher values for GCM-driven ensemble mean than for ERA40 ensemble mean, which is in accordance with what was seen in figure 4.1a. For ERA40, the maximum IQR areas correspond to the areas where there are maximum absolute differences of ERA40-OBS. GCM ensemble mean IQR shows higher values for along the northern west coast of the IP and the Pyrenees, where the GCM-OBS field also showed maximums. This shows that the areas where the GCM-driven ensemble mean is farthest from observations are also the areas where models tend to diverge in results, therefore producing high IQR.

For spring precipitation, the observed climatology (figure 4.4a) shows a similar spatial pattern as the winter one. However, the difference field between ERA40 ensemble mean and OBS - figure 4.4b - shows that the ERA40 ensemble mean overestimates precipitation in most of the IP. Furthermore,

there is a significant area where the ERA40-OBS difference is nearly null. However, once more, differences are larger over the Pyrenees. GCM-OBS field (figure 4.4c), shows a similar pattern to what is seen for winter, but with higher difference values.

The IQR associated with ERA40 and GCM ensemble mean of spring climatologies (not shown) shows higher uncertainty in areas such as along the north coast and in the Pyrenees, specially for GCM-driven ensemble. Furthermore, and in accordance with figure 4.1a, GCM-driven ensemble shows overall higher uncertainty than the ERA40-driven ensemble.

The observed climatology of summer precipitation - figure 4.5a - shows a clear pattern of very dry/dry south and north IP, respectively. All three difference fields (figures 4.4b, 4.5c and 4.5d) show very low values, with the exception of the north, especially the Pyrenees that show higher differences.

Precipitation is associated with highly complex dynamical systems and is, therefore, a difficult variable to simulate. On the other hand, the dynamical systems associated with the lack of rain are simpler and therefore, more easily simulated. Since summer is the season with least precipitation, as can be confirmed by comparing observed climatologies for all seasons, the small differences between ensemble means and observations are easily understood.

Not shown here is the IQR for ensemble means of summer precipitation climatologies. However, these fields are consistent with what is expected and therefore show low uncertainties for this season, except for the northern IP, where there are significant amounts of precipitation and thus, uncertainty is higher.

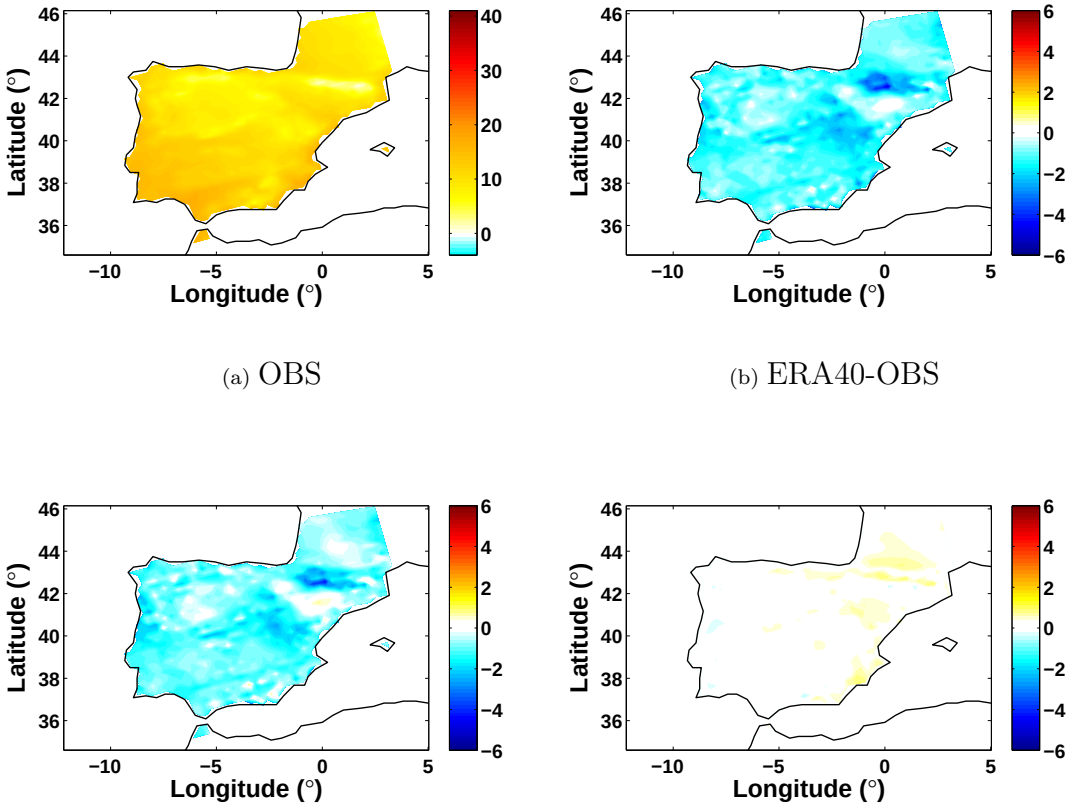
The observed autumn precipitation climatologies and ensemble mean differences can be seen in figure 4.6. Observed climatology is similar to that of spring. ERA40-driven ensemble mean shows a general underestimation of autumn precipitation, except in the more north-east area of the IP where it shows either null or low positive values. The maximum difference is found in the far north-west of the IP. Contrary to ERA40 ensemble mean, GCM shows an overestimation of the precipitation, with focus on the Pyrenees and the Meseta Central. Differences between GCM and ERA40 ensembles are higher in west coast, Pyrenees and Meseta Central, but show a consistent pattern of positive values, meaning that GCM ensemble mean overestimates precipitation also in relation to ERA40, which would be expected since in those areas ERA40 either underestimates or overestimates by low levels. IQR for autumn precipitation ensemble climatology (both ERA40-driven and GCM-driven) shows a very similar pattern to spring and are, therefore, not shown here.

4.1.2 Maximum Temperature

The observed winter climatology for maximum temperature, as well as difference the fields, are shown in figure 4.7. The observed climatology shows lower maximum temperatures for mountainous regions such as the Pyrenees; also, southern IP shows higher winter maximum temperatures. Differences between ensemble means and observations (figures 4.7b and 4.7c) show very similar patterns which, together with the fact that the GCM-ERA40 field shows mostly null values, indicates that there is very little difference between the two types of simulations (ERA40-driven or GCM-driven). Both these ensembles show an underestimation of the maximum temperature with higher negative values along the Pyrenees, where this underestimation is higher. This is also the area where the GCM-ERA40 field shows non-null values.

Similar to what happened with the difference fields, both ensemble means show very similar IQR fields (not shown), indicating that the uncertainty associated with the ERA40 ensemble is similar to that associated with the GCM ensemble. This is in agreement with figure 4.1b that shows an overall similar uncertainty for both ensemble means. In addition, the IQRs show higher values along the eastern coastline (for both ensembles) and near the Pyrenees (only for GCM-driven ensemble mean).

As seen before for winter, spring climatology differences (ERA40-OBS and GCM-OBS) - figure 4.8



(a) OBS

(b) ERA40-OBS

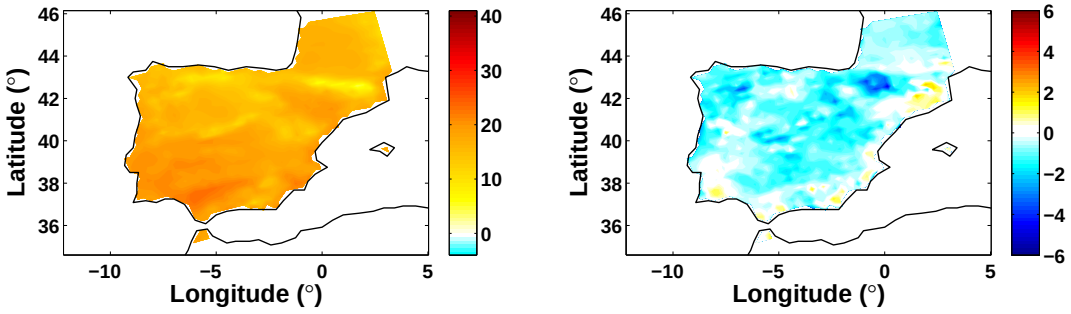
(c) GCM-OBS

(d) GCM-ERA40

Figure 4.7: Winter maximum temperature climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d), for the 1961-2000 period (°C).

- are very similar, showing a systematic underestimation of both ensembles with regard to observations. However, GCM-driven ensemble shows a higher underestimation of maximum temperature than ERA40 ensemble mean. Although not very evident, there are differences between the two ensemble means and which can be seen in figure 4.8d. In this figure, one can see that ERA40 ensemble mean simulates higher maximum temperatures than GCM ensemble mean and is therefore a less marked underestimation of observations (this was already concluded by comparing figures 4.8b and 4.8c). Uncertainty fields for both ensemble means (not shown) are very similar. These show higher uncertainty along the south-eastern coast of the IP, where the Sierra Nevada stands. On average, the IQR for maximum temperature spring climatologies is 2 °C and reaches no more than 3 °C.

Summer maximum temperature observed climatology can be seen in figure 4.9a. This field shows a maximum in the south middle area of the IP that reaches 35 °C, while on average, maximum temperature is only about 25 °C. The Pyrenees, as well as the northern coastal region show a minimum of maximum temperature (between 15 and 20 °C). The difference field for ERA40-OBS shows two clear areas where the ERA40 ensemble mean either over or underestimates maximum temperature. Mainly along the coastline (west, south and east) there is an overestimation by ERA40 ensemble mean, which varies between 1 and 3 °C. On the other hand, the remainder of the domain shows underestimation of the maximum temperature, but with lower values ($\sim$ 1 °C). GCM-driven ensemble shows a similar pattern having however higher area of underestimation by the models, specially in the far north where the difference values reach -3 °C. Not shown here is the IQR fields associated with each of the



(a) OBS

(b) ERA40-OBS

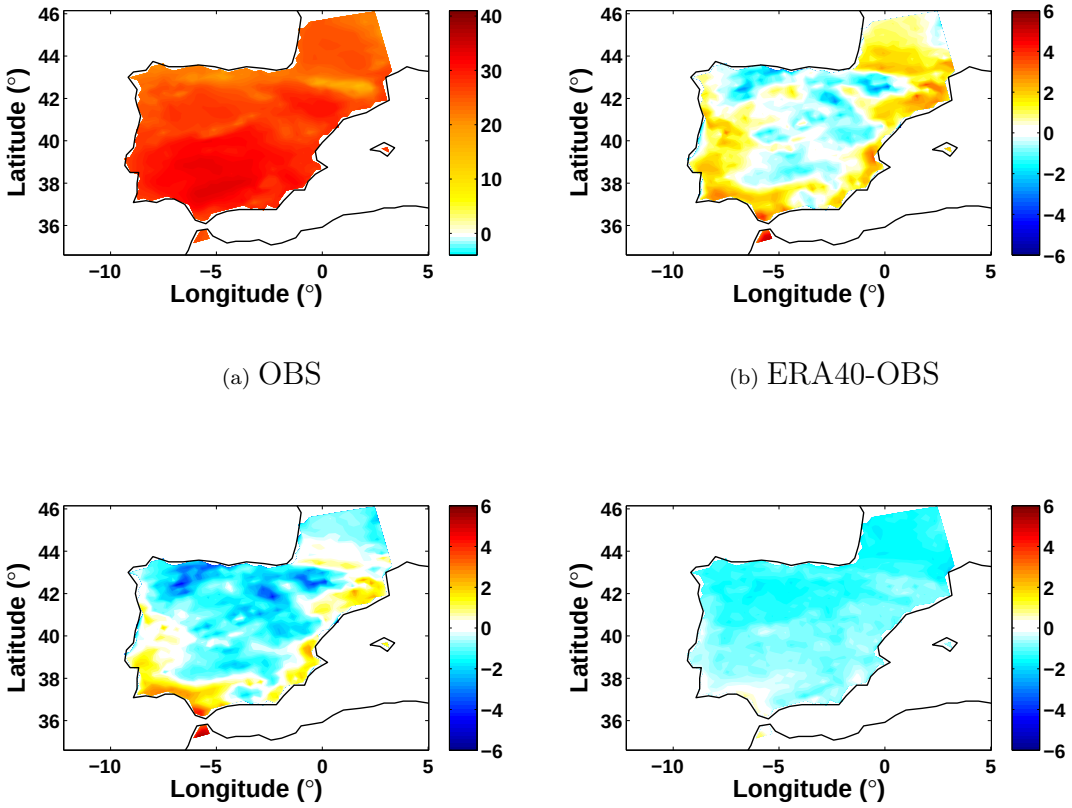
(c) GCM-OBS

(d) GCM-ERA40

Figure 4.8: Spring maximum temperature climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d), for the 1961-2000 period (°C).

ensemble means, which show higher uncertainty in areas of more complex topography, as seen before for other seasons and precipitation.

Figure 4.10 shows autumn maximum temperature climatologies and difference fields. As can be seen in the difference fields - figures 4.10b and 4.10c - both ensemble means show very similar differences from the observations. Largest differences between modelled data and observations take place for the Pyrenees area. Interestingly, this area shows overestimation by models on the east and underestimation to the west when compared to the observations. Even though ERA40-OBS and GCM-OBS fields show very similar patterns, when the GCM-ERA40 difference field is computed, some asymmetries arise. GCM-driven ensemble mean shows an underestimation of maximum temperature with relation to ERA40-driven ensemble mean. These differences are low (between 0 and 1 °C) and nearly constant over the entire domain. Although difference fields (ERA40-OBS and GCM-OBS) show low IQR values, fields for each of the ensemble means show uncertainties of ~ 2 °C that reach up to 5 °C in areas of more complex terrain.



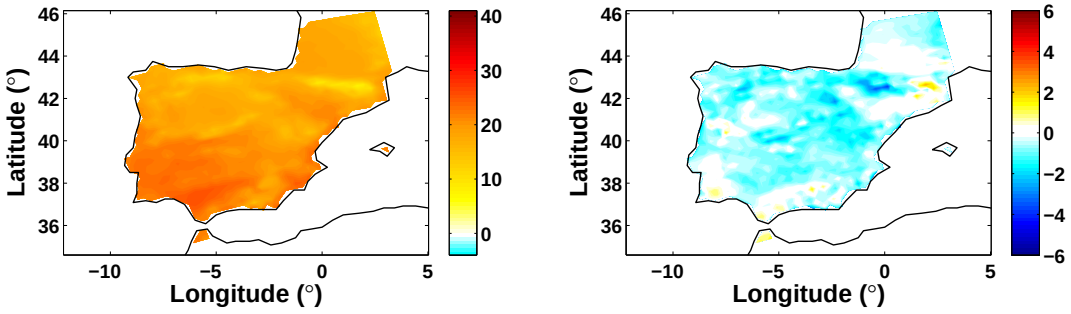
(a) OBS

(b) ERA40-OBS

(c) GCM-OBS

(d) GCM-ERA40

Figure 4.9: Summer maximum temperature climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d), for the 1961-2000 period (°C).



(a) OBS

(b) ERA40-OBS

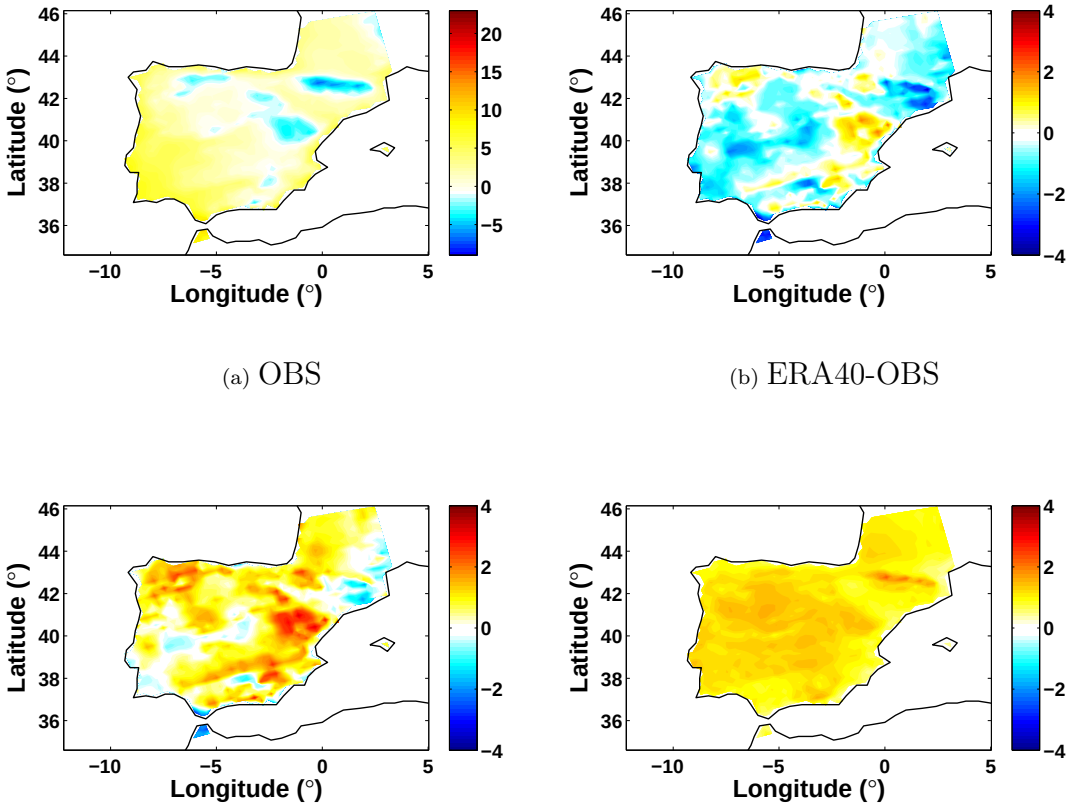
(c) GCM-OBS

(d) GCM-ERA40

Figure 4.10: Autumn maximum temperature climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d), for the 1961-2000 period (°C).

4.1.3 Minimum Temperature

In figure 4.11a, the observed minimum temperature winter climatology is shown. Temperatures in the Pyrenees region are low (−5 °C), as well as other high altitude areas (−2.5 °C) in the north of the IP. In the rest of the domain, minimum temperature winter climatologies show positive values, between 5 and 10 °C. ERA40-OBS field shows that ERA40 ensemble mean mainly underestimates minimum temperature, with the exception of the far north coast and east coast of the IP, where this field shows positive values. Underestimation is higher (with a 3 °C maximum) in the Pyrenees and in the centre of the IP, while overestimation is not higher than 2 °C and is, on average, only 1 °C. Mainly, the GCM-OBS field shows positive values, reaching a maximum of 4 °C and being, on average, between 1 and 2 °C. The two ensemble means show very different model reproduction of the minimum temperature winter climatology, being that the underestimation by ERA40 is lower than the overestimation by GCM ensemble mean. These results lead to a positive difference field between GCM-ERA40 in the entire domain, specially over the Pyrenees where the ERA40 underestimation of observations is higher. In agreement to what was seen in figure 4.1c, ERA40 shows an overall lower uncertainty than GCM ensemble mean. ERA40 shows higher IQR along the coastline while GCM ensemble driven only shows close to null values of uncertainty in the southern area of the IP, close to Sierra Nevada.



(a) OBS

(b) ERA40-OBS

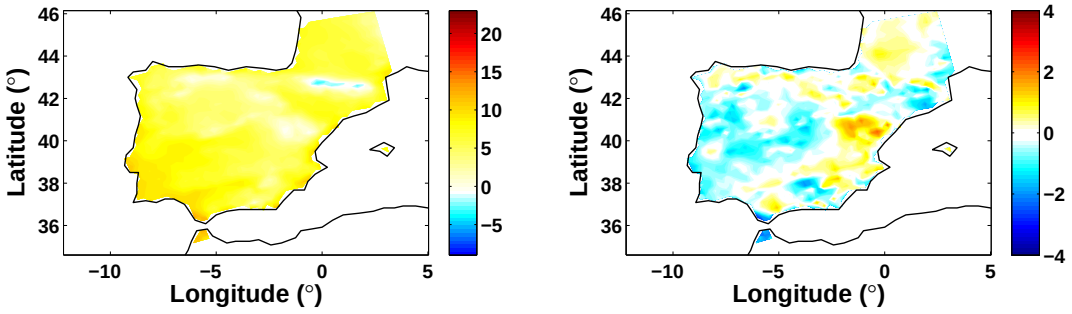
(c) GCM-OBS

(d) GCM-ERA40

Figure 4.11: Winter minimum temperature climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d), for the 1961-2000 period (°C).

Spring minimum temperature as well as difference fields can be seen in figure 4.12. Contrary to what was seen for winter, spring difference fields for ERA40 and GCM are similar in pattern although with distinct amplitudes. Both difference fields (figures 4.12b and 4.12c) show modelled minimum temperatures higher than observations by $\sim 2\,^{\circ}\mathrm{C}$ in the north-eastern area of the IP. The remainder of the domain shows negative differences for ERA40-driven ensemble mean and either negative or null differences for GCM-driven ensemble mean. These results show that, although both ensembles show similar results, GCM ensemble mean results are actually closer to observations than ERA40-driven ensemble mean. However different these two difference fields are, the actual differences between GCM and ERA40-driven ensembles are low and stay between zero and $1\,^{\circ}\mathrm{C}$, being higher in areas of more complex terrain.

Observed minimum temperature summer climatology can be seen in figure 4.13, together with the difference fields. It can be seen that ERA40 shows higher discrepancy from the observations than GCM-driven ensemble mean. Figure 4.13c shows several areas where the difference between GCM and observations is zero, while on the figure 4.13b. ERA40 and OBS differences are positive and close to $1\,^{\circ}\mathrm{C}$. Difference field between GCM and ERA40 shows only negative values, meaning that GCM ensemble mean produces overall lower minimum temperatures (by $1\,^{\circ}\mathrm{C}$) with lower values reaching zero, to the west. GCM ensemble mean has a higher uncertainty (not shown), specially in the south-east quadrant of the IP, where it reaches $6\,^{\circ}\mathrm{C}$. ERA40 ensemble mean uncertainty is also higher in this area, but with much lower values, reaching only about $3\,^{\circ}\mathrm{C}$.



(a) OBS

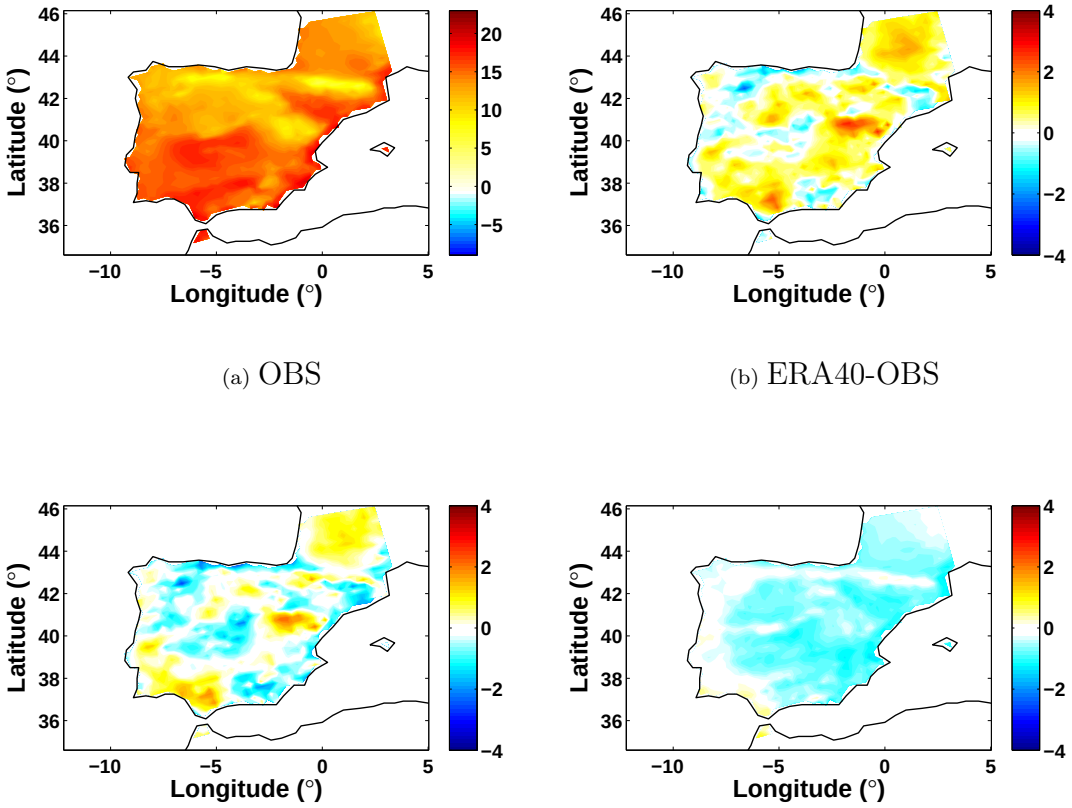
(b) ERA40-OBS

(c) GCM-OBS

(d) GCM-ERA40

Figure 4.12: Spring minimum temperature climatology (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d) for the 1961-2000 period (°C).

For both ensembles, difference with observations and IQR fields for autumn minimum temperature are similar to spring ones and are, therefore, not shown here.



(a) OBS

(b) ERA40-OBS

(c) GCM-OBS

(d) GCM-ERA40

Figure 4.13: Summer minimum temperature climatology for observations (a), and difference fields for: ERA40-driven ensemble mean climatology and observed (b), GCM-driven ensemble mean climatology and observed (c) and GCM-driven and ERA40-driven ensemble mean climatologies (d), for the 1961-2000 period (°C).

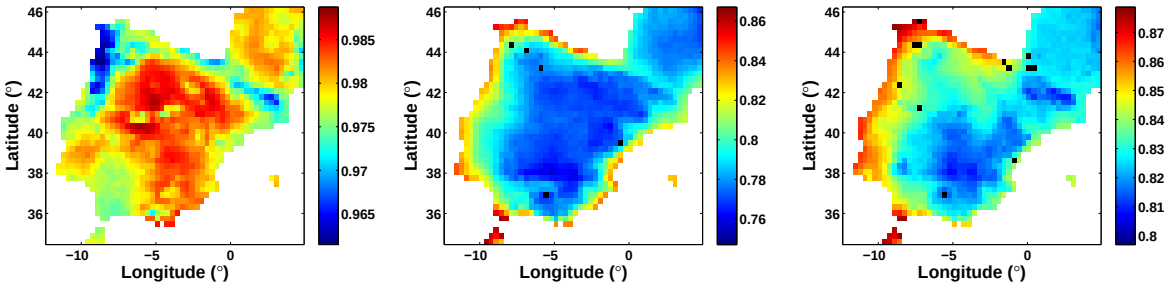
4.2 Inter-model comparison Part I - Regional Climate Model performance Evaluation

RCMs can be evaluated in their skill, by using all the ERA40-driven simulations and comparing them to observations. Furthermore, the ensemble mean of these simulations can also be compared to observed data.

In order to evaluate overall ERA40-driven ensemble mean performance, its distribution was compared to the observed distribution through the use of the K-S test in each grid point, for each of the three variables considered.

When analysing figure 4.14, where the maximum difference between observed and modelled CDFs (D_n) is shown, one can easily see that the areas where D_n is larger for precipitation tend to coincide with the areas where D_n is lower for both maximum and minimum temperatures. High values of D_n indicate that there is a large difference between the CDFs, at one point or another. The contrary is also true.

Firstly and foremost, the K-S test shows that in all grid points for precipitation and most of the



(a) Precipitation (b) Max. Temperature (c) Min. Temperature

Figure 4.14: K-S test statistics (maximum difference between the observed and ERA40- driven ensembles mean CDF - D_n) determined for the entire 40 year period between 1961 and 2000, for each of the variables being studied. Black pixels represent areas where the null hypothesis is rejected and therefore the distributions are found to be significantly different.

grid points for temperatures, modelled and observed CDFs can be considered equal. D_n determined for precipitation shows larger values inland in the IP, and in the eastern region. Along the northern Portuguese and Galitian coastal areas there is smaller difference between modelled and observed CDFs, showing that models have higher accuracy. Precipitation in the IP has three different forcings: frontal systems, convection and orography. Accuracy differences in model simulation of precipitation might be related to the different forcings in each area of the IP. In the interior area of the IP, where D_n is higher, precipitation is frequently of convective origin, due to topographic and geographic characteristics. Since the horizontal scale of these processes is significantly smaller than the model's horizontal resolution, they are solved empirically using parametrisations, leading to larger differences between observed and modelled data. On the other hand, in the northern areas of the western coastal region where D_n values are lower, precipitation is mainly due to the passage of synoptic frontal systems, which are of large scale. This might be the reason for local better model accuracy since these processes are explicitly resolved by the models. Mountainous regions such as Pyrenees and Sierra Nevada also show lower D_n values, although being a region of complex topography. Nevertheless, the local distributions of simulated and observed precipitation are not statistically different.

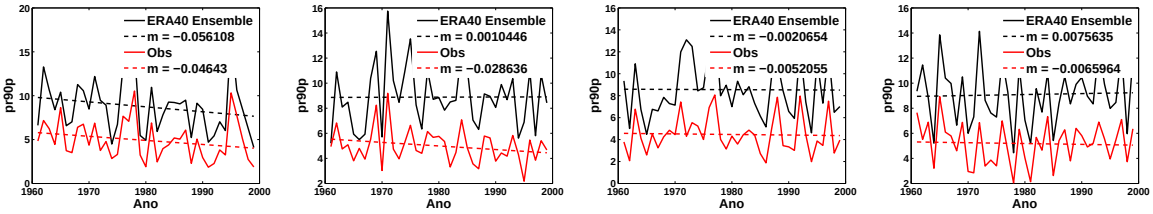
Maximum and minimum temperatures show a clear dichotomy of lower/higher D_n values in centre/ coastal IP, which is more evident for minimum temperature. High D_n values along the coastline might be associated with the poor resolution of the water-land border.

The fields presented so far represent the seasonal climatologies. However, equal climatologies can arise from series with distinct distributions. The extremes of a distribution are as important as the mean and together, these statistics provide valuable information. Therefore, variability should also be taken into account when studying model performance. Furthermore, as this study is taking into consideration long time periods, 40 years, there is the possibility of a trend in the variables that is also significant when studying model performance.

4.2.1 Trends of precipitation and temperature indices

In order to take variability into account when studying the RCM performance, the pr90p, tasmax90p and tasmin10p trends of yearly values for each season were determined. To provide insight into overall behaviour of these variables, time series of the spatial mean - together with the respective trend - was computed.

In figure 4.15, the time series of seasonal pr90p Iberian mean and respective trend, can be seen. Observations show that there is a negative trend in pr90p (number of days with precipitation amount

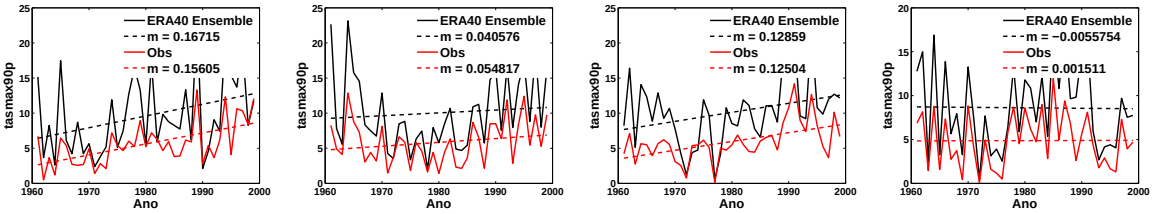


(a) DJF (b) MAM (c) JJA (d) SON

Figure 4.15: Time series of seasonal pr90p spatial mean in days/year (full line) and respective best fit line (dashed line) of observed values (in red) and ERA40-driven ensemble mean (in black). m is the slope of the best fit line and thus represents the trend in days/year^2.

---

above the $90th$ percentile for the control period) leading to a lower number of days in 2000 when compared to 1961 for all seasons. The highest absolute trend is found for winter months ~ -0.056 days/year^2). For the other seasons, this trend is on average ten times smaller with the lowest absolute value for spring (only 0.001 days/year^2). Furthermore, observed pr90p shows yearly values lower than ERA40-driven ensemble. The best fit line's slope represents the spatial mean average of pr90p trend. As can be seen, the observed and ERA40-driven ensemble trends are similar and have the same sign for the extreme seasons, winter and summer. On the other hand, spring and autumn show different sign trends for observations and ERA40-driven ensemble meaning that the overall Iberian trend is not being well reproduced by the ERA40-driven ensemble mean.

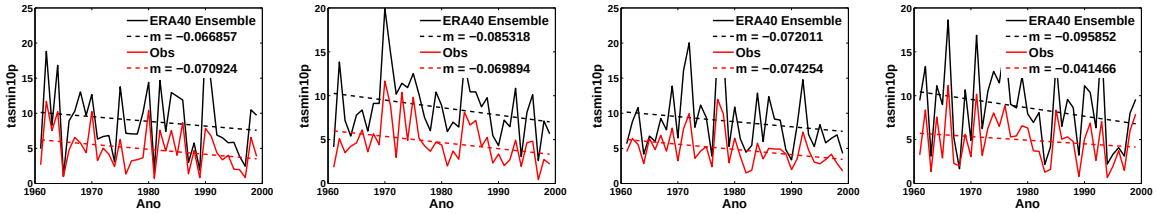


(a) DJF (b) MAM (c) JJA (d) SON

Figure 4.16: Time series of seasonal tasmax90p spatial mean in days/year (full line) and respective best fit line (dashed line) of observed values (in red) and ERA40-driven ensemble mean (in black). m is the slope of the best fit line and thus represents the trend in days/year^2.

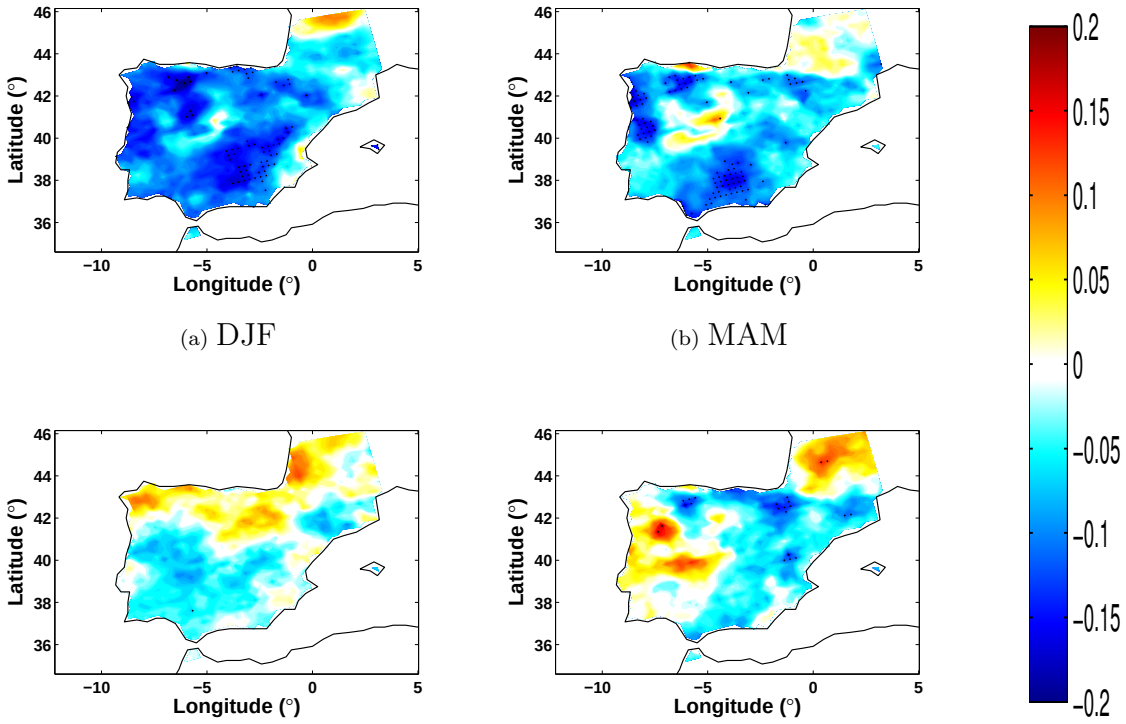
Time series of seasonal tasmax90p spatial mean, determined for both observations and ERA40 ensemble mean can be seen in figure 4.16. Observations show that the trend in days with maximum temperature higher than the $90th$ percentile for the control period is positive, meaning that in the year 2000 there were more warm days than in 1961. Overall, tasmax90p tends to be overestimated by ERA40 ensemble mean, with the ERA40 minimums closer to the observed minimums than the maximums. Winter and summer have the highest trends (~ 0.16 days/year^2) while autumn has the lowest observed trend of 0.0015 days/year^2. For all seasons except autumn, the ERA40 trend is positive like observations which shows negative ERA40 ensemble mean trend, and positive observed trend.

By analysing figure 4.17, that shows the tasmin10p seasonal time series, one can see that in all seasons observed trends are negative and of the same approximate magnitude. Much like observations, ERA40 ensemble mean trends are negative, with its highest differences during autumn when ERA40 overestimates the trend by ~ 0.05 days/year^2. Similar to what happened for the previously analysed variables, ERA40 ensemble mean consistently overestimates tasmin10p, with the difference between ERA40 ensemble and observations being larger for the maximums and smaller for the minimums with nicely reproduced overall variations by ERA40-driven ensemble mean.



(a) DJF (b) MAM (c) JJA (d) SON

Figure 4.17: Time series of seasonal tasmin10p spatial mean in days/year (full line) and respective best fit line (dashed line) of observed values (in red) and ERA40-driven ensemble mean (in black). m is the slope of the best fit line and thus represents the trend in days/year^2.

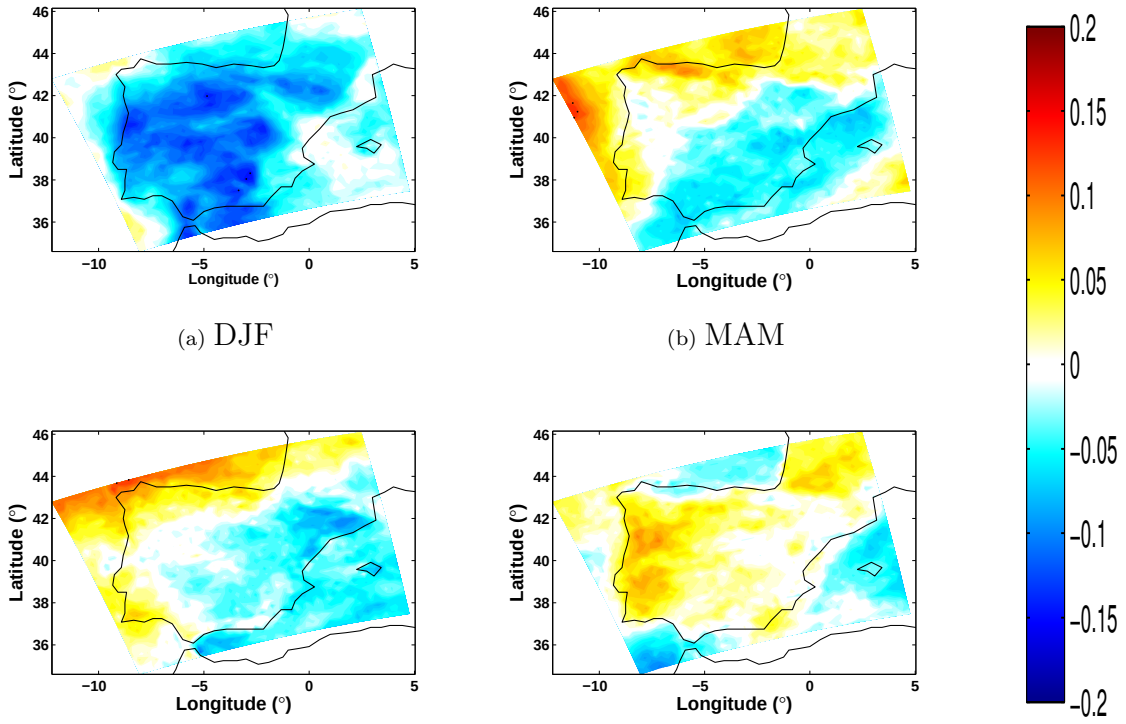


(a) DJF (b) MAM (c) JJA (d) SON

Figure 4.18: Observed trend in pr90p (days/year^2) determined for the 1961-2000 period (in colour) and grid points where the statistical significance is above 95% (in black dots).

The observed seasonal trend of pr90p shows very distinct seasonal and geographical patterns as shown in figure 4.18. During the winter months, the pr90p trend is negative, mainly without statistical significance showing a decrease in number of days with precipitation above the $90th$ percentile from 1961 to 2000 in the entire domain. However, this trend is lower in the centre of the IP and along the eastern coastline. Spring shows a less homogeneous negative trend, with positive values in the centre, where the trends were lower during the winter, reaching values nearly as high in magnitude as the negative trends. Summer presents an even more distinctive geographical pattern with positive trends along the northern part of the IP and negative trends in the remainder of the domain. The autumn field display an increase in rainy days in the western and central areas of the IP and decrease in the rest of the domain.

The trend in pr90p shows geographical patterns with seasonal specificities which should be well reproduced by models in order for them to be useful in future climate change studies.



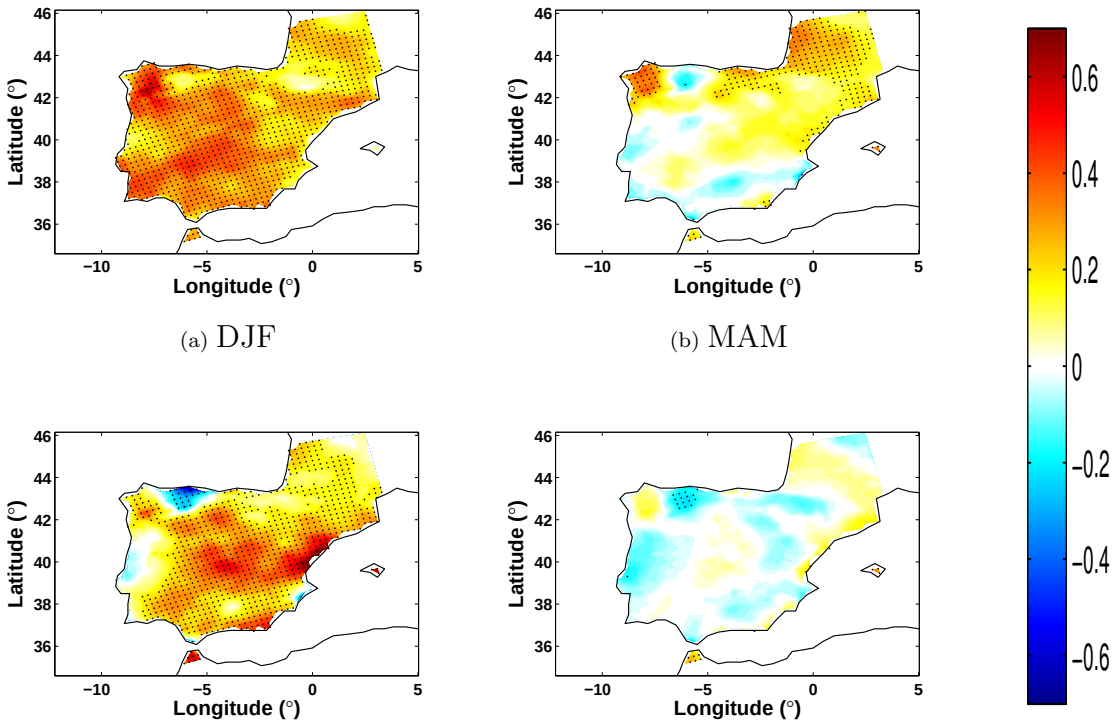
(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.19: ERA40-driven simulations' ensemble mean of the pr90p trend (days/year^2) determined for the 1961-2000 period.

The ensemble mean and observed trends of pr90p for ERA40-driven simulations can be seen in figure 4.19. When comparing the ensemble mean trend with the observed trend for winter months, one can see that the sign of the trend is being well reproduced. However, simulations seem to be underestimating this trend, with less negative values and larger domain with -0.06 or higher (but less than zero). The spring patterns are less well reproduced. Ensemble mean of ERA40-driven simulations shows positive trends in the north and south-west of the IP (between 0.05 and 0.1) and negative (~ -0.05) in a band aligned with the eastern coastline and extending to over 500 km inside the IP. This pattern is not a good representation of the observed one for this season. Summer simulated trends are closer to the observed ones, with a north/ south pattern of positive and negative trends respectively. However, in the far south-western part of the IP, simulated trends are also positive, contrary to what is observed. In addition, the simulated trends are lower in magnitude than the observed ones. The season that shows less consistency between observed and modelled results is the autumn. ERA40-driven simulations' ensemble mean shows a positive pr90p trend in most of the IP with null values on the eastern side. Not only is this pattern far from the observed trends, but also the maximum values are lower than the ones observed. One can thus conclude that the trend in pr90p is being underestimated by the models.

As was done for precipitation, the seasonal observed trend for tasmaxp90p was determined for the 1961-2000 period (figure 4.20). Furthermore, the statistical significance was determined for each grid point. It can be seen in figure 4.20 that the extreme seasons (winter and summer) show mainly positive trends, as high as 0.6 days/year^2. These are also the trend fields that show statistical significance. The summer field also shows a geographical pattern of negative trends in the northern coast of the IP



(a) DJF (b) MAM

(c) JJA (d) SON

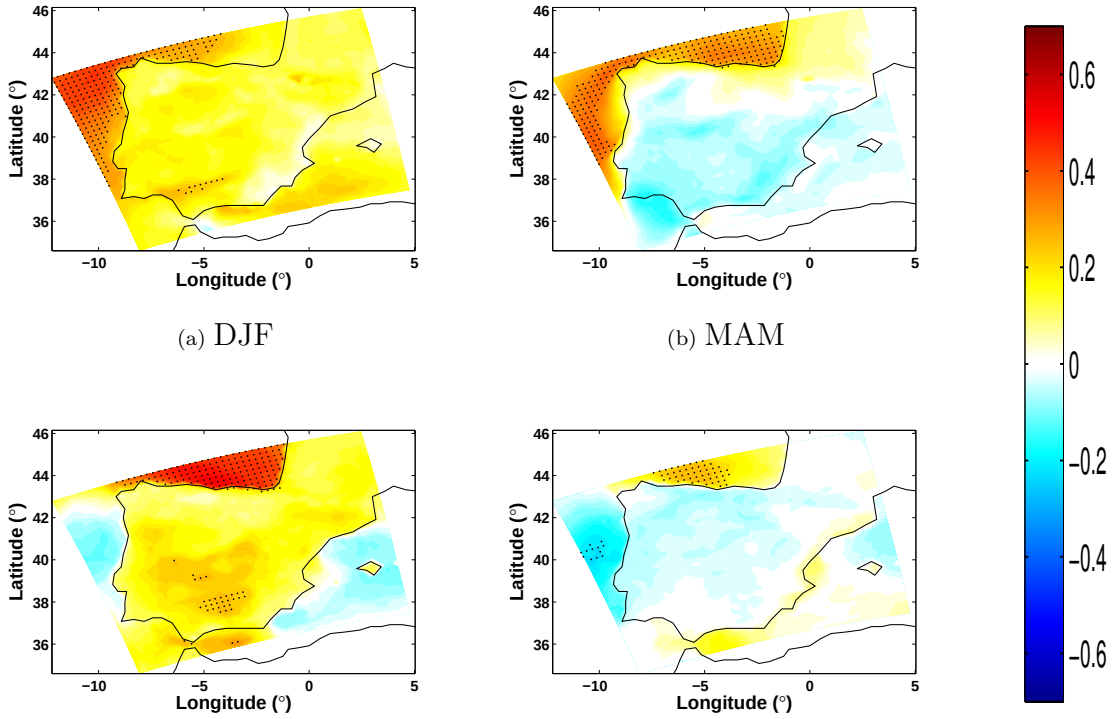
Figure 4.20: Observed trend of tasmax90p (days/year^2) determined for the 1961-2000 period (in colour) and grid points where the statistical significance is above 95% (in black dots).

---

and in a small coastal area north of Lisbon. Spring trends are positive in the eastern/north-eastern parts of the IP that reach a maximum of 0.3 days/year^2 and negative trends in the remainder of the domain reaching no more than -0.2 days/year^2. Autumn trends show a more monotonous pattern of trends varying mainly between zero -0.2 days/year^2.

The ERA40-driven simulations' ensemble of the tasmax90p is represented in figure 4.21. Comparing this figure with the observed one (figure 4.20), one can see that for winter and summer, the positive signal is represented. However, the maximum amplitude of the observed trend is larger and the simulations do not show the same spatial variability. Furthermore, the locally distinct negative trends in the observed summer data (north of the IP and along the west coast, north of Lisbon) are not reproduced by simulations. These poor representations of spatial variations and differences make a more localised study of climate change impacts highly inaccurate and therefore difficult to pursue, specially when for mitigation and adaptation purposes. Even tough the autumn negative trend is being simulated, its spatial variation is not. While observation fields show null or, slightly positive trends, in the centre of the IP, the ensemble only shows null values in the far south and along the eastern coastline. Spring also has an overall negative trend with the exception of the area along the northern coastline, where the trends are positive. Once more, this pattern is not a good representation of the observed trends.

Also, trends have been calculated for tasmin10p - figure 4.22. In general, tasmin90p observed trends are negative in the southern IP, extending north along the Portuguese west coast region for all four seasons and reaching -0.2 days/year^2 as the maximum negative trend. The aforementioned are also the grid points for which the tasmin10p trend shows statistical significance. The trends are locally positive in different areas for each season, reaching a maximum of 0.1 days/year^2. These positive trend maximums are located in the centre of the IP and north of Portugal for the winter, north-western part of the IP for the spring and summer and autumn, when they are extended more to the centre/



(a) DJF (b) MAM

(c) JJA (d) SON

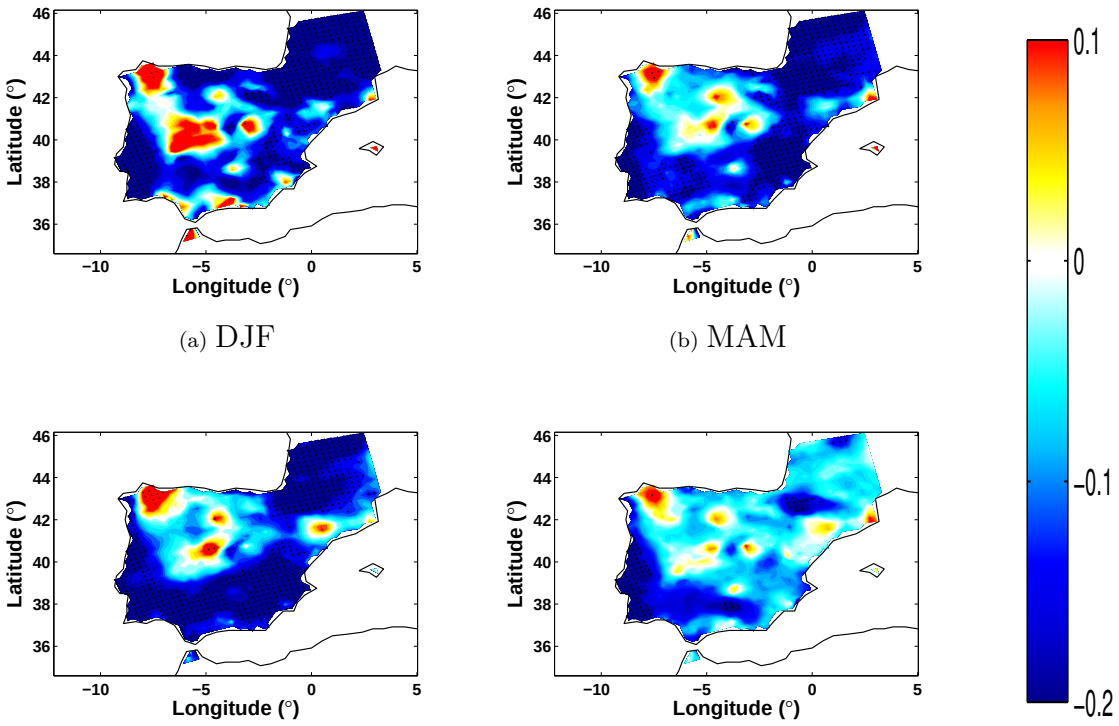
Figure 4.21: ERA40-driven simulations' ensemble mean of the tasmax90p trend (days/year^2) determined for the 1961-2000 period.

---

south of the IP.

The tasmin10p trend ensemble for ERA40-driven simulations is shown in figure 4.23. As happened for pr90p and tasmax90p trends, the overall sign of the tasmin10p trend is being reproduced by the ERA40-driven simulations. However, spacial patters of isolated areas of different trends and the amplitude itself of the signal are both not being well reproduced. Furthermore, winter ensemble shows an area of positive trends in the south-east that where observed trends are negative.

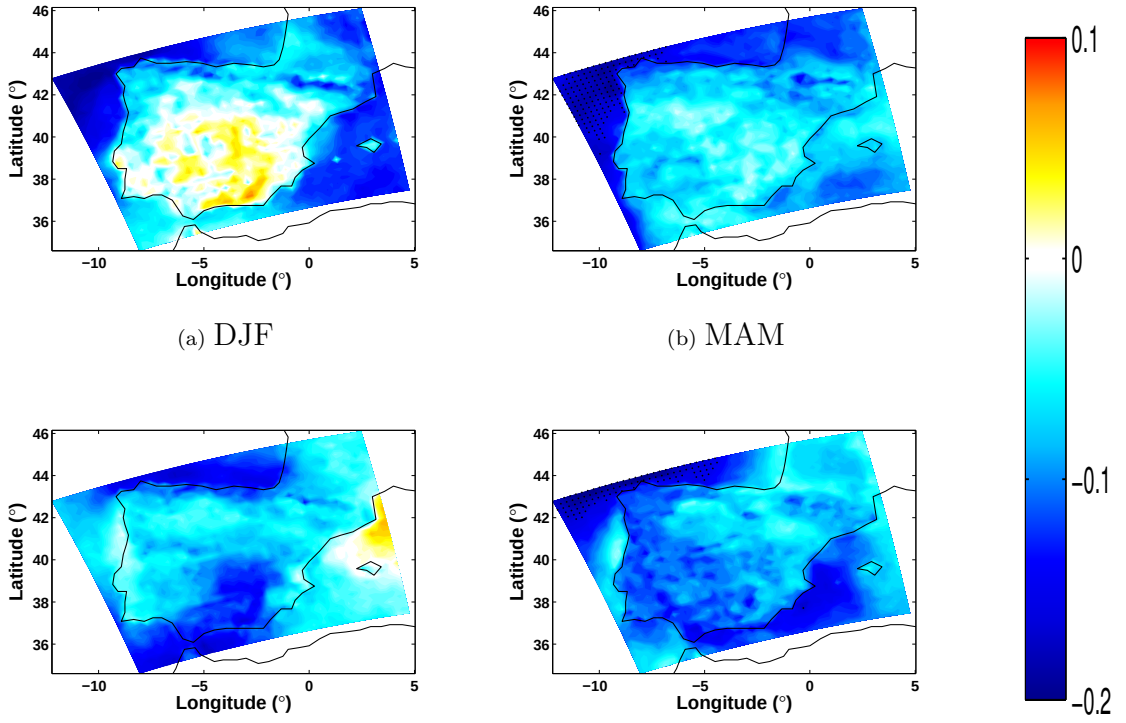
These results show that ensemble mean of the trends is an overall fair representation of the observed trends. Furthermore spatial patterns, specially those that are smaller and that present higher spatial gradient are not reproduced by simulations. This inability of spatial pattern reproduction presented by the simulations, creates difficulties in local studies of climate change. In order to perform them, one would have to use a higher spatial resolution, since some of the patterns are produced by a couple of grid points in these simulations.



(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.22: Observed trend in tasmin10p (days/year^2) determined for the 1961-2000 period (in colour) and grid points where the statistical significance is above 95% (in black dots).



(a) DJF (b) MAM

(c) JJA (d) SON

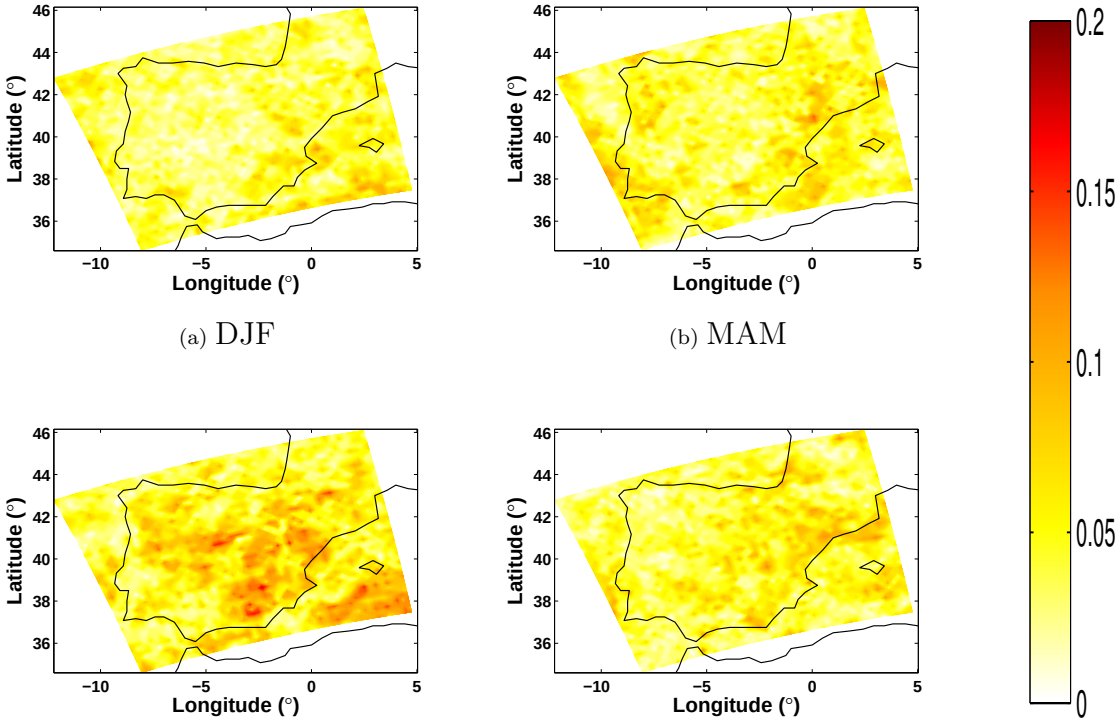
Figure 4.23: ERA40-driven simulations' ensemble mean of the tasmin10p trend (days/year^2) determined for the 1961-2000 period.

4.2.2 Uncertainties associated with the RCM ensemble

As done for the seasonal climatologies, for each ensemble, the interquartile range of the trends are also determined. The IQR associated to the ERA40-driven simulations' ensemble mean of the pr90p trend is shown in figure 4.24. Overall, the IQR has the same maximum magnitude of the original variable (the pr90p in this case). This indicates that there is great inconsistency among models when simulating the pr90p trend and therefore, great uncertainty associated to the ensemble mean determined. The season that shows larger ensemble uncertainty is summer, where the IQR values reach their maximum in the eastern and central parts of the IP, while the lowest uncertainty is for the winter months. The latter is in agreement with the previously seen of the fact that winter trend ensemble being is the closest to the observed pr90p trend.

The uncertainty (IQR) associated to the ensemble mean of the tasmax90p trend for ERA40-driven simulations can be seen in figure 4.25. The seasons that shows the larger IQR are spring and summer with values over land as high as 0.25 days/year^2. These are particularly high for the spring in the north/north-east of the IP and for the summer, in a band along the Iberian south-eastern coastline and in the south of Portugal. In the case of spring, this uncertainty coincides with an area of low ERA40-driven ensemble performance in simulating trends, showing that not only is the model ensemble a bad representation of the observed trends, but also the models have low consistency among themselves. On the other hand, autumn seems to be the season with lowest overall IQR, despite arguably being the season with the poorest spatial pattern reproduction by the ERA40-driven ensemble.

The IQR for the tasmin10p trend ensemble is shown in figure 4.26, with the highest IQR for the winter months. However, they show overall poor consistency between models, since the IQR is as high as the values of the trend. This shows that the simulations produce trends that vary largely among



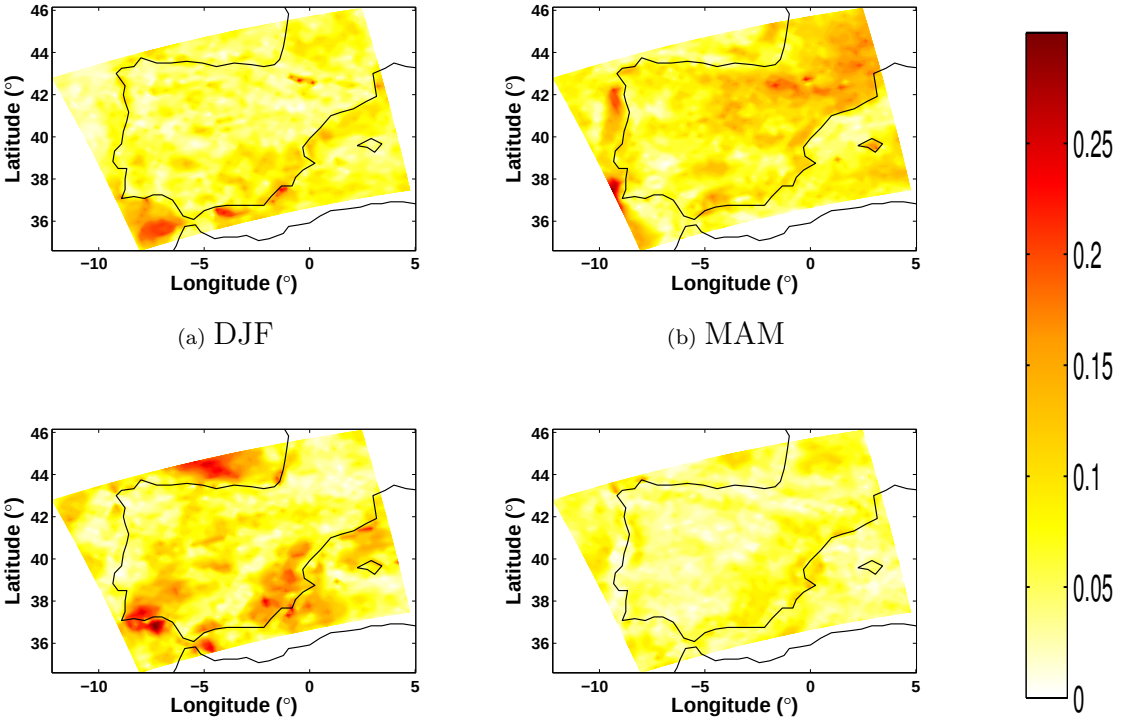
(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.24: Interquartile range associated with the ERA40-driven simulations' ensemble mean of the pr90p trend determined for the 1961-2000 period.

---

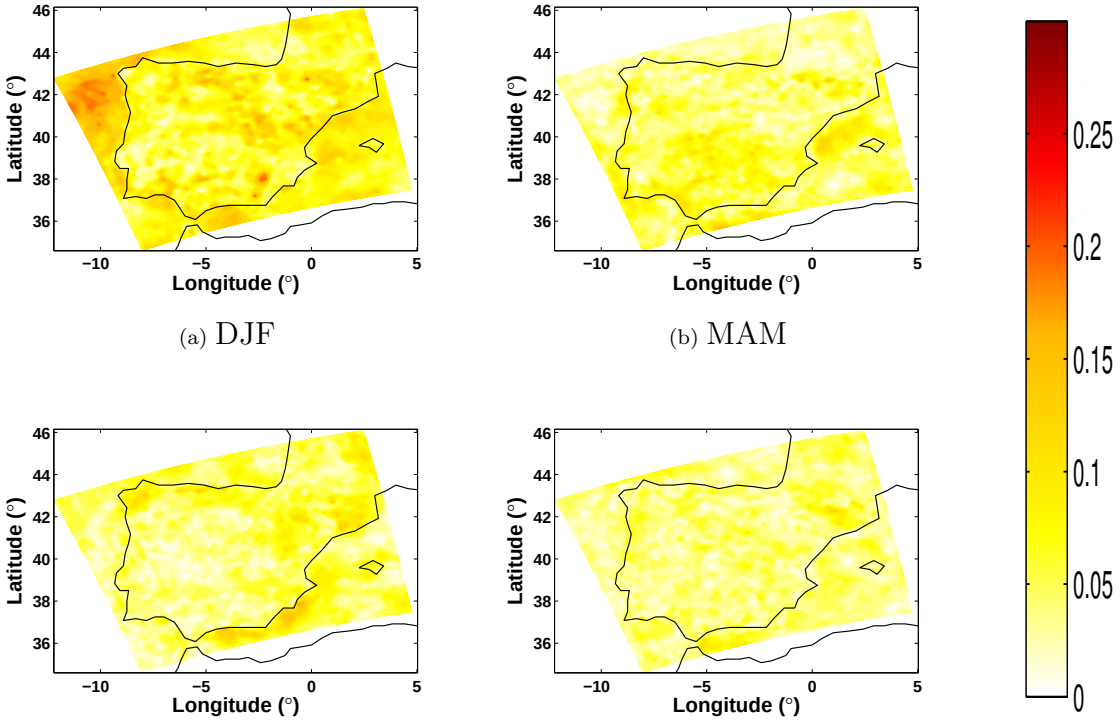
models.



(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.25: Interquartile range associated with the ERA40-driven simulations' ensemble mean of the tasmax90p trend determined for the 1961-2000 period.



(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.26: Interquartile range associated with the ERA40-driven simulations' ensemble mean of the tasmin10p trend determined for the 1961-2000 period.

4.2.3 RCM skill evaluation

In order to evaluate individual model performance in reproducing pr90p, tasmax90p and tasmin10p seasonal trends, the skill parameters in equations 3.2 through 3.5 were applied to the fields for each of the ERA40-driven simulations, as well as the ensemble mean (of which the fields were shown). These parameters were therefore determined for the trend fields. As can be seen in figures 4.27 through 4.29, which shows the values of the RMSE, BIAS, STDE and the CORR between observations and each of the models, one can conclude the ensemble of the simulations does not always outperform all models.

RMSE of pr90p trends shows low variation between models and ensemble mean, specially for summer and winter months , being that it is slightly lower for these than for autumn and spring (mean 0.04 against 0.06 mm/year^2). Overall, the lowest RMSE is for ensemble mean of all six models, even though for HadRM3Q16 - spring - and HadRM3Q0 - autumn - show lower RMSE than the ensemble mean. As for STDE, inter-model and inter-seasonal variation is much higher. The ensemble only shows the lowest STDE for summer months and on the remainder of the, most models outperform it. All models show approximately the same BIAS for each of the seasons, except for summer months, during which three models show negative values (HadRM3Q0/3/16) while the others are positive. Since there is a signal variation of BIAS between models, the ensemble shows a close to zero BIAS. CORR is in general highest for winter and lowest for spring months, when it is also less variable between models and ensemble. Only for winter is the correlation coefficient higher for the ensemble, reaching nearly 0.5. Summer and autumn have the highest inter-model correlation variability. For the first, the model with higher correlation is RCA while for the latter is HadRM3Q0 (both with ~ 0.4).

Skill measuring parameters determined for the tasmax90p trend fields show a lower performance than for pr90p. The season that shows lowest RMSE and STDE is autumn. RMSE shows lower

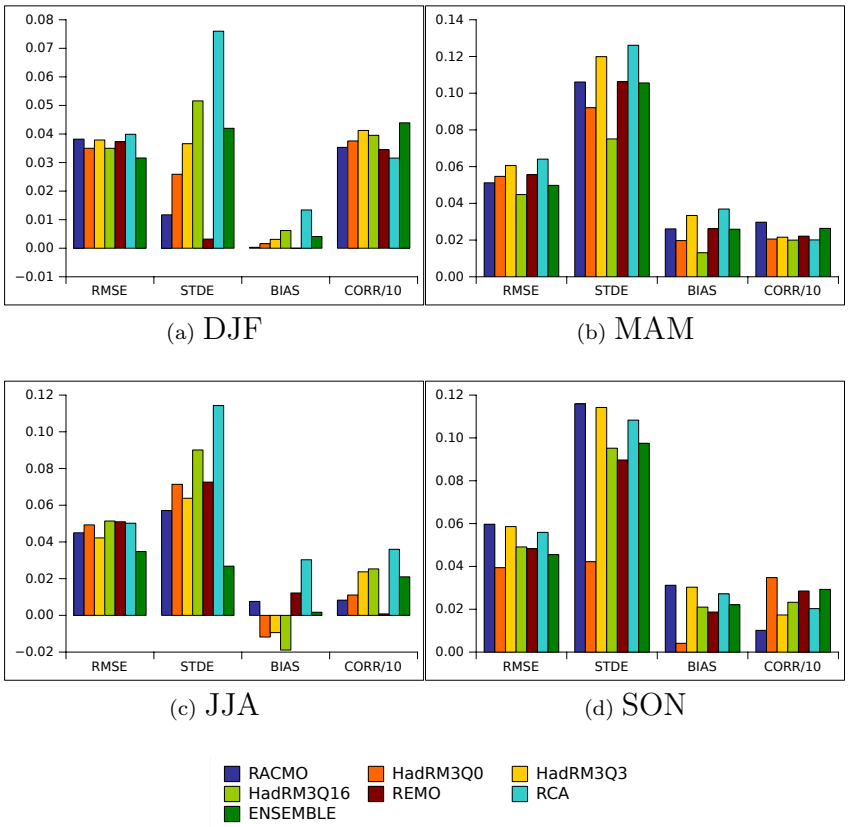


Figure 4.27: Skill Parameters - RMSE (days/year^2), STDE (days/year^2), BIAS (days/year^2) and CORR (divided by 10) - of the pr90p trend for each of the ERA40-driven simulations.

---

inter-model variability for autumn, with the lowest RMSE for HadRM3Q3 and ensemble mean, and highest for RCA. For winter months, the models with lowest error are HadRM3Q0 and REMO wiht RCA being the worst. For the other two seasons, HadRM3Q3 and HadRM3Q6 have the lowest RMSE, while the ensemble shows average performance. STDE is overall best for HadRM, with each sensitivity, showing higher skill for different seasons. In general, RCA is the model with the highest STDE. BIAS is negative for all seasons and models, with the exception of autumn HadRM3Q0 simulations that show positive BIAS. Furthermore, the lowest BIAS is for HadRM simulations, once more with a different combination of sensitivities for each of the seasons. CORR is higher during spring months, never surpassing 0.33 (REMO) and lowest during summer, reaching a maximum of only 0.17 (for HadRM3Q3).

Overall, one can say that the best performing model is HadRM3Q0, (low sensitivity run). On the other hand, the generally worst performing models are REMO and RCA, with higher RMSE, STDE and BIAS and lower correlation. RNMSE and STDE tends to be higher for winter months and lower, also with lower inter-model variability, for autumn. Winter, spring and summer have positive BIASes , except for REMO and RCA that have negative values. Autumn presents mostly negative BIAS, except for RACMO. Tasmin10p trend is the only variable that shows negative correlation values between observed and REMO and RCA modelled trends (~ -0.27). Higher correlations were found for winter - approximately 0.36 and lowest for spring, only ~ 0.19.

Taking into consideration these skill parameters, the model that presents lower RMSE and therefore lower amplitude errors, as well as lower phase errors (given by a lower overall BIAS) and higher CORR is HadRM3Q0.

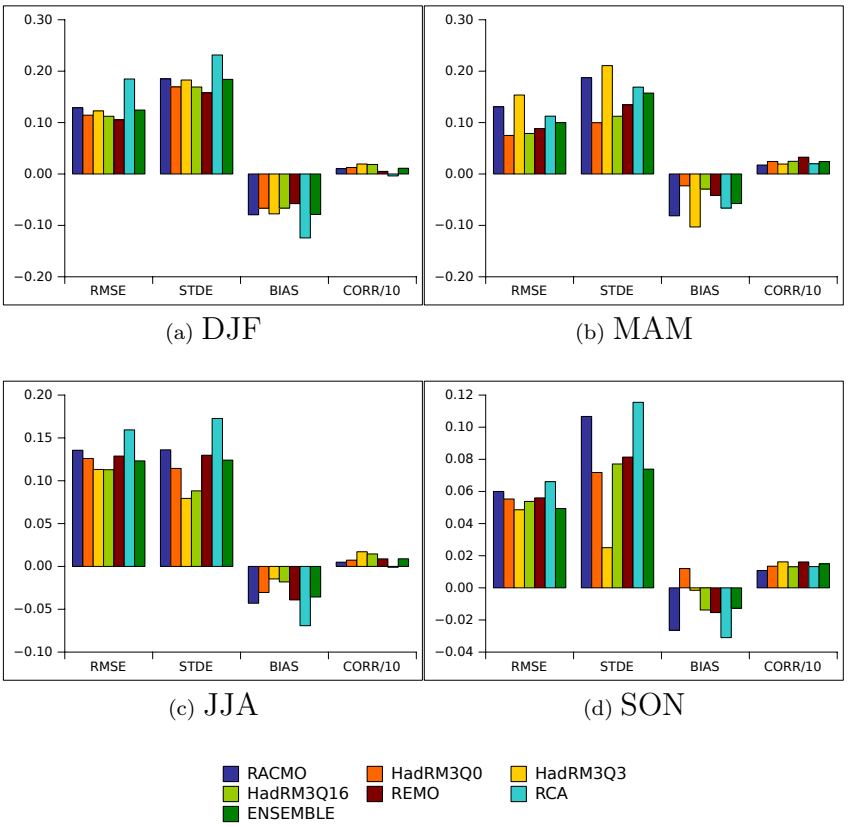


Figure 4.28: Skill Parameters - RMSE (days/year^2), STDE (days/year^2), BIAS (days/year^2) and CORR (divided by 10) - of the tasmax90p trend for each of the ERA40 driven simulations.

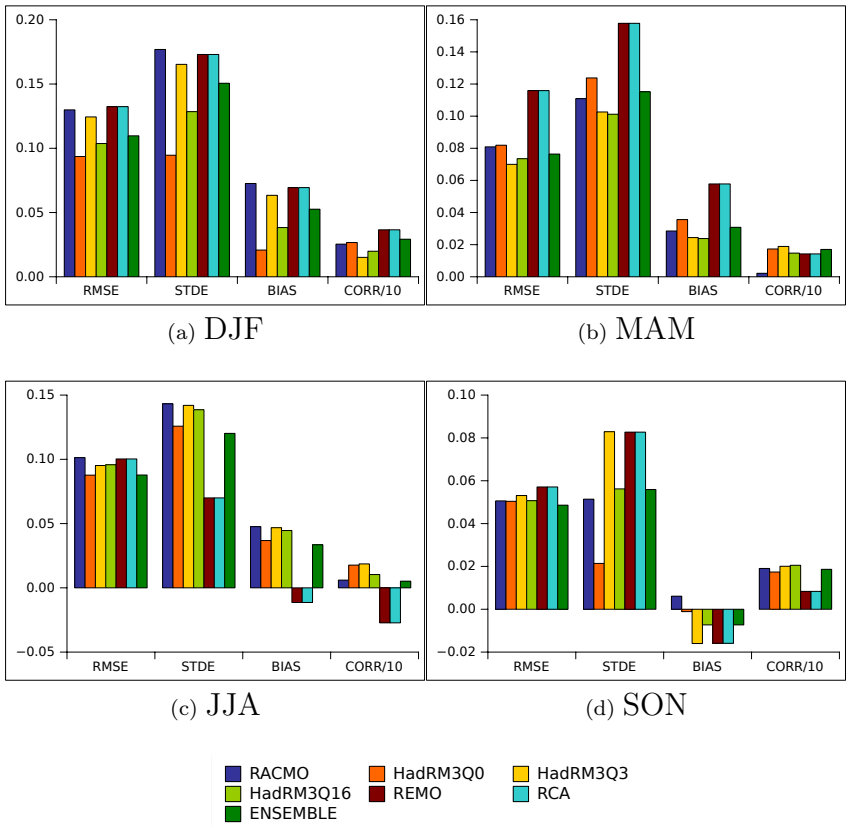
4.3 Inter-model comparison Part II - RCM-GCM combination performance Evaluation

This section is separated into 3 parts: in the first two, precipitation and temperature indices are analysed, while the third focuses on the GCM-driven skill.

4.3.1 Precipitation Indices

The indices studied here are the following: pr90p, CWD and CDD. Their description can be found in the previous chapter. As can be seen by analysing the spatial mean of precipitation indices seasonal climatologies (figure 4.30), GCM-driven simulations do not reproduce the seasonal climatologies of these indices. However, this misrepresentation is higher for pr90p than for the other two indices, although the ensemble uncertainty is lower. The pr90p index is highly underestimated by the GCM-driven simulations, which simulate an average of close to zero days per year with precipitation above the $90th$ percentile for the control period, while observations have approximately five days per year in every season.

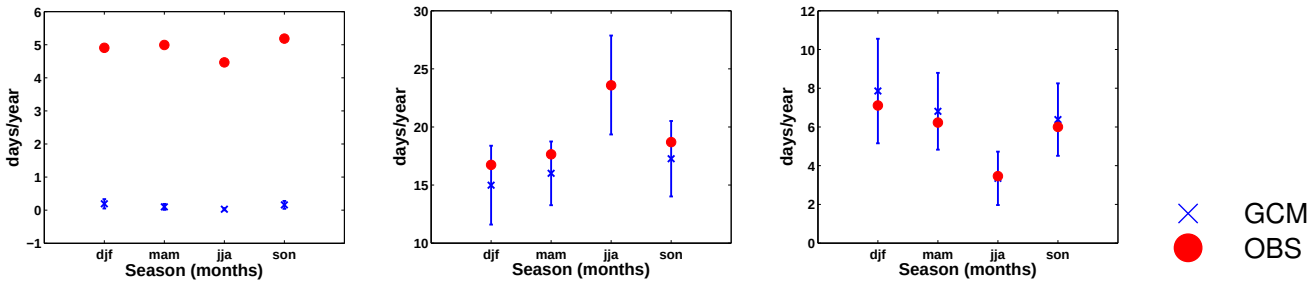
The number of dry days (CDD) and wet days (CWD), show high ensemble uncertainty, but modelled results are closer to observations, with the maximum difference between observed and modelled being only one day. During summer CDD, has modelled values close to the observed, but also has the highest ensemble uncertainty. As seen before for ERA40-driven ensemble, summer precipitation is better reproduced by models than for the other seasons. Furthermore, while CDD tends to be overestimated (higher for models than for observations), CWD has the exact opposite behaviour. This is to be expected, since models are in fact underestimating overall precipitation for the IP. From this figure



(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.29: Skill Parameters - RMSE (days/year^2), STDE (days/year^2), BIAS (days/year^2) and CORR (divided by 10) - of the tasmin10p trend for each of the ERA40 driven simulations.



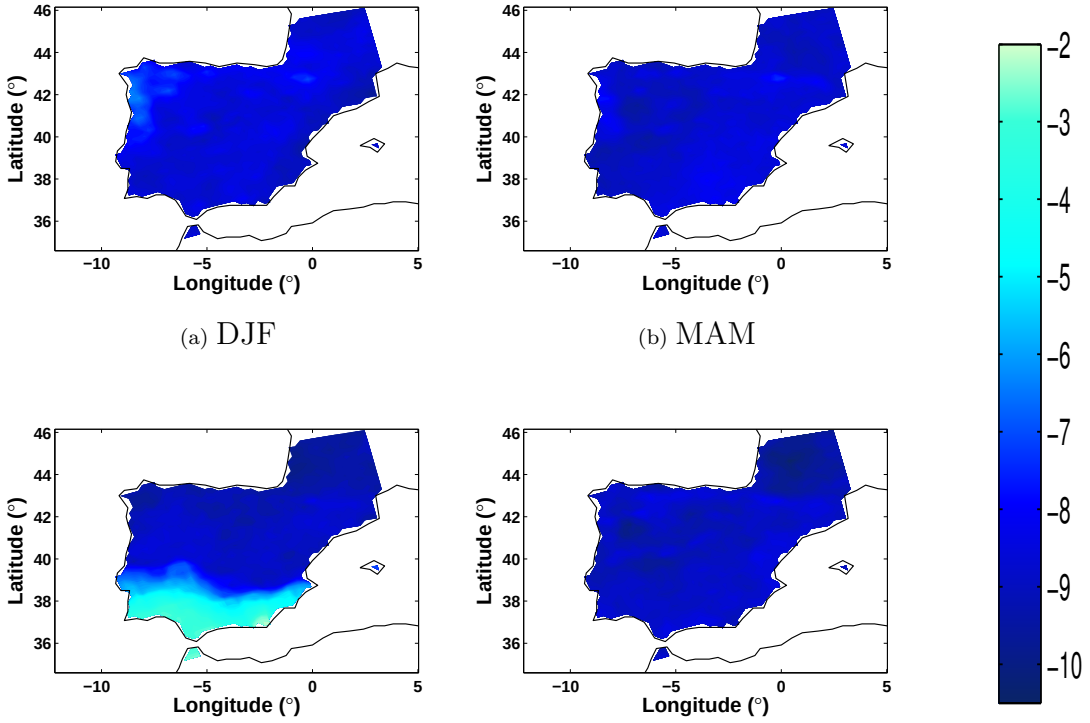
(a) pr90p (b) CDD (c) CWD

Figure 4.30: Spatial average of pr90p, CDD and CWD seasonal climatologies (days/year) over the IP for observations and GCM-driven ensemble mean (together with ensemble mean uncertainty bars).

---

one can further conclude that models tend to produce fair results on the wet/dry day ration but show difficulties in reproducing the precipitation amount (amplitude). Differences between GCM-driven ensemble and observed precipitation indices' seasonal climatologies are statistically significant at the 5% confidence level for all seasons.

For an in-depth understanding of model behaviour, there is the need to analyse the climatology fields of these indices. However, in order to more easily evaluate the differences between observations and modelled indices, the difference fields between GCM-driven ensemble mean and observations (GCM-OBS) was computed, as well as the uncertainty associated with each of the seasonal index ensemble mean.



(a) DJF (b) MAM

(c) JJA (d) SON

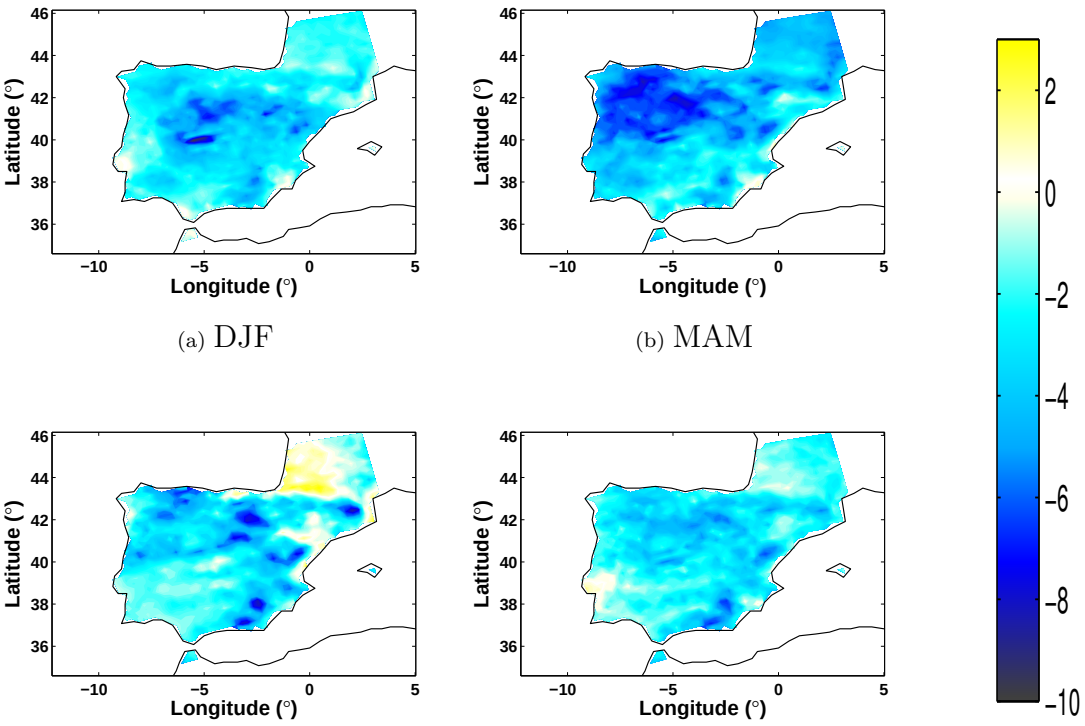
Figure 4.31: Difference field between GCM-driven and observed pr90p seasonal climatologies (GCM-OBS) over the IP for the 1961 to 2000 period (days/year).

The difference fields between GCM-driven ensemble and observed pr90p seasonal climatologies, can be seen in figure 4.31. The four fields have consistent negative values, meaning that models produce less days with high precipitation (precipitation above the $90th$ percentile for the control period). Furthermore, all seasons but summer show very similar fields, with values ranging from -8 to -10 days per year. On the other hand, the summer difference field shows a band of less negative values in the south of the IP that vary between -2 to -5 days per year. This is therefore the area and the season for which the GCM-driven ensemble mean better performs.

In order to understand the behaviour of GCM-driven simulations, there is the need to analyse ensemble mean uncertainty for the seasonal ensemble mean climatology for pr90p (not shown due to space constraints). Summer is the season with lower IQR. Together with the fact that the south of the IP has lower values of GCM-OBS during the summer, this shows that models tend to perform more consistently with each other and closer to observations during dry seasons, specially in the driest areas. In addition, it can be seen that the areas with higher uncertainty are the Pyrennes and the area along the northern west coastline (all seasons but summer) and the Cantabrian Mountains (spring).

Since RCM-GCM combinations will be used for future climate simulations, the analysis of their performance will take more indices into account. Two of those indices are considered: the number of dry days (CDD) and the number of wet days (CWD). Difference fields between GCM-driven ensemble mean and observations for CDD is shown in figure 4.32.

As happened for pr90p, there is also an underestimation of CDD, meaning that models simulate less consecutive dry days than those occurring in reality. Higher values of underestimation are more widespread throughout the IP during summer and spring months, while being only local in spring and



(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.32: Difference field between GCM-driven and observed CDD seasonal climatologies (GCM-OBS) over the IP for the 1961 to 2000 period (days/year).

---

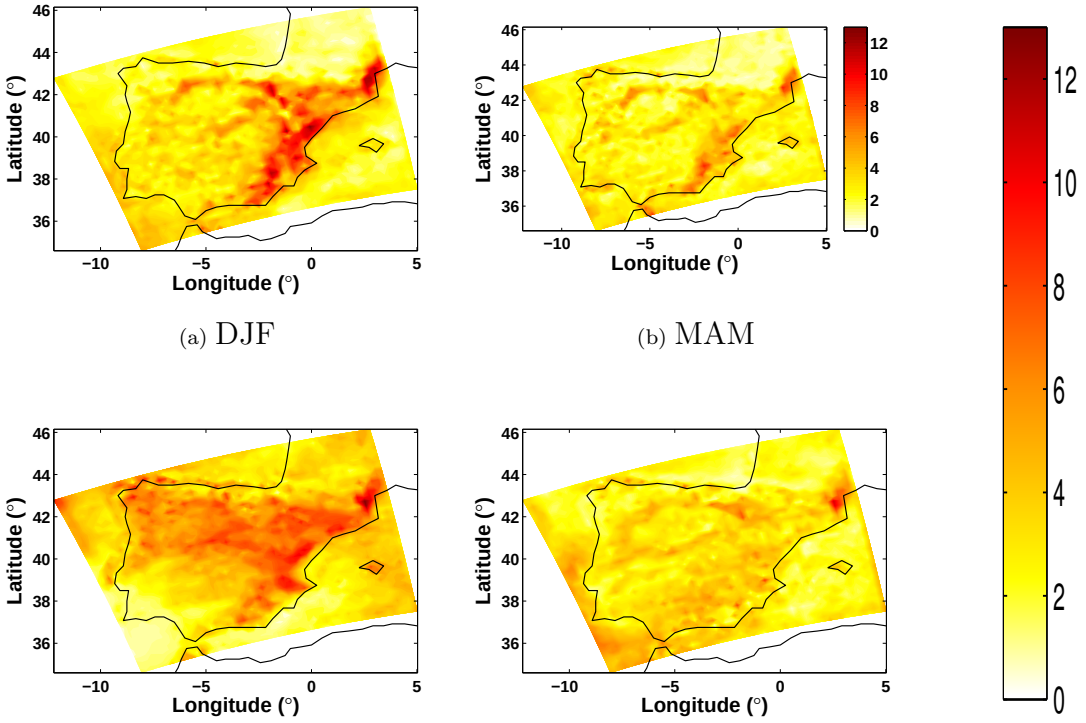
autumn. The latter presents an area around the Lisbon region and the former in the north-east coastal region where the difference field is null. The difference field is only positive north of the Pyrenees, already outside the domain under study.

The uncertainty field associated with each of the seasonal climatologies' ensemble can be seen in figure 4.33. This figure shows higher values along the eastern coastline extending to the west in the north, for the winter and summer months, with a maximum of 12 days per year and an average of 8 to 10 days per year. The other two months have lower GCM ensemble uncertainty, $\sim$ 4 days per year. In this case, there is no obvious topographic influence in the modelled results, since both the difference and the IQR fields show no marked differences along the highest and more complex peaks.

To complete the precipitation analysis, one can see the CWD climatology difference field between GCM-driven ensemble mean and observations - figure 4.34. Contrary to what happened for pr90p and CDD, this difference field shows a more complex spatial pattern of positive and negative changes. In general, north and eastern areas of the IP have positive GCM-OBS differences, while the rest of the domain shows negative values. This indicates toward an overestimation of the maximum number of consecutive wet days by the GCM ensemble mean. Along the Pyrenees, the difference field presents high positive values, specially for winter and spring, while in the west/ south-west region, the field is negative (with lower absolute value).

The uncertainty associated with the GCM-driven ensemble mean of seasonal CWD climatologies can be seen in figure 4.35. The IQR fields show higher uncertainty for winter and spring months, with focus on the north-west and the pyrenees. On average this maximum area reaches up to 8 days per year, while in the rest of the domain and for the remainder of the seasons, is mainly between 1 and 3 days per year.

Results for these three indices show how that, in general, precipitation patterns are being repro-

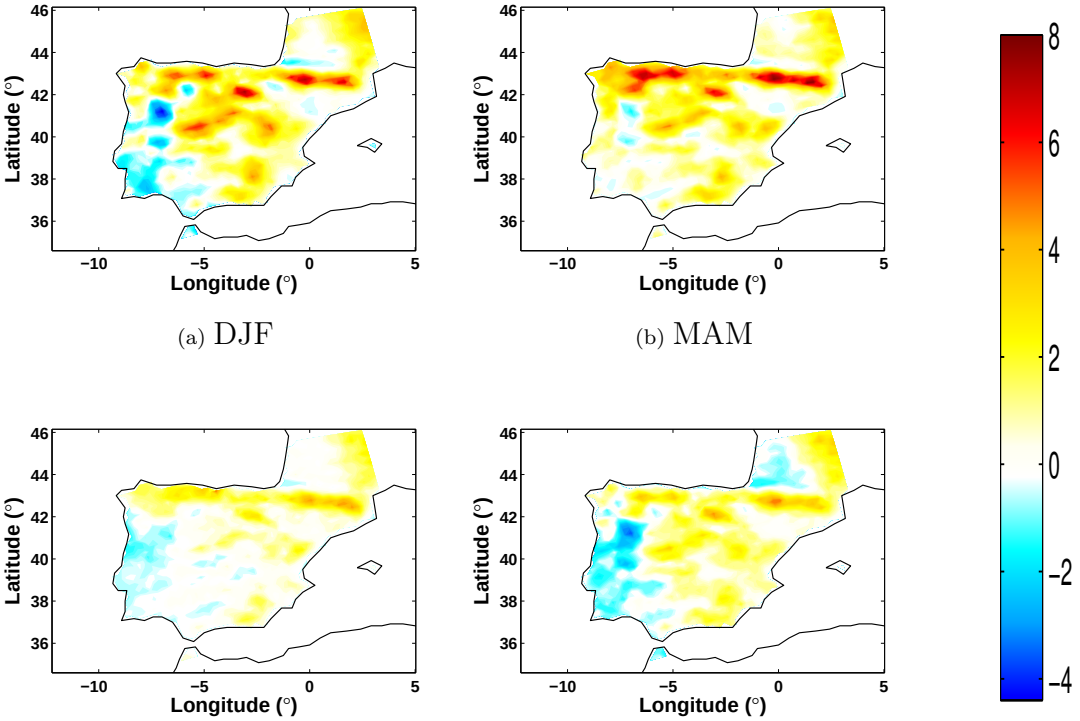


(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.33: GCM-driven ensemble mean of CDD seasonal climatology IQR (days/year).

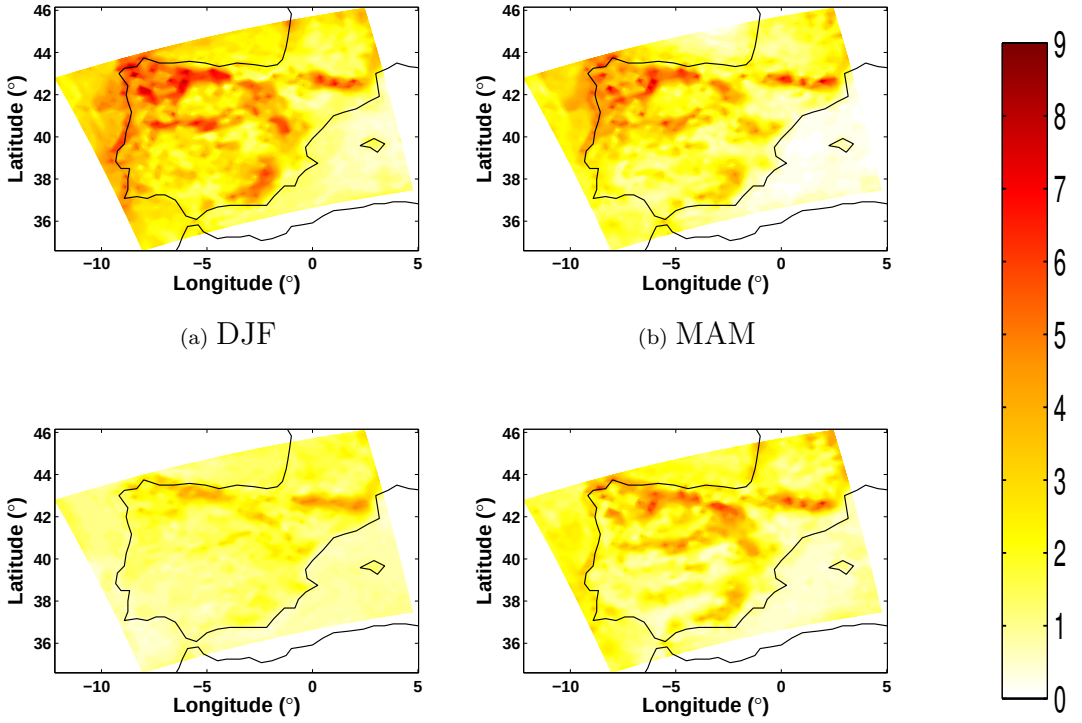
duced in the domain. Overall there is an underestimation of the number of days with precipitation amount $> 90th$ percentile an in the number of consecutive dry days. On the other hand, the number of consecutive wet days is being overestimated by GCM-riven ensemble. The conclusions on these indices point to models producing a higher number of consecutive rainy days, but with lower daily precipitation amounts. This, together with the fact that models are also producing lower number of consecutive dry days, might point to an overall similar total amount of precipitation, being distributed differently in time.



(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.34: Difference field between GCM-driven and observed CWD seasonal climatologies (GCM-OBS) over the IP for the 1961 to 2000 period (days/year).



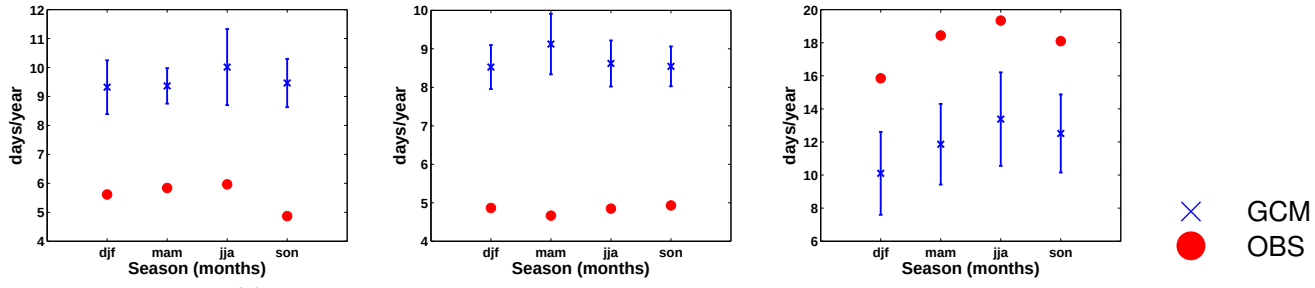
(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.35: GCM-driven ensemble mean of CWD seasonal climatology IQR (days/year).

4.3.2 Temperature Indices

As done for precipitation to the ERA40-driven indices (tasmax90p and tasmin10p), the ETR - Extreme Temperature Range - was also determined and will be analysed in this section. The spatial mean of their seasonal climatologies was computed for both observations and GCM-driven ensemble, together with the ensemble uncertainty.



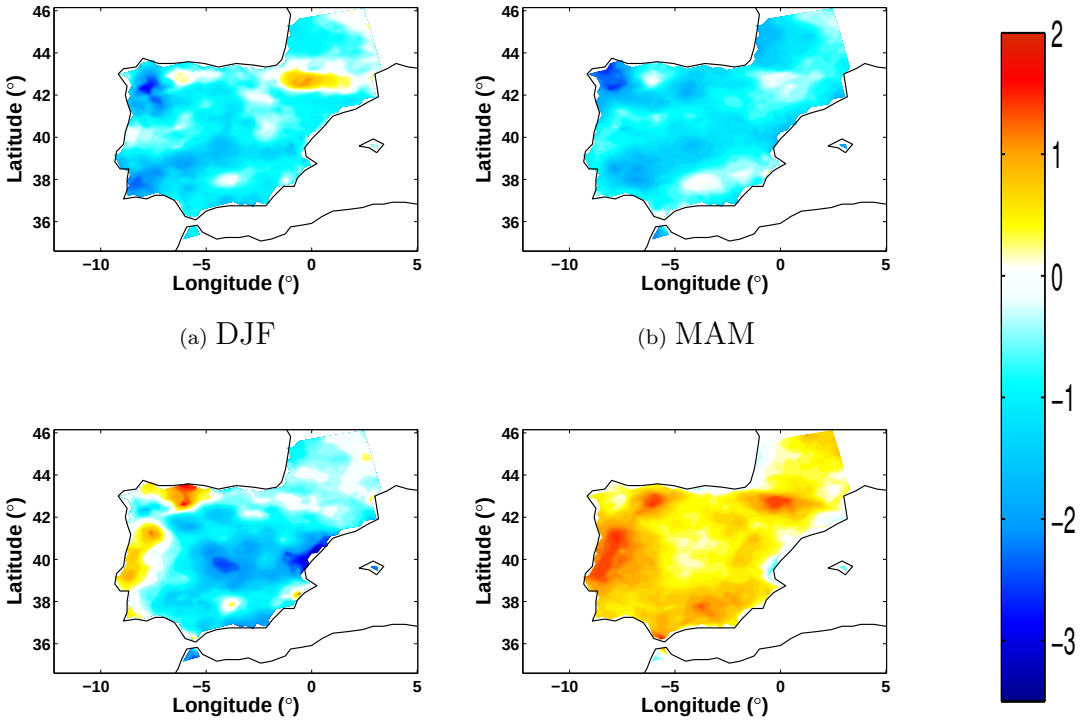
(a) tasmax90p (b) tasmin10p (c) ETR

Figure 4.36: Spatial average of tasmax90p, tasmin10p and ETR seasonal climatologies (days/year) over the IP for observations and GCM-driven ensemble mean (together with ensemble mean uncertainty bars).

Figure 4.36 shows the spatial mean of the temperature indices. Modelled mean seasonal climatologies results that are very different from observed indices. Both tasmax90p and tasmin10p are highly overestimated by the GCM-driven ensemble. This overestimation by the ensemble is similar for all

seasons and both indices. On the other hand, uncertainty is higher for summer tasmax90p and lower for the same index during the spring months. Contrary to the other two indices, the ETR shows model overestimation, similar for all seasons. This overestimation is higher than the underestimation for the other two indices. Furthermore, all observed indices are outside the range of uncertainty of the ensemble mean. Differences between GCM-driven ensemble mean and observed temperature indices' seasonal climatologies are statistically significant at the 5% confidence level for all seasons except for summer ETR.

The spatial pattern of these indices is also relevant and can provide more insight into the behaviour of the GCM-driven ensemble mean in different areas of the domain, which can be seen in figures 4.37 for tasmax90p, 4.38 for tasmin10p and 4.39 for ETR.



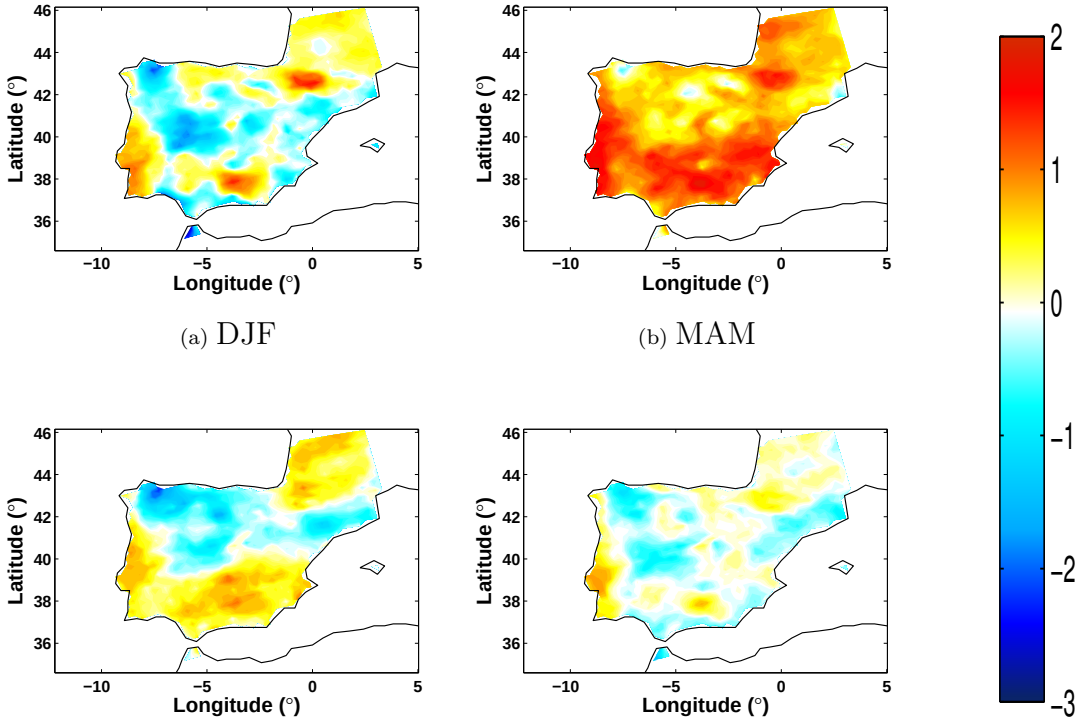
(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.37: Difference field between GCM-driven and observed tasmax90p seasonal climatologies (GCM-OBS) over the IP for the 1961 to 2000 period (days/year).

As can be seen in figure 4.37, for all seasons except autumn, the ensemble tasmax90p is lower than for observations and therefore the difference fields are mainly negative. However, there are areas for winter (the Pyrenees) and summer (north and west central coast) where this field is positive. Autumn difference field also shows positive differences which are higher in the three areas mentioned before for winter. Difference values have an absolute maximum of 3 days per year, which is high considering that, as can be seen in figure 4.36a, spatially averaged ensemble values are ~ 9 days per year. Uncertainty for tasmax90p seasonal ensemble mean climatology shows nearly constant IQR, between zero and 1 day per year.

In the difference fields for seasonal tasmin10p climatologies (figure 4.38), it is clear that the season presenting higher differences between the ensemble and the observed climatology is spring with, at least, half of the domain reaching between 1.5 and 2 days per year. Summer climatology difference field shows a clear negative/ positive pattern in the north/ south of the IP respectively, with the absolute



(a) DJF (b) MAM

(c) JJA (d) SON

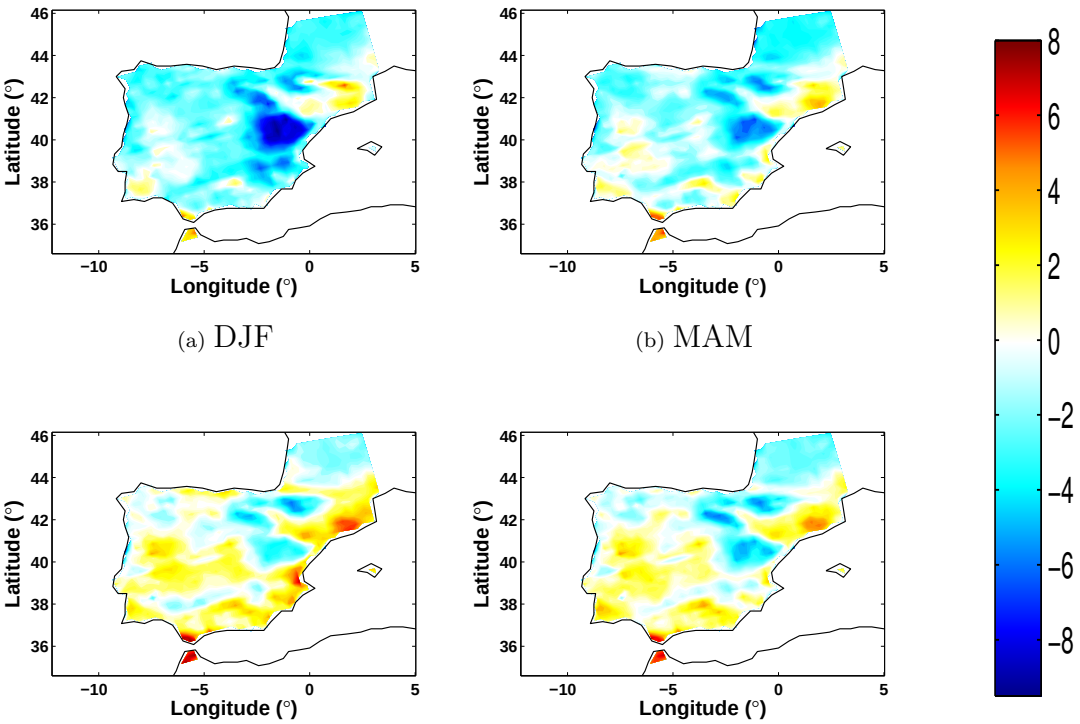
Figure 4.38: Difference field between GCM-driven and observed tasmin10p seasonal climatologies (GCM-OBS) over the IP for the 1961 to 2000 period (days/year).

maximum difference being only 1 day per year. Also in the far north of the Pyrenees, difference field shows positive values. Autumn is the season with lower overall difference between the GCM-driven ensemble and observed climatologies. Highest values do not surpass 0.5 days per year in the Lisbon area, and a good part of the domain shows null differences. Once more, since the uncertainty fields show a monotonous behaviour, and due to space constraints, they are not shown here. However, the uncertainty associated with tasmin10p ensemble mean seasonal climatologies is always 1 day per year and show little spatial variations.

Difference fields of seasonal climatologies for summer and autumn are very similar, with the amplitude of the differences being slightly larger for summer months. These fields are mostly positive, reaching up to 7 days per year. Negative differences are located in the north-east area of the IP and in a smaller area in the north west coastline. In these areas, models are generating a lower extreme temperature range value than observed. Winter and spring also have similar difference fields, with higher negative values for winter (minimum of ~ -8.5) and more area with positive difference for spring. Differences around Gibraltar are consistently higher (positive) for all seasons of the year. Uncertainty fields for the GCM-driven ETR seasonal climatologies (not shown here) have similar results for every season with values on average no higher than 4 °C and reaching a maximum of 6 °C. The results given by the temperature indices show that models are simulating a larger amount of both cold days (tasmin10p) and hot days (tasmax90p). However, the temperature range (ETR) is being underestimated and therefore modelled ETR is lower than observed.

4.3.3 RCM-GCM skill evaluation

In order to test the similarity of the yearly indices distributions, the K-S Test was applied to precipitation (pr90p, CDD and CWD) and temperature (tasmax90p, tasmin10p and ETR) indices, comparing



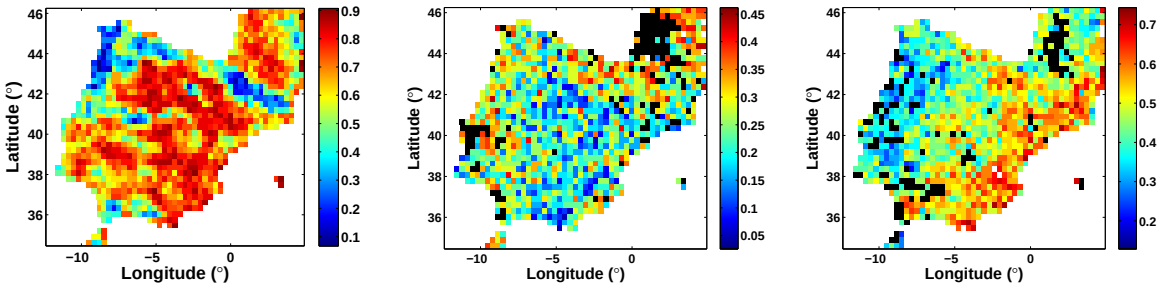
(a) DJF (b) MAM

(c) JJA (d) SON

Figure 4.39: Difference field between GCM-driven and observed ETR seasonal climatologies (GCM-OBS) over the IP for the 1961 to 2000 period (°C).

---

the GCM-driven ensemble and observed distributions. The K-S test statistics can be seen in figure 4.40, and for temperature in figure 4.41. Black pixels are for grid points where the null hypothesis has been rejected and therefore the distributions are considered to be statistically different.



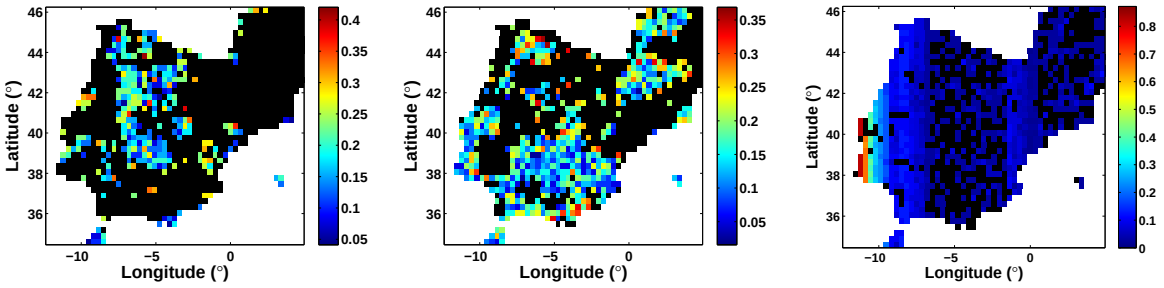
(a) pr90p (b) CDD (c) CWD

Figure 4.40: K-S Test statistics (maximum difference between the observed and GCM-driven ensembles mean CDF - D_n) determined for the entire 40 year period between 1961 and 2000, for each of the precipitation indices being studied. Black pixels represent areas where the null hypothesis is rejected and therefore the distributions are found to be significantly different.

K-S test statistics shows accordance between distributions for all three precipitation indices, although there are some grid points for the CDD and CWD where the distributions are considered to be statistically different (black pixels). However, these two indices show lower levels of D_n, meaning that the

difference between the observed and modelled distributions is lower. This difference is more evident for the CDD, which is also the index that has higher number of pixels where the null hypothesis is rejected. As mentioned before, there is better performance by models for dry days than for wet days, which explains the lower overall D_n values.

When analysing each index individually, it can be seen that, specially for pr90p and CWD, there is a clear spatial pattern arising. Pr90p shows higher accordance (lower D_n) between modelled and observed distributions in the north western coast and along the Pyrenees. On the other hand, CWD shows lower D_n to the west and increasingly higher to the western Iberian coast. The CDD field has less marked differences but presents slightly higher D_n towards the west coast and in the north, close to the Pyerenees.



(a) tasmax90p (b) tasmin10p (c) ETR

Figure 4.41: K-S Test statistics (maximum difference between the observed and GCM-driven ensembles mean CDF - D_n) determined for the entire 40 year period between 1961 and 2000, for each of the temperature indices being studied. Black pixels represent areas where the null hypothesis is rejected and therefore the distributions are found to be significantly different.

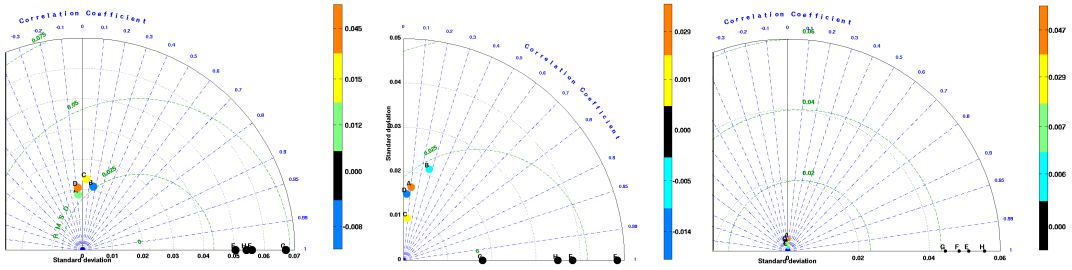
Table 4.1: Correspondence between the letters in the Taylor Diagrams and the seasons of modelled and observed indices.

Letter	Season	Data
A	DJF	GCM-driven ensemble mean
B	MAM	
C	JJA	
D	SON	
E	DJF	E-Obs
F	MAM	
G	JJA	
H	SON	

As can be seen in figure 4.41, tasmax90p and tasmin10p shows a large number of grid points in which the modelled distribution is considered to be statistically different from the observed one. However, the grid points where the null hypothesis is not rejected, have low values of D_n, never higher than 0.4 days. ETR K-S test also shows a high number of grid points where the modelled distribution is considered to be statistically different from the observed one. However, the grid points where the null hypothesis is not rejected, have low D_n (between 0 and 0.2 days) except for the southern area of the western coastline, where D_n reaches values as high as 0.8 days.

The evaluation of the ensemble in representing the spatial distribution of the seasonal trend of the indices was pursued by determining parameters such as the CORR, the RMSE, Standard Deviation

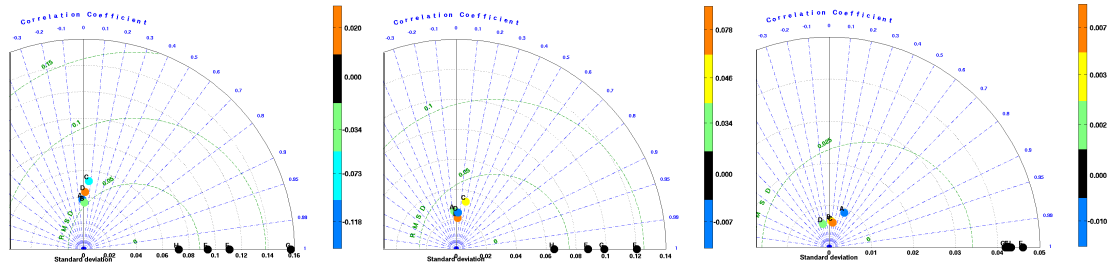
and BIAS. The first three can be summarized into a single Taylor Diagram, first described by Taylor (2001) which, when altered can integrate the BIAS as well. These diagrams were computed for each of the three indices, taking their seasonal trend and therefore comparing the modelled spatial patterns with the observed ones. In these diagrams, each letter corresponds to a different season of either observed or modelled indices, and their correspondence can be seen in table 4.1.



(a) CDD (b) CWD (c) pr90p

Figure 4.42: Taylor Diagrams summarizing the skill parameters of the seasonal precipitation indices trends of the GCM-driven ensemble mean and the observations.

The markers associated with observed data, have a null BIAS and are therefore black, while the others have different than zero BIAS. As can be seen in figure 4.42, which shows the Taylor diagrams for the trends of the three precipitation indices under study, there is low CORR between the seasonal modelled and observed trends. This correlation is always lower than 0.4. For all three precipitation indices, correlation is highest for spring, ranging from 0.12 for pr90p to 0.26 for CDD. Error is similar for seasonal CDD (~ 0.025 days/year) while it varies between 0.03 and 0.07 days/year for pr90p and 0.01 and 0.05 for CWD. BIAS is only negative for spring CDD and autumn CWD and is never higher than 0.05 days/year for all seasonal indices. From these results one can conclude that, although there is low spatial correlation between observed and GCM-driven indices, the modelled indices' standard deviation is always lower than the observed indices' standard deviation, meaning that the variability of the GCM-driven ensemble mean is lower than the observed.



(a) tasmax90p (b) tasmin10p (c) ETR

Figure 4.43: Taylor Diagrams summarizing the skill parameters of the seasonal temperature indices trends of the GCM-driven ensemble mean and the observations.

The Taylor diagrams for temperature indices' trends under study can be seen in figure 4.43. Tasmax90p trends have similar standard deviation, ranging from 0.020 to 0.035. However, RMSE and correlation show some differences for each season, being that the lowest value is found for spring. Tasmin10p has the highest correlation coefficient (0.19), albeit having the highest standard deviation and RMSE. On the other hand, summer has lower values of these two parameters and lowest correlation. The Taylor diagram for ETR shows less distinction between seasons, with all the models presenting nearly null RMSE and standard deviation. However, correlation coefficient is higher for winter months (~ 0.39) and lowest for summer (~ 0.06). Correlations are lower than 0.1 for all the markers, except for summer, which has a correlation coefficient close to 0.40.

These indices show an overall poor performance. Although modelled standard deviations are generally lower than observed ones, meaning that the variability of the modelled system is contained in the observed variability. The trends of these indices are not well reproduced. Temperature indices show higher performance than precipitation indices, which show higher ability to be reproduced during the summer.

Chapter 5

Conclusions

The goal of this work was to evaluate the performance of Climate Models in simulating precipitation, maximum and minimum temperature characteristics over the Iberian Peninsula. This goal was pursued by grouping the simulations into two different sets: ERA40-driven and GCM-driven. The first group was used to evaluate the performance of the RCMs, since they had the GCM forcing in common (ERA40 reanalysis), while the second enabled only the evaluation of the RCM-GCM performance.

The first step was to compare the seasonal climatologies of observations, ERA40 ensemble and GCM ensemble. Afterwards, the trend of precipitation and temperature indices was computed for the ERA40-driven ensemble, as well as observations. Overall, it was found that seasonal climatologies of the three variables are well reproduced by both the ERA40-driven and the GCM-driven ensembles, with the first one behaving more closely to observations. The spatial patterns of the difference fields - ERA40-OBS and GCM-OBS - show that, for the three variables there is a clear lower performance by both ER40-driven and GCM-driven ensembles in areas of more complex topography, with the Pyrenees being the most affected. Coastal regions also tend to show lower performance. The uncertainty associated with each of the ensembles is also larger in complex topography areas with the Pyrennes being, once more, one of the most affected areas. Since ERA40 reanalysis are based on observations, the generally lower uncertainty obtained for the ERA40-driven ensemble can be easily understood. Taking into account that temperature is a variable continuous in time and space, contrary to what happens for precipitation, the lower relative differences between observations and models obtained for temperature are as expected.

The first part of the Climate Models' performance evaluation was done for the Regional Climate Models RCMs) by comparing the ERA40-diven precipitation, maximum and minimum temperatures. The distributions of the ensemble yearly values of these variables were statistically compared to observed to observations using the K-S test. It was concluded that there are few grid points in which the observed and ERA40-driven ensemble distributions can be considered to be statistically different. Pr90p modelled results show that this index is better reproduced along the north western coastline, possibly due to the fact that in the remaining regions, the contribution of convective and orographic precipitation (smaller scale) is higher. Secondly, the trends of precipitation and temperature seasonal indices - pr90p, tasmax90p and tasmin10p - were computed. The time series of the spatial mean of these indices shows that they are consistently overestimated by ERA40-driven ensemble. Spatial patterns of the trend of these indices show that the overall signal of the trends is being reproduced by the ERA40-driven ensemble. However, spatial patterns such as isolated areas of different signal trends tend to either be underestimated or not even reproduced. When determining skill parameters - BIAS, error standard deviation, root mean square error and correlation - for each of the ERA40-driven simulations, as well as the ensemble of those models, no clear conclusion as to which model performs better was reached. Albeit not always being the best performing one, the ensemble is overall closer to observations.

The second part of the performance evaluation focuses on the GCM-driven simulations and therefore on the performance of the RCM-GCM combination. In order to do so, the previously mentioned seasonal indices (pr90p, tasmax900p and tasmin10p), as well as CDD, CWD and ETR were used. It was found that the spatial average of the seasonal climatology of temperature indices show larger deviation from observations than precipitation ones. Furthermore, the ensemble simulates results close to observations for the rain/ no rain indices (CWD and CDD), while showing difficulties simulating the amount of precipitation for extreme events (pr90p). While there is an overall underestimation of the ETR, both tasmax90p and tasmin10p are overestimated by models. When analysing the spatial patterns of the differences between observed and GCM-driven ensemble climatologies of these indices,

one can conclude that there is no clear area of worst or better model performance except for CWD that shows larger differences along the Pyrenees and the Cantabrian Mountains for all seasons. The K-S test performed to the distributions of yearly values of the indices shows that there are more grid points where the distributions are considered to be statistically different with a confidence level of 5% than what was found for ERA40-driven ensemble, for the same indices. Temperature indices show higher ratio of grid points statistically different than precipitation indices. These fields also showed that for precipitation indices that are related to the existence of rain, there aree areas where modelled indices perform well. For example, pr90p shows higher performance in the north west coast and along Pyrenees whereas CWD shows lower performance along the Iberian coastline. Although temperatures present large areas where the distributions are considered to be significantly different, the ones where they are similar have low values of D_n. Skill evaluation of the GCM-driven ensemble was also pursued by the drawing of Taylor diagrams which show that summer tends to be a better performing season than the others and that the Pearson correlation between modelled and observed indices is generally low.

Although not being the main goal of this work, there are also important conclusions on the behaviour of the indices during the 1961 to 2000 period which shed light on the changes occurring in the Iberian Peninsula. The observed trends of pr90p show that there were in 2000 less days with precipitation over the $90th$ percentile than in 1961. This is specially true for winter but less so for summer months. Also, the centre area of the IP also suffers lower decrease than the rest of the domain during winter and spring months. The number of days with maximum temperatures over the $90th$ percentile has a positive trend, specially for the extreme seasons (winter and summer). However, there are negative trends in the north coast during summer and east/north-east spring. Contrary to this, tasmin10p show averaged negative trends, with summer and spring having the largest negative trends and autumn the lowest. These results point to a a decrease in in number of rainy days (pr90p), while the number of hot days (tasmax90p) tend to rise and the number of cold days (tasmin10p) tend to fall. These results point to a warmer Peninsula, where rainy days have become more scarce, although no conclusion can be reached about the total amount of precipitation from this analysis.

Performance evaluation of climate simulations is an important tool in order to assess the degree to which models are accurate. This becomes specially important when these same models are used to simulate future climate projections and therefore there is the need to minimize the uncertainty. Since other sources of errors and uncertainties, such as the radiation forcing, cannot be easily reduced (one cannot determine exactly the emissions scenario that will be closer to reality in the future) it becomes important to minimize the other sources of uncertainty. Extreme temperature and precipitation events are two of the leading causes of concern when studying future climate change, therefore, it becomes important to understand, not only how the original variables are being simulated, but also these extremes.

In order to manage a deeper a understanding of model behaviour, this study should be extended to other indices, or even the complete combination of intensity and duration of both temperature and precipitation events. Furthermore, corrections could be developed and applied - for example a BIAS correction - in order to improve the performance of the ensemble and therefore decrease the uncertainty associated with climate projections.

Bibliography

Alexander, L., Zhang, X., Peterson, T., Caesar, J., Gleason, B., Klein Tank, A., Haylock, M., Collins, D., Trewin, B., Rahimzadeh, F., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *J. geophys. Res*, 111(D05109):22.

Allan, R. and Soden, B. (2008). Atmospheric warming and the amplification of precipitation extremes. *Science*, 321(5895):1481–1484.

Bucker, G. (2005). Vulnerable populations: lessons learnt from the summer 2003 heat wave in europe. *Euro Surveillance*, 10(7):147.

Déqué, M., Rowell, D., Lüthi, D., Giorgi, F., Christensen, J., Rockel, B., Jacob, D., Kjellström, E., De Castro, M., and van den Hurk, B. (2007). An intercomparison of regional climate simulations for europe: assessing uncertainties in model projections. *Climatic Change*, 81:53–70.

Déqué, M., Somot, S., Sanchez-Gomez, E., Goodess, C., Jacob, D., Lenderink, G., and Christensen, O. (2012). The spread amongst ensembles regional scenarios: regional climate models, driving general circulation models and interannual variability. *Climate dynamics*, 38(5):951–964.

Easterling, D., Meehl, G., Parmesan, C., Changnon, S., Karl, T., and Mearns, L. (2000). Climate extremes: observations, modeling, and impacts. *Science*, 289(5487):2068–2074.

Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G., Allen, S., et al. (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 Cambridge University Press, Cambridge, UK, and New York, NY, USA*.

Gallardo, C., Gil, V., Hagel, E., Tejeda, C., and de Castro, M. (2012). Assessment of climate change in europe from an ensemble of regional climate models by the use of köppen–trewartha classification. *International Journal of Climatology*.

Houghton, J., Ding, Y., Griggs, D., Noguer, M., van der LINDEN, P., Dai, X., Maskell, K., and Johnson, C. (2001). *Climate change 2001: the scientific basis*, volume 881. Cambridge University Press Cambridge.

Luna, T., Rocha, A., Carvalho, A., Ferreira, J., and Sousa, J. (2011). Modelling the extreme precipitation event over madeira island on 20 february 2010. *Nat. Hazards Earth Syst. Sci*, 11:2437–2452.

Maraun, D. (2012). Nonstationarities of regional climate model biases in european seasonal mean temperature and precipitation sums. *Geophysical Research Letters*, 39(6):L06706.

Maxino, C., McAvaney, B., Pitman, A., and Perkins, S. (2008). Ranking the ar4 climate models over the murray-darling basin using simulated maximum temperature, minimum temperature and precipitation. *International Journal of Climatology*, 28(8):1097–1112.

Prömmel, K., Geyer, B., Jones, J., and Widmann, M. (2010). Evaluation of the skill and added value of a reanalysis-driven regional simulation for alpine temperature. *International Journal of Climatology*, 30(5):760–773.

Roy, P., Gachon, P., and Laprise, R. (2011). Assessment of summer extremes and climate variability over the north-east of north america as simulated by the canadian regional climate model. *International Journal of Climatology*.

Taylor, K. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, 106:7183–7192.

Wilks, D. (1995). *Statistical methods in the atmospheric sciences*, volume 59. Academic press.