UNIVERSITAT DE BARCELONA

Improvement of seasonal forecasting techniques applied to water resources and forest fires

Raül Marcos Matamoros

Improvement of seasonal forecasting techniques applied to water resources and forest fires

UNIVERSITAT DE BARCELONA

Raül Marcos Matamoros

Department of Astronomy & Meteorology

University of Barcelona

Director :	Dr. M^{a} Carmen LLASAT	-	University of Barcelona
Co-Director :	Dr. Pere QUINTANA	-	Ebre Observatory

A thesis submitted for the degree of

Doctor Philosophiae

Barcelona, November 2015

Jury

President :	Dr. Jeroni LORENTE	-	University of Barcelona
Secretary :	Dr. Luis GARROTE	-	Madrid Polytechnic University
Chair :	Dr. Antonio PARODI	-	CIMA Research Foundation
Substitutes :	Dr. Mª Rosa SOLER	-	University of Barcelona
	Dr. Ana IGLESIAS	-	Madrid Polytechnic University
	Dr. Giorgio BONI	-	CIMA Research Foundation

Acknowledgements

Quan em miro a l'horitzó del temps hi veig la persona que un dia donà les primeres passes d'aquesta petita aventura. Quatre anys després, n'escric les darreres paraules. Ha estat un viatge emocionant i ple de contrastos: moments d'angoixa, passió, esforç, felicitat i sorpresa s'han anat alternant com els instants que compta el rellotge. Però abans que visiteu els paratges reocorreguts, voldria mostrar el meu agraïment a totes aquelles persones que al llarg d'aquest camí m'han donat els coneixements, l'experiència i els records que han contribuït a omplir aquestes pàgines del meu llibre dels dies.

En primer lloc voldria agrair a la Dra. Carme Llasat i al Dr. Pere Quintana haver dirigit i codirigit, respectivament, aquesta tesi. Gràcies a Carme per la seva càlida acollida al si del grup GAMA, així com per la seva dedicació, consells, comprensió i diàleg que m'ha permés completar aquesta empresa tot progressant a nivell acadèmic, personal i científic. Agrair-li també haver posat damunt la taula la possibilitat de realitzar una tesi sobre previsió estacional, un tema que se m'ha revelat fascinant i ple de futur. Gràcies a Pere per les fructíferes discussions científiques, els consells, el seguiment d'estiu a l'Observatori de l'Ebre i per mantenir una comunicació fluïda i propera. També voldria agrair al *Ministerio de Educación Cultura y Deporte* la beca FPU que m'ha permés realitzar aquesta tesi.

Moltes gràcies a Marco Turco, per haver-me iniciat en el camp de la investigació; per haver-me introduït al món del MatLab; per les discussions científiques apassionades; la seva ajuda desinteressada i els seus consells; la seva accolida a Torino i, en definitiva, per haver esdevingut un bon amic. Grazie mille Marco! Agrair també a Joan Gilabert haver fet de l'ambient al despatx una atmosfera amena, divertida i distesa; també per mostrar-me la potència del GIS com a eina i el paper de la geografia a la nostra societat. A Montse Llasat, per haver-me facilitat la integració al grup; per la seva atenció al detall; la seva ajuda en temes administratius; per haver-me demostrat la importància i la feina que hi ha al darrere del món de la divulgació científica; i per recordar que de l'eficiència participen també els descansos de mig matí. A Maria i Anna, per la seva contribució a fer pinya i al funcionament del grup com un tot integrat i agradable.

M'agradaria agrair a Jöhel Gailhard, Remy Garçon, Mathieu Le-Lay i, en general, a tots els membres d'EDF-DTG la seva acollida a Grenoble. Em van fer sentir com un més de l'equip i van contribuir a que el meu pas per la ciutat fos una experiència fructífera, grata i amena. Agrair també a la fundació Pedro i Pons el seu finançament per a la realització d'aquesta estança.

Una menció especial per a Enric i David, que per molt que ens separen continents o el temps sempre acabem trobant el moment per reviure instants i planejar noves aventures. Ja sabeu que l'absurditat pot ser la més gran virtut. A Dani, per les converses i dinars a peu de tesi, i a qui transmeto tota la meva força per al proper any i mig. A Cesc, per les xerrades, rialles i passeigs per Viena i Barcelona. I a tu, Albert, per haver caminat junts durant 15 anys plens d'història. En fa cinc vas encetar un repte ple d'il·lusió. Sé que també l'haguessis acabat i ho hauríem celebrat. Allà on siguis, va per tu.

Un agraïment sincer a la meva família, que sempre ha estat al meu costat i m'ha ensenyat la importància que tenen l'esforç i la constància a la vida. A la meva mare, per ser-hi sempre, un oasi dinàmic de moviment perpetu. Al meu pare, sempre a punt per donar un cop demà, resolt i ferm. Al meu germà, capità de mars i oceans, per totes les peripècies viscudes i les que vindran. Als meus *iaios* d'Alcanar, per fer de l'experiència i l'estima per allò que t'agrada una font d'il·lusió per al dia a dia. Als *iaios* de les Cases per ensenyar-me el valor de cada any viscut i omplir la tesi d'instants de mar i partits del Barça. Un record especial per al meu *iaio* Batiste que no em podrà veure finalitzar allò que em va veure començar. Ja he acabat la 'faena' que sempre et deia.

I gràcies a tu Gessamí. Per tots els moments viscuts; la passió de les nostres converses; la il·lusió dels nostres viatges; per la teva alegria i el teu positivisme; perquè ens ho qüestionem tot; per la maduresa que has mostrat i el suport que m'has donat; per ser qui ets, per ser com ets. Et vaig conéixer als inicis d'aquest viatge, com el Sol que escalfa el camí del vent. Junts hem descobert móns i paisatges, l'avui i l'endemà. Has fet d'aquest tram de vida un recorregut inoblidable. Somiem. Somiarem.

Resumen

En esta tesis se estudian los beneficios de distintos métodos de calibración sobre la previsión estacional y la mejora del pronóstico estacional de recursos hídricos e incendios forestales. Sequías e incendios son un problema consustancial a la región Mediterránea y es probable que puedan empeorar si el cambio climático continúa. Ambos riesgos son una fuente de costes económicos importantes, daños en el medio ambiente y, en el caso de los incendios forestales, incluso de pérdidas humanas. Estos impactos han fomentado el interés por el desarrollo de protocolos de decisión que permitan reducir la vulnerabilidad a través de medidas de adaptación y mitigación. En ese sentido la predicción estacional podría participar de esta tarea anticipando el comportamiento de los recursos hídricos y los incendios forestales a meses vista. Además, la previsión estacional también puede dar lugar a marcos operativos que puedan funcionar tanto en las condiciones climáticas actuales cómo en las futuras.

Sin embargo, en latitudes extra-tropicales la previsibilidad estacional generalmente se considera bastante limitada y, en consecuencia, las predicciones estacionales rara vez se utilizan en la toma de decisiones. No obstante, hay estudios que sugieren que los métodos de calibración estadística podrían ayudar a mejorar las previsiones de los modelos actuales. Por lo tanto, la brecha existente entre los objetivos de los usuarios finales y la investigación teórica requiere trabajos que permitan demostrar la utilidad de las predicciones estacionales. Así pues, siguiendo esta línea de investigación y para lograr dicho objetivo este proyecto se ha estructurado en tres sub-objetivos: evaluación del rendimiento de un modelo de previsión estacional operativo y de sus calibraciones, el pronóstico estacional de los recursos hídricos y el pronóstico estacional de los incendios forestales.

Las salidas del modelo de predicción estacional ECMWF System-4 se han calibrado mediante las técnicas MOS-analog, regresión lineal y corrección media del sesgo sobre cuatro dominios: Europa, España, Cataluña y la cuenca del río Muga. La validación de estos resultados y su comparación con la climatología, persistencia y las salidas sin calibrar del S4 se ha desarrollado desde una perspectiva determinista y probabilística mediante parámetros como el MAE, los diagramas de Taylor, las curvas ROC, el diagrama de atributos o las curvas de valor económico.

Una vez determinado que el embalse de la Boadella es un buen caso de estudio tanto por sus características geomorfológicas así cómo por la longitud de sus series de datos (1971-2013) se procede a la selección física de los predictores para la construcción del modelo de regresión múltiple (MLR) que permita reproducir las anomalías mensuales de caudal de entrada, salida y volumen de dicha infraestructura hidráulica. El proceso que se sigue es una previsión de tipo *out-of-sample perfect prognosis* con *total screening* de las variables conduciendo a un elevado número de modelos que son filtrados de acuerdo con la variabilidad explicada, el MAE y el coeficiente de información de Akiake. Finalmente, de entre estos modelos se escoge aquel que con el mínimo número de predictores sea capaz de maximizar la variabilidad explicada.

Con cada uno de los modelos seleccionados en el apartado anterior se ha realizado la previsión hasta un horizonte de 7 meses con los datos sin post-proceso del ECMWF-System 4 y las observaciones E-OBS. También se ha evaluado la bondad de las previsiones en comparación con los valores climáticos y las observaciones antecedentes mediante la metodología desarrollada en la tesis. Después, la previsibilidad estacional del caudal y volumen de la Boadella ha sido evaluada a través de varios métodos de pronóstico estacional.

Por último, con respecto al pronóstico estacional de incendios forestales el primer paso ha sido modelizar el área quemada durante el verano (JJAS) en Cataluña a través de un MLR considerando las condiciones de sequía presentes y antecedentes (mediante los índices SPI/SPEI). Posteriormente, el ajuste ha sido evaluado bajo distintas configuraciones de previsión estacional.

Entre los resultados obtenidos se ha observado que la mejor predicción del System-4 ECMWF, ya sea calibrado o sin calibrar, se concentra en el primer mes de previsión. A ese horizonte las predicciones deterministas mejoran la climatología y la persistencia en la mayoría de meses y para todas las variables consideradas. La evaluación probabilística, por su parte, se ha mostrado especialmente positiva en los meses de invierno. Además, también se ha comprobado el valor añadido de las técnicas de post-proceso estadístico (MOS-analog, regresión lineal y corrección media del sesgo). Dichas técnicas siempre mejoran las salidas del modelo original mediante la corrección de los sesgos del modelo de primer orden. Sin embargo, el resultado de la calibración MOS-analog insinúa la posibilidad de ir más allá de estos resultados si se pudiese aumentar suficientemente la cantidad de meses dónde buscar situaciones análogas.

En referencia a las aplicaciones para las variables del embalse (anomalías en el caudal de entrada, salida y volumen) nuestro estudio ha revelado que las tres variables se pueden modelizar mediante regresión lineal múltiple (MLR). En los tres casos los meses de verano mostraron una previsibilidades más allá del primer mes, un resultado significativo para los gestores hídricos. Por otra parte, los resultados muestran también que en el caso de las anomalías de volumen los pronósticos estacionales a través de MLR podrían empezar a sustituir operativamente las previsiones climatológicas actuales. Esto también es cierto para algunos meses del caudal de salida (especialmente los de mayor demanda hídrica, JA). Para las anomalía de caudal de entrada, sin embargo, aunque Julio muestra un comportamiento muy bueno, hace falta investigar mejor las relaciones del caudal de salida con los diversos predictores para intentar mejorar el rendimiento de los modelos y así poder llegar a sustituir la climatología cómo método de previsión.

Por último, en relación con los incendios forestales, el vínculo existente entre el área quemada en verano (JJAS) en Cataluña y las condiciones de sequía anteriores nos ha permitido modelizar satisfactoriamente el área quemada a través de un modelo MLR. Dicho modelo, además, proporciona una estimación de la anomalía positiva o negativa del área quemada antes de la estación de incendios y puede ser fácilmente adaptado a otras regiones de tipo Mediterráneo.

Abstract

This thesis studies the benefits of different calibration approaches on seasonal forecasting and the improvement of seasonal prognosis of water resources and forest fires. Droughts and wildfires are an inherent problem to the Mediterranean and are likely to worsen if climate change continues. Both hazards are a source of important economic costs, environmental damage and, in the case of wildfires, even life losses. These impacts have encouraged policy- and decision-makers to reduce vulnerability by placing great efforts in the development of mitigation and adaptation strategies. Seasonal forecasting could help with this task by foretelling the behaviour of water resources and wildfire with months in advance. Furthermore, it has the capacity to provide operational frameworks that can work both in present and future climate conditions.

However, seasonal predictability in extra-tropical latitudes is usually considered rather limited and, consequently, seasonal forecasts are seldom used in decision-making. There are studies, though, suggesting that calibration methods could help improving current model's output. Thus, the existing gap between end-user goals and theoretical research needs more work to demonstrate the utility of seasonal forecasts. To achieve this objective this study has been divided in three sub-objectives: skill assessment, seasonal forecast of water resources and seasonal forecast of forest fires.

The *skill assessment* comprises an evaluation of the skill of the raw ECMWF System-4 output in Europe, Spain, Catalonia and the Muga river basin; and the study of the impact on the ECMWF System-4 performance of the MOS-analog and linear regression calibrations in comparison to mean bias correction. As for the *seasonal forecast of water resources* the application began with the modelling of the Boadella reservoir in-flow, out-flow and volume anomalies through a

Multiple Linear Regression (MLR) procedure. Afterwards, the seasonal predictability of the Boadella predictands has been evaluated through several seasonal forecast approaches. Finally, regarding the *seasonal forecast of wildfires* the first step has been to model summer (JJAS) burned area in Catalonia through a MLR with antecedent and current year drought conditions. Subsequently, the MLR performance has been tested under different seasonal forecast configurations.

Among the results obtained it has been found that most of the skill of the ECMWF System-4 is focused in the first lead. At this horizon deterministic forecasts improve climatology and persistence in the majority of months and for all the variables considered. The probabilistic assessment, on the other hand, showed this skill was mainly centred in the winter months. Also, the added value of calibration post-processing techniques has been checked. These techniques always ameliorate the skill of the original model output by correcting first order biases. Nevertheless, the MOS-analog outcome has also hinted the possibility to go beyond these results if the analog pool was sufficiently increased.

In reference to the reservoir's applications the perfect prognosis approach revealed that in-flow, volume and out-flow anomalies can be modelled through multiple linear regression (MLR). In all three cases summer months showed enhanced predictability way beyond the first lead, a significant result for water managers. Moreover, the results proved that volume anomaly seasonal forecasts could begin the operational switch from customary climatology to another forecast strategy based on MLR models. This is also true for some months in the outflow's modelling. For the in-flow case, though, there is still further research needed before reaching that sate. Finally, regarding forest fires, exploiting the relationship between summer burned area and preceding drought conditions can lead to MLR models that provide a seasonal estimate of the expected above/below-normal summer fire burned area in Mediterranean-type regions.

Un camí de molts móns

Contents

List of Figures

CHAPTER 1

Introduction

1.1 Motivation

The recent IPCC Working Group II reports (IPCC, 2014) warn of the potential growth of natural risks around the world due to the steady increase in temperatures (IPCC, 2013a). These comprise a broad range of different phenomena such as heat waves, droughts, flooding, cold spells, cyclones or wildfires. In this context the Mediterranean area, being amidst Africa and the Eurasian continents, has been identified as a climate change hot-spot for its possible impacts on the population's safety and way of life (Diffenbaugh *et al.*, 2007; Diffenbaugh and Giorgi, 2012).

However, among the previous threats droughts and wildfires are already an inherent problem to the Mediterranean and are likely to worsen if the warming continues (see i.e. Flannigan *et al.*, 2009; Dai, 2011; Turco *et al.*, 2014). Droughts, for example, have been documented in the region for many centuries (Barriendos *et al.*, 2003; Brewer *et al.*, 2007) and are still a major trouble in modern societies (see i.e. Nola *et al.*, 2008). On the other hand, Mediterranean summer fires comprise the greatest part of the 500000 hectares and the 50000 fires that each year burn in Europe (San-Miguel-Ayanz *et al.*, 2013b). Regularly then, we find that both hazards are a source of important economic costs, environmental damage and, in the case of wildfires, even life losses. (Carroll *et al.*, 2007; Moreira *et al.*, 2011).

Therefore, great efforts are placed on developing mitigation and adaptation strategies to reduce our vulnerability to these hazards. Some of them are detailed in the hydrologic administration plans set up in Spain and Catalonia to coordinate management response to water scarcity (see i.e. MIMAM, 2007; ACA, 2009a); and others even have noticeable achievements (see i.e. Plana, 2011) such as the observed decreasing trends of summer number of fires and burned area attained in Catalonia (Turco *et al.*, 2013b). Nevertheless, there is a high potential mitigation tool that despite being in operational stage for several years, still has not been widely adopted: *seasonal forecasting*. Theoretically, this field could help reducing drought and wildfire impacts by foretelling their behaviour with months in advance (Mavsar *et al.*, 2013; Dutra *et al.*, 2014). Furthermore, it also has the capacity to establish operational frameworks that can work both in present and future climate conditions.

Seasonal forecasting tries to foretell climate anomalies at time-scales comprised between one month and one year (see chapter 2). This places seasonal forecasts in a temporal framework of great interest for many societal sectors and economic users (Doblas-Reyes *et al.*, 2013). However, seasonal predictability in extra-tropical latitudes is often deemed as complex and limited to some seasons or events and, consequently, it has not been widely adopted in decision-making (i.e. Brands *et al.*, 2012; Iglesias and Garrote, 2014). The resolution to undertake this thesis, though, was triggered by the works of Palmer and Anderson (1994); Llasat *et al.* (2010) and the technical memorandum of the ECMWF System-4 (S4, hereafter; Molteni *et al.*, 2011). These studies suggest that there could be coherent signals in the models coming from the observed general circulation and other inertial climatic factors (see chapter 2). Further research confirmed that calibration methods could help to improve the model's output (i.e. Johnson and Bowler, 2009; Boer *et al.*, 2013). Additionally, the existence of a gap between end-user goals and theoretical research and the subsequent need for applications demonstrating seasonal forecast utility offered a window of opportunity for this project (see i.e. Johnston *et al.*, 2004; Schneider and Garbrecht, 2006; Garbrecht and Schneider, 2007; Kumar, 2009; Coelho and Costa, 2010).

1.2 Challenges

In spite of the steady advances, seasonal forecasting has to face many theoretical and practical challenges to become an everyday tool in our societies. These challenges are related either to accurate skill assessment, the identification of fundamental uncertainties and knowledge gaps, development and implementation of improvement strategies or the demonstration of useful applications in different fields. Works such as Schwierz *et al.* (2006), Koenigk and Mikolajewicz (2009), Doblas-Reyes *et al.* (2013), Weisheimer and Palmer (2014) and many others suggest the most important to be,

- Establish the skill of each new forecast system and/or configuration so as to clearly identify the progress achieved and its predictive limits. Actually, with every new model instalment it is vital to clearly assess the improvements with respect to its predecessors and/or their contemporaries. This can be also related to the evaluation of uncertainties and predictability limits that might derive in the identification of mechanisms that were not previously considered in the models.

- Determine an optimal number of verification techniques to evaluate the desired aspects of an issued forecast so as the information could be communicated clearly both to the academic world and end-users. In this sense it is vital to test and introduce any new method or strategy that allows us to look at the forecast from different perspectives and assess whether it can be used to better characterize the predictions.

- Improve the ocean-atmospheric and land-surface coupled models by identifying the main sources of uncertainty and the most influencing predictability features. This can be achieved through the implementation of physical relationships not considered before, parametrization improvement, reduction or elimination of physical equations' approximations, enhancement of initial condition assimilation and the increase of model resolution.

- Characterize the skill of the models at different scales from planetary to local domains. Since most of the seasonal forecasting models offer coarse resolutions this involves the application of downscaling techniques. Often these methods are used to resolve higher resolution features that are not well represented in the raw model output.

- Improvement of model output through calibration. The raw model forecasts contain biases and the application of linear and non-linear calibration techniques can be used to correct them. However, the identification of the best techniques it is not straightforward and might depend on the domain and variable considered as well as in the nature of the calibration technique itself.

- Prove that the information provided by current seasonal forecast systems can be already useful for certain applications. Nowadays most end-users do not rely on seasonal forecasts because there are still few implementations that combine users needs and vulnerabilities to state the advantages of using the limited skill of current forecasts or the predictability hidden in antecedent observations. This often requires a fluid knowledge exchange between researchers and users to elaborate specific verification solutions to ease the transferability of the message and maximize its utility.

1.3 Antecedents & Objective

This thesis is continuation of a seasonal forecast research line within the GAMA group (Meteorological Hazards Analysis Team) and it has the support of the Catalan Water Agency (ACA) and it is also considered in the HYMEX international program. Some of the ideas carried out in this thesis were firstly shaped in the CENIT-SOSTAQUA project where the use of seasonal data extracted from low resolution coloured charts showed some potential in the forecast of volume anomalies (Llasat *et al.*, 2010). Others implied the adaptation and development of downscaling calibration methods based on past and present collaborations with Électricité de France (EDF-DTG) and the University of Grenoble (Llasat and Puigcerver, 1997; Gibergans-Báguena and Llasat, 2007; Llasat *et al.*, 2010); the inspiration from the integration of meteorological and hydrological models at different scales (Quintana-Seguí *et al.*, 2008, 2009, 2010); the conceptual influence from projects AMPHORE and ESTCENA (Altava-Ortiz *et al.*, 2006; Herrera *et al.*, 2011; Turco *et al.*, 2011); and the promising results of wildfire modelling in our group (Turco *et al.*, 2013a,b, 2014). Hence, considering the contemporary challenges of the field and the aforementioned antecedents our main objective for this thesis is:

> **The study of the benefits of different calibration approaches on seasonal forecasting and the development of strategies to improve the seasonal prognosis of water resources and forest fires.**

To achieve this objective we have divided our work in three other sub-objectives with different embedded tasks:

1. Skill assessment

- **a)** Evaluation of the skill of the raw ECMWF System-4 (S4) output in Europe, Spain, Catalonia and the Muga river basin.
- **b)** Impact on the S4 performance of the MOS-analog and linear regression calibrations in comparison to mean bias correction Europe, Spain, Catalonia and the Muga river basin.

2. Seasonal forecast of water resources

- **a)** Modelling of the Boadella reservoir in-flow, out-flow and volume anomalies through a Multiple Linear Regression (MLR) procedure.
- **b)** Evaluation of the seasonal predictability of the Boadella reservoir predictand anomalies through several seasonal forecast approaches.
- **c)** Performance comparison of the considered seasonal forecast strategies with climatology.

3. Seasonal forecast of forest fires

a) MLR modelling of summer (JJAS) burned area in Catalonia taking into account antecedent and current year drought conditions with the Standardized Precipitation Index and the Standardized Precipitation and Evapotranspiration Index (SPI/SPEI).

b) Performance of the MLR model under different seasonal forecast configurations.

1.4 Organisation

This thesis is structured in ten chapters. Particularly, this is the last section of **chapter 1**, devoted to put forward our work's outline. Previously, we have presented our motivating force, the current challenges in the seasonal forecasting knowledge, our study choice and the objectives set to progress in their resolution. **Chapter 2** is where we introduce the seasonal forecasting field in accordance with the existent literature. Therein we deal with its nature, its predictability sources, the methodological approaches and its potential applications. In this last aspect we centre our attention in the two topics that focus our dissertation's seasonal forecast implementation: Mediterranean water scarcity and summer wildfires. Afterwards we find **chapter 3**, which embodies the characterization of the different datasets used. In **chapter 4** we analyse the climatology and relief of the studied regions and domains. This is done for Europe, Spain, Catalonia, the Muga river basin and the Boadella reservoir. **Chapter 5** contains the description of the mathematical techniques applied in the calibration, modelling and verification processes. Each of its sections refers to the chapters where the corresponding methodology is implemented.

The next three chapters develop the objectives raised at the beginning of this thesis. Each of these chapters is presented in a self-contained structure which also includes a brief reference to the data, region and the methods used. In case a more expanded view of these contents was needed the reader is referred to their respective extended versions of chapters 3, 4 or 5. In **chapter 6** we study the performance of the ECMWF System-4 monthly forecasts in four domains:

Europe, Spain, Catalonia and the Muga river basin. Afterwards, we compare these results with three calibration approaches and climatology and persistence controls. These calibration are: mean bias correction (linear), linear regression (linear) and MOS-analog (non-linear). Once their main characteristics are identified we move to the first application of seasonal forecasting, **chapter 7**, studying the prediction of in-flow, volume and out-flow monthly anomalies in the Boadella reservoir (Muga river basin). We begin by constructing multiple-linear regression (MLR) models for each month identifying the leading predictors for each variable. Subsequently we test every forecast horizon with different forecast systems that combine antecedent observations with the S4 calibrations and comparing them with persistence, climatology and the corresponding MLR models in perfect prognosis conditions. **Chapter 8** consists on testing the predictability of summer burned area in Catalonia with antecedent and forecast values of SPI and SPEI indices through a two term MLR model with antecedent and current drought conditions. In this case we test the skill of the found MLR model with antecedent observations, persistence and bias corrected ECMWF System-4. Finally, **chapter 9** recaps the overall conclusions and original contributions of this work; and **chapter 10** hints the possible paths for future research.

CHAPTER 2

Seasonal forecasting

2.1 What is seasonal forecasting?

Seasonal forecasting is a discipline that can be placed somewhere between weather an climate forecasting (see figure 2.1). Its purpose is the estimation of meteorological anomalies at monthly or multi-month level for lead-times ranging from one month to one year. However, knowing that the meteorological skill decreases for forecasts of, approximately, one or two weeks ahead (Lorenz, 1969b) one might ask about the feasibility of issuing successful forecasts with months in advance. In principle, the answer to this question is affirmative for in seasonal forecasting one seeks departures from the climatological mean making these forecasts more responsive to boundary conditions (i.e. Barnston *et al.*, 1994; Palmer and Anderson, 1994; Carson, 1998; Murphy *et al.*, 2001). Actually, seasonal forecasting was not thought to be possible until the late seventies when several studies centred on the Indian monsoon showed some predictability at time-scales larger than a month (see i.e. Charney and Shukla, 1981). This is due to the fact that some systems with which the atmosphere is bounded and interacts evolve more slowly than the atmosphere itself, favouring some states instead of others. Often, these boundary conditions are also known as *external* or *inertial forcings* and the most important can be summarised in the following list,

a) SST

b) Soil moisture

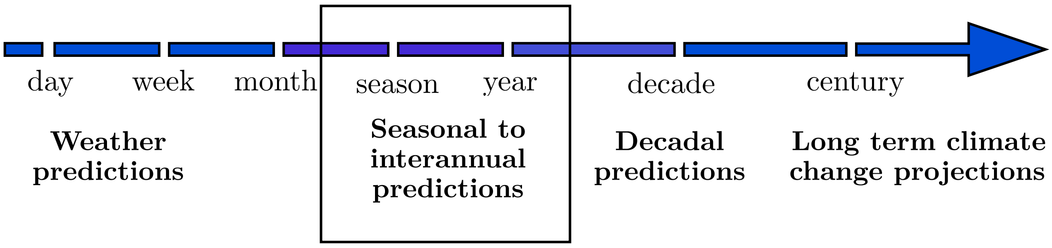


Figure 2.1. Weather to climate forecasting time-scales. Elaborated from Boer *et al.* (2013).

c) Snow and sea-ice cover

d) Stratospheric circulation and stratospheric thermal anomalies

Some of these relations were firstly identified as teleconnections, that is, changes in one variable in one part of the world that affected other variables farther apart in the globe without apparent spatial connections among them. Examples of teleconnections are pressure indices such as the North Atlantic Oscillation (NAO) or the Eastern Atlantic (EA) affecting rainfall in Western Europe (see i.e. Castro *et al.*, 2011; el Kenawy *et al.*, 2012); as well as other indices such as the *El Niño* Southern Oscillation (ENSO), SST anomalies in the eastern part of the equatorial Pacific that influence rainfall and temperature regimes in many parts of the planet (see i.e. Ropelewski and Halpert, 1986; Wang *et al.*, 1999, and section 2.2.1).

Actually, in Europe, these predictors, their relationships and the nature and implications of seasonal forecasting are or have been studied in several projects such as the European PRediction of climate Variations On Seasonal to interannual Time-scales (PROVOST; Doblas-Reyes *et al.*, 2000); the Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER; Palmer *et al.*, 2004); ENSEMBLES (Weisheimer *et al.*, 2009); or the on-going Seasonal-to-decadal climate Prediction for the improvement of European Climate Services (SPECS; Manzanas *et al.*, 2013) and the EUropean Provision Of Regional Impact Assessment on a Seasonal-to-decadal timescale (EUPORIAS; Falloon *et al.*, 2014).

Nowadays, seasonal forecasting has become a field in expansion, thanks to the advances in the Global Climate Models and the increase in computational power. The World Meteorological Organization coordinates the efforts of a selection of international centres in charge of the release of the operational long range forecasts. They are called Global Producing Centres (GPC) and their seasonal forecasts are then transferred to Regional Climate Centres as well to the National Meteorological and Hydrological Services to adapt and communicate this information to the end-users (see i.e. Doblas-Reyes *et al.*, 2013). This coordination tasks are mainly undertaken by two institutions: the WMO Lead Centre for Long Range Forecast Multi Model Ensemble and the WMO Lead Centre for Long Range Forecast Verification System.

2.2 Seasonal Predictability

Predictability as a concept refers to the possibility to foretell future states of a given system. Its characterization is achieved by identifying elements or mechanisms that connect the system's past states with its future behaviour. Applied to seasonal forecasting, it speaks of all the Earth system features, normally tropospheric boundary conditions, that can contribute to elaborate anomaly predictions of meteorological variables with months in advance (see i.e. Froude *et al.*, 2013). Yet in 1994, during the first steps of seasonal forecasting models, Palmer and Anderson (1994) thoroughly reviews the existing results in the field and finds that multiple components of the general atmospheric circulation might present predictability beyond the usual synoptic patterns thanks to boundary conditions such as the SST. However, he also states that this characteristic would be rather concentrated in the tropics for in the extra-tropics the internal instabilities of the system as well as the effects of non-linearities are much stronger.

Nowadays, though the existing studies confirm the idea that predictability in the extra-tropics is reduced in comparison to the tropics (see i.e. Doblas-Reyes *et al.*, 2013), this does not mean that it is missing at all (see i.e. Stockdale, 2000). For example, several studies have observed that winter circulations in northern latitudes can be grouped in quasi-stationary states that are preserved during weeks with relatively rapid transitions among them (Rodwell and Doblas-Reyes, 2006; Guéremy *et al.*, 2012). Since the objective of seasonal forecasting it is not to pin-point the exact moment of these transitions it would suffice to identify the driving factors of these states to increase extra-tropical predictability. Actually, winter is deemed to be the more predictable season in Europe and North-America (see i.e. Barnett and Preisendorfer, 1987; Carson, 1998; Quan *et al.*, 2006; Folland *et al.*, 2012; Scaife *et al.*, 2014), probably due to the prevalence of westerlies (Fil and Dubus, 2005). However, this is something linked to the intrinsic variability of the atmospheric system in every region (Palmer *et al.*, 2004) and this is only one aspect of the more general notion that the capacity to anticipate the atmospheric behaviour depends on the season because the annual cycle affects the evolution, stationariness and interaction between the atmosphere and its multiple boundary conditions (Boer *et al.*, 2013).

Hereafter we will present five of the principal sources of seasonal predictability, namely: sea surface temperature (SST), soil moisture, snow cover, sea-ice cover and stratospheric thermal anomalies. Other factors such as variations in the atmospheric chemical composition, aerosol concentration, land cover change or volcanic eruptions can modify the radiative balance of the climate system and, consequently, influence its seasonal predictability but, strictly, they are not sources of seasonal predictability themselves (Doblas-Reyes *et al.*, 2013).

2.2.1 Sea Surface Temperature (SST)

Sea surface temperature is one of the major influencing factors on the general atmospheric circulation. Its enormous heat capacity slows down the thermal response to external forcings and, as a result, the atmosphere can be biased towards specific weather regimes. Therefore, a great amount of current research on seasonal predictability has been focused on trying to foresee changes in the SST and/or its interactions with the atmosphere (see i.e. Yang and Lau, 1998; Trenberth, 2005; Haren *et al.*, 2012). One way to ease this task is the construction of ocean indices such the El Niño Southern Oscillation (ENSO) which, to date, is the main source of seasonal predictability at a planetary level (i.e. Palmer and Anderson, 1994; Wang *et al.*, 1999; Ye and Hsieh, 2008). That said, in the extra-tropics this influence it is not as clear as in the tropics (Palmer and Anderson, 1994; Stockdale, 2000). For instance, authors such as Rodwell and Doblas-Reyes (2006) note that even though the role of the SST as a seasonal predictor for Europe it is not well established there are some observational evidences that particular SST anomaly patterns can have important impacts on the continental temperatures.

Finally, the SST can be included in Global Climate Models (GCM) in two forms. Tier-two models assume the ocean in a fixed state and allow the atmosphere to evolve under its influence (see i.e. Bengtsson *et al.*, 1993). Tier-one models, on the other hand, couple the atmosphere to the ocean in a way that both systems can impact on the evolution of each other (see i.e. Blumenthal, 1991). Nowadays the latter approach is generally preferred for it better represents the natural behaviour of the atmosphere and the ocean (Kug *et al.*, 2008).

2.2.2 Soil moisture

Soil moisture can modify the Bowen ratio of a given surface (quotient between sensible and latent heats). As a consequence, it affects its temperature (Dai *et al.*, 1999; Hirschi *et al.*, 2010). This is generally a local effect with little long-term structure. However, under some circumstances it can influence temperature and precipitation with months in advance even at continental scales (see i.e. Beljaars *et al.*, 1996; Stockdale, 2000; Kanamitsu *et al.*, 2003; Orth and Seneviratne, 2012). Palmer and Anderson (1994) attributes this predictability to the fact that the hydrological state of the soil could act as an stabilizer of certain circulation patterns turning them more prevalent than they would be without these conditions.

In fact, authors such as Rodwell and Doblas-Reyes (2006) state that soil moisture was determinant in the European heat wave of 2003 for this same reason. Others, like Wilks (2008) suggest that drought indices as the Palmer Drought Severity Index (PDSI) can be used to improve seasonal forecasts. Also, since soil moisture influences the structure and abundance of vegetation it can be directly (or indirectly through precipitation) a source of predictability for wildland fires (i.e. Nepstad *et al.*, 2004; Turco *et al.*, 2013a). Nevertheless, the driving mechanisms of this kind of land-atmosphere dynamics (evaporation, heat-fluxes, . . .) are still not well represented in current dynamical models and so it is expected that as soon as this question is addressed its role in seasonal forecasting will be clarified.

2.2.3 Snow and sea-ice cover

The snow cover influence on the atmospheric long-range predictability comes from the albedo and its role as a water reservoir (Fukutome *et al.*, 2001; Boer *et al.*, 2013). Until recently, it was believed that its impact was restricted to regional scales (Yang and Lau, 1998; Céron *et al.*, 2010; Brands *et al.*, 2012). However, the latest studies suggest that the snow pack in the northern hemisphere has an influence in the NAO and the northern Pacific pressure regimes (Shongwe *et al.*, 2007). Still, since it is a non-stationary predictor more studies are needed to exactly pin-point its influence domains. As for the sea-ice, it seems that the Arctic-sheet can have some influence in the summer-storm track and winter circulation in the northern Atlantic (Balmaseda *et al.*, 2010), but it is still not clear whether these effects are robust enough to be systematically used in operational seasonal forecasting (Wielicki *et al.*, 2002). Even so, the initialization of sea-ice in the models is still subject to intense research (Orsolini *et al.*, 2011; Sigmond *et al.*, 2013a) and has to be improved before assuming or discarding its role in seasonal predictability.

2.2.4 Stratosphere

Recent studies have proposed that changes in the stratospheric thermal distribution and circulation can propagate through the troposphere in form of anomalies in a process that can take as long as two months (see i.e. Baldwin and Dunkerton, 2001; Cohen *et al.*, 2007; Ineson and Scaife, 2008; Orsolini *et al.*, 2009). It is thought that the predictability associated to the stratosphere comes from two sources: the quasi-biennal oscillation (QB; Marshall and Scaife, 2009) and sudden stratospheric warmings (SSW; Sigmond *et al.*, 2013b). The general mechanism of action remains unknown and it is still under study (Boer *et al.*, 2013).

2.3 Methodological approaches

2.3.1 Dynamical

Dynamical seasonal models are Global Circulation Models (GCM) aimed to reproduce the physics of the ocean-atmosphere system and its interactions. Initially, they were limited to the study of tropical domains and the ENSO (Cane *et al.*, 1986), but the teleconnections found between these areas and extra-tropical regions suggested that the use of coupled ocean-atmospheric models could also be implemented to middle-latitudes. Hence, many experiments were set to study seasonal forecasting in the extra-tropics and, before long, they obtained a rather positive outcome (see i.e. Molteni *et al.*, 1987; Branković *et al.*, 1990; Ji *et al.*, 1994).

At that point in time, due to the inherent complexities of the interactions between the ocean and the atmosphere, there were two strategies to tackle dynamical forecasts. The so-called tier-one, which takes into account the physical interactions and mutual feedbacks between the atmosphere and the ocean (see i.e. Blumenthal, 1991); and tier-two, which establishes SST boundary conditions in a particular moment that affect the subsequent evolution of the atmosphere (see i.e. Bengtsson *et al.*, 1993). Tier-two models are less computer demanding but, in return, they normally offer worse results than tier-one models (Kug *et al.*, 2008). That being said, there are still great uncertainties on the results obtained by these models. This is due to knowledge gaps in the atmospheric and ocean understanding, along with the need for several parametrizations and computational approximations during calculations (see i.e. Palmer, 1999; DelSole and Shukla, 2010). However, as long as these gaps are progressively filled, and the computational power increases, it is thought that these uncertainties will decrease in the future (Doblas-Reyes *et al.*, 2013).

Additionally, the forecast output of dynamical models can be of two types: *deterministic* or *probabilistic*. Deterministic forecasts provide a single exact value for the variable whereas probabilistic forecasts describe the uncertainty inherent to the forecast in the form of an ensemble of different predictions driven by perturbations in the initialization of the mode (Palmer, 1999; Gneiting and

Raftery, 2005). Theoretically, the purpose of any ensemble is to characterize the *pdf* function corresponding to all the possible outcomes driven by the uncertainties affecting the initial conditions and the model itself.(i.e. Gutiérrez *et al.*, 2004; Doblas-Reyes *et al.*, 2009). In practice, since the ensemble of a single model does not always represent well the uncertainties of the model, nowadays it is becoming widespread the use of a multi-model ensemble or an ensemble of different model ensembles (see i.e. Barnston *et al.*, 2003; Hagedorn *et al.*, 2005; Doblas-Reyes *et al.*, 2005; Sirdas *et al.*, 2007). Ultimately, the more similar the ensemble *pdf* is to the observed, the more *reliable* the probabilistic forecasts will be. Likewise, a better *resolution* will be achieved as long as the variability of the ensemble approaches the inherent of the observed climatology (see i.e. Wilks, 2006).

2.3.2 Statistical

This strategy is based on the analysis of historical data in order to identify relationships among variables, atmospheric configurations or teleconnections that can be used to forecast the same or other variables (i.e. Barnston *et al.*, 1994; Murphy *et al.*, 2001; Diez *et al.*, 2005; Wilks, 2008; Kumar, 2009). Every statistical technique that allows us to identify such relationships can be used in this approach. Among the most common we can find: CCA (canonical correlation analysis), SVD (singular value decomposition), the analog method, MOS (Model Output Statistics), the MLR (multiple linear regression), and the Bayesian and neural networks (Gutiérrez *et al.*, 2004; Wilks, 2006; Benestad *et al.*, 2008). As of today its maximum advantage is to attain results similar to the raw output of dynamical models but at a much smaller computational cost (see i.e. Hastenrath *et al.*, 2009). On the contrary, there is the difficulty of establishing the stationariness of the relationships found and the limited number of past situations available. In reality, though, statistical strategies are no longer used alone. That is because the best results are attained when the dynamical model outputs are post-processed with statistical methods (Schepen *et al.*, 2012; Yoon *et al.*, 2012). This post-processing of the dynamical output can be of two forms:

a) **Regionalization**: also known as *statistical-downscaling*, it consists on the amelioration of the dynamical model output by filling the gap between the coarse resolution of the model and the real world. In this way we can increase the forecast spatial resolution and, thus, recover meteorological (climatological) features that otherwise would be hidden or directly missing (Maraun *et al.*, 2010; Turco *et al.*, 2011; Vrac *et al.*, 2012; Gutiérrez *et al.*, 2013). This statistical approach can add information not present in the dynamical output as long as it uses observation reference fields.

b) **Calibration**: it is the process by which statistical methods are applied to correct the biases and/or systematic errors of the model output by using antecedent series of forecasts and observations (Weigel *et al.*, 2009; Piani *et al.*, 2010; Peng *et al.*, 2014). These errors are consequence of the inherent limitations of the physical model related to parametrizations, equation simplification and uncertainties in the initialisation procedure.

Moreover statistical methods can be used, independently or combined, to isolate and characterize the uncertainties coming from different aspects of the dynamical modelling process (Doblas-Reyes *et al.*, 2009; Weisheimer *et al.*, 2011, 2014). This conception puts to an end the passionate controversy about the superiority of implementing either statistical or dynamical approaches for it combines the strengths of the two worlds. This is highly important in the field of seasonal forecasting applications, for it eases its main objective, which is to provide the end-user with the best seasonal information available.

2.4 Applications

This section refers to the potential developments that can make use of seasonal forecasting information to suit end-user needs. This is a sort of *holy grail* in the seasonal forecasting discipline, for the possibility of anticipating seasonal anomalies in the long-term is a clear advantage to deal with decision-making in many economic and strategic areas. Nowadays it is part of what is commonly known as *climate services* (Coelho and Costa, 2010; Graham *et al.*, 2011; Hewitt *et al.*, 2012) and embraces a number of fields:

i) Agriculture (Schneider and Garbrecht, 2006; Fraisse *et al.*, 2006; Garbrecht and Schneider, 2007; Davey and Brookshaw, 2011)

ii) Water resources (Chowdhury and Sharma, 2009; Céron *et al.*, 2010; Wang and Robertson, 2011; Robertson *et al.*, 2012; Dutra *et al.*, 2013)

iii) Energy (Damrongkulkamjorn and Churueang, 2005; García-Morales and Dubus, 2007; Block, 2011)

iv) Risk and disaster management (Goddard *et al.*, 2010; Tall *et al.*, 2012; Dutton *et al.*, 2013)

v) Forest fires (Roncoli *et al.*, 2012; Turco *et al.*, 2013a; Harris *et al.*, 2014; Spessa *et al.*, 2014)

vi) Health (Pascual and Dobson, 2005; Thomson *et al.*, 2006; Tompkins and Di Giuseppe, 2014)

vii) Hurricane activity (see i.e. Wang *et al.*, 2009; LaRow *et al.*, 2010; Kim and Webster, 2010; Leckebusch *et al.*, 2015)

viii) Financial markets (Gaushell, 2002; Campbell and Diebold, 2005; Brands, 2013)

In the following subsections we will center our attention in the two areas where we focus the application of seasonal forecasting in this thesis: Mediterranean water scarcity and summer wildfires.

2.4.1 Mediterranean water scarcity

Water scarcity in the Mediterranean is usually linked to the relationship between droughts, water supplies and water demands. Actually, droughts are a recurrent problem in southern Europe (see i.e. Lloyd-Hughes and Saunders, 2002; Brewer *et al.*, 2007; Nola *et al.*, 2008). Historical reports speak of such events arising in multiple moments of the past with great impacts to the affected communities (Martín-Vide and Barriendos, 1995). Examples of these conditions were the *Maldà* oscillation, characterized by an anomalous succession of intense drought and flood periods between 1760-1800 in the western Mediterranean (Barriendos *et al.*, 2003; Barrera *et al.*, 2006) or the long-lasting drought of 1812-1818, which affected the Mediterranean Coast and the Balearic Islands (Álvarez *et al.*, 2008). At present the Mediterranean keeps on being affected by these situations and it is presumed that the influence of climate change can strengthen these episodes in the future (see i.e. Hill *et al.*, 2008; Dai, 2011; MedCLIVAR, 2012; Singla, 2012). In Spain, at the midst of the XXth century, the authority carried out a plan to build multiple dams and canalizations to ease the effects of droughts and regulate water supply for the increase of agriculture and urban uses. Even though, hydraulic administration in Spain and Catalonia still has to deal with problems caused by droughts (MIMAM, 2007). In this context the optimization of dam management becomes a feasible need (i.e. Iglesias *et al.*, 2009; Iglesias and Garrote, 2014; Bianucci *et al.*, 2015).

The critical effects which dry and wet periods have on numerous aspects of the society and economy have turned the possibility to foretell rainfall or river-stream flow anomalies with months in advance an engaging subject of research in many parts of the world. In fact, in the late 2000s many studies began to show the feasibility of hydrological seasonal forecasting in middle latitudes, both in North America and Europe (see i.e. Wedgbrow *et al.*, 2002; Gobena and Gan, 2010; Céron *et al.*, 2010; Shukla and Lettenmaier, 2011; Zalachori *et al.*, 2012). However, the translation of seasonal forecasting advances to actual water-resource applications is still a matter of intense investigation (Soubeyroux *et al.*, 2010; Shafiee-Jood *et al.*, 2012). Some authors like Rodwell and Doblas-Reyes (2006) argue that the ultimate goal of a forecast is to influence decision making:

> "If a forecast, however skilful, has no impact on decision making, one can argue that it is pointless to make it. The *value* of such forecasts to a particular user depends on their vulnerability to weather or climate anomalies and on what actions they can take to mitigate against any loss."

But determining the *usefulness* of a forecast is a complex issue because it not only depends on the performance of the predictions themselves but also on the needs and vulnerabilities of the end-user (Watkins and Wei, 2008). Stockdale

(2000) and Steinemann (2006) state that working closely with decision-makers offers the possibility to tailor indices based on seasonal forecast information that, even with the limited skill of current models can have advantages for them.

2.4.2 Mediterranean summer wildfires

In Europe, approximately 500,000 hectares burn in about 50,000 fires each year (San-Miguel-Ayanz *et al.*, 2013b). Most of these are Mediterranean summer fires that lead to damage to the natural environment and property, causing loss of lives and important economic losses every year (Moreira *et al.*, 2011). Estimating the fire risk a few months in advance may thus allow fire protection agencies to devise timely reactions and adequate provision of human and material resources (Mavsar *et al.*, 2013). The predictability of fires is a complex issue, due in part to the fact that fire activity is closely related to both natural factors and human action whose relative importance on different scales may be challenging to estimate (Bonan, 2008; Bowman *et al.*, 2009; Dube, 2009; Macias Fauria *et al.*, 2011; Moritz *et al.*, 2012). Human activities influence fires either directly, via ignition and suppression, or indirectly, via fuel management (Moreira *et al.*, 2011; Ganteaume *et al.*, 2013). Another important driver of fires on a regional scale are climatic processes (Dwyer *et al.*, 2000; Meyn *et al.*, 2007; Flannigan *et al.*, 2000, 2009; Pechony and Shindell, 2010; Hessl, 2011). In particular, although most fires are ignited by human activity, year-to-year changes in the ease of ignition and in the burned areas are mainly related to interannual climate variability. Several studies support the hypothesis that in Mediterranean-type ecosystems, droughts are a primary driver of the interannual variability of fires, controlling fuel flammability and fuel structure (see, e.g. Pausas, 2004; Pereira *et al.*, 2005; Meyn *et al.*, 2007; Gudmundsson *et al.*, 2014). That is, drought conditions may lead to high levels of fuel flammability, but fire activity is also favoured by the presence of the fine fuels produced during antecedent periods with favourable climate conditions (see e.g. Turco *et al.*, 2013a; Koutsias *et al.*, 2013; Bedia *et al.*, 2014).

The dependence of wildfires on weather and climatic conditions therefore means that fire risk has a certain level of predictability (Bonan, 2008; Dube, 2009; Macias Fauria *et al.*, 2011). Fire risk predictability has been thoroughly addressed in the case of fire danger forecasts from 1 day to 2 weeks, through fire weather indices (e.g. Canadian Forest Fire Weather Index van Wagner and Pickett, 1985; Bedia *et al.*, 2013; de Groot and Flannigan, 2014, for a review). On longer time scales, however, studies addressing seasonal prediction of fire danger are still relatively scarce with most of them following an empirical approach, statistically linking antecedent climatic variables used as predictors with observed fire activity in Australia (Harris *et al.*, 2014) and in North America (e.g. Chu *et al.*, 2002; Westerling *et al.*, 2002, 2003, 2006; Preisler *et al.*, 2004; Preisler and Westerling, 2007; Preisler *et al.*, 2008; Chen *et al.*, 2011; Shabbar *et al.*, 2011). For southern Europe, a recent study (Gudmundsson *et al.*, 2014) has shown that above-normal summer wildfire activity can be forecasted several months in advance by using drought indices, through the effect on fuel flammability, while a few studies have used global climate models (GCM) to seasonally predict fire danger (Roads *et al.*, 2005, 2010; Spessa *et al.*, 2014).

CHAPTER 3

Datasets

This chapter presents a comprehensive description of the various datasets used in this thesis. For each database it establishes the reasons behind our choice as well as its main features. Thus, it poses the basis to better understand the framework of the applied methodologies and the results obtained.

Its organization is as follows: the first section describes the observational dataset, E-OBS (3.1); the next one corresponds to the seasonal forecasting model, ECMWF System 4 (3.2); the third contains the hydrological data (3.3); and the last one is for the fire dataset (3.4) .

3.1 European Observational Dataset (E-OBS)

The observational dataset is the *truth* against which we verify our findings. Here, the word *truth* is very important because, philosophical considerations aside, it is not as unique as one might think. In fact, there is an intrinsic limitation in the measurement of spatially dependent continuous variables; for having an exact image of such systems also needs of a continuous (infinite) observational network. Since this is not possible, we will always have to rely on approximations to this *truth*. In our case we have chosen E-OBS, for the following reasons:

i. **It is an observational gridded dataset**. We prefer this approach to any reanalysis to avoid the possibility that the assimilation method could resemble to the one used by the ECMWF System 4. With this strategy we try to eliminate one hypothetical source of uncertainty in the verification process (see i.e. Kim *et al.*, 2012, as an example of the effect of different reanalysis in a verification process).

ii. **It covers the European domain**. Other observational datasets such as Spain02 (Herrera *et al.*, 2012) or the Mesoscale Alpine Programme (MAP, Bougeault *et al.*, 2001) have a larger station density but, on the other hand, are restricted to regional areas. To maintain homogeneity on the verification and calibration processes we decided it was worth having a slightly reduced number of stations if in return we got a more homogeneous coverage for Europe.

iii. It provides values in a **daily basis** from **1950 to present**, has extensive **metadata** with **interpolation error assessment** and it is **updated in a regular basis** (almost two times per year).

iv. **It offers a resolution of $\mathbf{0.25^\circ \times 0.25^\circ}$**. This resolution is the minimum encountered for the European domain that can approximately resolve river basins such *la Muga* with the aforementioned features.

In the ideal case, an observational interpolated gridded dataset of $0.25^\circ \times 0.25^\circ$ resolution would have one station per cell (at least). In the European domain, with an area of approximately 10^7 km^2, this would require a network of around 16000 stations. Unfortunately, current European observational network is both far from these numbers and from an optimal uniform distribution (see Haylock *et al.*, 2008, and figure 3.1).

Initially, E-OBS comprised 2200 stations with daily observations for precipitation, minimum, maximum and mean surface temperature, mostly spanning the period 1950˘2006 (Haylock *et al.*, 2008; Klok and Klein-Tank, 2009). Since then, almost two times per year there has been an update with a fix for minor bugs and the inclusion of new stations and/or the expansion of temporal ranges of the already existing datasets. Thus, this continuous revision made E-OBS a highly valuable database for verification purposes.

More specifically in this thesis we have used E-OBS v8.0. In this version most of the interpolated stations cover the period 1950-2012, with daily observations in more than 4000 sites for maximum and minimum temperature, and 7000 regarding precipitation. Figure 3.1 depicts E-OBS observation network distribution for maximum temperature. For the other variables, though the absolute number of stations might differ, the overall disposition is very similar.

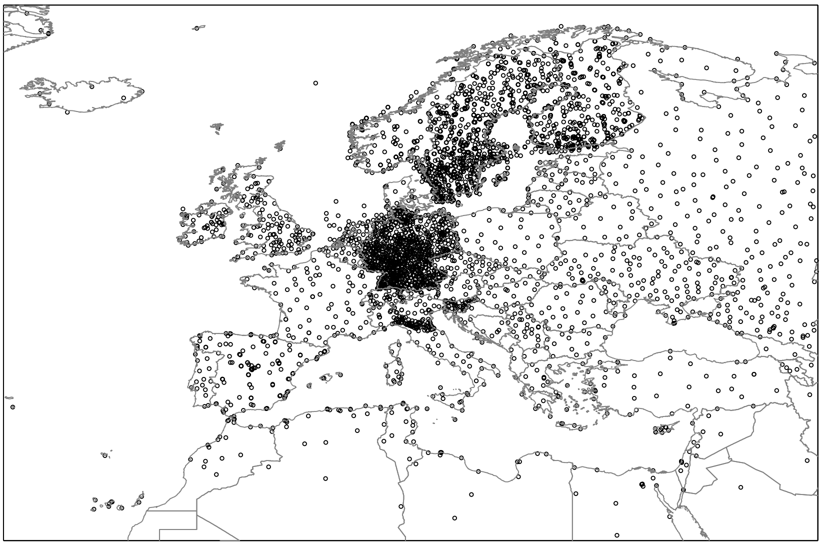


Figure 3.1. E-OBS v8.0 maximum temperature station cover map for Europe. Note network's heterogeneous distribution. Elaborated from ECA&D.

The methodology applied to interpolate the aforementioned observations is comparable to universal kriging (Journel and Huijbregts, 1978). In this case, however, Haylock *et al.* (2008) follow a three-step process to homogenize the climatic differences of the underlying daily data, a required condition for the correct application of a regular kriging process (further details can be found in Haylock *et al.*, 2008; Hofstra *et al.*, 2008).

In the presentation paper Haylock *et al.* (2008) speak of two levels of quality assessment. The first one dealt with raw data, identifying and correcting potential non-homogeneous series (i.e. Peterson *et al.*, 1998; Ducré-Robitaille *et al.*, 2003; Reeves *et al.*, 2007; Turco *et al.*, 2012). It involved quality tests to detect and remove doubtful values, i.e. outliers, and the identification of other obvious problems such as shifting in date assignment. The second level involved the characterization of global uncertainty, showing that the interpolation fraction was probably larger than all of the other sources combined. Hence, efforts were focused in the determination of its magnitude. As a result ECA&D supplies gridded standard error files for the entire domain and for each of the variables (see i.e. 3.2).

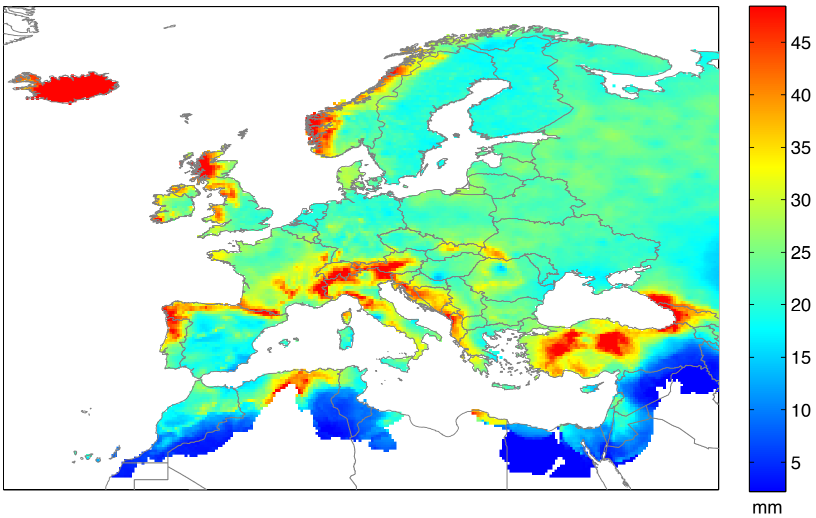


Figure 3.2. E-OBS v8.0 monthly median interpolation standard error map for Europe's precipitation (1950-2012). The biggest errors are mostly centred on high mountain ranges and in the rainiest areas of the continent. Elaborated from ECA&D.

3.2 ECMWF System-4 (S4)

Dynamical seasonal forecasting models did not become operational until they outperformed simple statistical approaches such as *climatology* and *persistence*. This did not happen until the mid 90's when dynamical seasonal forecasts progressively turned from experimental to fully operational frameworks (see Molteni *et al.*, 2011, and chapter 2).

S4 is today's most recent instalment of ECMWF in the seasonal forecasting field. It consists of a coupled suite of atmospheric and oceanic models reproducing the general oceanic and atmospheric circulations as well as their entangled feedbacks. This is achieved through the numerical resolution of the thermodynamic and motion equations corresponding to each grid box in which the S4 divides our planet's atmosphere and ocean.

The atmospheric component is the CY36R4 version of ECMWF's weather forecasting model IFS (Integrated Forecasting System) with a land surface initialization driven by ERA Interim (Dee *et al.*, 2011). The horizontal resolution is of approximately 80 km whereas for vertical distribution the atmosphere has 91 levels, up to 0.01 hPa. As for the ocean model, it uses NEMO (Nucleus for European Modelling of the Ocean) version 3.0 and NEMOVAR (Mogensen *et al.*, 2012), a state-of-the-art modelling and analysis frameworks. Horizontally, NEMO grid boxes have an approximate length of 1° (110 km) while, vertically, they includes 42 levels. For further technical details we refer to Molteni *et al.* (2011).

We have preferred S4 because it is the evolution of the well considered ECMWF System 3 (see i. e. Stockdale *et al.*, 2011) and for its full potential has not been totally assessed (Molteni *et al.*, 2011; Rodrigues *et al.*, 2014). In comparison with S3 and regarding our studied region, some of its achievements involve: *a*) enhancement of deterministic and probabilistic scores in extra-tropical regions; *b*) improvement of natural ocean/atmosphere variability simulation and forecasting in tropical Atlantic and nearby regions; and *c*) increase of the anomaly correlation of ENSO indices for most regions and seasons. Thus, some of its interesting features include:

i. **It is an operational fully coupled GCM**. We choose fully coupled atmospheric-ocean models (tier-one, see chapter 2) for their best results in comparison with simpler models such as tier-two (Kug *et al.*, 2008).

ii. **It spans the European domain with a resolution of $0.75° \times 0.75°$**. Though there was the possibility to get 0.25° interpolated S4 data we decided to download the minimum native resolution available to take control of the interpolation process (see section 5.2).

iii. **It provides monthly mean values re-forecasts from 1981-2010 including a 15-members ensemble**. To have a climatic period reforecast with its associated ensemble is a neat advantage for the verification process.

iv. The **long experience** of the institution, ECWMF, **in the evaluation, updating and improvement of the inner structure of the model**. In fact, it is the fourth version of ECMWF seasonal forecast system since its first implementation in 1997 and his predecessors have been already evaluated in DEMETER and ENSEMBLES projects (Palmer *et al.*, 2004; Weisheimer *et al.*, 2009).

Hence, in this thesis we have worked with the 30 years S4 monthly mean values re-forecasts (1981-2010) with each run consisting of 7 month forecasts issued on the first day of every month. This means that each month of the year (January to December) has been forecasted seven times, ranging from the issued month (m-0) to 6 months in advance (m-6). In figure 3.3 we can find an schematic example of the S4 different forecast horizons for April.

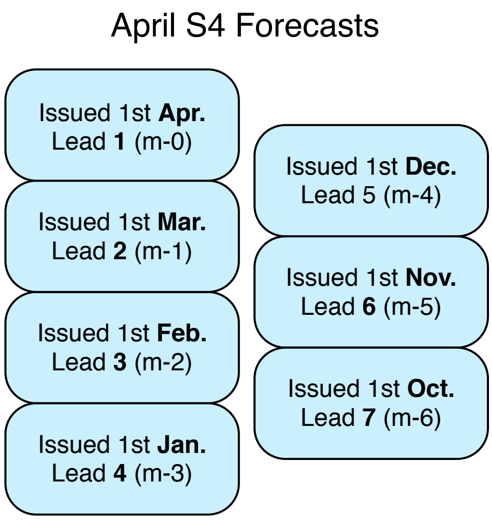


Figure 3.3. Schematic S4 forecast horizons for April. Each cell includes the day and month of the issued forecast along with the forecast horizon in two nomenclatures: *lead*, which corresponds to the number of months between the month of the most recent observations included in the model and the forecast month; and *m-number* in which *m* referes to the forecast month and *number* the months between this month and the month in which the forecast was issued.

Every forecast comes along a 15-members ensemble at a spatial resolution of 0.75°. Most of the spread in the ensemble is internally generated and, even though there are initial perturbations and stochastic physics that contribute to its generation, the role of initial perturbations is rather limited (Molteni *et al.*, 2011).

Finally, it has to be mentioned that we could eventually obtain S4 full resolution re-forecast issued at every month thanks to AEMET and ECMWF authorization. Prior to that, however, most of the work of this thesis had been done through the CHFP (Climate-System Historical Forecast Project) which publicly provides 4 initializations of the *hindcast* (February, May, August and November) in a resolution of 2.5° but for the same period and forecast attributes as the complete re-forecast.

3.3 Hydrological data

The hydrological data comprises mean daily values of in-flow, out-flow and total water volume measured by ACA (*Agència Catalana de l'Aigua*) in the Boadella reservoir, on the upper part of the Catalan *Muga* river basin (see section 4.4 for its detailed description). This particular choice responds to the accomplishment of a number of requisites to check the usability of seasonal forecast information:

i. **It is within the studied area and has in-flow, out-flow and water volume records for the period 1981-2010.** This gives us the possibility to depict the performance of S4 seasonal forecasts with the three variables that mostly determine the availability of stored resources.

ii. **There are no up-stream influencing human infrastructures**. This is important for it assures that the quantities measured by the in-flow gauge are determined only by natural factors.

iii. **It satisfies water demands from agriculture and urban areas with observed discontinuous drought periods**. The existence of such periods can influence the provision for water and, therefore, it offers the opportunity to assess the potential benefits of seasonal forecasting information in such circumstances.

Though we have used data for 1981-2010, these time-series go from 1971 to 2011. They have been recorded with automatic gauges in the reservoir system (entrance and exit) which offer direct (in-flow and out-flow) or indirect (volume from water level) measurements. There is also an undefined period in which in-flow data was indirectly computed through variations in water volume and out-flow measures (ACA personal communication). That being said, all datasets underwent an automatic checking through the verification tool of the *Cicle de l'Aigua al Territori* based on the software HEC-DSS Vue of the Hydrologic Engineering Center.

3.4 Fire data

The fire data is obtained from the Forest Fire Prevention Service of the *Generalitat de Catalunya*, SPIF, and it consists of monthly series of burned area (BA) in Catalonia for the period 1983-2010. This selection has been favoured for:

i. **It spans the period 1983-2010 with comprehensive metadata support.** After seeking in the European Forest Fire Information System (EFFIS) at NUTS3 resolution we found that Catalonia fire series is one of the few that covers a period similar to the S4 re-forecast. Particularly, this choice is in coherence with the former team-works by Turco *et al.* (2013a, 2014) and the hydrological case studied.

ii. **It represents fire behaviour of a typical Mediterranean environment (see section 4.3).** Hence, the results will have the potential to be transferred to other areas with similar characteristics.

We have also established a lower limit of 0.5 ha below which fire records are not considered to guarantee the homogeneity of fire series (see figure 3.4). This condition has been adopted in accordance with SPIF because the lower threshold for wildfire report has been changing over time. Fortunately this restriction, if any, has a very limited effect on BA since already up to 74% of it comes from fires exceeding 500 ha (Turco *et al.*, 2013a).

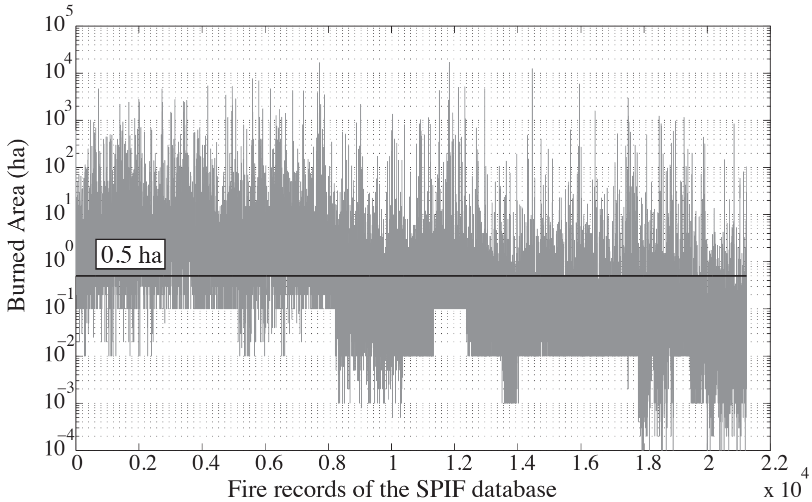


Figure 3.4. Burned area (BA) in each record of the SPIF database in chronological order (BA in logarithmic scale, log10). Note the evolution of the lower limit of fire recording along the series. From (Turco *et al.*, 2013b).

CHAPTER 4

Regions of study

This chapter compiles the geographical and climatological characterization of the regions studied in this thesis. In the forthcoming chapters this will help us understand the importance of seasonal forecasting in different periods of the year and its dependence on the areas considered. The structure of this unit goes from larger to local domains as follows: the first section is devoted to Europe (4.1); the next one refers to Spain (4.2); the third is for Catalonia (4.3); the fourth comes for the Muga river basin (4.4); and the last one is for the Muga's river Boadella reservoir (4.5).

4.1 Europe

Europe is a large peninsula limiting to the north with the Arctic Ocean; to the south, with the Black, Caspian and Mediterranean seas; to the West, with the Atlantic Ocean; and that it is connected to the east with the Eurasian continent. As we can see in figure 4.1 its relief is greatly varied, ranging from vast plains to high mountain areas of more than 2500 m. Often, these height transitions occur in relatively small distances, leading to an abrupt succession of different climatologies. This can be observed in the Mediterranean shores, where barely a few kilometres inside we can encounter mountain ranges of more than 500 m that can surpass 2000 m in just 50 or 100 kilometres more.

In fact, Europe's northern extra-tropical situation determines some meteorological features that, combined with every region's particular relief, leads to

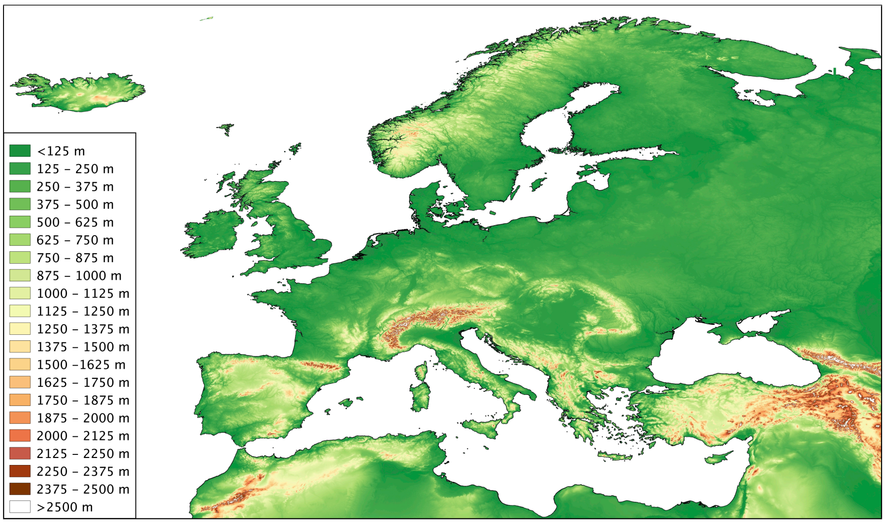


Figure 4.1. Europe's relief. Elaborated from GMTED2010 (Danielson and Gesch, 2011).

a distinct mix of climatic regions. Firstly, we should highlight the prevalence of western oceanic circulations, a characteristic that regulates much of the prevailing weather over the continent. For instance, common yearly precipitation amounts above 1100 mm in the occidental coasts of Great Britain, Ireland, Norway, France or Spain can be directly linked to these westerlies. Mountainous ranges such as the Pyrenees, Alps, Massif Central or the Apennines are also regions of high annual rainfall with total cumulates of more than 1100 mm. Finally, other maxima can be found in the north-eastern Adriatic coasts and in the Genoa gulf, and are linked to the presence of the Alps and Mediterranean cyclone configurations (see figure 4.2).

The driest regions, with yearly values below or around 400 mm coincide with central and southern Spain and regions limiting to the east with the Black and Caspian seas. The vast plains of eastern Europe have values ranging from 500 mm to 700 mm, with local registers above/below these numbers. In the Mediterranean, the eastern coasts of Italy and Greece have total values around 500 mm. The other regions have annual amounts between 600 and 900 mm. In figure 4.3 we can observe precipitation behaviour throughout the four seasons. In the Mediterranean shores the driest season coincides with summer, with values around or well below 100 mm. In the western coasts of the continent, although the precipitation

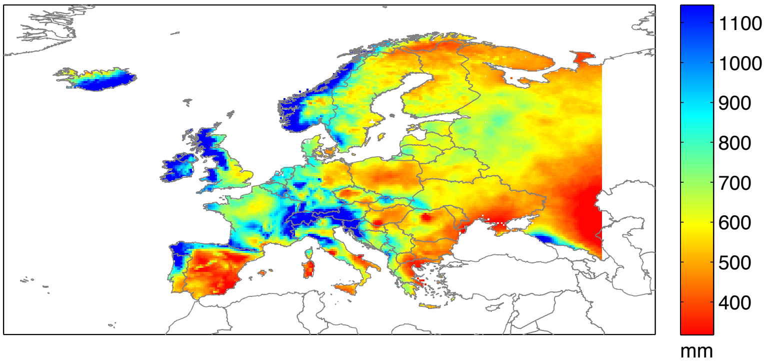


Figure 4.2. Europe annual precipitation climatology for the period 1981-2010. Elaborated from E-OBS.

amount is higher compared to the aforementioned area, it is also reduced regarding to the other seasons. Turning to the Alps and central and eastern Europe plains, the maximum in rainfall occurs during summer months and it is linked to prevalence of daily convection. In the eastern part of the continent the minimum occurs during spring and winter whereas in the Atlantic shores the maximum is extended from autumn to spring. In the Mediterranean, on the other hand, the wet season is autumn. In this region there is also a secondary maximum that shows up in spring in western Mediterranean, and during winter in the central and eastern area.

Europe's situation, surrounded by seas and oceans, is what mostly determines the thermometric regime of the continent. Latitude and altitude are the other factors that play a key role in the final characterization of Europe's temperatures. Hence, all coastal areas have relatively warmer temperatures than the inner regions of the continent and also warmer compared to regions that are in the same latitudes but surrounded either by terrain or colder oceans. In fact, the Mediterranean is the warmest sea around the continent and, therefore, it contributes to give the areas near it the warmest annual median maximum and minimum temperatures. Temperature magnitude also varies through the continent in accordance with latitude and altitude. The highest regions are normally cooler whereas the lower are hotter. The differences between the highest mountainous ranges and the nearest plains is about 10°C regarding median maximum temperatures and 6°C for median minimums. As for latitudinal differences, the gap

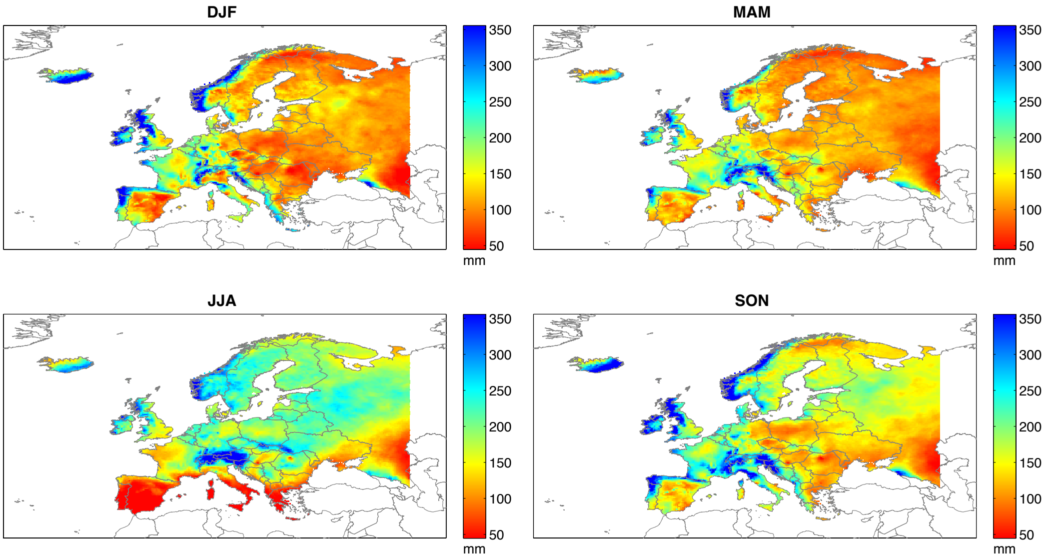
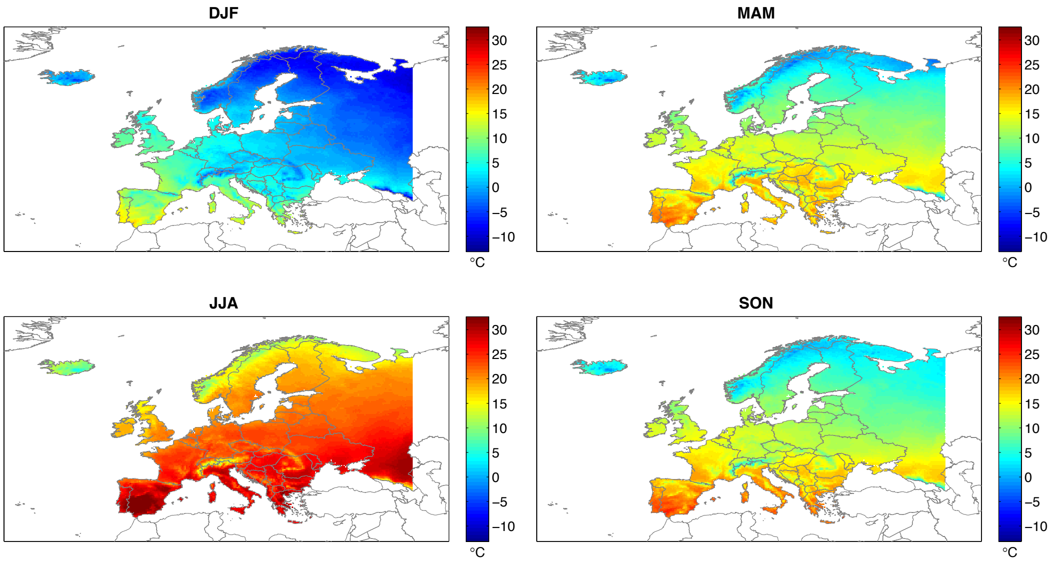
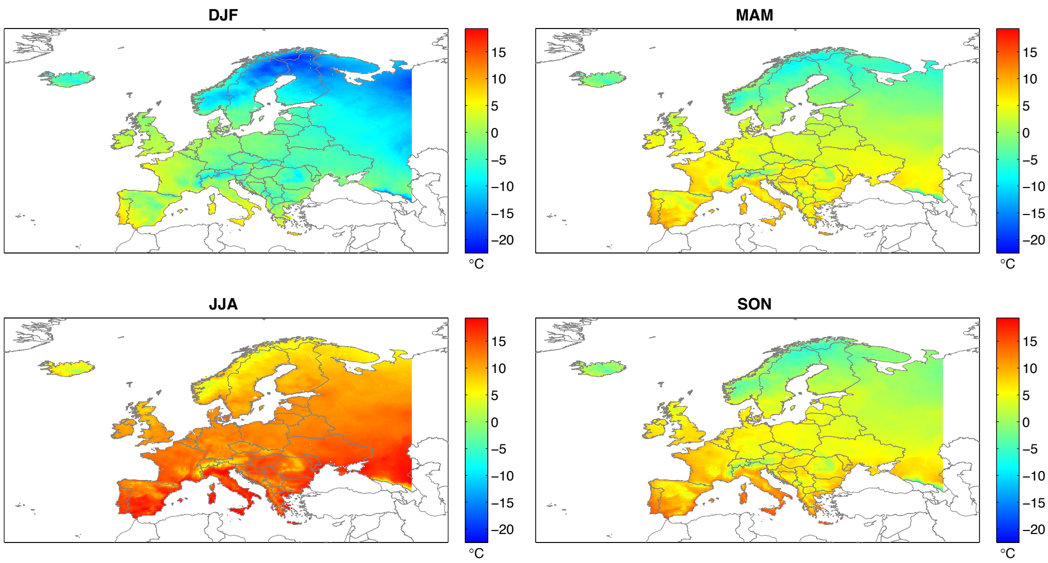


Figure 4.3. Europe seasonal precipitation climatology for the period 1981-2010. Elaborated from E-OBS.

between the southern and northern regions is usually around 15°C for minimum temperatures, and 25°C for maximums. More specifically, differences also progressively increase from north-western Europe to south-western Iberian Peninsula and south-eastern Europe, near the Caspian sea. Turning to seasonal behaviour, the annual cycle is present in all regions, with progressive winter-summer and summer-winter transitions. Additionally, oceanic sectors have milder differences among seasons than continental areas, where these contrasts are stronger (see figures 4.4a and 4.4b).



(a) Maximum temperature



(b) Minimum temperature

Figure 4.4. Europe's seasonal maximum and minimum temperature climatologies for the period 1981-2010 (a) Median seasonal maximum temperature (b) Median seasonal minimum temperature. Elaborated from E-OBS.

4.2 Spain

Spain is the largest country in the Iberian Peninsula, in south-western Europe (see figures 4.1 and 4.5). Although it includes some regions which are not in the Peninsula, like the Balearic and Canary Islands, hereafter we will use *Spain* to refer to Peninsular Spain. It is mainly surrounded by water bodies: the Mediterranean to the east and south-east; and the Atlantic Ocean to the north, north-west and south-west. To the west it limits with Portugal and, in the north-east, it is connected to the rest of Europe through the Pyrinnees. The former and *Sierra Nevada* are its highest mountain ranges, with summits above 2500 m.

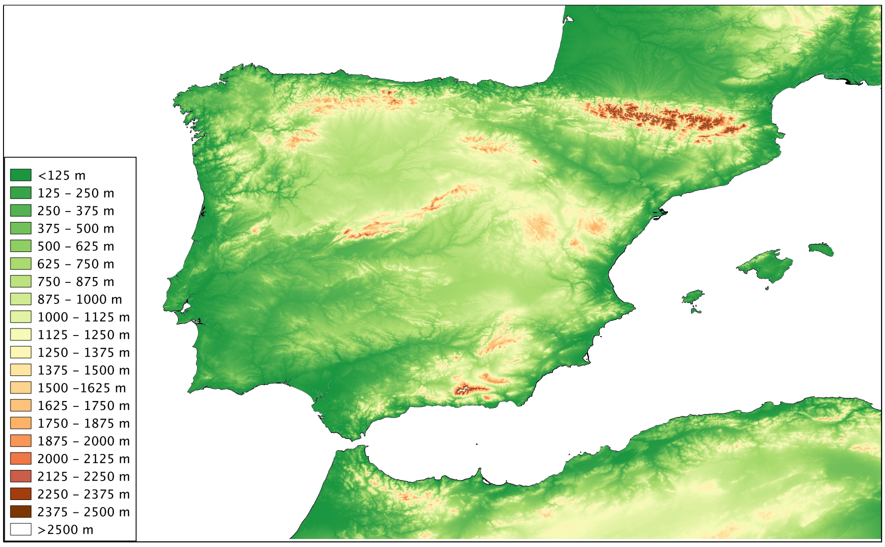


Figure 4.5. Iberian Peninsula's relief. Elaborated from GMTED2010 (Danielson and Gesch, 2011).

Its geographical situation, relief and the influence of surrounding water bodies determines a variety of climatic regimes. Although it is also influenced by westerlies, only the north-west is completely defined by them. Annually, the precipitations are maximum in the north, with median total amounts above 700 mm (see figure 4.6). In fact, the maximum is reached in the north-West, were year amounts are beyond 1100 mm. A secondary maxima can be found in the Pyrennees and in some regions of the northern coasts, with values around or above 1100 mm. The driest zones are in the south-east, the Ebro Valley and the interior plateaus, with yearly rainfall below 400 mm. still, mountain ranges in these areas have higher precipitation numbers, normally between 500 and 700 mm. In the rest of the country rainfall is comprised between 500 and 800 mm. Taking a deeper look into the seasonal distribution of this rainfall (see figure 4.7) we find that the driest season appears in summer with median values below or around 100 mm in the most part of the Peninsula. The rainiest season is autumn, with a secondary maximum that comes in winter for the south-west, and in spring for the east.

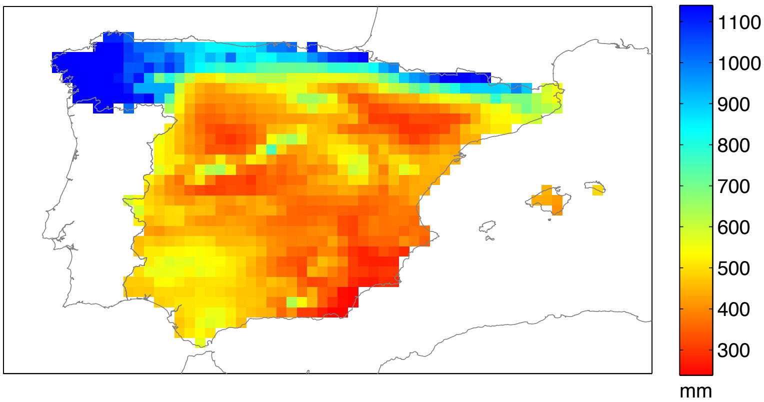


Figure 4.6. Spain annual precipitation climatology for the period 1981-2010. Elaborated from E-OBS.

If we turn our attention to maximum and minimum temperatures, their median annual behaviour is very similar. Additionally, in these images we can also observe the effect of latitude and height, which give generally colder temperatures in northern and higher domains. Another features shown in these maps are: *a*) the effect of water bodies, giving higher temperatures in coastal sites in comparison to their surroundings; *b*) the increased temperature in Ebro and Guadalquivir valleys; and *c*) the lower temperatures of the Spanish interior plateaus. These properties are also maintained when going seasonally (see figures 4.8a and 4.8b. Particularly, we can highlight a well defined annual cycle with the hottest season in summer, the coldest in winter and with milder temperatures in spring and autumn. In this last case we can see that autumn is warmer than spring, probably consequence of the summer thermal inertia thanks to the presence of the sea.

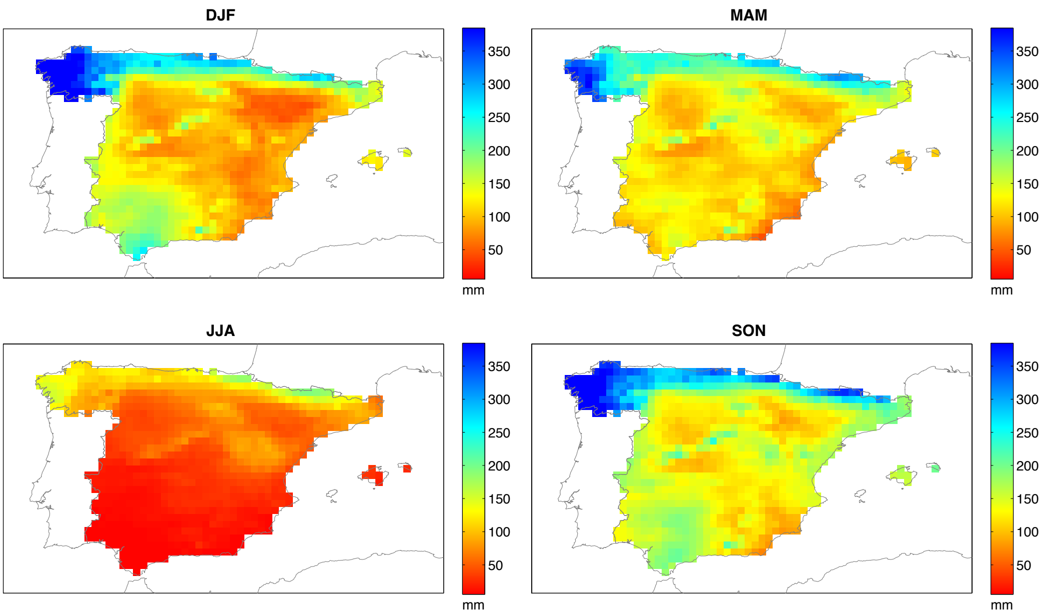
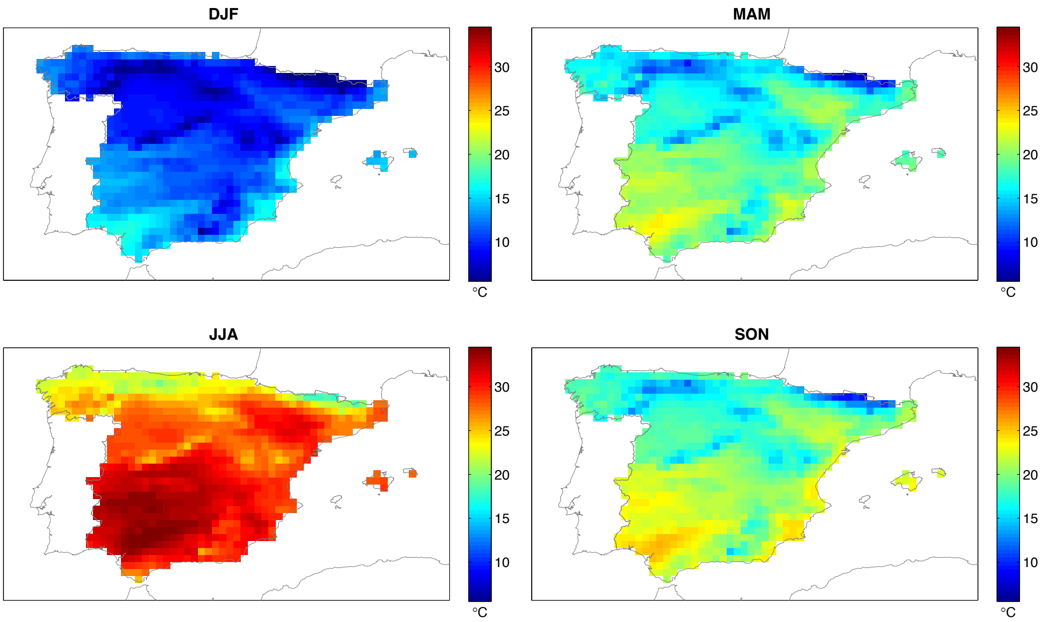
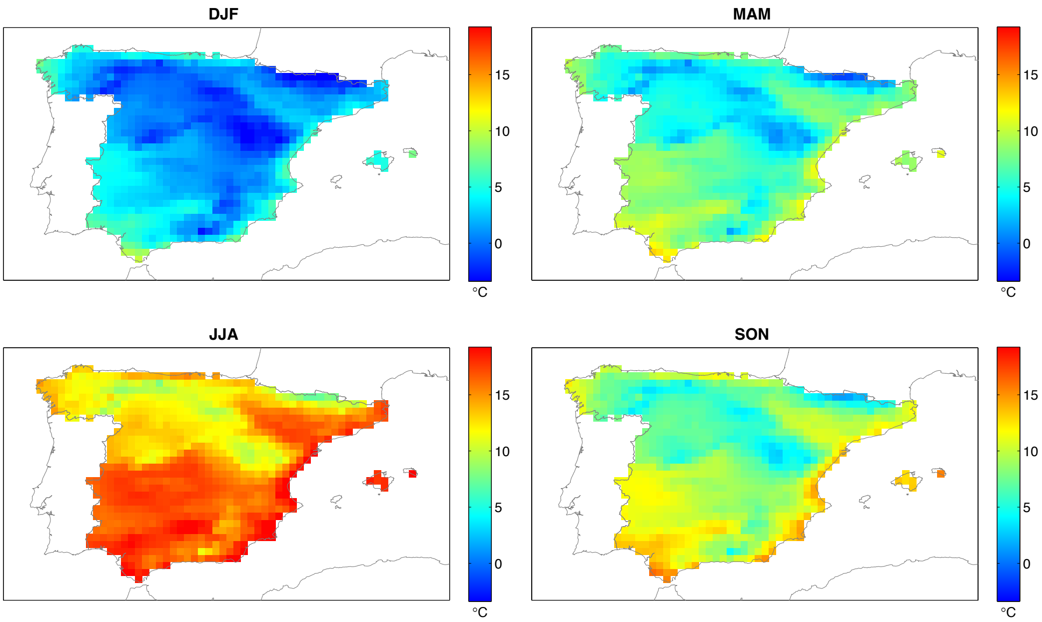


Figure 4.7. Spain seasonal precipitation climatology for the period 1981-2010. Elaborated from E-OBS.



(a) Maximum temperature



(b) Minimum temperature

Figure 4.8. Spain seasonal maximum and minimum temperature climatologies for the period 1981-2010 (a) Median seasonal maximum temperature (b) Median seasonal minimum temperature. Elaborated from E-OBS.

4.3 Catalonia

Catalonia is located in the north-east of the Iberian Peninsula. Its surface area of 31930 km^2 is bounded on the north by the Pyrenees (summits above 2500 m) and on the east by the Pre-coastal range (1700 m) and the Coastal Range (700 m), both parallel to the Mediterranean coast (see figure 4.9). Catalonia's yearly precipitation decreases from more than 1100 mm in the northern Pyrenees to less than 500 mm in the south. The minimum, around 400 mm, is located in the Central plain. The thermometric regime is characterized by warmer temperatures near the Mediterranean which progressively decrease as we move to the west and to the north. Seasonally, the driest and hottest season coincides with summer (specially in the interior). The rainiest season, conversely, falls in autumn. In the northern area winter and spring are also important periods of rainfall. Speaking of temperature, the coldest season is winter. Spring and autumn are milder, with the latter being a bit warmer than the former. Finally, summer is the warmest.

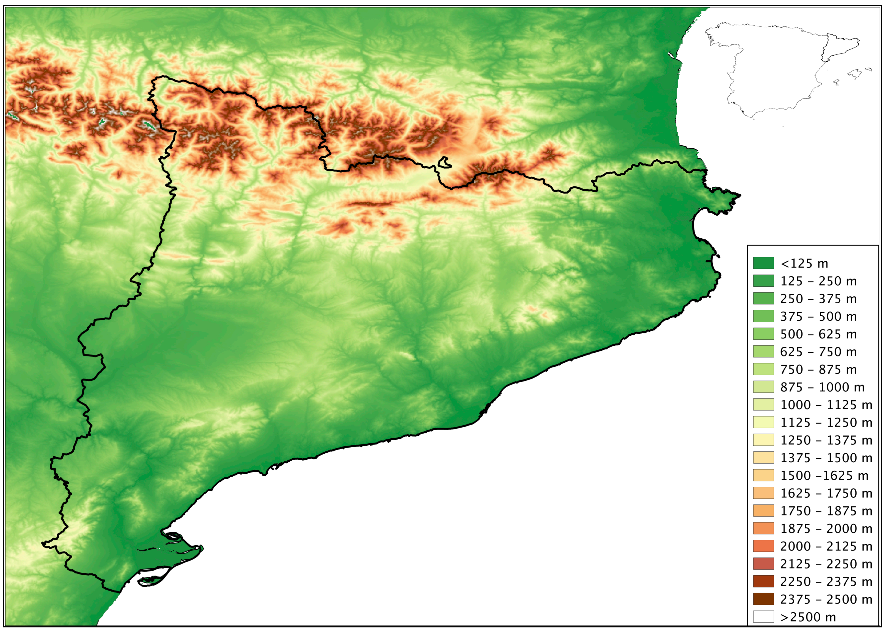


Figure 4.9. Catalonia's relief. Elaborated from GMTED2010 (Danielson and Gesch, 2011).

4.4 Muga river basin

The Muga river basin is located in the north-eastern part of the Iberian Peninsula. It covers a catalan region delimited to the north by the Pyrenees, and the Mediterranean to the east (see figure 4.10). Although its surface is relatively small, about 854 km^2, there are high contrasts between the mountainous region, with altitudes around 1100 m (summits about 1400 m) and the lower sedimentary plains. The river Muga, which gives the basin's name, is 64 km long and in its upper flow is regulated by the Boadella reservoir (see section 4.5. The mountainous area is mainly covered with forests whereas the lower heights are devoted to agriculture. Main urban areas lie in the lower stream. Its climogram shows that summer is the hottest and driest season (see figure 4.11). On the contrary, the wettest season is autumn, with a secondary maximum in late spring (April-May). The coldest season is winter, whereas spring and autumn are milder and show some inertia from the precedent seasons. Temperatures along the year are always positive and tend to be cool in winter (max. ∼13°C min. ∼3°C), mild in spring-autumn (max. ∼17°C min. ∼10°C) and hot in summer (max. ∼27°C min. ∼16°C).

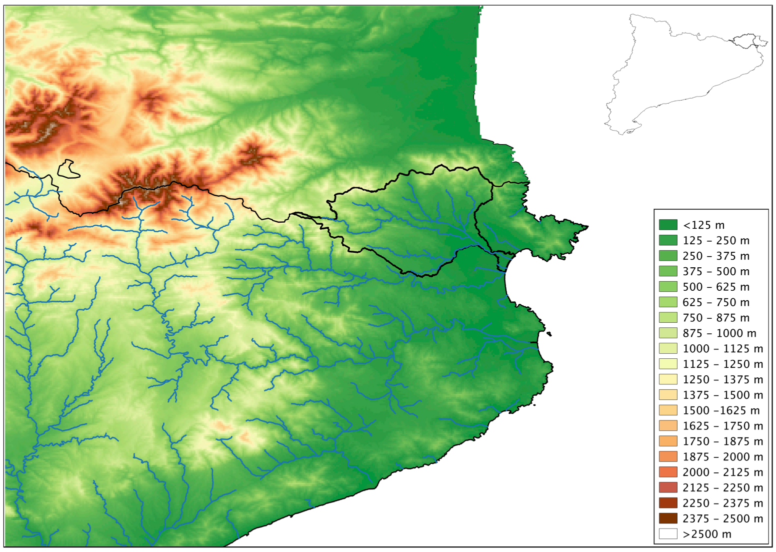


Figure 4.10. Muga basin's location and relief. Elaborated from GMTED2010 (Danielson and Gesch, 2011) and Spanish river cover from the *Ministerio de Agricultura, Alimentación y Medio Ambiente.*

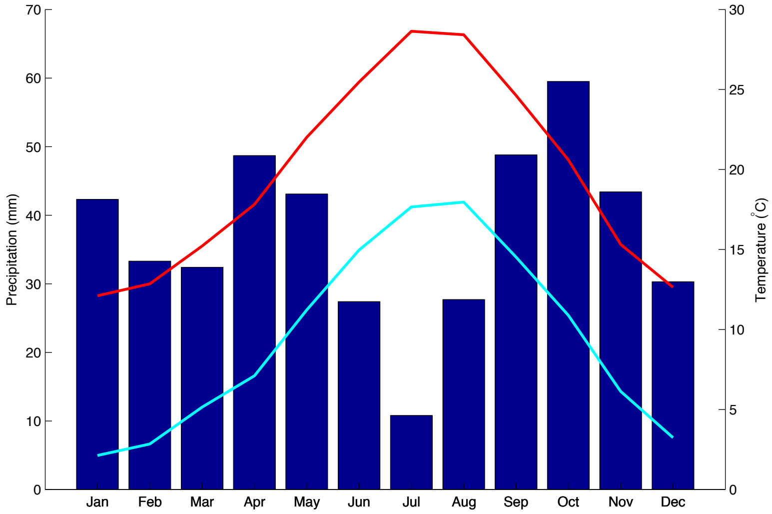


Figure 4.11. Muga basin's climogram. Elaborated from E-OBS.

4.5 Boadella reservoir

The Boadella reservoir, built between years 1959 and 1969, is located to the south of Darnius municipality, in Catalonia's Muga river basin. It collects water from a sub-basin of approximately 182 km^2 with summits ranging from the 1373 m of the Bassegoda mount to more than 1400 m in the Salines massif (see figures 4.10 and 4.12). The sub-basin is overgrown with vegetation: oaks in the low lands and chestnut and beech in high areas. The dense undergrowth also determines a high level of water retention. The reservoir accomplishes four objectives: *a*) flood lamination; *b*) irrigation; *c*) urban water supply; and *d*) electricity production. Its maximum length and depth are, respectively, 8.5 km and 54 m. The occupied area is 364 ha, and though the highest volume retained could be of 62 hm^3 for security reasons the maximum volume allowed is 61 hm^3 (Pavón, 2001a,b; Colomer *et al.*, 2004).

The reservoir is fed by the Muga river. Its hydrological regime is typical of the Mediterranean, with important fluctuations during the year. Winter and spring are the seasons of maximum contribution, whereas the annual minimum

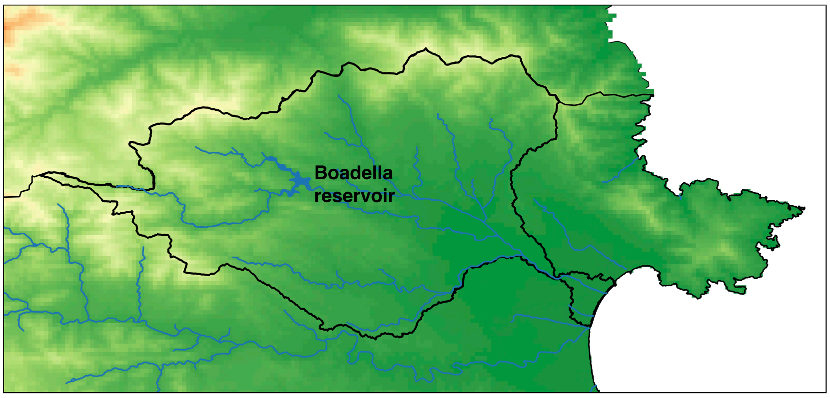


Figure 4.12. Boadella reservoir location in the Muga basin. Elaborated from GMTED2010 (Danielson and Gesch, 2011) and Spanish river cover from the *Ministerio de Agricultura, Alimentación y Medio Ambiente.*

coincides with the summer dry season. Figure 4.13 depicts this behaviour for the Muga river at the reservoir's entrance. In general the reservoir reaches its highest capacity in spring, after the rainfall and snow-melting periods. As for its minimum, it shows up in september, when the intensive irrigation after the dry season ends (see figure 4.14). Figure 4.15 shows the climatological distribution of the reservoir out-flow. On average, the water removed is distributed as follows (Colomer *et al.*, 2004):

a) 79% for electricity production and irrigation.

b) 4% for urban supply.

c) 5% for ecological flow.

d) 9% in peak flow periods or for other reasons.

The annual in-flow oscillates between 60 and 70 hm^3. However, the highly variable Mediterranean behaviour of Muga flow together with the intensive evaporation rates and water uses drive important fluctuations in the inter-annual water storage. Taking a closer look to the monthly series (see figure 4.16) we will not only see the annual cycle but also the alternation of high and low storage periods.

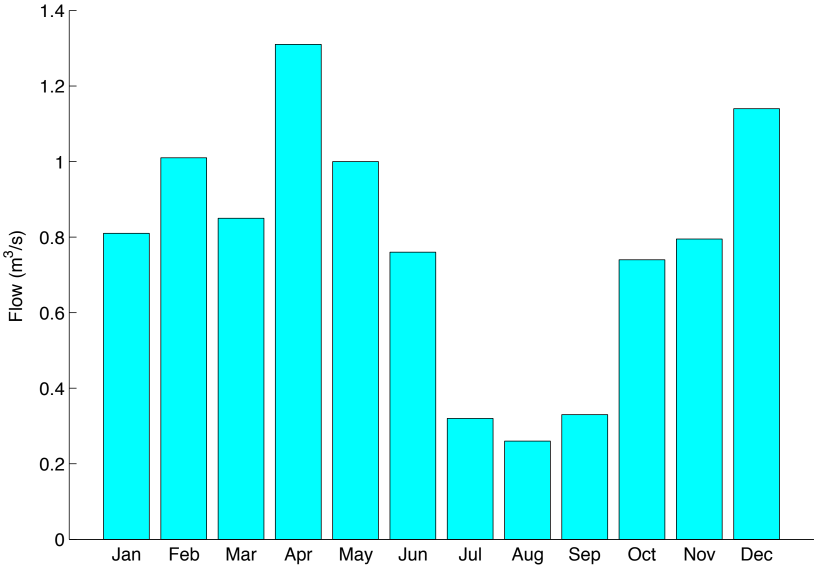


Figure 4.13. Climatology of the Muga river Boadella in-flow for the period 1981-2010. Note the summer minimum and the winter-spring maximum. Elaborated from ACA Boadella in-flow dataset.

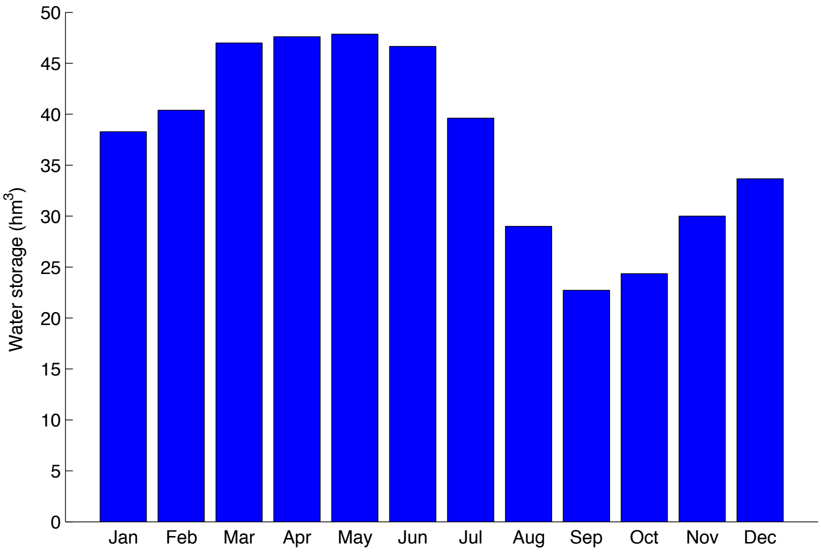


Figure 4.14. Climatology of the Boadella reservoir volume storage for the period 1981-2010. Note the maximum of late spring and the minimum of September. Elaborated from ACA volume Boadella dataset.

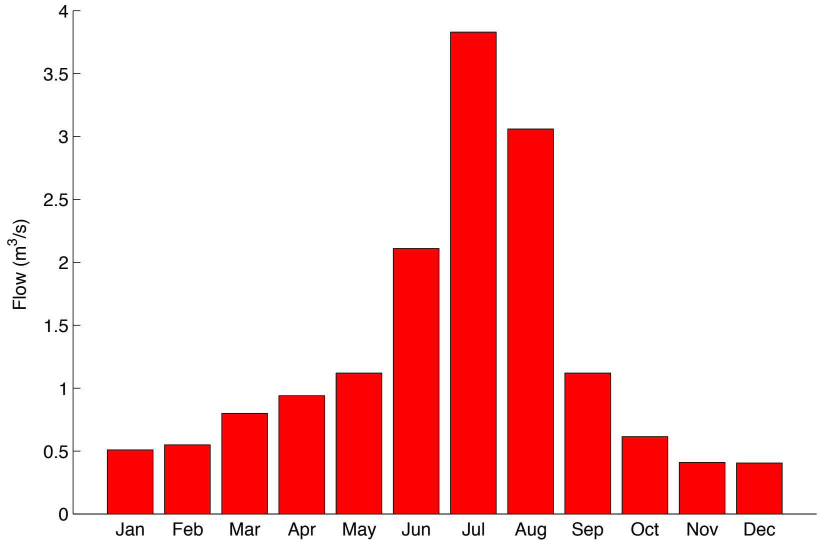


Figure 4.15. Climatology of Boadella's reservoir out-flow for the period 1981-2010. Note the spike in summer. Elaborated from ACA out-flow Boadella dataset.

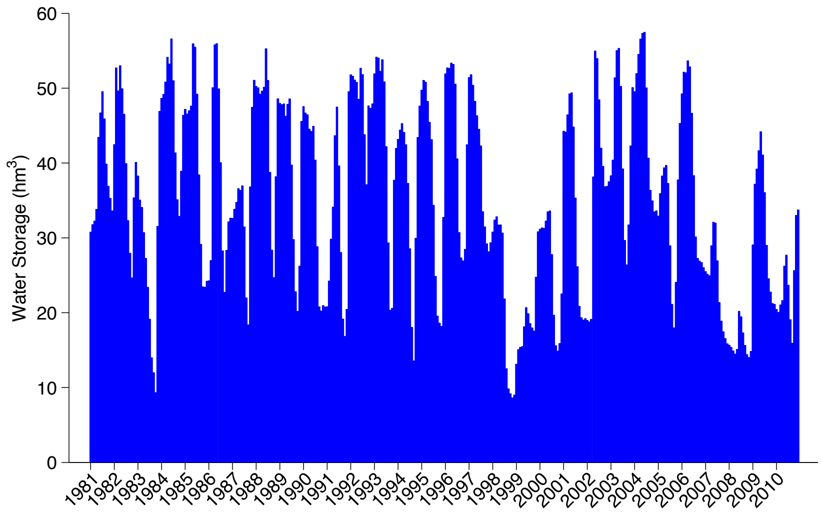


Figure 4.16. Boadella storage series for the period 1981-2010. Note the year cycle and the multiple stress periods. Elaborated from ACA volume Boadella dataset.

Figure 4.17. Boadella reservoir. Picture taken by Jordi Verdugo.

CHAPTER 5

Common methods

Nowadays we can find a myriad of computational tools that in a fairly straightforward way enable us to carry out complex statistical tests and post-processing techniques. This chapter is centred on the mathematical description of the common methods implemented in the calibration, verification and modelling strategies of this thesis. Further considerations and particular insights on the applied approaches are left to be specifically developed in chapters 6, 7 and 8.

5.1 Quantiles

Quantiles, *fractiles* or also, *percentiles*, are one of the more elementary methods to describe any kind of dataset. They are widely used in the characterization of meteorological and climatological variables due to their *resistance* to outliers (their value it is not susceptible to them) and their *robustness* (their value is independent of the underlying distribution). Each quantile, q_p, is a number q which represents the fraction of data in the dataset, p, that is smaller than this number. From this definition it is clear that to determine any quantile there is a need to arrange the members of the dataset in order, in what is known as the *order stastics* (Wilks, 2006).

The most frequent quantile to work with is the $q_{0.5}$ or the median. It can also be recognized as the 50th percentile and its familiarity comes from the fact that lies in the center of the dataset leaving the same number of elements above and below. There is a detail to be born in mind when obtaining the median from the *order statistics* since its computation varies a little if the number of elements in the dataset, n, is odd or even. In (5.1) we summarise how to work out this difference,

$$q_{0.5} = \begin{cases} x_{([n+1]/2)} & , \quad n \text{ odd} \\ \frac{x_{(n/2)}+x_{([n/2]+1)}}{2} & , \quad n \text{ even} \end{cases} \tag{5.1}$$

To sum up we can say that the median is the middle order statistic if the number of elements is odd, and the average of the two middle statistics if it is even. Other common quantiles are the *quartiles*, $q_{0.25}$ and $q_{0.75}$, because they are the basis to obtain the *Interquartile Range*, IQR.

$$IQR = q_{0.75} - q_{0.25} \tag{5.2}$$

Additional less frequent but also used quantiles are the *terciles*, $q_{0.333}$ and $q_{0.667}$; the *quintiles*, $q_{0.2}$, $q_{0.4}$, $q_{0.6}$ and $q_{0.8}$; and the *deciles*, $q_{0.1}$, $q_{0.2}$,... $q_{0.9}$. In this thesis we have only worked with the median, the IQR and terciles.

Quantile-Quantile plots

A quantile-quantile (Q-Q) plot is a scatter-plot to compare any pair of datasets in order to see whether they follow the same distribution. These datasets can range from direct observations to modelled (or extracted) from an analytical distribution. The mathematical function that corresponds to the distribution giving the more similar structure to the current scatter-plot is the *quantile function*. Two datasets following the same distribution would have their points on the $1:1$ diagonal line (see figure 5.1).

This conception is the basis for the so-called *Quantile-Mapping* methodology used to correct the modelled distribution with the observed *pdf* (see i.e. Déqué, 2007; Wilcke *et al.*, 2013). In this thesis it is applied in chapter 8 as a substitute to mean bias correction in the construction of SPI and SPEI series from the mixing of S4 and E-OBS datasets.

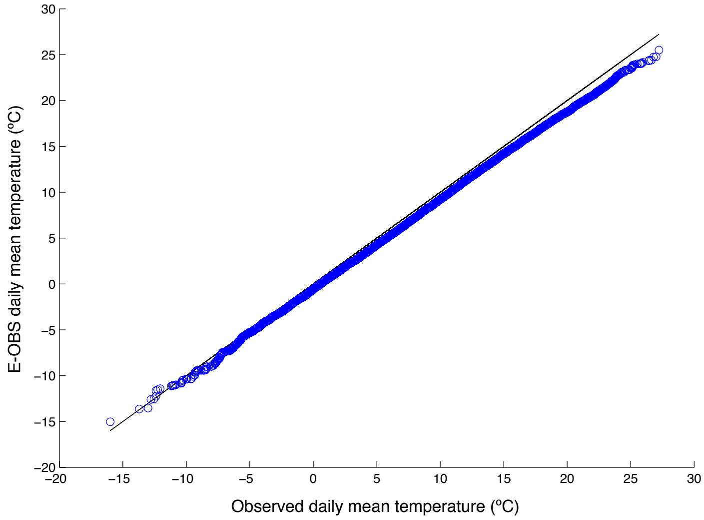


Figure 5.1. Quantile-Quantile plot example for Montbrun mean temperature (Tarn river basin, France). In this case both datasets approximately follow the same distribution (except for the highest values).

5.2 Bilinear interpolation

Bilinear interpolation can be qualitatively conceived as the process to recover an approximate value of a 2-variable function, $f(x,y)$ in a location of a 2D grid that does not belong to any of the regular points of the grid for which we already have the values of the function $f(x,y)$. To do so we begin by performing a linear interpolation in one direction that is followed by an identical interpolation in the remaining dimension. This concatenation of linear applications turns the method quadratic and, therefore, no linear. Mathematically, to perform a bilinear interpolation at an arbitrary point $s=(x,y)$ we need to know the values of the function $f(x,y)$ at the four nearest grid points: $S_{11}=(x_1,y_1)$, $S_{12}=(x_1,y_2)$, $S_{21}=(x_2,y_1)$ and $S_{22}=(x_2,y_2)$. Once this first condition is accomplished we can perform the first linear interpolation without regarding direction because, in the end, we obtain the same result. Hence, let us begin by the x-direction,

$$\begin{aligned} (x,y_1) &\approx \frac{x_2-x}{x_2-x_1}f(S_{11})+\frac{x-x_1}{x_2-x_1}f(S_{21}) \\ f(x,y_2) &\approx \frac{x_2-x}{x_2-x_1}f(S_{12})+\frac{x-x_1}{x_2-x_1}f(S_{22}) \end{aligned} \tag{5.3}$$

Then, it is time to perform the second interpolation along the y dimension,

$$f(x,y) \approx \frac{y_2 - y}{y_2 - y_1} f(x, y_1) + \frac{y - y_1}{y_2 - y_1} f(x, y_2) \tag{5.4}$$

Finally we only have to introduce (5.3) in (5.4) and compute the approximate value of the function at the desired point. We have used this methodology to increase the resolution of S4 forecasts from 0.75° to 0.25° in section 6.2.1 to match E-OBS resolution and grid-point position. Afterwards, this interpolated dataset has also been used in chapter 7.

5.3 Principal Component Analysis (PCA)

The PCA analysis is a mathematical technique that, in summary, allows us to build a minimum set of n independent dimensions from a N dimensional field, with the possibility to reconstruct the N field from these n dimensions at any moment (with $N > n$). Each member of the n dimensional field is a linear combination of the elements of the N field. The size of the n field is determined by the fraction of the variability of the original data that we want to retain and by the existent dependencies among the dimensions of the N domain. In practice this means that all the complex calculations can be applied on the reduced n domain and, if needed, recover the N field at the end of the process. This is really helpful in large dimensional fields ($N >> n$) with multiple dependent dimensions, as it is usually the case in atmospheric maps (Gutiérrez *et al.*, 2004).

We have applied this PCA in section 6.2.4 to ease the search for analogs in the MOS-analog calibration of S4 forecasts. Since this thesis only uses this technique as a tool, and PCA is an extensive topic, for a deeper mathematical insight the reader is referred to Wilks (2006) or Benestad *et al.* (2008).

5.4 Leave-one-out cross validation

The *leave-one-out cross validation* (LOOCV) is a particular application of the *cross validation* method (Michaelsen, 1987). It consists on leaving aside one member of the n dataset and fitting our model with the remaining $n-1$ samples.

The resulting model is then used to predict the value of the unused element and, therefore, to test its performance. This process is repeated n times and, as a result, we end having n different models that give an overall picture of the chosen predictor/s functioning (see figure 5.2).

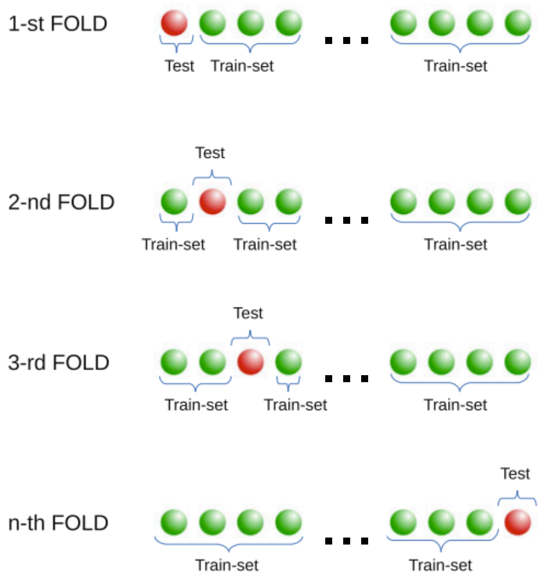


Figure 5.2. A schematic view of an out-of-sample test that follows a leave-one-out cross-validation approach. Iteratively, all the single samples from the original dataset are used as the test data with the remaining $n-1$ elements as training data. Adapted from (Turco *et al.*, 2012).

One important condition to be met before relying on its results is that there should be no underlying relationships among the members of the n dataset, for then this would lead to overconfidence when testing the performance of the model (DelSole and Shukla, 2009). Nevertheless, providing that our dataset accomplishes this condition and all of its members are independent of each other, we can shed light on the advantages on the use of this technique. The first one arises when we are dealing with a model approximation that has an open number of possible predictors, i.e. multiple-linear regression model, since we can use it to make visible the existence of over-fitting with our chosen predictor set. Besides, it puts the constructed model in an operational context, which can give us an idea of its true performance. Some authors such as Barnston *et al.* (1994) highlight that this approximation gives us a sub-estimation of precision because with

LOOCV we are not using all the data during the verification process. However, since the final model would be constructed with all the available data and the testing process is equal for all the predictor sets, its impact in the general outcome is rather limited.

One of its main drawbacks, however, is that it can be computationally expensive. Fortunately enough, our applications have been quite affordable and we have been able to apply it in all the verification approximations of chapters 6, 7 and 8.

5.4.1 Uncertainty estimation

In chapters 7 and 8 when applying the out-of-sample test we have also estimated the associated uncertainty to the outcome following the methodology proposed by Calmanti *et al.* (2007). The practical implementation of this method can be summarised as: *a*) computation of the residual variance for the study period; *b*) generation of 1000 white stochastic series with variance equal to that calculated previously; and *c*) addition of the stochastic series to the predicted model values, generating an ensemble of 1000 downscaled series.

5.5 Calibration

Calibration of model forecasts is an extensive topic spanning from the most straightforward strategies to the most mathematically demanding. In this thesis we have chosen three approaches: mean bias correction, MOS-analog and linear regression. In the following lines we will present a brief description of each technique along with the reasons for its selection.

5.5.1 MOS-Analog

This subsection introduces the MOS adaptation of the popular analog methodology, the MOS-analog technique. It is based on the concept of *analog* which was firstly introduced in Meteorology by Lorenz (1969a) and extensively applied ever since (see i.e. Duband, 1981; van Den Dool, 1989; Zorita and von Storch, 1999; Benestad *et al.*, 2008). Nowadays it is a popular and widely used technique in calibration and downscaling of weather and climate forecasts. It is built on the hypothesis that similar atmospheric patterns lead to similar atmospheric effects. This idea is the premise to set up a simple algorithm to search past situations that are similar to the one that is forecast. The main advantages of this method are: *a*) it is able to reproduce non-linear relationships between predictors and predictands; *b*) it is easy to implement with a low computational cost; and *c*) it is able to reproduce realistic and spatially coherent patterns.

Conversely, its main drawback is that it cannot simulate unobserved weather phenomena. Even so, it can produce accumulated values, means or frequencies larger (or smaller) than the historical ones if we consider the union of several of the analog independent outcomes. This limitation is related to the assumption of *stationarity* (Wilby *et al.*, 2004), a common weakness in many statistical methods and dynamical models, since the statistical relationships of the former and the parametrizations of the latter must hold in all the time series (Trenberth *et al.*, 2003). This limitation, that should be cautiously considered in climate change projections, it is less problematic when employed in calibration processes of current forecasts with known past datasets. Even so, this weakness can be mitigated by using a long and varied historical database (i.e. van den Dool, 1994; Zorita and von Storch, 1999; Diomede *et al.*, 2006; Barrera *et al.*, 2007) and with statistical relationships based on a small number of parameters and/or robust physical predictor-predictand connections (Benestad *et al.*, 2008; Maraun *et al.*, 2010). This is particularly the case of the MOS analog calibration, where the same variable acts as predictor and predictand. Hence, given an historical training period (with known predictors and predictands) and a test period (with known predictors), the MOS-analog calibration consists of three main steps (see also figure 6.2):

1. Selection of the study region.

2. For each uncalibrated forecast in the test set (acting as predictand) search for the nearest analog/s pattern in the training period (formed by the remaining forecasts of the same variable acting as predictors). These patterns are chosen considering the Euclidean distance between the fields (Matulla *et al.*, 2007).

3. The corresponding observation fields are then picked to compose the final calibrated forecast.

Finally, it is also worth noting that the selected analogs should correspond to similar periods of the year because otherwise the inherent variations in the annual cycle can lead to different predictand outcomes even if the predictor patterns were similar (Lorenz, 1969a). In this thesis we have applied this method in chapters 6 to calibrate raw S4 forecasts and in chapter 7, where the calibrated S4 forecasts have been used to analyse the performance a seasonal forecast model of reservoir in-flow, out-flow and volume anomalies.

5.5.2 Linear regression

The method of *linear regression* seeks the linear relationship between two variables, one acting as the *dependent* or *predictor* variable, y, and the other as the *independent* or *predictand* variable, x. This is achieved through the depiction of the line that minimizes the error between the values of y obtained from x in its corresponding *scatterplot*. The line is obtained through the expression (5.5),

$$\hat{y} = a + bx \tag{5.5}$$

The circumflex accent means that (5.5) determines the predicted value of y. The customary error criterion is the minimization of the sum of the squared errors. That is why this technique is also called *least-squares regression*. Unfortunately, the priority to minimize the distance between the points and the line means that this method is not resistant to outliers. The differences between the points and the line are called *residuals* and are computed as,

$$e_i = y_i - \hat{y}(x_i) \tag{5.6}$$

Where y and x refer to the actual values of data and the $\hat{y}$ to the predicted values through equation (5.5). Thus, if we combine equations (5.5) and (5.6) we get equation (5.7),

$$y_i = \hat{y}_i + e_i = a + bx_i + e_i \tag{5.7}$$

which states that the true value is the sum of the predicted value plus the residual. The expressions for the intercept of the line, a and its slope, b, are:

$$b = \frac{\sum_{i=1}^{n} [(x_i - \bar{x})(y_i - \bar{y})]}{\sum_{i=1}^{n} (x - \bar{x})^2}$$
$$a = \bar{y} - b\bar{x} \tag{5.8}$$

Where $\bar{x}$ and $\bar{y}$ are the respective mean values of the x and y datasets. For more information on the derivation of (5.8) equations see i.e. Wilks (2006). In the MOS approximation of linear regression the pair of x and y variables are, respectively, the forecasts and observations of the re-forecast period that will serve us to calibrate the line that will be used to correct the present forecast. We have used this procedure in chapter 6 to correct S4 forecasts.

5.5.3 Mean bias correction

Perhaps the simplest method available, this approach consists on seeking for the *mean error* in the model forecasts and adding (subracting) it from every single element of the dataset in order to have zero mean error after the calibration. In a linear regression, the independent term accomplishes the same function. Mathematically the mean error can be written as,

$$ME = \frac{1}{n}\sum_{k=1}^{n}(f_k - o_k) \tag{5.9}$$

Where f_k accounts for the forecast and o_k for the observation. This approach has been used in chapter 6 to correct the raw S4 forecasts and compare its performance with the more complex MOS-analog and linear regression calibrations. Afterwards, the resulting dataset has been applied in chapter 7.

5.6 Multiple Linear Regression modelling

The linear regression presented in the previous section 5.5.2 is a particular case of multiple linear regression (MLR). The main difference between them is that while in linear regression there is only one possible predictor, in MLR they can be several. In this case, however, our objective rather than calibrating a dataset is to model monthly in-flow, out-flow and volume anomalies in a reservoir (see chapter 7) and summer burned area anomalies in a region (see chapter 8). The most simple form of MLR is the one of equation (5.10).

$$Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{i,j} + \epsilon_i \tag{5.10}$$

Where for each element i, Y is the predictand; the $X_{i,1}, \ldots, X_{i,p}$ are the p predictors; ϵ_i is the residual term; and the $\beta_0, \ldots, \beta_p$ are the *regression parameters* which are found by minimizing the sum of the squared residuals (see i.e. Wilks, 2006).

The preference for this methods lies in the hypothesis that monthly anomalies of the aforementioned predictands can have approximate linear relationships with monthly anomalies of the meteorological predictors we work with, namely, maximum and minimum temperatures, precipitation and SPI/SPEI indices. Besides, this method is conceptually accessible and requires a limited number of computing resources, which means that the methodology might be transferred to other regions or situations with ease (see i.e. Huth, 2002; Benestad *et al.*, 2008). However, it is important to bear in mind that the open number of possible predictors can lead to over-fitting so their inclusion must be considered case by case with caution. To overcome this pitfall we have used the *Akaike Information Criterion*. Eventually though, to progress in the understanding of any problem, each predictor inclusion should respond to a physical connection with the predictand.

5.6.1 Akaike Information Criterion

The Akaike information criterion (AIC) is an information theory tool used in model selection (Akaike, 1974). For a given statistical model it can be computed from equation (5.11) as,

$$AIC = 2k - 2\ln(L) \tag{5.11}$$

where L is the maximized value of the likelihood function of the model; and k refers to the number of estimated parameters. Hence, for a set of different models calibrated with our dataset the preferred would be the one with the minimum AIC. The main advantage of this criterion is that it is a trade-off between the goodness-of-fit and the risk of over-fitting, since it includes a penalty for the number of included parameters (predictors). It is important to note that the AIC only accounts for the relative quality among the models, it tells nothing about the absolute quality of them, for it can happen that all the models fit badly and still find a minimum AIC. In our case it is used to aid the discrimination of the best models in chapters 7 and 8.

5.7 Forecast Verification

When facing any kind of forecast there is always a question to ask: How good is it? *Forecast verification* (or also *validation*) is the tool that lets us give an answer. It deals with the comparison of past forecasts with the respective observations in a range of ways. However, there are multiple aspects of a forecast that can be assessed and so are the possible responses to that question. The *Anscombe's quartet* is a good example of the importance of analysing data from a diversity of viewpoints (see figure 5.3).

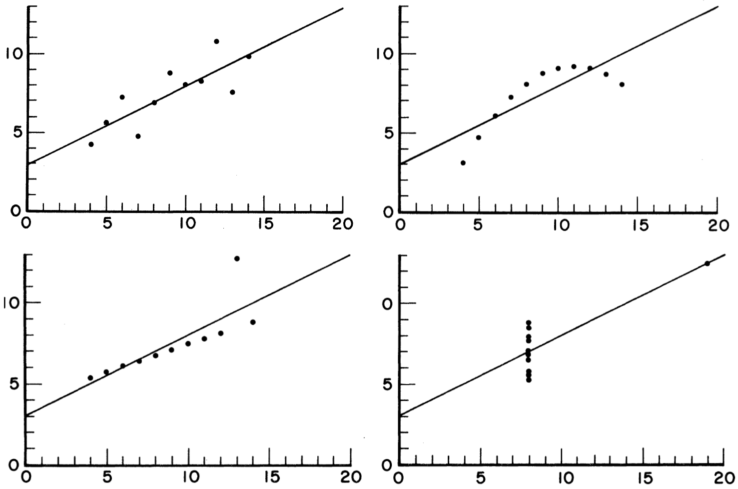


Figure 5.3. Anscombe's quartet. All the depicted scatterplots have the same Pearson correlation coefficient, mean and variance. The differences among datasets highlight the need to look at them from a diversity of approaches. From Anscombe (1973).

That is why, before issuing a verdict, there is always a need to look at a forecast from different perspectives taking into account not only the needs of the end-user but also the inherent complexities of the forecast itself (Doblas-Reyes *et al.*, 2008). Therefore, among all the possible methods to work with, we have to bear in mind that there is no *universal* metric that can account for the quality of all the aspects of a model (see i.e. Barnston *et al.*, 1994; Carson, 1998; Wilks, 2006). Hence, only a combination of them can provide us with a precise idea of its performance. Consequently, in the following lines we will present the metrics that in accordance to our objectives will help us to assess the behaviour of the models in the forthcoming chapters.

5.7.1 Validation from deterministic forecasts

5.7.1.1 Goodness-of-Fit

When assessing the goodness of fit of a regression model there are two common metrics in use. There is also a third one, the F ratio, but its application has some restrictions that are not usually met with the MLR and, therefore, it will not be considered (see i.e. Wilks, 2006). The other two are the *Mean Squared Error* (MSE) and the *Coefficient of Determination* (R^2). The MSE indicates the uncertainty around the regression line, providing us with an evaluation of the accuracy of the model. It can be computed from equation (5.12).

$$MSE = \frac{1}{n-k-1}(SST - SSR) = \frac{SSE}{n-k-1} \tag{5.12}$$

Where k refers to the number of regression parameters and n to the total number of compared elements. The SST, SSR and SSE expressions are summarised in (5.13).

$$\begin{gathered} SST = \sum_{i=1}^{n}(y_i - \bar{y}) \qquad SSR = \sum_{i=1}^{n}[\hat{y}(x_i) - \bar{y}] \\ SSE = \sum_{i=1}^{n} e_i^2 = SST - SSR \end{gathered} \tag{5.13}$$

SST is an acronym for the *sum of squares total* which stands for the sum of squared deviations of the predictand, y, around its mean value; SSR is the *regression sum of squares* this time accounting for the squared differences between the regression predicted values, $\hat{y}$ and the observed mean of the predictand, $\bar{y}$; finally, SSE accounts for the *sum of squared errors*, which is the sum of squared differences between the residuals (5.6), e_i, and their mean, $\bar{e} = 0$ (for further information see i.e. Wilks, 2006). In the perfect case, $MSE = 0$, SST equals SSR and the $SSE = 0$. The second usual measure to evaluate the fit of a regression is the *coefficient of determination*, or R^2. This can be computed from,

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{5.14}$$

The R^2 accounts for the fraction of the predictand variability (proportional to SST) that is reproduced in the regression SSR. In a perfect situation $R^2 = 1$; while in the worst, $R^2 = 0$. It is important to remark that R^2 value does not imply *causality* for it only checks the existence of a relationship between the predictand and predictors. In this work we have used the MSE and R^2 as discriminant factors in the choice of MLR models in chapter 7.

5.7.1.2 Pearson correlation coefficient

The *Pearson correlation coefficient* between two variables x and y is a tool to asses the linear relationship between them and can be applied to any pair of matching datasets, namely time-series or map fields. It can be regarded as the proportion of the covariance between the two variables to the product of the their respective standard deviations,

$$r_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i - \bar{x})(y_i - \bar{y})\right]}{\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^{1/2}\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}} \tag{5.15}$$

Where x_i and y_i refer to each element pair; $\bar{x}$ and $\bar{y}$ are the mean values; and n is the total number of elements (it is the same for both variables). One advantage of the Pearson correlation is that it is bounded: $-1 \leq r_{xy} \leq 1$. Thus, the best possible values are 1 or -1, depending on the sign of the linear relationship between x and y. When there is no kind of linear relationship between them, its value is 0. Another important property is that in linear regression its square value coincides with the *Coefficient of Determination* (5.14). Again, we have to remember that this does not imply *causality* between the two variables, just the existence of a linear relationship that can either be *casual* or *causal*. Finally, it is worth noting that Pearson correlation coefficient is neither robust nor resistant and so has to be used with caution. In our work we have used it in the verifications sections of chapters 6, 7 and 8.

5.7.1.3 MAE, MSE and RMSE

In accordance to (Wilks, 2006) there are three well-established accuracy measures in use. All of them can be applied upon time-series or field maps. The first one is the *mean absolute error* (MAE),

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|f_k - o_k| \tag{5.16}$$

Where f_k and o_k refer to each of the forecast-observation pairs in the series. As it can be inferred from equation (5.16) the outcome of a perfect model would equal the observations and, consequently, $MAE = 0$. The second accuracy measure, the *mean squared error* (MSE) has been already introduced in section 5.7.1.1. We reproduce here its expression for completeness.

$$MSE = \frac{1}{2}\sum_{m=1}^{M}(f_m - o_m)^2 \tag{5.17}$$

Clearly, for a perfect forecasts $MSE = 0$. Often, the MSE is transformed into the third accuracy measure, the *root mean squared error* ($RMSE$) by means of a squared root (5.18). This is done to retain the units of the variable and, thus, ease its comprehension as an error magnitude.

$$RMSE = \sqrt{MSE} \tag{5.18}$$

The $MSE/RMSE$, being squared averages of differences tend to penalise most the bigger errors than the MAE. We have extensively used them in chapters 6 and 7.

5.7.1.4 Taylor diagram

The Taylor diagram (see Taylor, 2001) is a graphical device to depict in a single figure the centred RMSE (cRMSE), the standard deviation and the Pearson correlation coefficient. The main difference of the RMSE and the cRMSE is that the latter accounts for the RMSE of the anomalies instead of the direct difference between forecasts and observations. It is based on the fact that the centred RMSE can be written in the form of equation (5.19).

$$cRMSE = \sqrt{\sigma_f^2 + \sigma_o^2 - 2\sigma_f\sigma_o r_{fo}} \tag{5.19}$$

Where σ_o and σ_f correspond, respectively, to the standard deviations of observations and forecasts; and r_{fo} is the Pearson correlation coefficient between forecast and observed values. (5.19) is directly analogous to the *law of cosines.*

$$c^2 = a^2 + b^2 - 2ab\cos\theta \tag{5.20}$$

If we then associate each of the elements of equations (5.19) and (5.20) we can describe a triangle in which two of the legs correspond to the standard deviations σ_f and σ_o and the third, to cRMSE. Besides, if we now focus our attention to the angle, θ, between σ_f and σ_o we will see that its cosine is the correlation r_{fo}. This construction can be directly described with polar coordinates: the forecast field is plotted at the farthest part of the radius σ_f at an angle $\cos^{-1}(r_{fo})$. On the other hand the distances to the reference point indicate the cRMSE (see figure 5.4). We have used Taylor diagrams in the forecast verification sections of chapters 6 and 7.

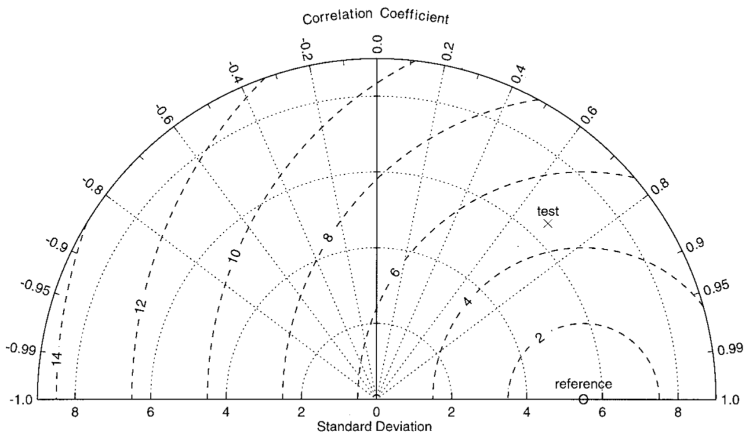


Figure 5.4. General structure of a Taylor diagram. From Taylor (2001).

5.7.2 Validation from probabilistic forecasts

5.7.2.1 Contingency tables

In many occasions we need to collapse probabilistic forecasts to non-probabilistic yes/no predictions by means of threshold considerations. In fact, some verification techniques have to screen all of these thresholds to obtain the necessary information to build graphical displays such as the *ROC* or the *Economic Value* curves (see subsections 5.7.2.6 and 5.7.2.7).

For each case we can build what is commonly known as a 2×2 *contingency table* which is a useful summary tool for verification of non-probabilistic yes/no forecasts. It is also the basis for computing multiple *quality* scores, some of them fundamental in the aforementioned verification constructions (Wilks, 2006). Figure 5.5 depicts the general structure of one of such tables. We can see that the event occurrence was successfully forecast a times out of n total forecasts, that is, the analysed model had a *hits*. Conversely, this model forecast the event on b instances in which it was not finally observed, namely, the model gave b *false alarms*. Similarly, there are c *misses* in which the model did not forecast the event when it actually occurred. Finally, the d term accounts for the *correct rejections* or situations in which the model did not forecast the event and it did not happen. Outside the table it is usual to include the *marginal totals* and the sample size, $n = a + b + c + d$.

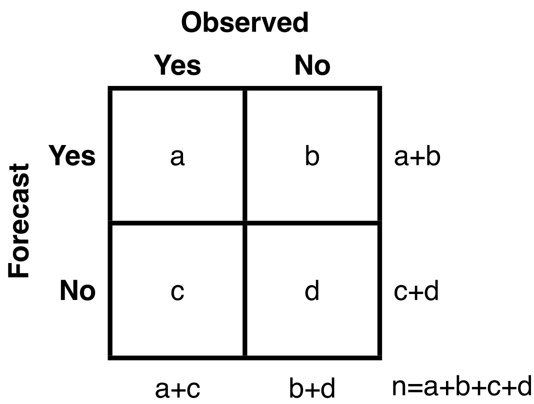


Figure 5.5. Relationship between counts (letter a-d) of forecast/event pairs for the dichotomous noprobabilistic verificaiton situation as displayed in a 2×2. Also shown are the marginal totals, indicating how often each of the two events was forecasted and observed in absolute terms and the sample size n. Adapted from (Wilks, 2006).

Two scores, namely the *False Alarm Rate* (F) and the*Hit Rate* (H), are widely employed to characterize a dichotomous forecast system that can be described by a 2×2 contingency table. F can be computed as,

$$F = \frac{b}{b + d} \tag{5.21}$$

That is, F is the ratio of *false alarms* with respect to the total number of the event non-occurrences. The best possible value of F is zero and the worst is one. As for the H, it is defined as,

$$H = \frac{a}{a + c} \tag{5.22}$$

Which is the fraction of *hits* with reference to the number of event occurrences. It is positively oriented being 1 its best value and 0, the worst. Equations (5.21) and (5.22) are the conceptual and geometrical basis for ROC and EV curves of chapters 7 and 8.

5.7.2.2 Verification Rank Histogram

The *verification rank histogram* is a graphical device to assess whether an ensemble includes the possible observations as equiprobable forecast members of the ensemble. If we have an ensemble of n_{ens} members it is build as follows. Consider n possible forecast situations. For each one of them we will have n_{ens} forecasts and the final observation. If both the n_{ens} and the observation come from the same distribution the rank of the observation would be any of $n_{ens} + 1$ values, that is, $i = 1, 2, 3, \ldots, n_{ens} + 1$. If, for example, the observation is larger than any other of the ensemble, the assigned rank would be $n_{ens} + 1$; conversely, if it is smaller than any other ensemble value, the assigned rank would be 1. If we annotate the ranks of the n corresponding observations and plot them in the form of a histogram we will end having the *verification rank histogram*. In figure 5.6 we can see what a normal *verification rank histogram* looks like and the possible departures from normality. In our case it has been applied as a preliminary analysis in chapter 6 to verify that a 15-member ensemble produces and under-dispersed verification rank histogram (see figure 5.6).

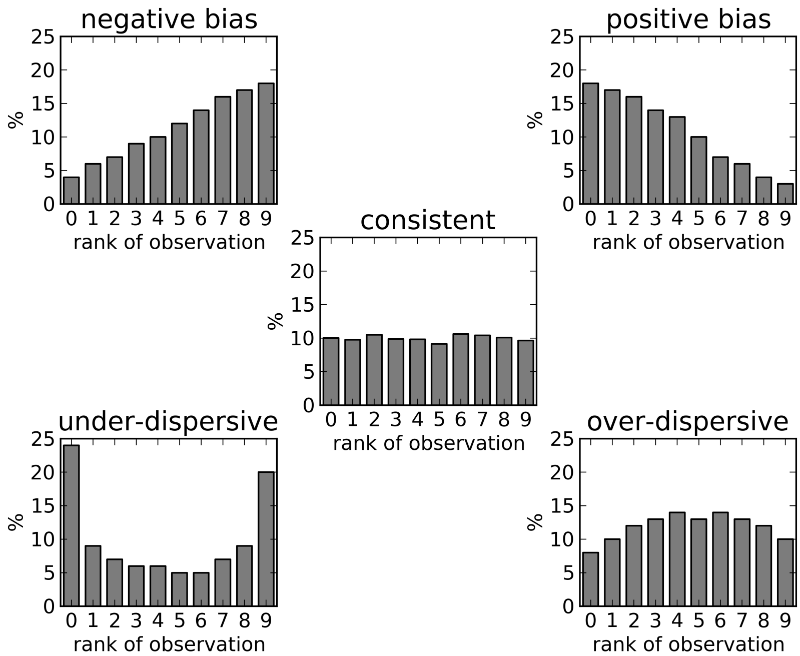


Figure 5.6. General form of a verification rank histogram and its departures from uniformity. From (Callado *et al.*, 2013).

5.7.2.3 Brier Score

The *Brier Score* (BS) is, basically, the mean squared difference between the forecast probability of a given event, f_k, and the occurrence, $o_k = 1$, or not, $o_k = 0$, of the event. It is a scalar accuracy measure for probabilistic forecasts.

$$BS = \frac{1}{n}\sum_{k=1}^{n}(f_k - o_k)^2 \tag{5.23}$$

The index k refers to each pair of the n forecast-observation events. It can take values between 0 and 1, $0 \leq BS \leq 1$, being 0 the BS of a perfect forecast. Additionally, if we would like to compare the BS of a given model with a reference forecast (climatology in our case) we can use the *Brier Skill Score* (BSS) that is computed as follows,

$$BSS = \frac{BS - BS_{cl}}{BS_{prf} - BS_{cl}} \tag{5.24}$$

Since the BS of a perfect forecast equals 0, $BS_{prf} = 0$, equation (5.24) yelds,

$$BSS = 1 - \frac{BS}{BS_{cl}} \tag{5.25}$$

When the probabilistic forecast comes from a relatively small ensemble ($n_{ens} < 40$) there is an intrinsic negative bias in the BSS (Weigel *et al.*, 2007). This is corrected in the so-called *diescrete Brier Skill Score* which can be obtained from equations 5.26 and 5.27,

$$BSS_D = 1 - \frac{BS}{BS_{cl} + D} \tag{5.26}$$

$$D = \frac{1}{M} p\,(1-p) \tag{5.27}$$

Where M is the size of the ensemble and p the climatological probability of the event. In this thesis we have used the *discrete Brier Skill Score* referenced to *climatology* in chapters 6 and 7.

5.7.2.4 Attributes diagram

The *attributes diagram* is a reliability diagram that includes the components of a singular decomposition of the Brier Score (see i.e. Wilks, 2006),

$$BS = \underbrace{\frac{1}{n}\sum_{i=1}^{l} N_i (f_i - \bar{o}_i)^2}_{\text{Reliability}} - \underbrace{\frac{1}{n}\sum_{i=1}^{l} N_i (\bar{o}_i - \bar{o})^2}_{\text{Resolution}} + \underbrace{\bar{o}\,(1-\bar{o})}_{\text{Uncertainty}} \tag{5.28}$$

Where I refers to the total number of probability categories (e.g. 5 categories would mean that the probabilities can only take one of these five values: 0%, 25%, 50%, 75%, 100%); N_i is the number of occasions each probability category has been forecast; n is the total number of forecast-observed pairs; $\bar{o} = \frac{1}{n}\sum_{k=1}^{n} o_k$ is the overall occurrence frequency of the event observed; and $\bar{o}_i = \frac{1}{N_i}\sum_{k \in N_i} o_k$ is the relative frequency of the event occurrences with respect to the ith probability forecast category. It is worth to remember that if the event occurs $o_k = 1$ and 0, otherwise. In figure 5.7 we can see an example of one of such diagrams. The diagonal line corresponds to a *perfect reliability*, and is the line that the outcome of a perfect forecast would follow. The horizontal dashed lines refer to the resolution term in equation (5.28) and represent the limit of *no-resolution*. This limit is established by the climatological probability of the forecast event. On these lines the model is unable to distinguish if the event occurrence is more or less probable than its climatological probability (Wilks, 2006). The shaded region is determined by the line of no-skill. It is halfway the lines of perfect-reliability and no-resolution and defines the forecast elements that contribute positively in the overall skill in 5.28. It is important to note that lying outside the shaded region does not imply that our forecast is useless (see i.e. Weisheimer and Palmer, 2014, and the introduction to section 5.7)).

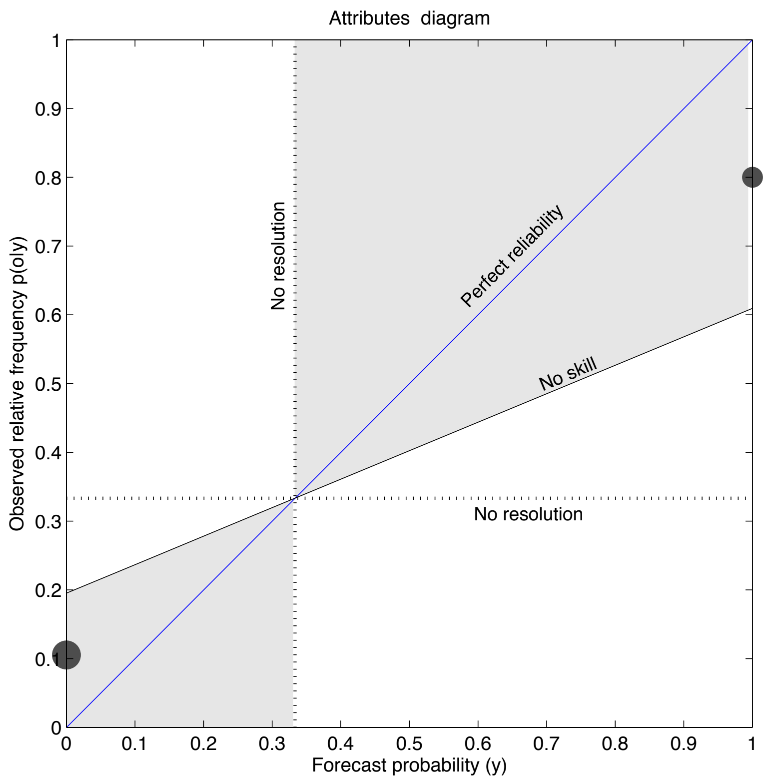


Figure 5.7. Example of an attribute diagram. Elaborated with data from chapter 7.

5.7.2.5 Ranked Probability Score

The *Ranked Probability Score* (RPS) is a tool to assess the performance of multi-categorical probabilistic forecast systems. These systems issue, for every forecast,

J probability values corresponding to each of the J pre-established categories where the final observation could be. It is computed as the squared difference between the forecast, F, and observed, O, cumulative vectors. So for each forecast-observation pair we have,

$$RPS_i = \sum_{m=1}^{J} (F_m - O_m)^2 \tag{5.29}$$

Where F and O are computed as,

$$\begin{aligned} F_m &= \sum_{i=1}^{m} f_i \,, \qquad m = 1, \ldots, J \\ O_m &= \sum_{i=1}^{m} o_i \,, \qquad m = 1, \ldots, J \end{aligned} \tag{5.30}$$

J refers to the number of established categories and, consequently, to the F and O vector dimensions; f_i is the forecast probability; and o_i corresponds to the occurrence, $o_i = 1$, or not, $o_k = 0$, of the event. For instance, the cumulative vectors for F would be, $F_1 = f_1$, $F_2 = f_1 + f_2$, $F_3 = f_1 + f_2 + f_3$,. . .,$F_J = f_1 + f_2 + f_3 + \ldots + f_J$. It is clear that $F_J = 1$ and $O_J = 1$ since their components represent the probabilities of a fixed set of categories. If we combine equations (5.29) and (5.30) we will get,

$$RPS_i = \sum_{m=1}^{J} \left[\left(\sum_{j=1}^{m} f_i \right) - \left(\sum_{j=1}^{m} o_j \right) \right]^2 \tag{5.31}$$

The RPS has a negative orientation, the best possible value is $RPS = 0$. Its extension to n forecasts is direct,

$$\overline{RPS} = \frac{1}{n} \sum_{k=1}^{n} RPS_k \tag{5.32}$$

Finally, we can compute the *Ranked Probability Skill Score* with respect to climatology as,

$$SS_{RPS} = \frac{\overline{RPS} - \overline{RPS_{clim}}}{0 - \overline{RPS_{clim}}} = 1 - \frac{\overline{RPS}}{\overline{RPS_{clim}}} \tag{5.33}$$

For situations in which the ensemble size is smaller than 40 members the RPSS suffers from a negative bias (Weigel *et al.*, 2007). The way to proceed is the same as in section 5.7.2.3,

$$RPSS_D = 1 - \frac{RPS}{RPS_{cl} + D} \tag{5.34}$$

$$D = \frac{1}{M}\frac{J^2 - 1}{6J} \tag{5.35}$$

Where M is the size of the ensemble and J is the number of forecast categories. In this thesis we have used the *discrete Ranked Probability Skill Score* referenced to *climatology* in chapters 6 and 7.

5.7.2.6 ROC curve

The *ROC curve* is an acronym for *Relative Operating Characteristic curve* (see i.e. Gutiérrez *et al.*, 2004; Wilks, 2006) that we have applied in the verification sections of chapters 7 and 8. It is built by screening the H and F performance of a forecast system (see section5.7.2.1) through all the probability threshold range, that is, $0 \leq p \leq 1$. Afterwards, these values are depicted in a diagram where the H is represented in the y-axis and the F, in the x-axis. In both cases the axis' limits are 0 and 1.

From these considerations we can infer that in the lowest threshold, $p = 0$, the H and F are maximum because our system always forecasts the occurrence of the event. On the other hand, in the upper threshold, $p = 1$, our system will never predict the event, and the H and F are minimum. These points constitute, respectively, the ending and beginning of every ROC curve (see i.e. figure 5.8). The diagonal line that can be observed in this figure corresponds to the ROC curve generated by an aleatory forecast system based on the climatology of the event (see i.e. Wilks, 2006). Therefore, the more the curve of a forecast system falls to the left of the climatology line, the better the system is. This can be quantified through the so-called *ROC Area* which is the area under the ROC curve. Considering that the climatological forecast system would give a ROC area of 0.50, values over this threshold will generally represent better forecast systems. Nevertheless, in every case it is also necessary to look at the ROC curve because for each threshold the system may still be falling to the right or left of the diagonal line.

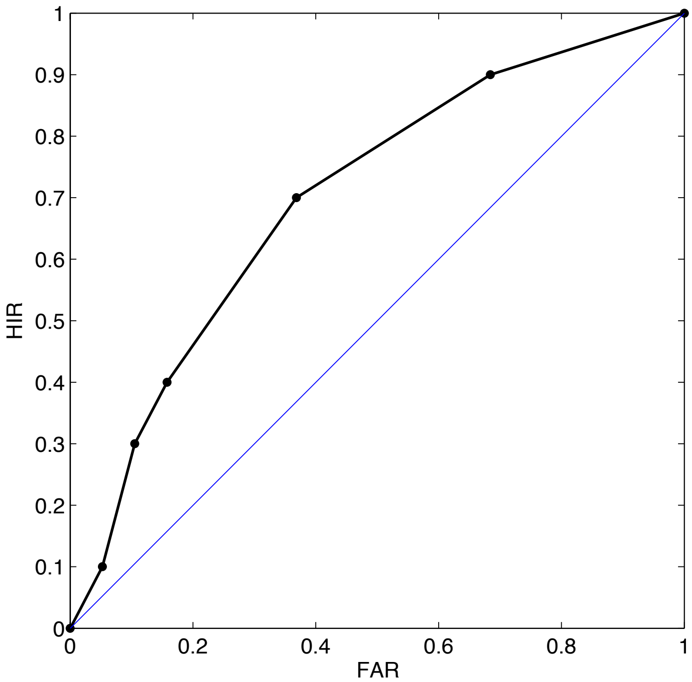


Figure 5.8. Example of a ROC curve. Elaborated with data from chapter 7.

5.7.2.7 Economic Value

The *Economic value*, EV, is a verification parameter for probabilistic forecasts of dichotomous events to summarise their quality from a cost/loss perspective (Gutiérrez *et al.*, 2004). Each EV can be obtained from a contingency table (see figure 5.9) taking into account that any preventive action has a cost, C, whereas taking no action has the risk of having a loss, L. From figure 5.9 the total expense of the probabilistic system can be computed as,

$$TE = \alpha C + \beta C + \gamma L \tag{5.36}$$

Where α, β, γ and δ refer to the relative frequencies obtained by dividing the elements a, b, c and d of figure 5.5 by the sample size, n. If we now consider (5.21) and (5.22) and re-write (5.36) in function of H and F yields,

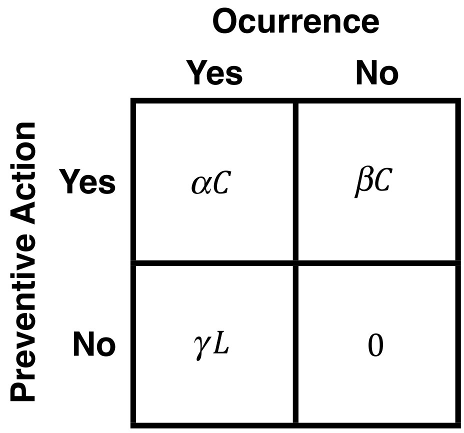


Figure 5.9. Contingency table for a dichotomous cost/loss approximation. Adapted from (Gutiérrez *et al.*, 2004).

$$TE = H\ p_c\ C + F\,(1 - p_c)\,C + (1 - H)\,p_c\ L \tag{5.37}$$

Where p_c refers to the climatological probability of the event occurrence,

$$p_c = \alpha + \gamma = \frac{a + c}{n} \tag{5.38}$$

Let us now introduce the total expense of the *perfect* and *climatological* forecast systems. In a perfect prediction $H = 1$ and $F = 0$ and thus,

$$TE_{perf} = p_c\ C \tag{5.39}$$

Moving on to the climatological forecast, it would have a global expense of,

$$TE_{\text{ clim}} = \min\{C, p_c C\} \tag{5.40}$$

Because the expense generated by the preventive action should never surpass the limit $L\ p_c$. If we now take into account all the aforementioned developments, the EV of a prediction can be conceived as the difference between what we could get with our model with respect to the climatology and what we could obtain with a perfect model,

$$EV = \frac{TE - TE_{\text{clim}}}{TE_{perf} - TE_{\text{clim}}} \tag{5.41}$$

Now if we insert the ratio $R = C/L$ in (5.41) we will have the expression of the EV with respect to the cost/loss ratio and the H and F.

$$EV = \frac{H\ p_c\ R + F\,(1 - p_c)\,R + (1 - H)p_c - \min\{R, p_c\}}{p_c R - \min\{R, p_c\}} \tag{5.42}$$

Since H and F depend on the probability threshold, p_t, that converts our probabilistic forecast to a dichotomous prediction (Wilks, 2006), we have that the EV is both function of p_t and the cost/loss ratio, R. Consequently, for every threshold, p_t, we will have a function $EV = f(R)$. Additionally, since these functions are continuous, for every value of R it will exist a p_t that maximizes the EV of the system. Particularly, this is reached when $R = p_c$ for each p_t. In that situation (5.42) becomes,

$$EV = \frac{H\ p_c^2 + F\,(1 - p_c)\,p_c + (1 - H)p_c - p_c}{p_c^2 - p_c} \tag{5.43}$$

Which simplified, is equal to the *Hanssen-Kuipers Score* or HKS,

$$EV = H - F = HKS \tag{5.44}$$

Often, since every p_t has its own $EV(R)$ function, the final diagram only depicts the envelope of these maximum EV values (see figure 5.10). In this thesis we have applied the EV in chapter 7.

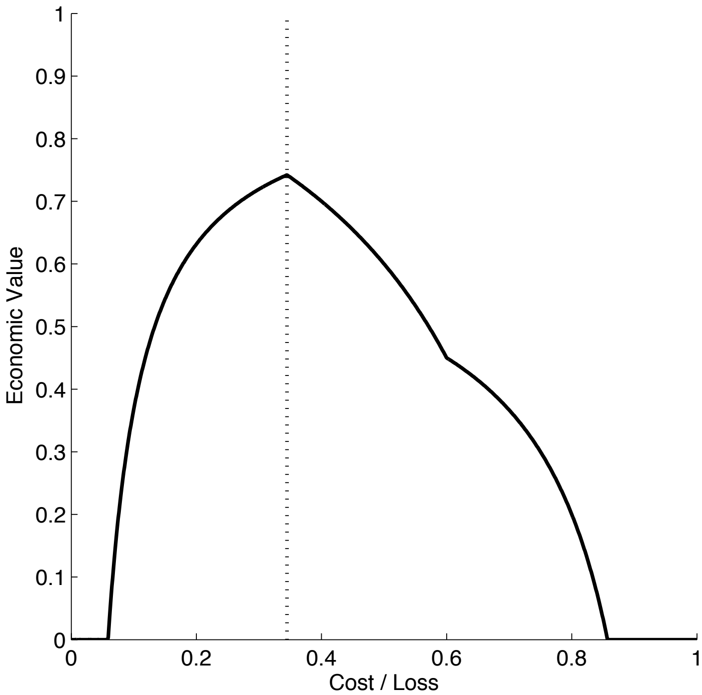


Figure 5.10. Example of an economic value envelope curve. Usually, to avoid cluttering, economic value plots only depict the envelope curve of the maximum economic value for each Cost/Loss ratio and probabilistic threshold. Elaborated with data from chapter 7.

CHAPTER 6

Seasonal forecasting: calibration for improvement

6.1 Overview

The following lines present an effort to explore seasonal predictability and its possible improvement by means of calibration of S4. Particularly, we will focus our attention on a multiple-step approach, ranging from continental to regional domains, in a process through four areas (see chapter 4):

i.	Europe	**iii.**	Catalonia
ii.	Spain	**iv.**	Muga river basin

The reason behind this choice is to provide a thorough comparison of S4 performance before/after calibration, not only to test the former hypothesis but to cover a broader spectrum of end-user needs. Likewise, our objective is also to analyse the effect of domain size on the performance of our hypothesis and methods. In this regard, the results obtained will be applied in chapters 7 and 8 to show the potential benefits of seasonal forecasting in the fields of water resources and summer fire predictability. But before, we will have to answer some questions:

1. Which is the current skill of S4 in the aforementioned domains? Does it depend on them?

2. Can S4 forecasts be improved by linear/non-linear calibrations? At which scales? At what lead-times?

3. Are our calibrations able to improve the results of a simple-bias correction? At what cost?

Each response will pave the way to the contextualization of our framework as well as to the establishment of prospects for the upcoming applications. We will see, for example, that the answer to the first question provides us with a robust starting point, a picture of the model behaviour depending on the area studied, as well as its performance at different lead-times and for different variables. The second issue, on the other hand, leads us to investigate the possibility of adding information to the forecasts seeking for hidden signals through calibration. Finally, the last inquiry will reveal if the complexity added by calibration is justified or not.

Consequently, the success of this chapter also revolves around the validation process. It involves deterministic and probabilistic verification metrics that will be matched with climatological and persistence controls to test overall S4 performance (original and calibrated). This will consist of an out-of-sample verification of the issued S4 forecasts against the observational dataset, E-OBS. In both cases the variables considered are: *a*) maximum temperature; *b*) minimum temperature; and *c*) precipitation amount. More details on these datasets can be found in sections 3.1 and 3.2. Hence, the organization of this chapter is as follows: the *Methodology* section presents and justifies the selection of different calibration and data-treatment strategies; then, *Results* sums up the outcome of the analysis performed; and finally, in *Discussion* & *Conclusions* there are the answers to the questions previously raised.

6.2 Methodology

6.2.1 Region selection and interpolation

In section 3.2 we have seen that the S4 *hindcast* (1981-2010) is composed of a 15-member ensemble of 1 to 7 monthly lead-time predictions issued every month in a global grid resolution of 0.75 degrees. In our case, after extracting European forecasts, the next step has been to interpolate its resolution from 0.75 to 0.25 degrees through an interpolation method. This is meant to ease the verification process by matching S4 and E-OBS resolutions. As for the interpolation itself, it has been based on the application of a bilinear approach (see section 5.2) because its mathematical simplicity ensures that changes on the original information (Accadia *et al.*, 2003) would be minimized through the calibration approach.

Regarding the smaller domains, they have been drawn from the European region extraction. For Spain and Catalonia it has been quite trivial, since it is enough to superimpose the region shape on the Europe dataset and identify the grid-points that fall within. However, for a smaller area such the Muga river basin it can be the case that no grid-point falls inside the polygon. Consequently, we had to develop a protocol to select the N closest points to a given region (see figure 6.1),

i. Take the coordinates of the region that define its polygon and compute the median values of its latitudes and longitudes.

ii. Enlarge the region's polygon by adding or subtracting 0.1° to every pair of shape coordinates (latitude and longitude) depending on whether they are above or below the aforementioned median values. Repeat until the enlarged polygon encompasses the minimum required N points.

iii. As a result the region shape expands isotropically until it reaches the minimum of N inside points. These points will always be the closest to the original region.

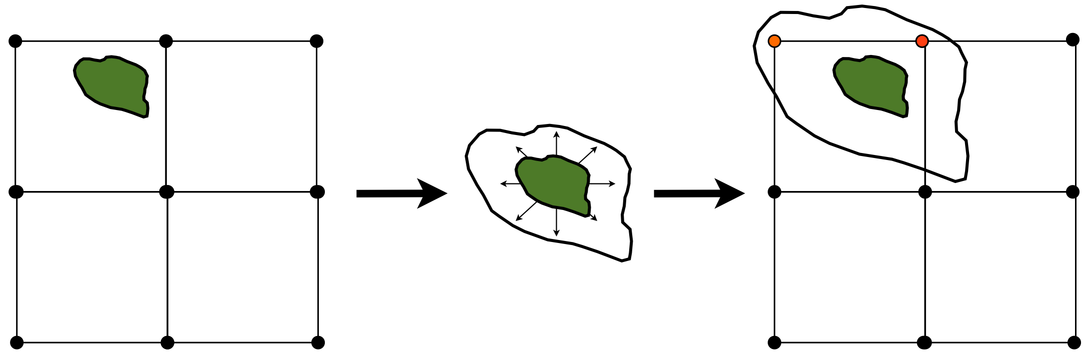


Figure 6.1. Small region expansion-selection scheme. The colored grid-points falling inside the expanded shape are the two closest.

6.2.2 S4 ensemble recombination

The analysis of the ensemble through the verification rank histogram has shown that with 15 members we get an under-dispersed histogram, with its characteristic U-shape (see section 5.7.2.2). Although this is something that it is inherent to our hindcast it is important to bear it in mind when analysing the results obtained. Even so, the available computing resources have been not enough to work independently with each one of the 15 members of the ensemble in the deterministic approximation.

To overcome this situation we have restructured the re-forecast ensemble through what we have called the *generalized ensemble* (GE). It is, plainly, a reorganization of the information of the ensemble in a smaller number of maps constructed with values corresponding to a given percentile. In this way we could maintain the characterization of the *pdf* of the ensemble but without the need to work with its every single member when going deterministically. To our knowledge, this approximation in literature is mainly limited to the work with mean/median maps or global values (i.e. Doblas-Reyes *et al.*, 2005; Johnson and Bowler, 2009). Here, we also retain the other percentiles to develop a deterministic approach (see next section) and visualize the variability of the *pdf*. The process to build the GE is as follows:

1. For each grid-point and lead-time retrieve the values corresponding to all the forecasts of the ensemble for that grid-point. In our case,

$$z_j = \{x_1, \ldots, x_{15}\} \tag{6.1}$$

2. Set percentiles of interest to define the general distribution. To cover the IQR and a lower and upper region of the distribution we have chosen,

$$z_j = \{p10, p25, p50, p75, p90\} \tag{6.2}$$

3. For each grid-point, draw the values corresponding to those pre-set percentiles to build a new ensemble matching the percentile thresholds defined. In our case,

$$z_j = \{x_{p10}, x_{p25}, x_{p50}, x_{p75}, x_{p90}\} \tag{6.3}$$

The choice of the p10 and p90 instead of the more common p95 or p99.9 to represent the extremes pursues the reduction of the induced error coming both from the size of the ensemble and the coarse E-OBS resolution. That said, p10 and p90 are the characteristic measures of moderate climate extremes (i.e. Moberg and Jones, 2005). As for the resulting GE, it has some advantages beyond the obvious reduction of hindcast size. Firstly, it represents the *pdf* of the original ensemble as a whole no matter if this corresponds to a single grid-point or a region field. In the latter this avoids the underlying complexity of dealing with the forecast variability (standard deviation), σ_{fc}, as combination of random variables with possible entangled dependencies,

$$\sigma_{fc} \propto \{\sigma_t, \sigma_{spc}, \sigma_{ens}\} \tag{6.4}$$

Where σ_t represents the variability linked to time; σ_{spc}, the variability linked to the spatial distribution; and σ_{ens}, the variability linked to the original ensemble members. Thus it also provides information about deficiencies in the original ensemble such as limited variability and the existence of systematic biases. Moreover, statistical downscaling and calibration methods can act directly on the general *pdf* independently of the nature of the underlying distributions and the dependence or independence of the random variables considered. Finally, spatially organized differences in any of the members of the generalized ensemble can give information about the general *pdf*. For example, imagine that we look at a map corresponding to the lowest 10th percentile of lead 1 forecast monthly precipitation anomaly for Europe and we observe an organized structure with positive values in the south and negatives in the north. This means the model ensemble expects a wetter than normal month for the south. It is worth noting that we can say nothing about the north because the drier than normal 10p is what we should expect for this percentile's behaviour.

6.2.3 S4 ensemble unification

Current models usually comprise the association of deterministic control output along with an ensemble of perturbed runs. The former is used to embody a deterministic forecast whereas the latter is the base for probabilistic predictions. However, one may ask whether there is a way to take advantage of the benefits of the ensemble to issue a deterministic forecast instead of sticking to the unperturbed run of the model. Though there are a number of strategies to face the problem, the most common approach deals with the ensemble mean and variance (Wilks, 2006). However, here we will offer an approach based on the construction of a single time series from the GE. To do so we have to follow these steps: :

1. Create a GE from the original ensemble (see section 6.2.2). In our case, the GE is formed of 5 member-series corresponding to percentiles 10, 25, 50, 75 and 90.

2. For each element of the grid-point series of the GE evaluate the Mean Absolute Error (MAE) for all the other elements following a Leave-One-Out Cross-Validation (see section 5.4). Here we have a hindcast period of 30 years (1981-2010) so each monthly MAE will be evaluated with 29 elements for each of the five members of GE.

3. The forecast for that particular element will be the value of the one in five GE member that showed a lower MAE in the LOOCV.

6.2.4 Calibration

As we have stated in section 5.5 we have applied three statistical calibration techniques on the interpolated and reorganized output of the S4 model:

i. MOS-analog

ii. Linear regression

iii. Mean bias correction

In this way we can evaluate the efficiency of both non-linear (MOS-analog) and linear (linear regression and mean bias correction) approximations. In the MOS-analog the search for analogs is made considering simultaneously the entire number of region grid-points, whereas in the linear regression and mean bias correction each grid-point series is independently calibrated. In the three cases predictors and predictands are referred to the same variables, predictors coming from the S4 previsions and predictands, from E-OBS. When determining the usefulness of calibration procedures we first compare scalar values against raw GE S4 forecasts. Afterwards, we work with unbiased S4 GE[1] in order to to better assess whether these calibration techniques can go beyond the skill recovered by correcting any existing first order systematic errors. Regarding the protocol to assess the performance of the calibrated forecasts it consists of a LOOCV. Hence, each year is calibrated with all the available years except from itself. Then the same process is repeated for every year of the 1981-2010 series and the validation metrics are computed.

MOS-Analog

The MOS-Analog technique has been chosen, among other alternatives (see 2.3.2), because it is a non-linear method that has already offered interesting results in RCM downscaling (Turco *et al.*, 2011) and for it can be transferred easily to model calibration. In this case it consists on seeking, for each member of the GE-S4, N

[1]This correction consists on the computation, for each pixel forecast, of the median anomaly in reference to E-OBS climatology. Afterwards, this quantity is added or subtracted to each forecasted value.

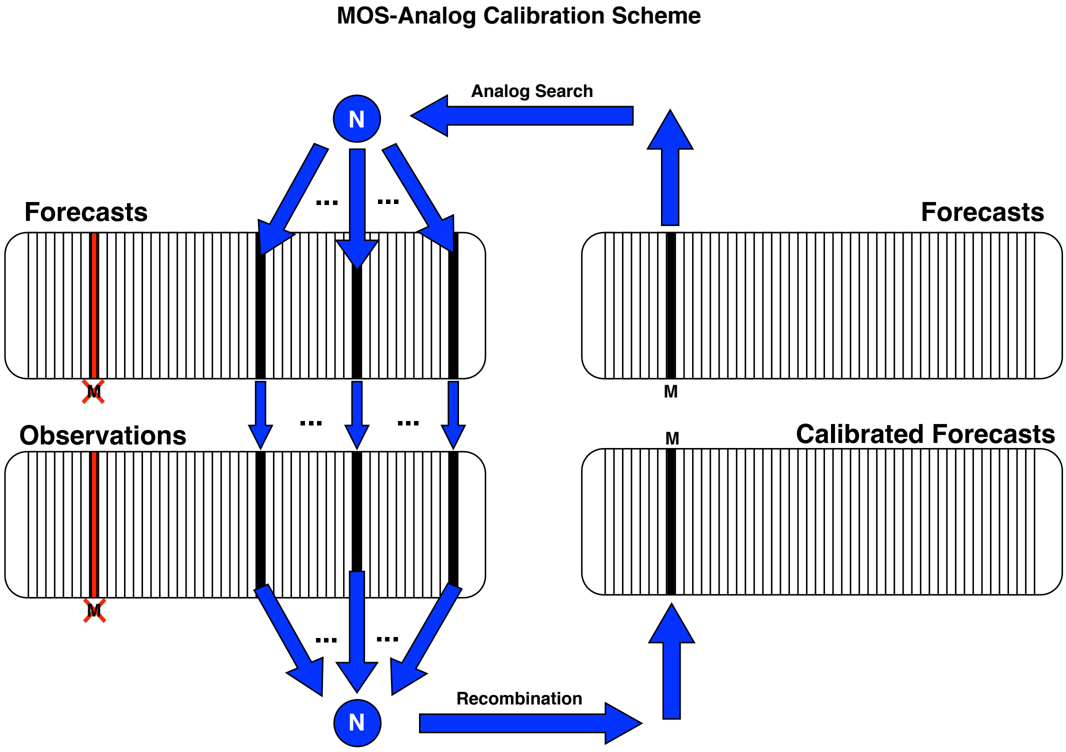


Figure 6.2. MOS-analog calibration scheme. For each lead, GE member, variable and forecast month we search the N analogs in the analog pool formed by the same month forecasts taken from the hindcast (excluding the forecast month that we want to calibrate). Afterwards, we take the N E-OBS observation fields corresponding to the N analogs and merge them to form a single output that will become our calibrated forecast.

same member analogs in a pool consisting of forecasts for the same month and horizon coming from the rest of the hindcast. The analogs are obtained through minimization of the euclidean distance among forecast maps of the same variable. Once found, we select the corresponding E-OBS observation field to be the *true* forecasts. Afterwards, we merge the N analogs through a simple *pixel mean* procedure to obtain a single *calibrated* forecast map (see figure 6.2). Finally, once the entire GE has been calibrated we perform the unification process of section 6.2.3. It is worth noting that the calibration process cannot be applied to the unified original forecast because otherwise we would have an over-fitting problem. For more details on the method see section 6.2.

The MOS application is done upon the principal components computed from every member of the GE maintaining a variability explained of the 99% to restrict the influence of its associated error (Gutiérrez *et al.*, 2004). The decision of using principal components instead of direct forecasts in the search for analogs is to reduce the computational burden of the calculations performed. To sum up, the MOS-analog calibration protocol can be summarised in the following steps,

1. Calculation of the principal components of the S4 *generalized ensemble* monthly forecasts preserving a minimum variability of 99%.

2. Application of the LOOCV MOS-analog process upon the principal components of each variable and for every GE member with the identification of the 5 nearer analogs and their corresponding observed fields.

3. Computation of the mean of the 5 observation fields to obtain a single calibrated forecast.

Regarding the domain, aside from working with multiple-sized dominions we also performed different experiments to assess the relationship between the analog pool and the domain size. We found that if we search for analogs directly upon a smaller region such Catalonia we get much better results than extracting values for the same domain but from a bigger area such as Spain or Europe. A possible explanation for this result could be that the bigger number of potential situations that can arise in the bigger domain cause that for an equal size of the analog pool, the representativeness of each member for the smaller portions of the domain decreases. Consequently, if the analogs for the smaller region are obtained from a bigger domain we would need a larger analog pool than if we directly searched them upon the smaller area. In table 6.1 there is a compilation of the MOS-analog experiments performed to select the optimal configuration (exp. 7).

Linear regression

When going for a linear calibration technique we have chosen the simplest one available: a linear regression of the form,

$$Y = bX \tag{6.5}$$

Table 6.1. MOS-analog experiments

Exp.	N° Analogs	Agg. Method*	Analog pool	PC var.
1	1	-	29	95%
2	3	WM	29	95%
3	5	WM	29	95%
4	10	WM	29	95%
5	5	M	29	95%
6	5	M	209	95%
7	5	M	29	99%
8	5	R	29	99%

* Weighted Mean (WM) / Mean (M) / Random (R)

The choice of a regression form without independent term is explained because it is the expected reaction of a perfect model. In fact, without biases or systematic errors, the perfect forecast should respect the value and magnitude of the observed changes. The calibration is performed through a LOOCV for each variable and for every GE member (percentiles 10, 25, 50, 75 and 90). In each LOOCV approach we have computed the b parameter taking forecasts as predictors, X, and observations as predictands, Y. Afterwards, the forecast to be calibrated is taken as predictor and with the parameter previously computed we have obtained the final calibrated forecast. It is important to remember that since this process is repeated for every single year we end having N calibrating parameters.

Mean bias correction

This strategy consists on searching for the *mean error* in the model forecasts and adding (subracting) it from every single element of the dataset to have zero mean error. Mathematically the mean error can be written as,

$$ME = \frac{1}{n} \sum_{k=1}^{n} (f_k - o_k) \tag{6.6}$$

Where f_k accounts for the forecast and o_k for the observation.

6.2.5 Verification

The skill of the original S4 forecasts and the calibration techniques has been evaluated using a LOOCV approach, considering E-OBS observations for the *hindcast* period 1981-2010. In that way we can simulate the operative way of working since we calibrate the forecasts with all the available past years, except from the current one. Hence, in the verification process the deterministic and probabilistic skills have been assessed (both kind of metrics are fully described in section 5.7). To evaluate the deterministic skill of the S4 and its calibrations the corresponding parameters have also been computed for *climatology* and *persistence* (at the corresponding lead) in order to identify their added value. The selected metrics have been: *a*) Spearman correlation; *b*) standard deviation; *c*) MAE and RMSE.

All the metrics have been computed for each grid-point series, and for the 30-years hindcast. Since the analysis is aimed to study individual months, this means that for each forecast horizon we have 12 field maps with 30 temporal values for each grid-point. One advantage of this approach is that it prevents the influence of the annual cycle in the verification process, for we compare observations and forecasts corresponding to the same month of the year, and therefore, the same radiative influx. However, the final verification metric has been presented through a single summarising value. This has been achieved by firstly computing the mean of the 30-element series at each grid-point of the re-forecast and, eventually, calculating the spatial median among the grid-points. The reason to use the *median* instead of the *mean* is because the former is more robust to outliers and represents better the overall performance over a region. Besides, in case there are no departures from normality, both parameters should give the same results.

Additionally, we have used Taylor diagrams to depict in a single figure the centred RMSE (cRMSE), the standard deviation and the Spearman correlation. To do so we have rearranged all the grid-point time series for each region to form a single series concatenating one after the other (Taylor, 2001). This strategy has one main advantage. For regions with two or more grid-points this lets us include climatology spatial correlation in the diagram. This is important because climatology does not have temporal correlation and, thus, any forecast correlation bigger than the associated to climatology hints the existence of temporal skill.

Turning to the probabilistic parameters, we have compared the performance of the S4 and its calibrations against *climatology* assuming percentiles 33 and 66 as thresholds for below, normal or above normal conditions which is a common format for seasonal forecasts (see i.e. Mason *et al.*, 1999; O'Lenic *et al.*, 2008). We have used: *a*) Discrete Brier Skill Score (dBSS); *b*) Discrete Ranked Probability Score (dRPSS). The differences between the standard BSS (RPSS) and the dBSS (dRPSS) can be found in subsections 5.7.2.3 and 5.7.2.5 but, synthetically, the dBSS and dRPSS are corrections of the BSS and RPSS when working with small ensembles. In the results chapter we only present the analysis of the dBSS because it depicts an independent assessment of the three studied categories (see 6.2.5). The dRPSS evaluation, on the other hand, acts rearranging all of them in a single index and since one or two categories often function much worse than climatology, the index becomes degraded and seldom shows global improvements beyond climatology. Finally, there is one remark to be made and it is that since we do not work with the entire *pdf*, just only 5 percentiles, we had to decide which approach to take when attributing a probability to a forecast. There were a couple of options: the conservative (collapse towards the lower probability) or the brave (collapse towards the higher). In overall, the differences between the two approximations were small to non-existent, and we stayed with the conservative conception.

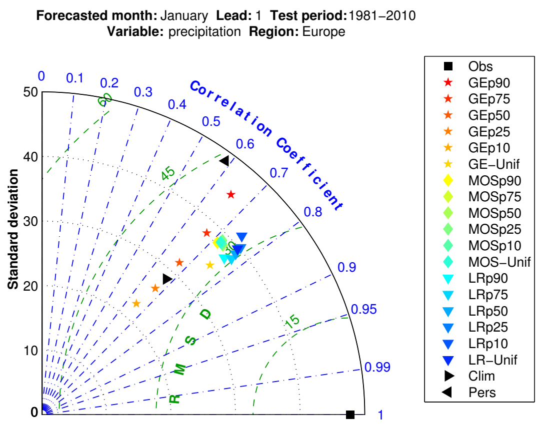
6.3 Results

The preceding section introduced the way in what we have rearranged the original S4 ensemble to characterize its *pdf* and to issue a unified deterministic forecast from itself. Here we will show the performance of this new set of forecasts with its corresponding MOS-analog, linear regression (LR) and mean bias correction calibrations, in comparison with observations and *climatology* and *persistence* controls.

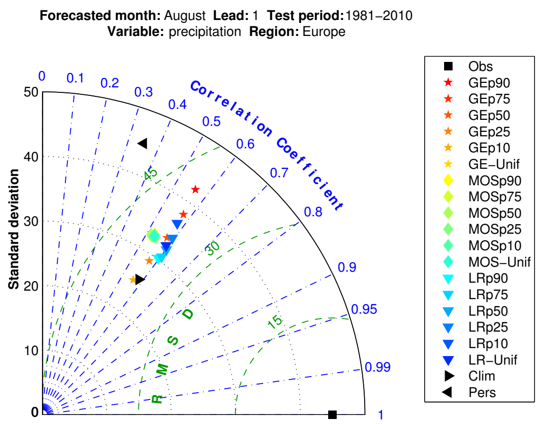
6.3.1 Europe

Precipitation

The analysis of lead one Taylor diagrams shows that the LR performs generally better than the MOS-analog and the original S4 forecasts. In fact, LR surpasses climatology from September to March whereas the MOS-analog only performs better than climatology in January. The raw S4 forecasts always work worse than climatology. The best forecast period goes from September to March whereas the worst forecast months comprise April to August. Figure 6.3 depicts the Taylor diagrams for the best and worst forecast months. After the mean bias correction of the original S4 forecasts we find that this is the most efficient way of calibration. Actually, when analysing the MAE graph (see figure 6.4) we see that the median of the distribution shows not only lower error than climatology but also lower error than the LR and MOS-analog approaches. It is worth noting that this improvement can be seen in all months. Again, the LR reveals to be a better option for calibration than the MOS-analog. The lowest errors are found in winter, whereas summer shows the highest. This might be due to the central Europe summer mesoscale convection that is not resolved by the original S4 resolution and confers an increased spatial and temporal rainfall variability to this season. Taking the analysis to the probabilistic field, the dBSS reveals that the improvement seen with the MAE does not suffice to go beyond any of the thirtile climatic forecast categories. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.

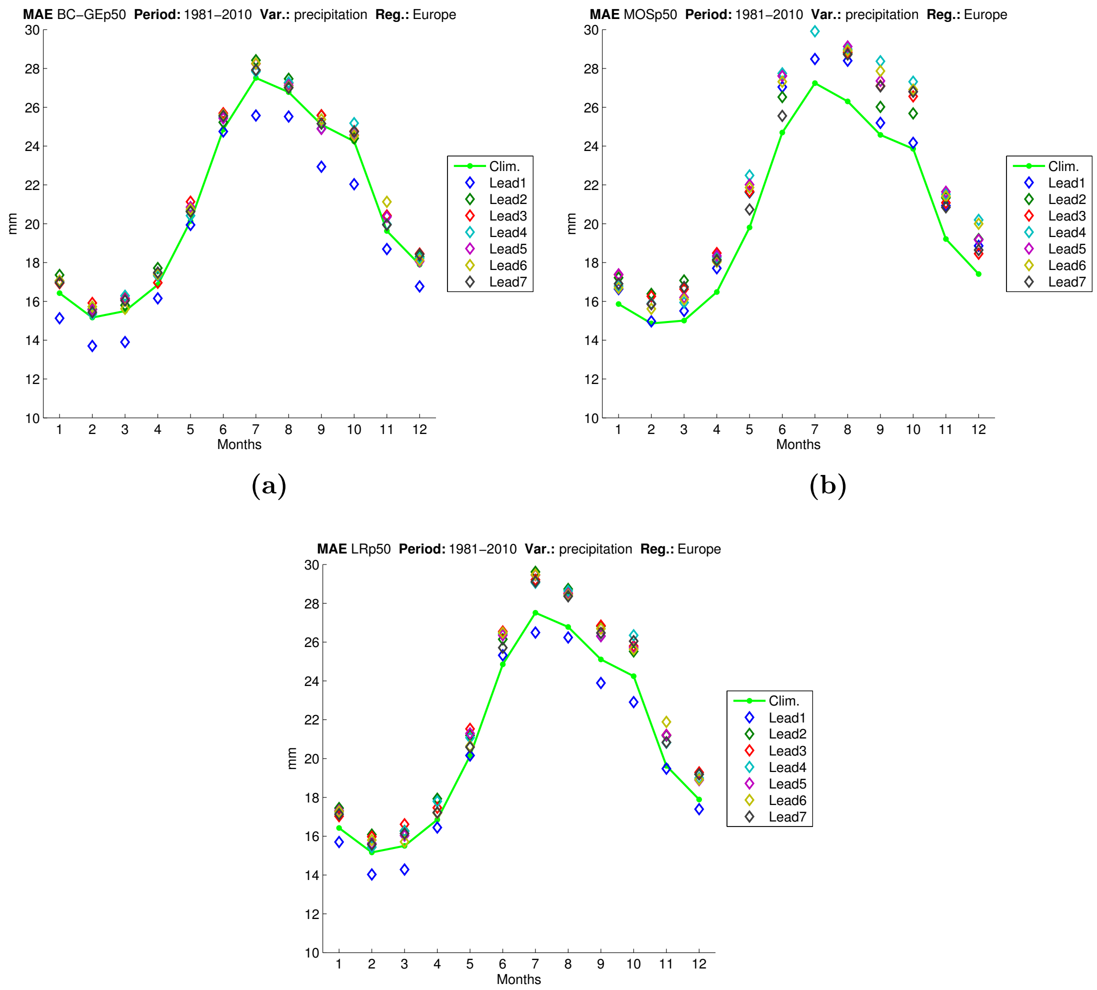


(a) Best



(b) Worst

Figure 6.3. Europe's Taylor diagrams for precipitation at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: January (b) The worst performing month: August.

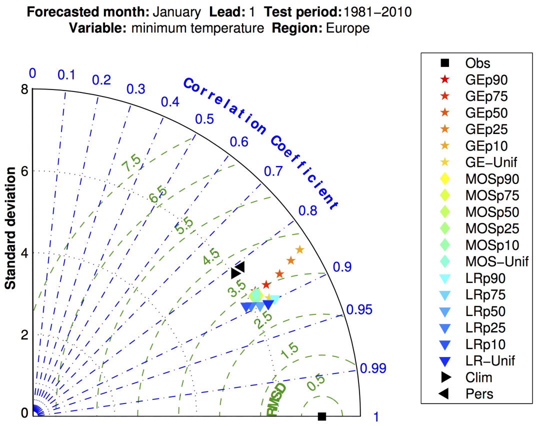


(a) (b) (c)

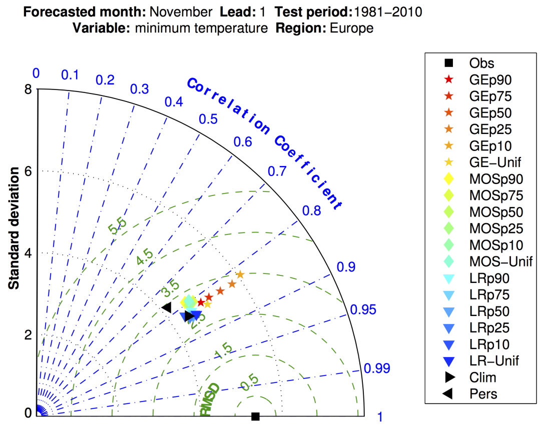
Figure 6.4. Europe's precipitation MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

Minimum temperature

In this case lead one Taylor diagrams identify the best forecast period from December to March and the best method, LR calibration. The period April to November is the worst, with the latter being the exponent of a fruitless forecast. On the other hand, January is the most successful (see figure 6.5). Once the original S4 is mean bias corrected this option becomes the most advantageous. In fact, in figure 6.6a we can see a general MAE reduction corresponding to this method that goes below climatology in all months. The greatest amelioration is attained in winter and, secondarily, also in the summer months. On the contrary, spring and autumn are the seasons with the modest improvement. Turning to the probability verification through the dBSS the best results are obtained with the mean bias correction, which in practice, are the same results that would be obtained with the original S4. The lower and upper thirtile report positive values in winter with also modest improvements in the summer months (see figure 6.7). The central tercile, however, is not improved. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.



(a) **Best**



(b) **Worst**

Figure 6.5. Europe's Taylor diagrams for minimum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: January (b) The worst performing month: November.

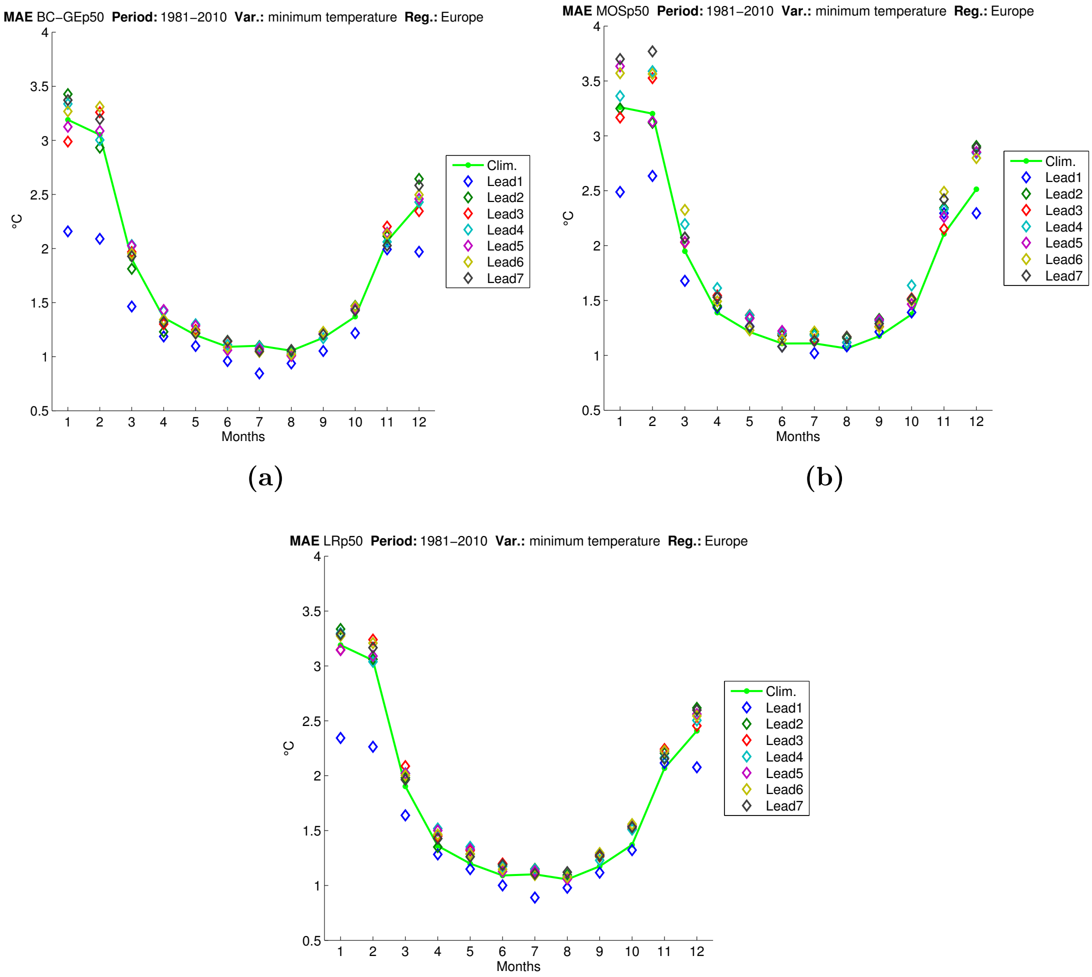
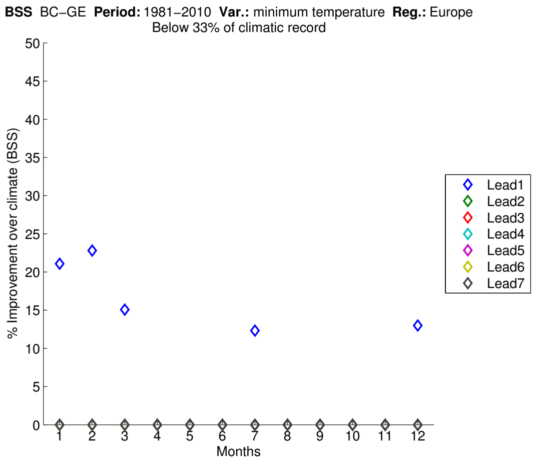
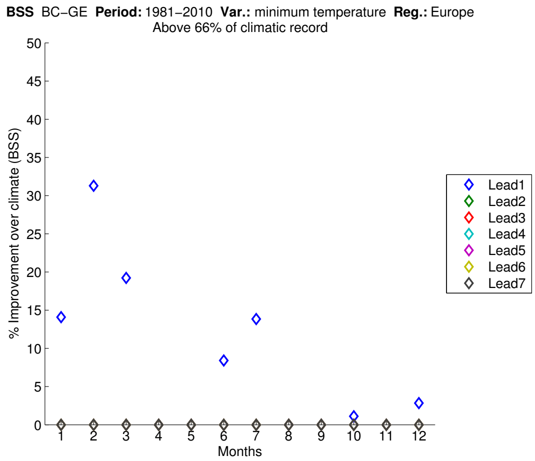


Figure 6.6. Europe's minimum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.



(a)

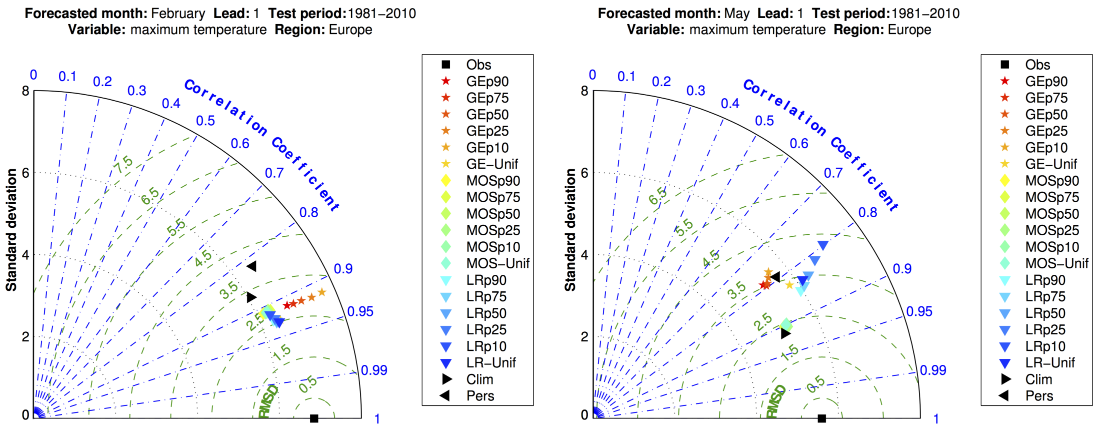


(b)

Figure 6.7. Europe's minimum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

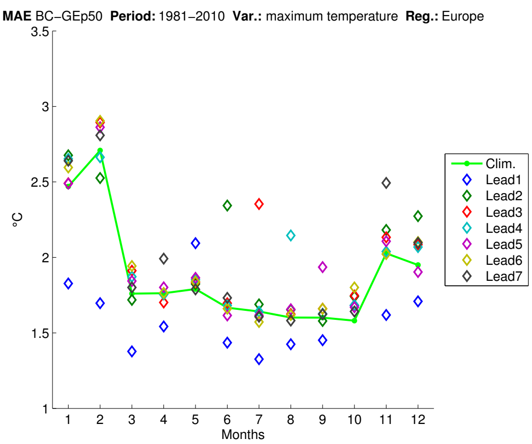
Maximum temperature

The lead one Taylor diagram comparison among the original maximum temperature S4, LR and MOS-analog calibrations shows that the LR gives the best results: it improves climatology in all months except May and October. Figure 6.8 depicts the best and worst forecast diagrams, namely, January and May. The amplitude of the correction, however, is increased when calibrating the original S4 forecasts through a mean bias correction. In the MAE study we have found that the largest error correction takes place in winter months whereas the other seasons, spring to autumn, see a smaller amendment with May and October maintaining their no-improvement with respect to climatology (see figure 6.9). Moving on to the dBSS, the mean bias correction of S4 keeps on showing the leading results with positive values in the lower tercile for December-March and July. In the upper tercile there is an enhancement beyond climatology for January-April, June, July and October. No improvement is detected for the middle tercile (see figure 6.10). The other calibrations perform generally worse. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.

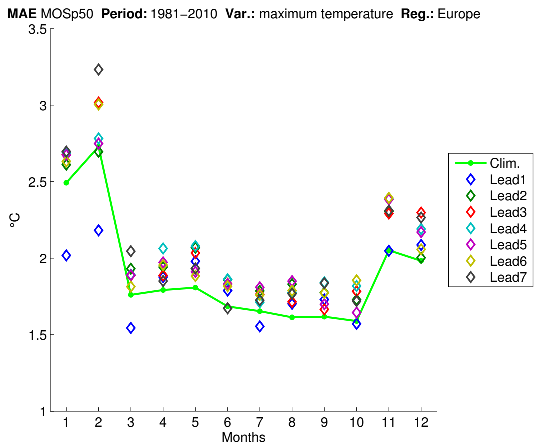


(a) Best **(b) Worst**

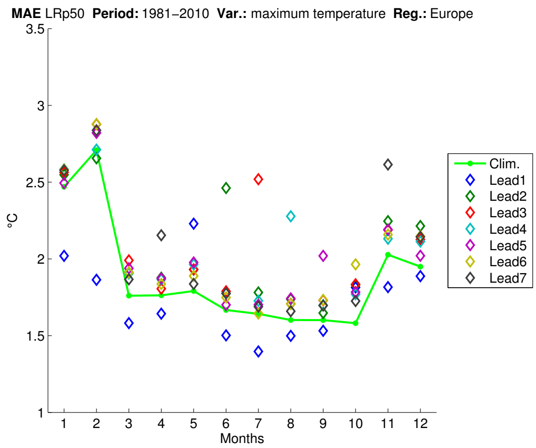
Figure 6.8. Europe's Taylor diagrams for maximum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: February (b) The worst performing month: May.



(a)



(b)



(c)

Figure 6.9. Europe's maximum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

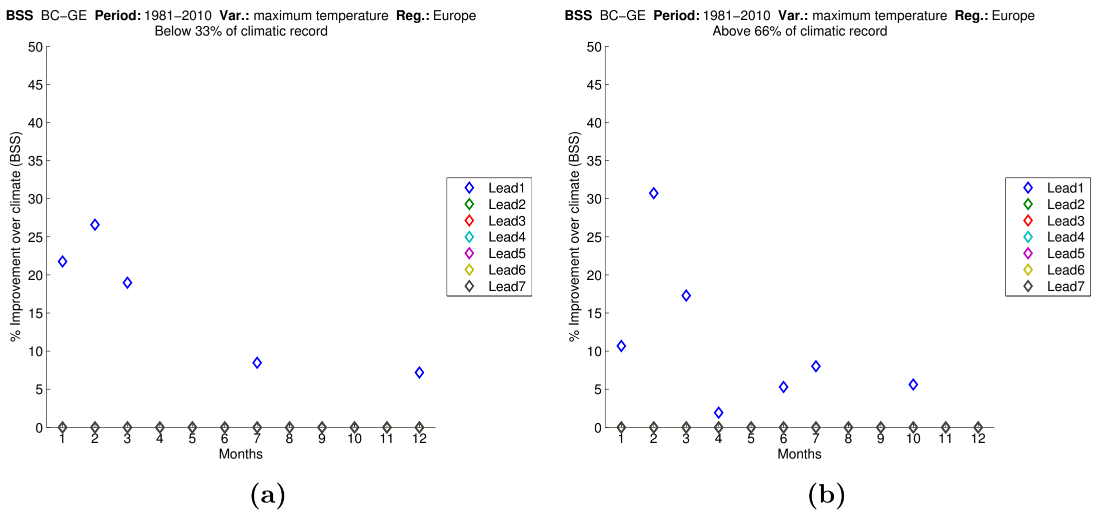
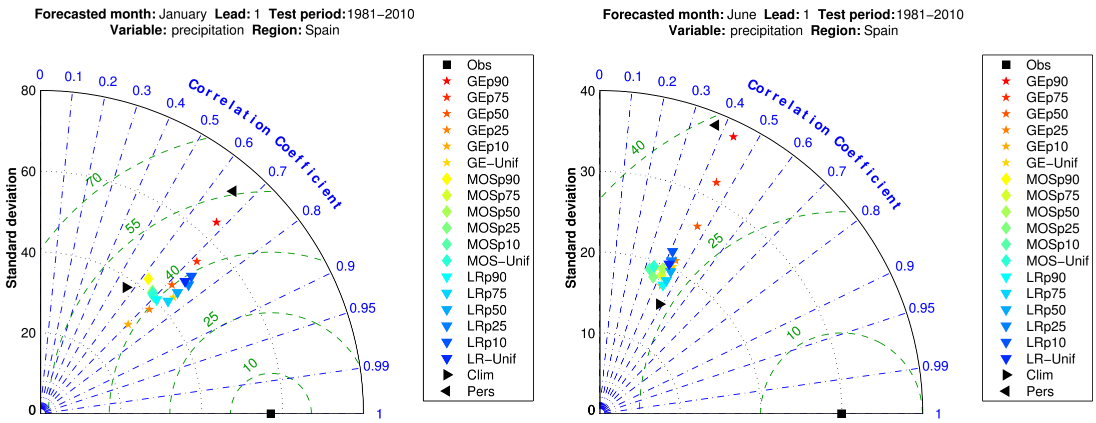


Figure 6.10. Europe's maximum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

6.3.2 Spain

Precipitation

In this case lead one Taylor diagrams show that the LR calibration gives the best results, going beyond climatology in all months except April, June, September and October. That being said spring and summer show the modest ameliorations. The MOS calibration seems to be the worst option and the original S4 lies somewhat in the middle. The best performing month is January and the worst, June (see figure 6.11). Once the original S4 is corrected through mean bias this option becomes the preferable. As we can see in figure 6.12 it reduces MAE below climatology for all months except June. Winter is the interval with larger reductions, whereas summer and autumn see the smaller enhancements. When turning to the dBSS we see that mean-bias corrected S4 only displays positive values in the lower and upper terciles for December and January. Besides, in the latter tercile there are also humble improvements in May and August. For the middle tercile, on the contrary, the climatology is not upgraded by any of the calibration systems. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.



(a) Best **(b) Worst**

Figure 6.11. Spain's Taylor diagrams for precipitation at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: January (b) The worst performing month: June.

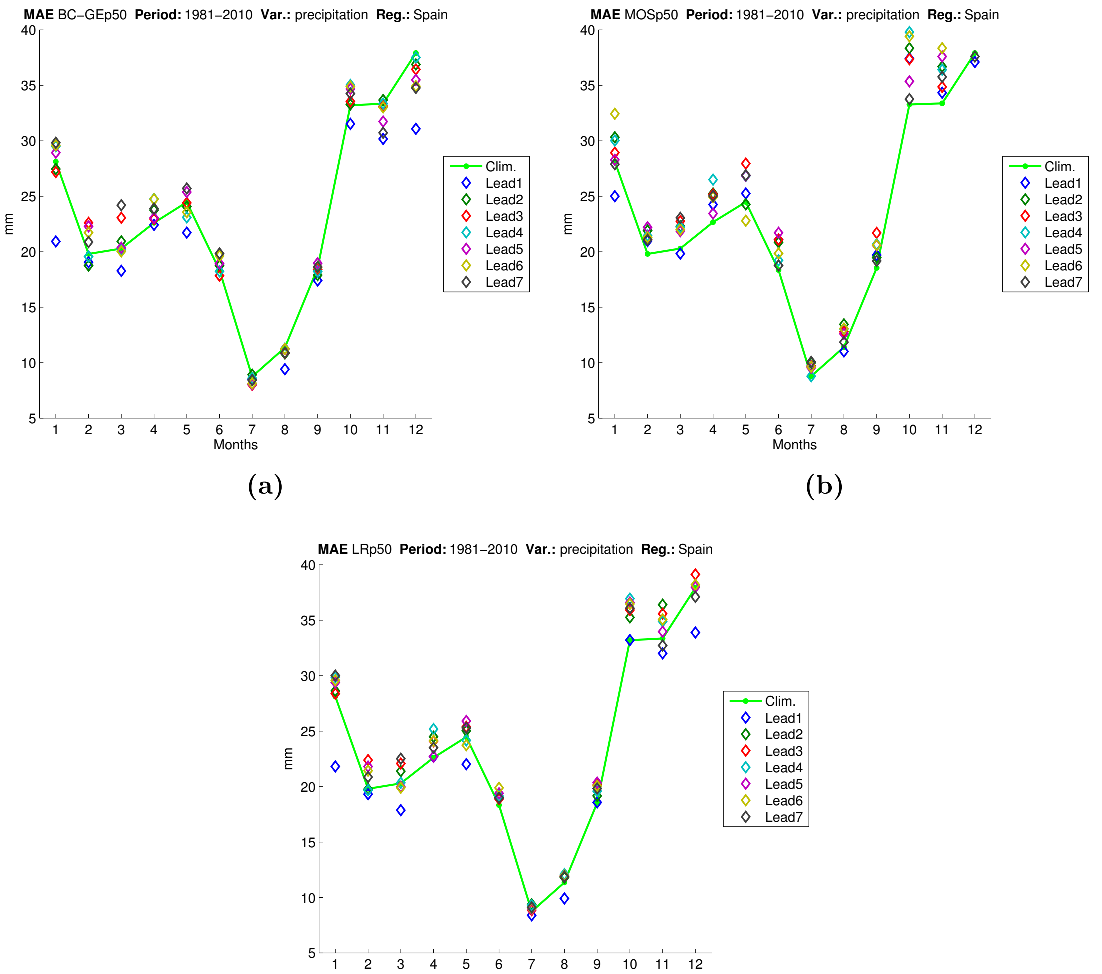


Figure 6.12. Spain's precipitation MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

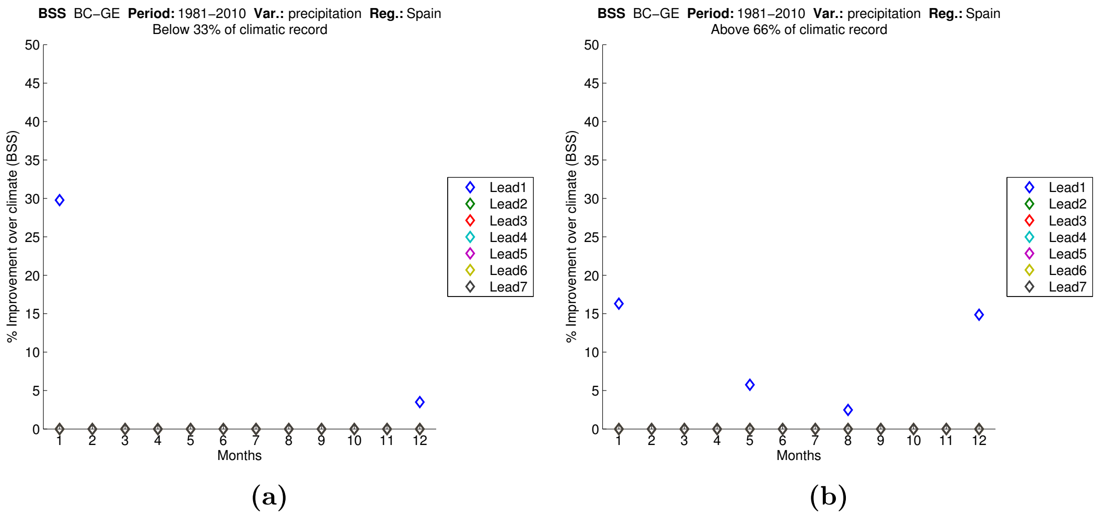
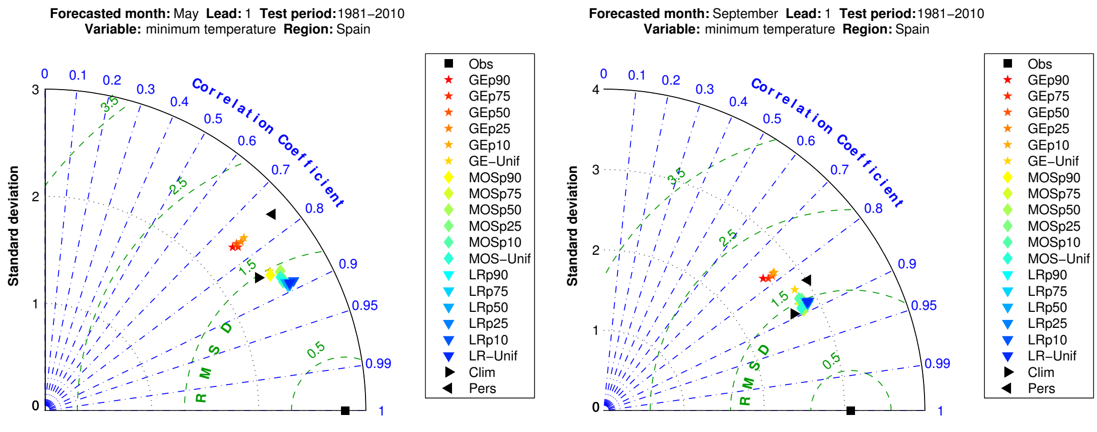


Figure 6.13. Spain's precipitation dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

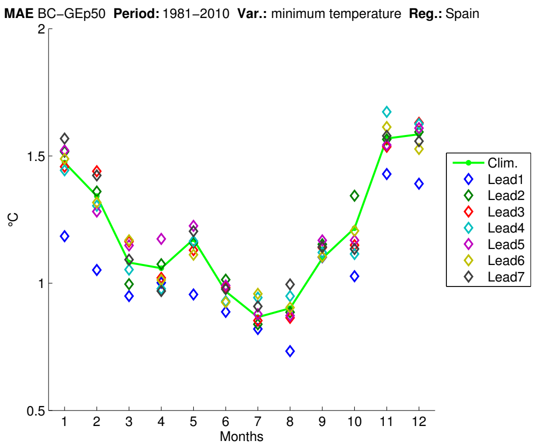
Minimum temperature

The analysis of lead one Taylor diagrams among the original S4 and the MOS-analog and LR calibrations reveal that the best results come with the latter, surpassing climatology in all months except March, September and November. The best performing month is May and the worst, September (see figure 6.14). Once the mean bias correction is applied the S4 becomes the best option followed by the LR and the MOS-analog calibrations. In the MAE graphs it depicts the largest upgrade with respect to climatology in all months except September. The magnitude of the improvement is similar throughout the year, with perhaps larger values in the winter months (see figure 6.15). Moving on to the dBSS analysis of the three options (see figure 6.16), we find the same behaviour, with the mean bias correction as the best performing method. In fact we see an amelioration beyond climatic values in the lower and upper terciles in January, February, May, August and October. Besides, in the lower thercile there is also an improvement in June. There are other leads that show modest upgrades in the lower tercile but they are not consistent and might be regarded as noise. As for persistence control, it is always the worst option to choose.

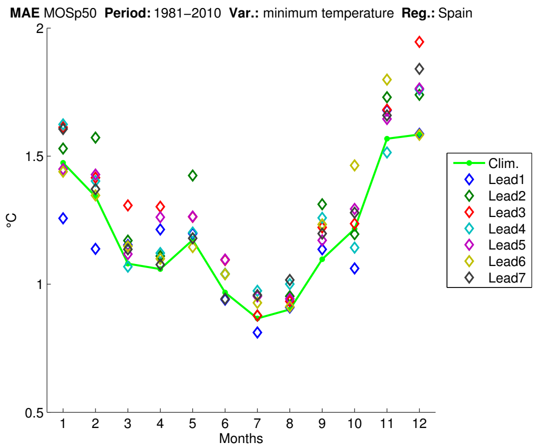


(a) Best (b) Worst

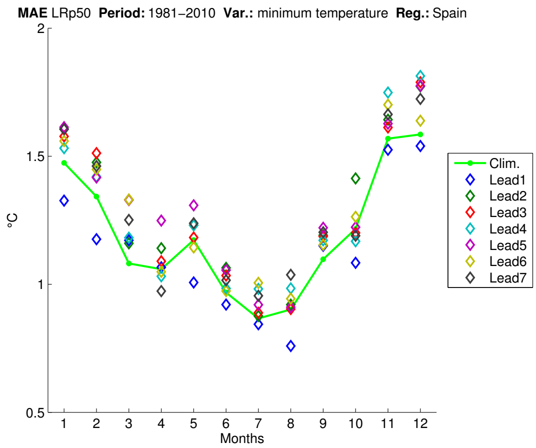
Figure 6.14. Spain's Taylor diagrams for minimum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: May (b) The worst performing month: September.



(a)

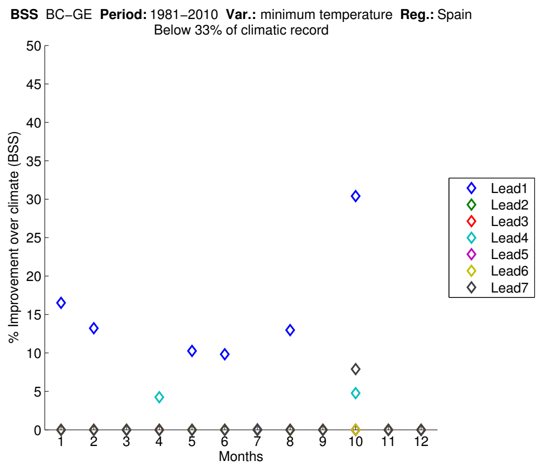


(b)

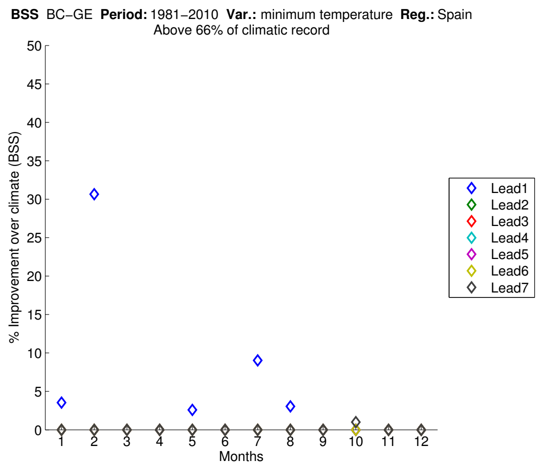


(c)

Figure 6.15. Spain's minimum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.



(a)

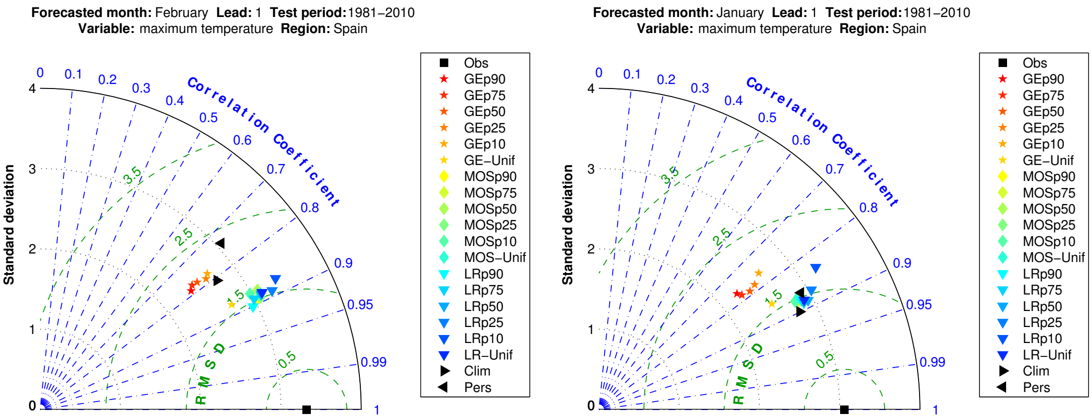


(b)

Figure 6.16. Spain's minimum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

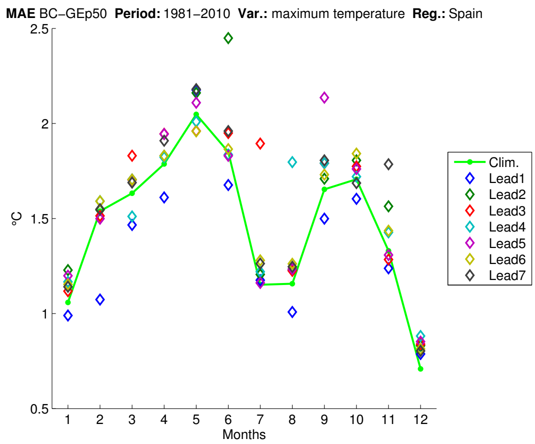
Maximum temperature

Maximum temperature Taylor diagrams for lead one show that the MOS-analog and LR calibrations are better than the original S4 forecasts. The LR gives the best results going beyond climatology in February, March, April, June, August and September (see figure 6.17). The comparison with S4 mean bias correction reveals that the former is the best choice, with a neat MAE upgrade in all months, as it can be observed in figure 6.18. The only months that do not surpass climatology are May, July and December. The largest improvement is found in spring, whereas the modest can be located in autumn. When looking at the dBSS (figure 6.19) the mean bias corrected S4 presents the best results. In the lower thirtile we find positive values in Feburary, June, July and October. The middle thirtile, conversely, does not show any improvement. Finally, the upper thirtile reports an upgrade beyond climatology in February, August and October. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.

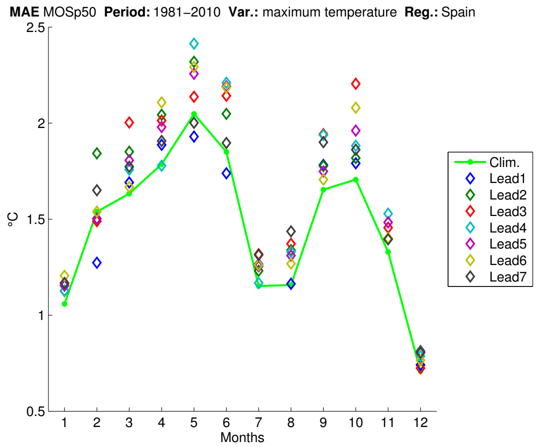


(a) Best **(b) Worst**

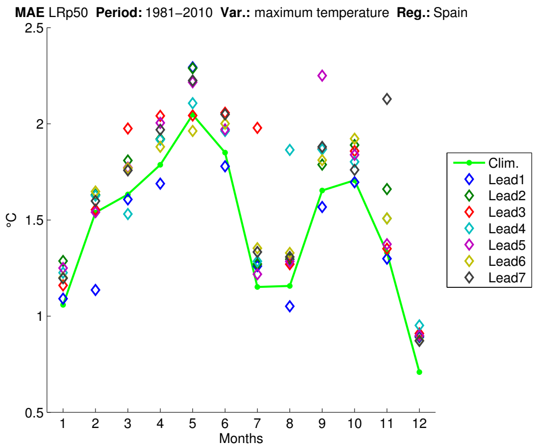
Figure 6.17. Spain's Taylor diagrams for maximum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: February (b) The worst performing month: May.



(a)

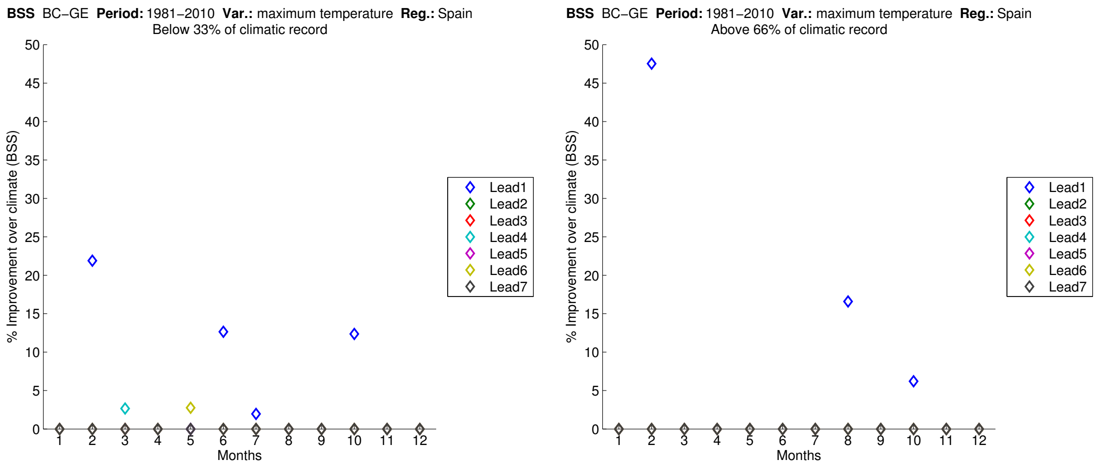


(b)



(c)

Figure 6.18. Spain's maximum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.



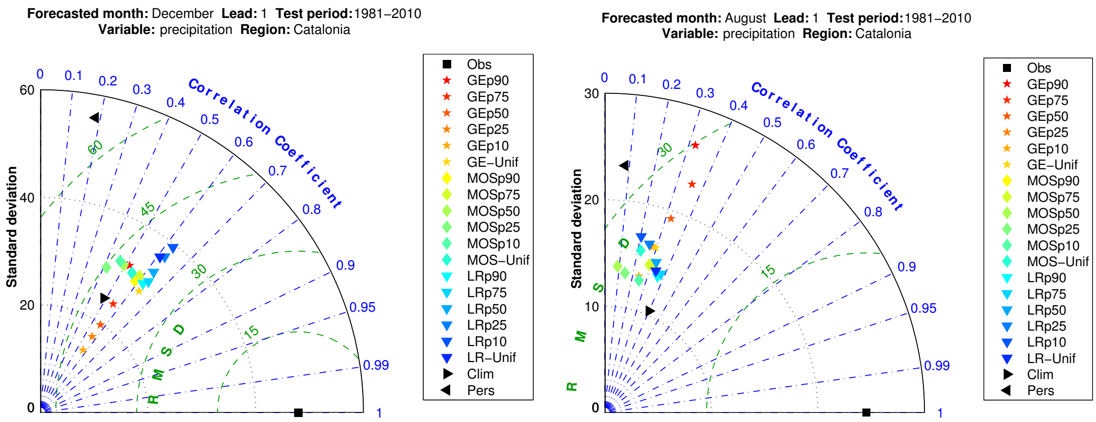
(a) (b)

Figure 6.19. Spain's maximum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

6.3.3 Catalonia

Precipitation

Lead one Taylor diagrams for precipitation in Catalonia show a similar behaviour as in the previous domains. When comparing original S4 with MOS-analog and LR calibrations, the best results are obtained with the latter. LR goes beyond climatology in all months except July, August, September and November. Figure 6.20 depicts the best and worst forecast months (December and August, respectively). Once the original S4 forecasts are mean bias corrected we find that this is the configuration which attains the best results. In the MAE plot (see figure 6.21) it improves all monthly records with respect to the LR. The only months that are worse than climatology are August and September. It is also worth noting that in February and March almost all seven leads show minor MAE than climatology. However, when turning to the dBSS we only find positive values in January for the lower thirtile; and in April, October and December in the upper tercile. The middle thirtile does not show any amelioration (see figure 6.22). At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.



(a) Best **(b) Worst**

Figure 6.20. Catalonia's Taylor diagrams for precipitation at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: December (b) The worst performing month: August.

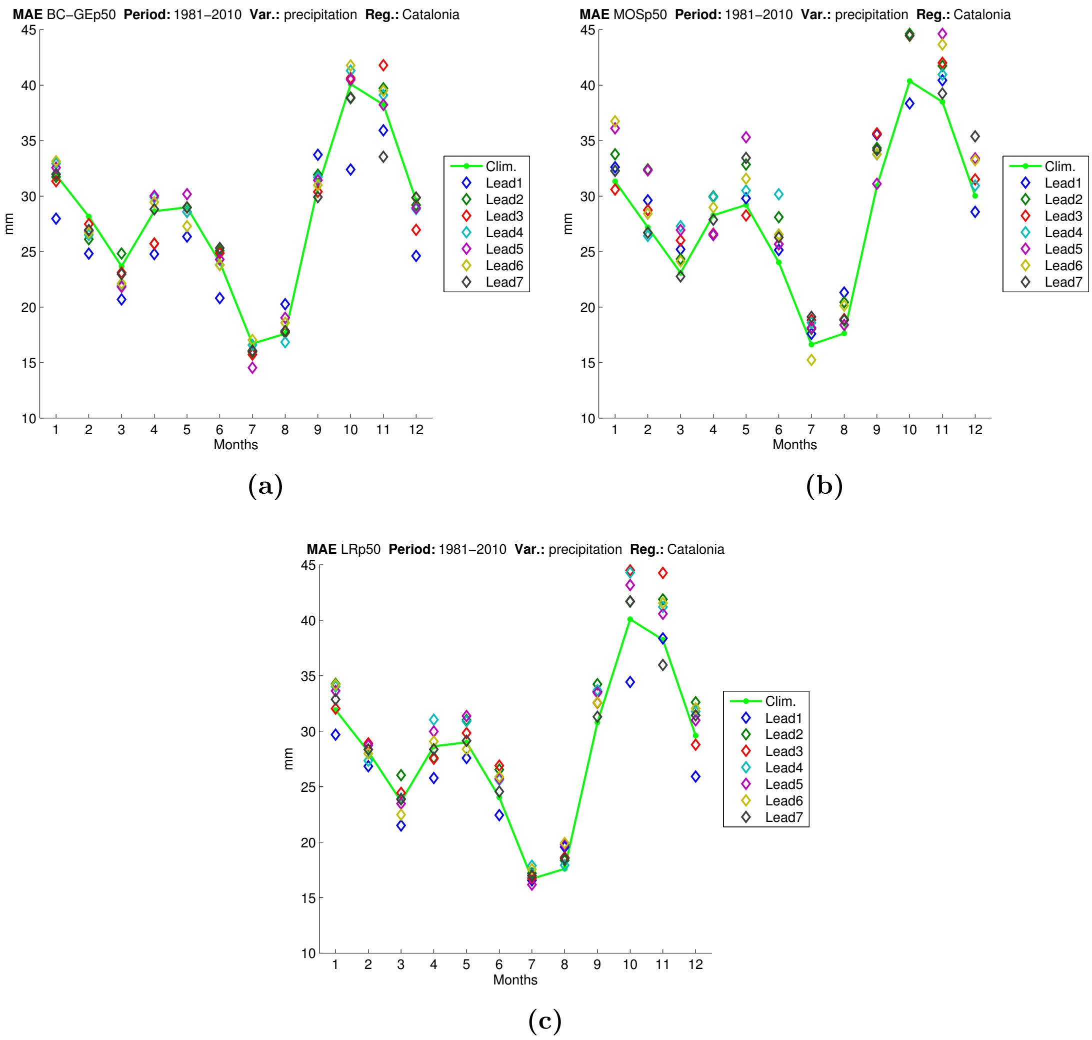
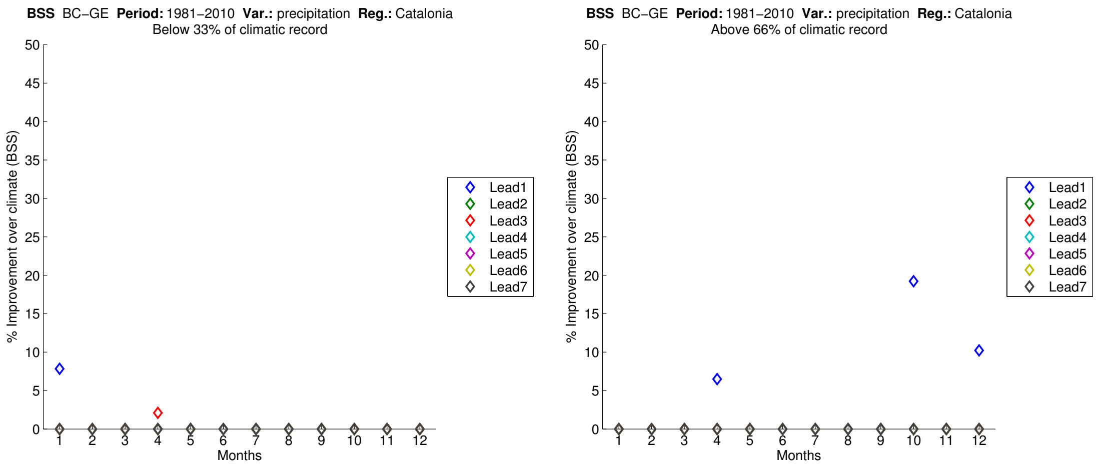


Figure 6.21. Catalonia's precipitation MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

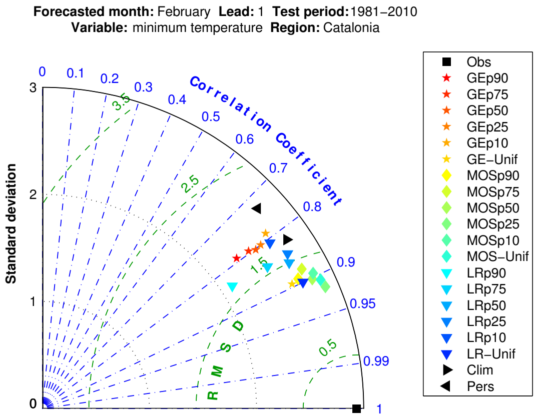


(a) (b)

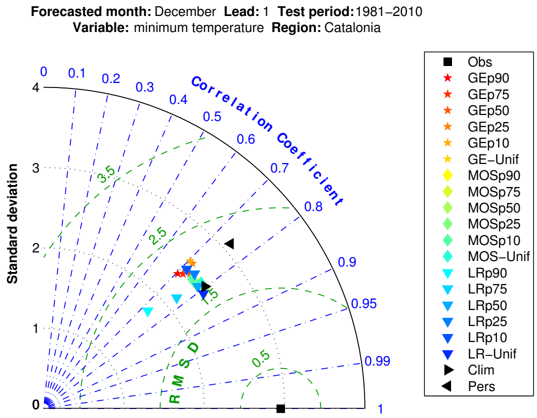
Figure 6.22. Catalonia's precipitation dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

Minimum temperature

Minimum temperature lead one Taylor diagrams exhibit the LR as the finest approach among the MOS-analog calibration and original S4 forecasts. It surpasses climatology in January, February, May, June, July, August, October and November. However, the MOS-analog calibration ameliorates the LR in February and March. Figure 6.23 depicts February and December as the best and worst performing months. Nevertheless, when the original S4 is mean bias corrected it surpasses the performance of the previously considered calibrations. The MAE plot (see figure 6.24) reports an improvement in all months which now show lower errors than climatology (except in September). The major MAE reductions are attained in winter and the lowest, in autumn. The dBSS for lead one reveals that the mean bias correction also offers the best results, with positive values in the lower thirtile for February, March, May, June, July, August and October. In the middle thirtile there is no improvement and for the upper one there is an enhancement only for January and March. Other positive values arise also at other leads but since they are not systematic we regard them to be the effect of noise (see figure 6.25). As for persistence control, it is always the worst option to choose.

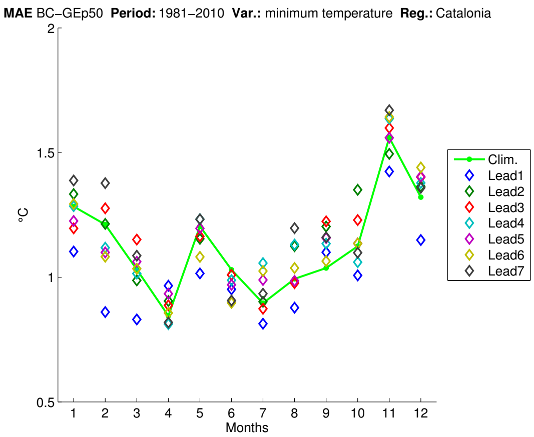


(a) Best

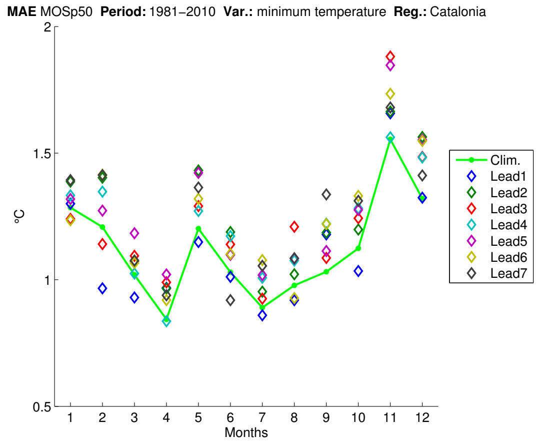


(b) Worst

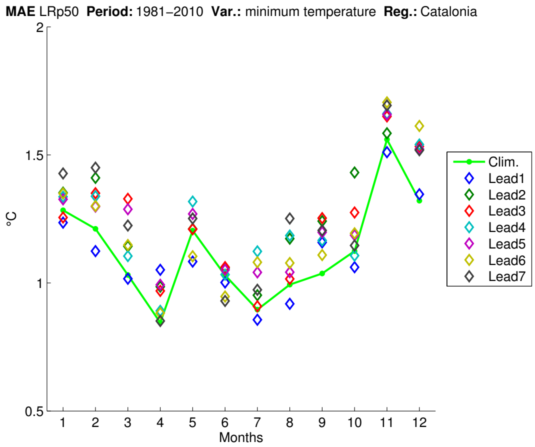
Figure 6.23. Catalonia's Taylor diagrams for minimum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: February (b) The worst performing month: December.



(a)

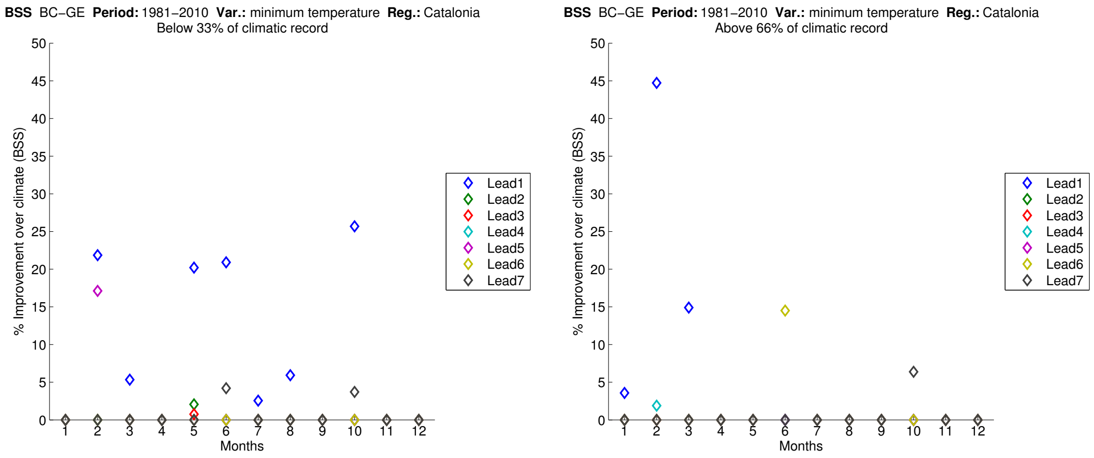


(b)



(c)

Figure 6.24. Catalonia's minimum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

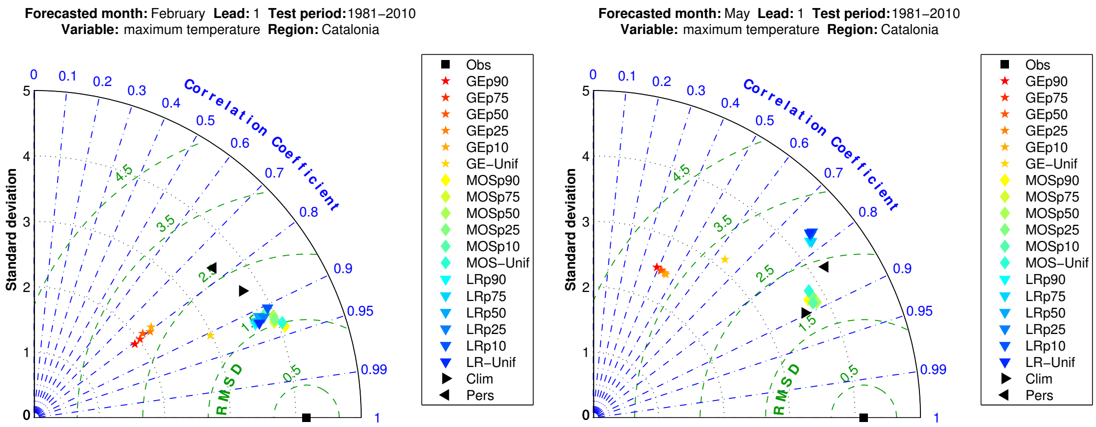


(a) **(b)**

Figure 6.25. Catalonia's minimum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

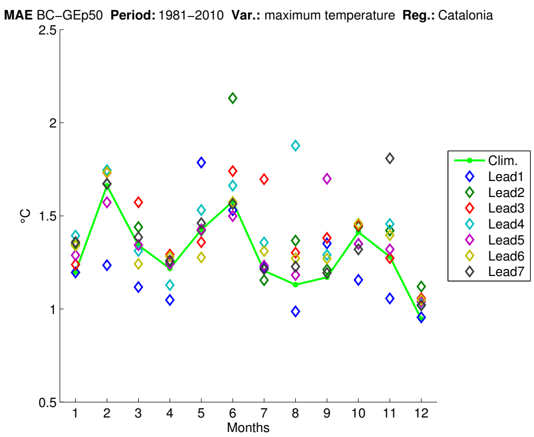
Maximum temperature

Moving on to the study of the maximum temperature lead one Taylor diagrams comparing the original S4 forecasts versus its MOS-analog and LR calibrations we find that the LR is the better option enhancing climatology results in February, March, April, August, October and November. The MOS-analog, on the other hand, is generally worse than the LR except in October. The best forecast month is February and the worst, May (see figure 6.26). The application of the mean bias correction upon the original S4 provides much better results than the LR and MOS-analog calibrations. As we can see in figure 6.27 there is a general reduction in the absolute errors. However, this is not enough to surpass climatology in the months that the LR could not improve. An interesting feature that can be observed in the same figure is that the MOS-analog shows lower error than mean bias correction in October. Mean bias correction is also the best option when analysed through the dBSS. It exhibits positive values in the lower thirtile for February, July, September, October and November. In the upper tercile there is an amelioration at lead one for February, March, August and October (see figure 6.28). In the middle thirtile there is no improvement. At further leads any approach works better than climatology. As for persistence control, it is always the worst option to choose.

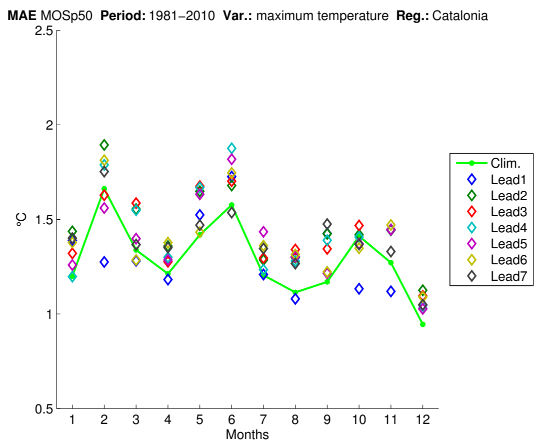


(a) Best **(b) Worst**

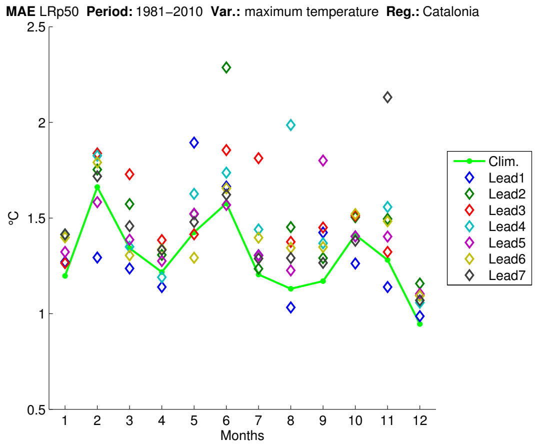
Figure 6.26. Catalonia's Taylor diagrams for maximum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: February (b) The worst performing month: May.



(a)



(b)



(c)

Figure 6.27. Catalonia's maximum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

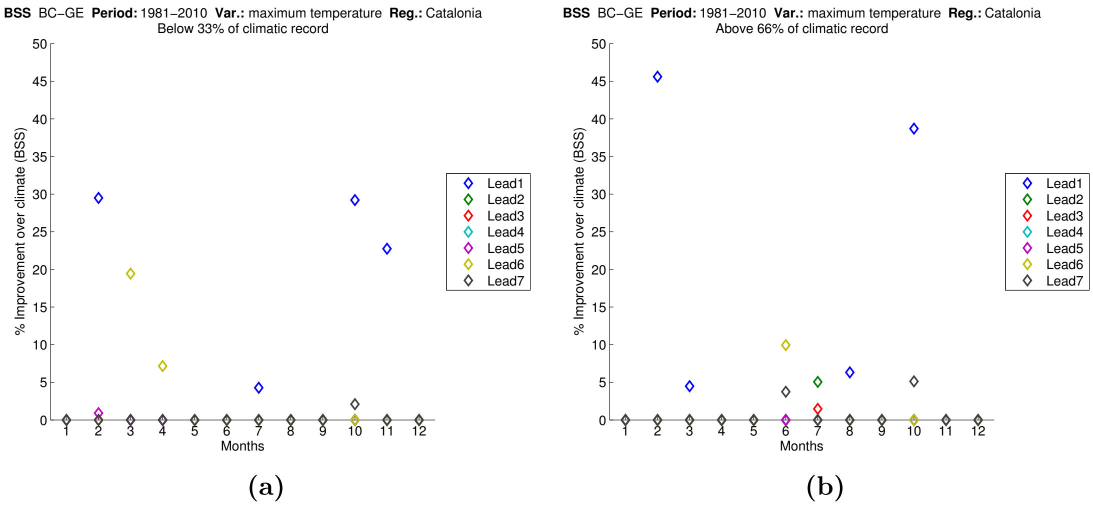
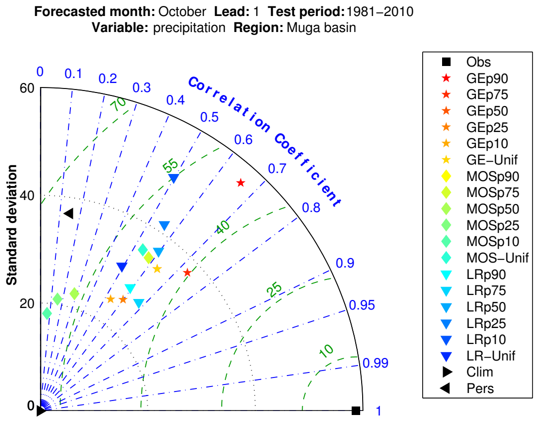


Figure 6.28. Catalonia's maximum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) The lower tercile (b) Upper tercile.

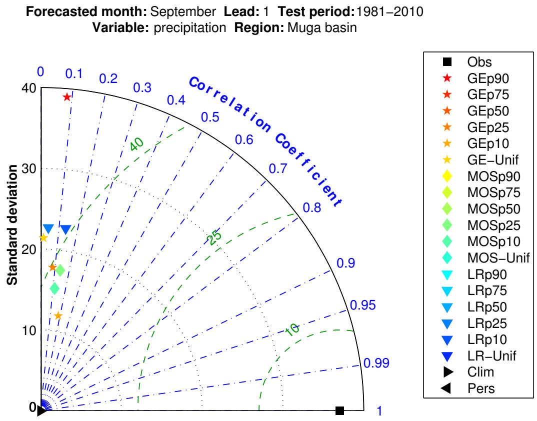
6.3.4 Muga river basin

Precipitation

Lead one Taylor diagrams for precipitation in the Muga river basin show that, in most cases, the original S4 offers best results than its LR and MOS-analog calibrations. Figure 6.29 depicts the best and worst months (October and September, respectively). The mean bias correction of the original S4 forecasts exhibits the greatest upgrade, far beyond the LR and MOS-analog calibrations (see figure 6.30). The largest ameliorations are found in March-June and October-December, although it surpasses climatology in all months but September. Some positive results are also observed at other leads, but their lack of consistency made us attribute the skill to the weak signal to noise signal found at lower scales. As for persistence control, it is always the worst option to choose. When analysing the dBSS we find that for lead one and the lower thirtile the mean bias corrected S4 does not show any positive value. For the middle thirtlile, on the other hand, October is enhanced. Eventually, in the upper thirtile the mean bias correction improves February, March, May, June, October, November and December (see figure 6.31).



(a) Best



(b) Worst

Figure 6.29. Muga basin's Taylor diagrams for precipitation at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: October (b) The worst performing month: September.

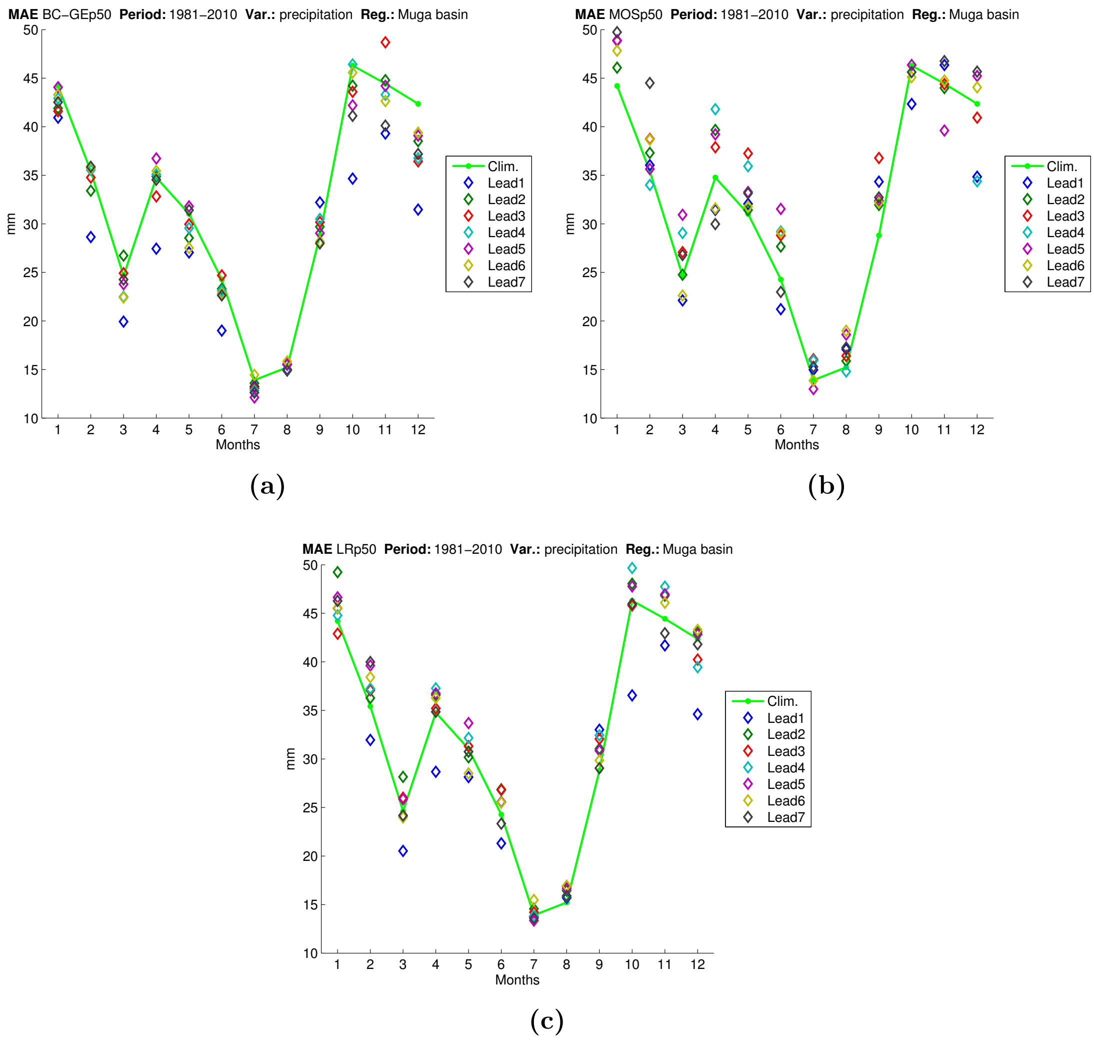
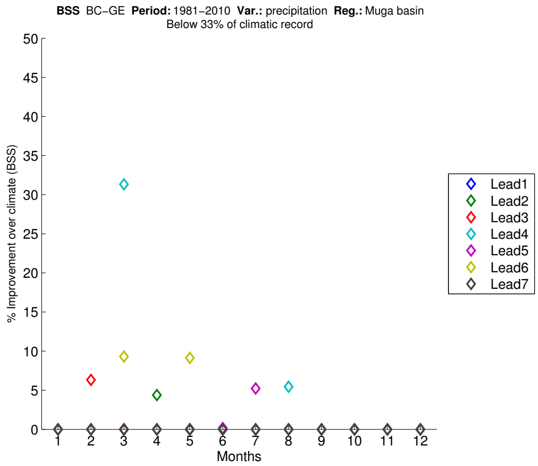
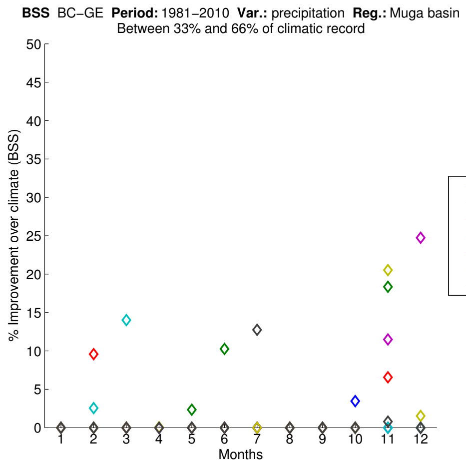


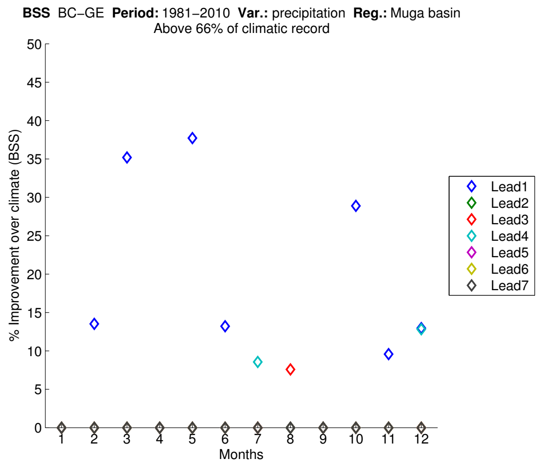
Figure 6.30. Muga basin's precipitation MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.



(a)



(b)

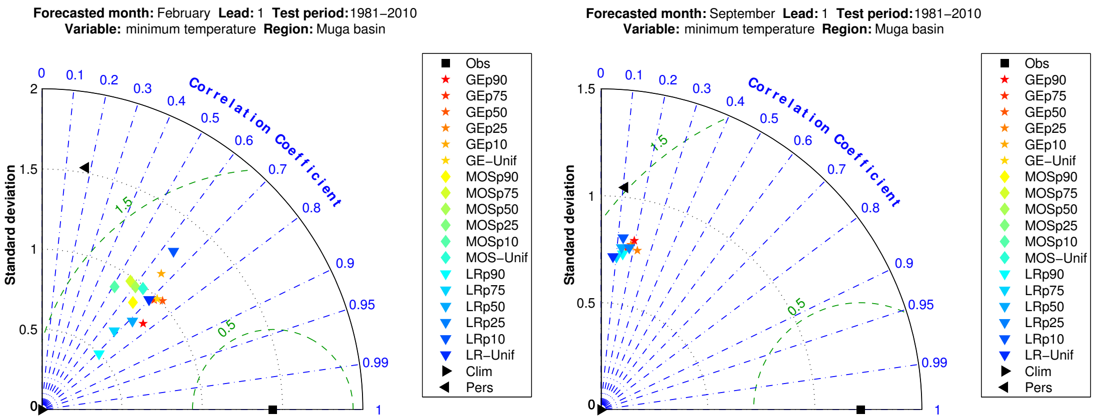


(c)

Figure 6.31. Muga_basin's precipitation dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) Lower tercile (b) Middle tercile (c) Upper tercile.

Minimum temperature

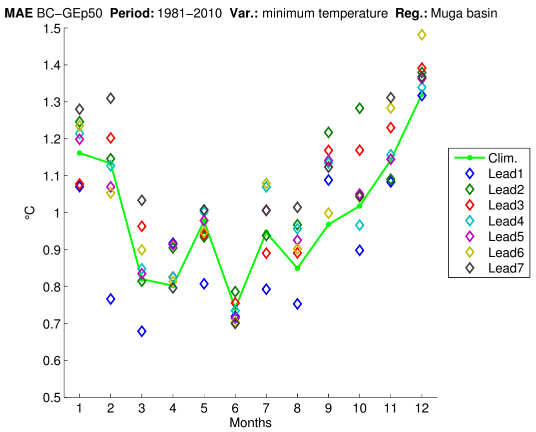
Minimum temperature lead one Taylor diagrams show that the original S4 forecasts exhibit better performance than the LR and MOS-analog calibrations. The best predicted month is February whereas the worst is September (see figure 6.32). Once the original S4 forecasts are mean bias corrected the MAE is improved beyond the original S4 and MOS-analog and LR calibrations. Figure 6.33 displays a surpassing of climatology skill in all months except April, September and December. In some particular cases, such as March, the MOS-analog gives better results. The biggest amelioration is observed in the January-March period, while the smallest is found in June and November. The dBSS assessment displays the best results for the mean bias corrected approach. In the lowest tercile we have positive values for February, May, June, August and October. In the middle thirtile only February is improved. Finally, for the upper thirtile we find positive values in January, February, March, October and June. As for persistence control, it is always the worst option to choose.



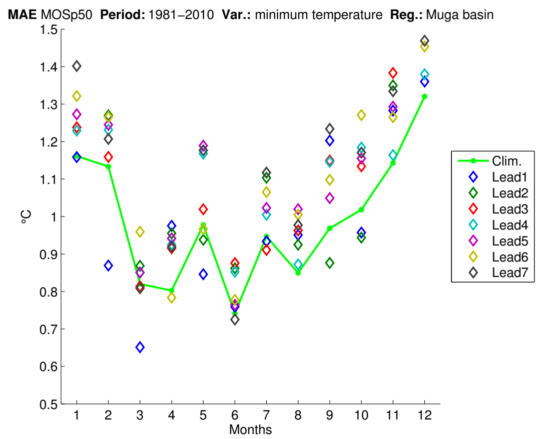
(a) Best

(b) Worst

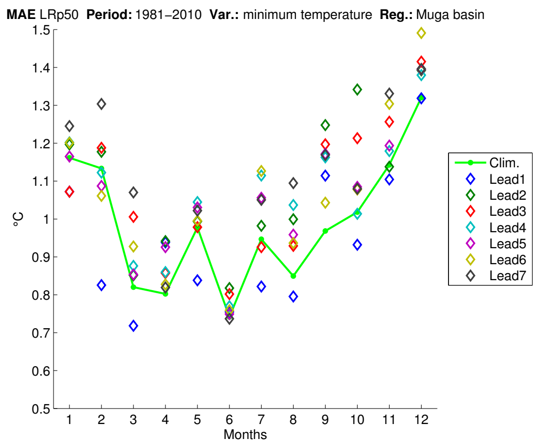
Figure 6.32. Muga basin's Taylor diagrams for minimum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: February (b) The worst performing month: September.



(a)

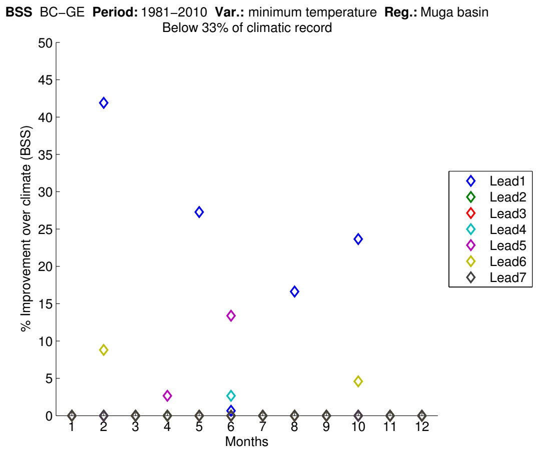


(b)

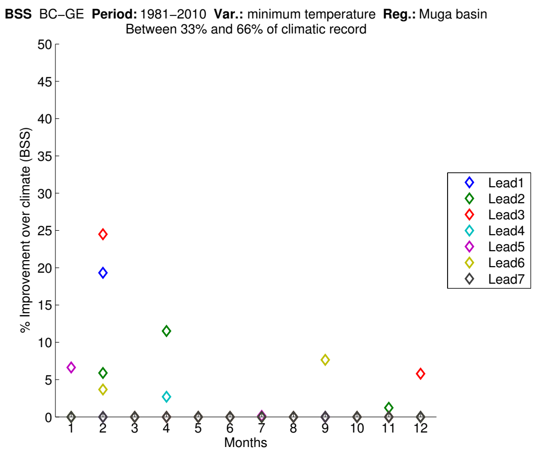


(c)

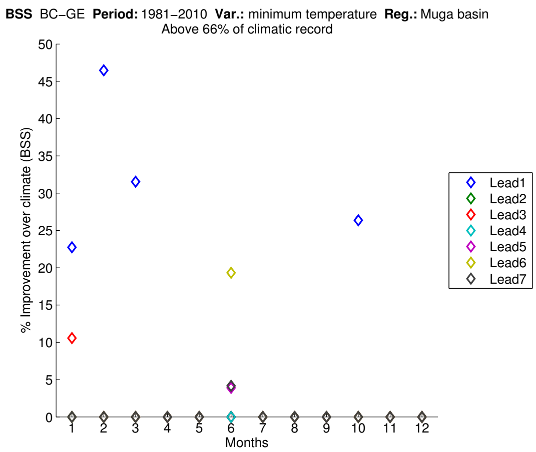
Figure 6.33. Muga basin's minimum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.



(a)



(b)

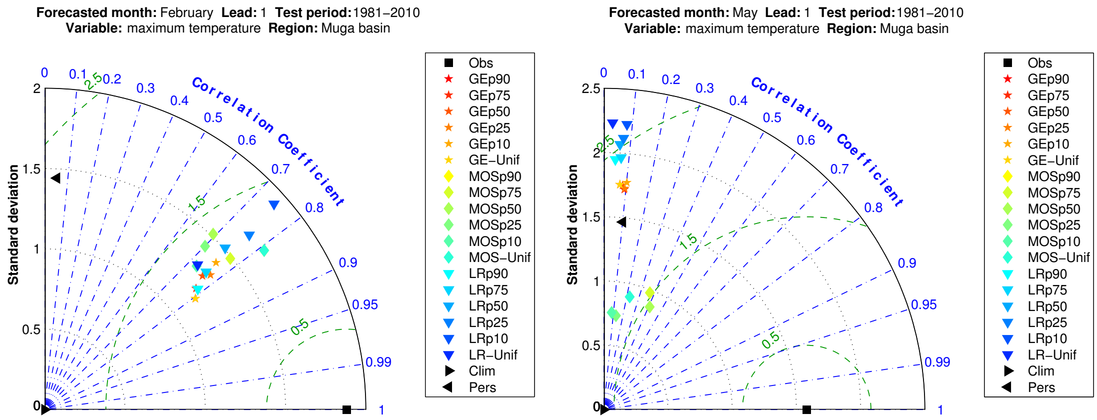


(c)

Figure 6.34. Muga_basin's minimum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) Lower tercile (b) Middle tercile (c) Upper tercile.

Maximum temperature

The analysis for maximum temperature is similar to the other variables for the same area. The comparison of the lead one original S4 forecasts with its LR and MOS-analog calibrations through Taylor diagrams displays the finest results for the former. February is the best predicted month, whilst May is the worst (see figure 6.35). When the original S4 is mean bias corrected it becomes the preferable approach (see figure 6.36). It improves climatology in all months except May and September. In this case we can also observe some consistency in the amelioration at lead-6 in March-June. Turning to the analysis of the BC/original S4 we can see positive values at multiple leads in the three categories (see figure 6.37). However, the best results are always attained at lead one. In the lowest thirtile the amelioration can be observed in the January-March period, June, August, October and November. In the middle thirtile lead one improvement is seen in October. As for the upper category there is an enhancement between December-March and August. An interesting feature of this assessment is the apparent consistency of lead-6 predictability since we have found positive values for the period March-May in the lower tercile; April-May in the middle; and May-June in the upper. As for persistence control, it is always the worst option to choose.



(a) **Best** (b) **Worst**

Figure 6.35. Muga basin's Taylor diagrams for maximum temperature at lead one for the best and worst performing months. Each diagram contains the GE for raw S4 forecasts, MOS-analog and MOS-LR S4 calibrations, their unified deterministic forecast as well as persistence and climatological controls. The period of study is 1981-2010 (a) The best performing month: February (b) The worst performing month: May.

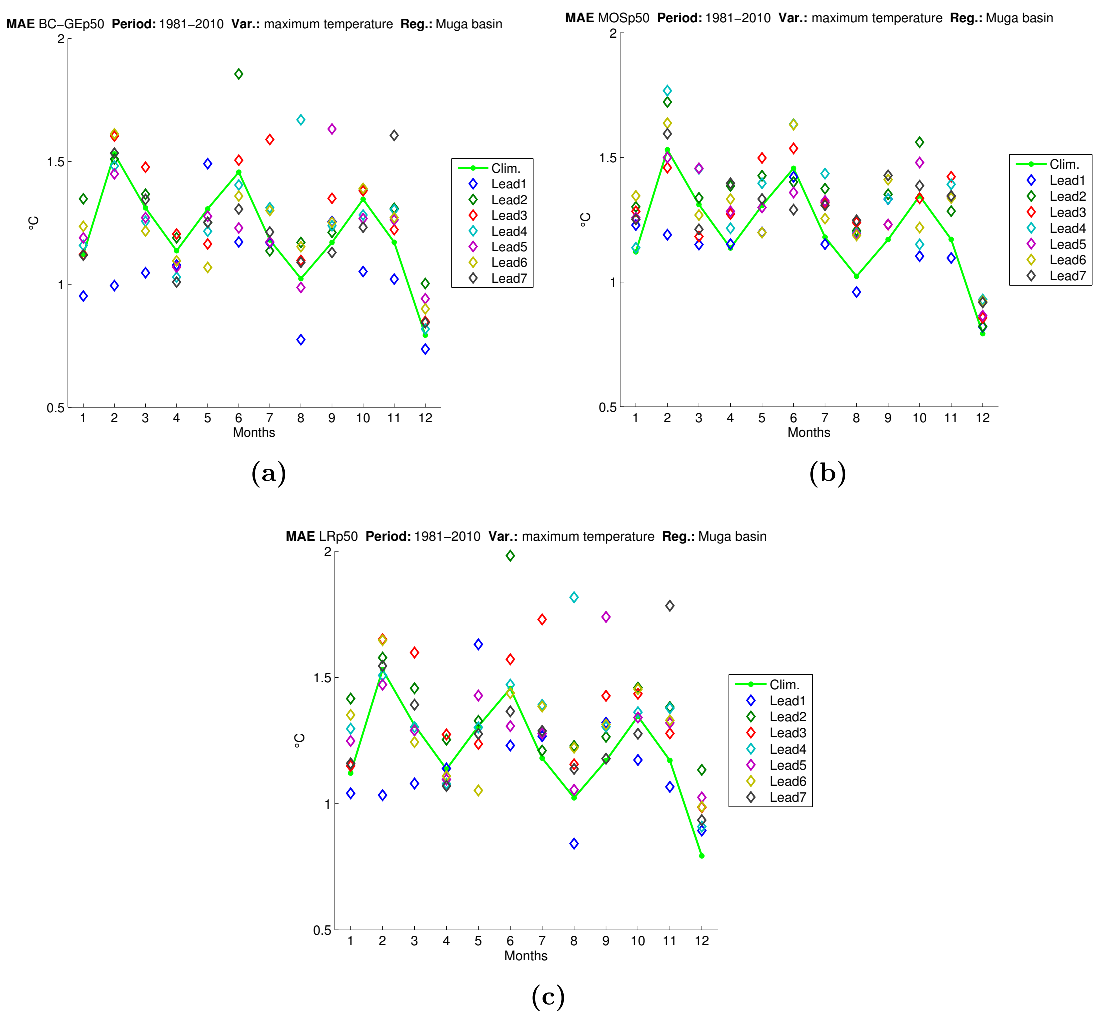


Figure 6.36. Muga basin's maximum temperature MAE plot for every month and lead including climatological control. The period of study is 1981-2010 (a) GE-50th percentile of bias corrected S4 (b) GE-50th percentile of MOS-analog calibrated S4 (c) GE-50th percentile of MOS-LR calibrated S4.

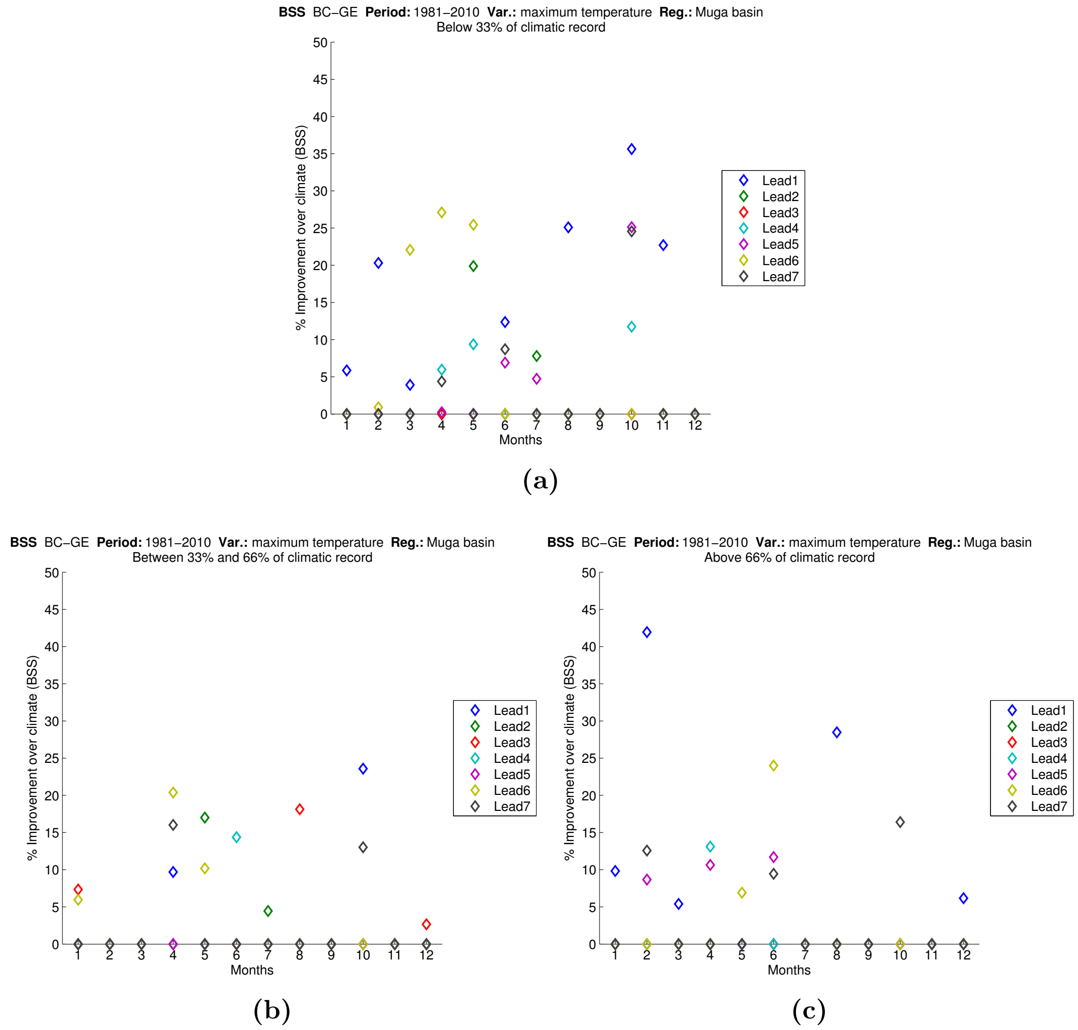


Figure 6.37. Muga_basin's maximum temperature dBSS percentage improvement over climate for all months and leads for BC-GE S4 (a) Lower tercile (b) Middle tercile (c) Upper tercile.

6.4 Discussion & Conclusions

In this chapter we have assessed current S4 skill in *a*) monthly precipitation, *b*) minimum temperature and *c*) maximum temperature as well as its possible improvement through the implementation of three calibration strategies: MOS-analog, linear regression and mean bias correction. They have been tested in four extra-tropical domains from lead-1 to lead-7 to evaluate their capabilities in a diversity of targets for seasonal forecast applications. These areas range from continental to grid-point scales and they are: Europe, Spain, Catalonia and the Muga river basin. The rearrangement of the original S4 information has been achieved through the so-called *generalized ensemble*, GE, and the construction of a *unified* deterministic forecast from itself. Besides, all the forecasts have been compared with E-OBS observations as well as E-OBS *climatology* and *persistence* as benchmark controls. This comparison has been eased by bi-linearly interpolating the original S4 to match the E-OBS grid resolution.

In the MOS-analog case we have also analysed the effect of the domain upon the analog search and subsequent calibration and we have found that for each region there is a certain number of analogs that optimize the representativeness of each member of the analog pool. For example, when working with the Catalonia domain we have found that the results were worse when extracting the Catalonia area from the Spanish or European domains than if we directly searched for analogs upon the Catalonia region. This might mean that a certain number of historical analogs may suffice or not depending on the number of possible situations that can arise on the studied region, that is to say, on the inherent variability of the variable field map. This result is in accordance with van den Dool (1994) findings and may be used to derive an expression to relate the inherent variability of a region with the optimal number of analogs needed to find similar monthly situations within an arbitrary error band. Conversely, this could also serve to state the maximum variability for which an analog pool is able to offer useful analogs. However, this hypothesis is beyond the objectives of this thesis and it is left for future studies.

Additionally, we have conducted seven MOS-analog experiments to take into account the influence of the number of analogs chosen, the methodology of analog aggregation, the variability threshold in the principal component selection and the total samples of the analog pool. The best results showed up with 5 analogs, mean aggregation, 29 analog-pool (corresponding to the same month for which we searched the analogs) and a retained variability of 99%.

The linear regression (LR), on the other hand, presented the initial controversy of considering its calibration equation with or without the independent term. We decided to follow the approximation that a perfect model would show a $1:1$ correspondence with the observations so the independent term should not appear. Some authors, like Benestad *et al.* (2008), state that the inclusion of this term is both necessary and useful because it accounts for the mean bias of the model. In practice, we have found that pushing the linear calibration through the origin inherently corrects part of the bias, as we have seen in the MAE plots. Notwithstanding, even if we had decided to include the independent term to correct the mean bias after the linear regression calibration, we can anticipate that it would add little value to the direct mean bias correction of S4 forecasts. This can be observed in the dBSS graphs where the mean bias has no effect (we work with anomalies) and the mean bias correction shows better or equal results than linear regression calibration. However, it is also true that the correction through the independent term might lead to slightly better MAE results and, hence, it is something that we would like to check in a near future. Now we can ask ourselves whether we have been able to answer the questions put forward at the beginning of this chapter. To the first one,

> Which is the current skill of S4 in the aforementioned domains? Does it depend on them?

We have seen that the original S4 forecasts excel *climatology* generally at lead one and in winter months. In Europe this is the case for maximum and minimum temperature, but not for precipitation. Spain and Catalonia improve precipitation but not temperatures. Finally, in the Muga river domain the raw lead one forecasts can surpass climatology for the three variables in winter, but also in autumn and the early spring. The observed first-lead winter predictability can be related to the stability of the winter general circulation anomalies (see i.e. Boer *et al.*, 2013). In the first horizon cases with no upgrade beyond climatology, the study of the bias seems to confirm that first order model biases are responsible for this skill reduction. At further leads climatological improvements are scarce. As for *persistence*, it is systematically surpassed by the S4 in most horizons and regions and so it cannot be considered as a functional monthly seasonal forecasting system. Hence, the next step has been to check the performance of its MOS-analog, linear regression and mean bias calibrations.

> Can S4 forecasts be improved by linear/non-linear calibrations? At which scales? At what lead-times?

After calibrating the S4 forecasts with the MOS-analog and LR approaches we have noticed a neat improvement of skill with both methods. However, the LR has given generally better results than the MOS-analog. Still, they are centred at lead one, but now with an increased number of months upgrading the climatology outcome. Occasionally, the MAE plots have also shown some ameliorations beyond climatology at other leads. This is specially perceptible in the smaller domains, Catalonia and the Muga river basin, but it also happens seldom in the larger ones. These enhancements are scarcely reflected in the dBSS graphs but are a sign that, under some particular and region-dependent circumstances, there can be foreseeable reductions of the MAE with months in advance. Yet, this predictability is not as systematic as the corresponding to lead one and, therefore, to discard the possibility of being noise, it has to be independently studied case by case (i. e. lead 6 maximum temperature forecasts for the Muga river basin for March-May).

As for the under-population of the centre members of the ensemble distribution (see figure 5.6) we have seen that it probably causes a recurrent decrease in skill when dealing with the middle tercile of each variable (with climatology giving the best results). However, since we have found positive performance in the lower and upper thirtiles we do not know if these results are enhanced or handicapped by this feature. For this reasons we plan to increase the number of ensemble members and clarify this aspect of our study at some point of the near future. This discussion leads us to the final question,

> Are our calibrations able to improve the results of a simple-bias correction? At what costs?

Following our analysis, even though the MOS-analog and LR calibrations have improved the original S4, they could not surpass the performance of mean bias corrected S4. Again, this improvement is rather limited at lead one but with visible exceptions in all domains. Incidentally, it is important to highlight that in all regions the first lead MAE of the mean bias calibrated S4 forecasts (50th percentile) improves climatology in practically all months. However, this correction seems to be more effective on temperatures and larger domains suggesting that these variables are more affected by first order biases and that these departures tend to arise more consistently on larger regions. When looking to the dBSS of the lower and upper terciles, though, these improvements are more restricted to winter and some scattered months in other seasons.

Going more specifically, and still focusing in the first lead, in the European domain all the variables are best forecast in winter (though June and July are also good months). In Spain, this is also the case for precipitation and minimum temperature, where winter is the best forecast period and August, July and June also show good performance. For maximum temperatures, on the other hand, February, June and October are the highest upgrade exponents. In Catalonia, February and October are the months which show the best improvements for the three variables. Besides, in the case of minimum temperature, winter is also a good lead one forecast period. Finally, in the Muga river basin we find that for precipitation the February-March and October-November periods show the largest ameliorations. The same months are found for minimum temperature only exchanging November by January. Finally, for maximum temperature, the best months are January-March, August and October-November.

The recurrent predictability of February and October in the Spanish Mediterranean (sometimes even beyond lead one) is highly valuable because both are key months in water management (ACA personal communication). Eventually, it is worth remarking that there are occasions in which the MOS-analog calibration gives larger ameliorations than the mean bias correction (and linear regression calibration). This points to the fact that, even if the linear calibration method offered a general better behaviour than MOS-analog calibration, the introduction of a larger analog pool could upgrade the MOS-analog performance. This is a recurrent problem when working with analogs but in this case, since our calibration pool is a re-forecast, it is feasible to think that somewhere in the future this might be further extended in the past. In such a case we would be able to check whether the MOS-analog calibration is improved by the pool expansion or not.

Ultimately, there are a couple of remarks to be made. The first deals with the unification strategy presented in section 6.2.3, since it showed a somewhat erratic behaviour, alternating good and bad results. Two reasons may explain this outcome. The first one is concerned with the possibility that no specific percentile of the *pdf* gives better results at each of the grid-points. The second is that the differences between the discriminating parameter (MAE) among the candidate members are too small and the method cannot efficiently distinguish the best option. One solution could be, either changing the discriminating parameter or increasing the re-forecast sample. We plan to address this in a future study, which simultaneously, will help us to establish which one of both hypothesis is correct. The second, and most important, is the role that our GE together with the Taylor diagrams has had in the study of the ensemble variability and skill. Particularly, their use showed to be very practical to assess the variability obtained through the ensemble in comparison to the observed and also to judge whether our calibrations or the model itself were able to add temporal skill beyond the spatial climatic characterization of the variable.

CHAPTER 7

Seasonal forecasting applications: water reservoirs

7.1 Overview

In chapter 2 we have seen that one of the most important applications of seasonal forecasting is the prediction of droughts. Humid and dry periods are, in fact, the main drivers of water resource distribution and, therefore, the main actors in its management. In the Mediterranean type ecosystems, for example, summer is typically a season of major hydrological stress. This causes a critical increase in water demands (specially for agriculture) that often poses strain on the available supply systems. Regions such as Spain, and particularly Catalonia, usually rely on the provision of reservoirs to guarantee the fulfilment of these demands. If we consider the growth of demographic and environmental pressures on water demands (Iglesias *et al.*, 2006) and the Mediterranean climatic tendency to drought (see i.e. Brewer *et al.*, 2007; Nola *et al.*, 2008; Turco and Llasat, 2011; Quintana-Seguí *et al.*, 2011), we can infer that Mediterranean regions are areas of increased vulnerability towards water scarcity (see i.e. Garrote de Marcos and Cubillo, 2008).

Studies such as Rossi *et al.* (2012) or Chavez-Jimenez *et al.* (2014) show that establishing warning thresholds to a reservoir can improve its management by enabling its administrators to adopt measures when the water amount descends below certain levels. For instance, if in May we have our reservoir at a 60% of its capacity and by looking to the historical record we know that in this situation there is a probability of 10% that by September we cannot satisfy water demands we can activate a protocol to adapt the decisions of the forthcoming months to prevent or minimise the hypothetical restrictions. Our window of opportunity arises because instead of looking into the climatological record in search for analogue events, we can try to take advantage of the memory of the system and the S4 seasonal information. To investigate whether we can improve the use of climatology with this strategy we have to answer a set of questions:

1. Can we model monthly anomalies for the in-flow, out-flow and volume stored in a reservoir through a multiple linear regression (MLR) approach?

2. Can we issue seasonal forecasts of these anomalies? Can these forecasts be more useful than climatology?

To do so we will develop monthly MLR models based on combinations of different predictor anomalies: *a*) maximum temperature, *b*) minimum temperature, *c*) precipitation, *d*) observed in-flow and *e*) observed volume; to foresee anomalies in the three main variables of the reservoir, *in-flow*, *out-flow* and *stored volume*. For further details on these datasets see sections 3.1, 3.2 and 3.3. These forecast configurations comprise *persistence* and *climatology* controls and MLR models combination of the antecedent observed conditions with *a*) climatology; *b*) ECMWF System-4 (S4); *c*) MOS-analog calibrated S4; and *d*) Linear regression calibrated S4 (LR). Since we work with anomalies, the S4 is already mean bias corrected. We also maintain the MOS-analog and LR calibrations because at a grid-point level there were some months that surpassed the S4 skill (see chapter 6).

To test the applicability of this methodology we have chosen the Boadella reservoir in the upper Muga river basin because it accomplishes the requirements for a case study of this kind (see sections 4.4 and 4.5). It satisfies water demands from urban areas and agriculture uses and it does not have any human up-stream structures that can interfere in its modelling. Besides, it also spans the 30-year period of the S4 re-forecast. Finally, the verification process will help us determine the best option among the six approaches.

Thus, we have organized this chapter as follows: firstly, *Methodology* section presents and justifies the selection of the model strategies and techniques; then, *Results* synthesises the outcome of the research performed; and finally, in *Discussion & Conclusions* there is an analysis of the answers to the questions raised.

7.2 Methodology

The statistical models built to foresee monthly in-flow, out-flow and volume anomalies in the Boadella reservoir are based on a multiple linear regression approach (MLR; see section 5.6 and i.e. Adamowski and Karapataki, 2010; Bao *et al.*, 2012; Adamowski *et al.*, 2012; Fan *et al.*, 2015). We have worked upon forecast and observed anomalies because we wanted to avoid hypothetical interferences of the year cycle (i.e. Wilks, 2006). Therefore, for every grid-point we have computed S4 and E-OBS monthly median values for the period 1981-2010 to subtract them from each forecast/observed value. Before any further step, though, we had to establish which were the potential predictors for each of our predictands. In the first place we have considered that the anomalies of the Muga in-flow are due to anomalies in precipitation, maximum and minimum temperatures because the response of the sub-basin is quite linear with precipitation (the observed rainfall and in-flow series have a Pearson correlation coefficient of 0.65, pval < 0.05), there are no up-stream human structures that can affect the river, and maximum and minimum temperatures act as proxies for evapotranspiration (see i.e. Thornthwaite, 1948; Xu and Singh, 2001; Hobbins *et al.*, 2008). In this way we keep the models simple, away from derived quantities and/or parametrizations. Secondly, volume anomalies have been regarded as dependent on the same variables as before but with the addition of in-flow anomaly observations. Finally, the out-flow monthly anomaly models should rely on E-OBS rainfall, maximum and minimum temperature anomalies as well as volume anomaly observations. The reason not to consider volume anomalies as out-flow dependent is because from a decision-maker perspective the ultimate driver of the out-flow anomalies are the existing anomalies in volume and not the other way around (EDF-DTG personal communication).

After this initial verification we have proceeded to build multiple MLR models for each month and variable to seek for the model with the greatest explanatory power and the smallest number of predictors. These predictors comprise monthly to several-month anomaly accumulations and its inclusion in each MLR model has been done through a total screening process. Furthermore, since our objective is to work with antecedent observations as well as forecasts we have considered predictors ranging up to one year in the past (with respect to the month that we wanted to model). This one year threshold has been established in accordance to the *spin-up* stage of hydrological modelling, because the system normally takes one-year to lose the memory of its initialisation conditions (see i.e. Ajami *et al.*, 2014; Rahman and Lu, 2015; Seck *et al.*, 2015). Additionally, we have limited ourselves to a maximum of five predictors to prevent model over-fitting as well as to only retain the predictors with greater physical meaning. The choice of total screening instead of a step-wise regression is because we have developed an algorithm for the former that allows us to easily take a deeper control of the selection process by setting a maximum number of predictors and separately analysing the relative importance of them (i.e. in groups of two, three, four, etc.). Nevertheless, we have also searched for the best models of one, two and three predictors with step-wise regression and found the same models as with total screening. That said, it is important to bear in mind that the total number of possible combinations for an established number of predictors can be computed through the binomial coefficient without repetition from (7.1),

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \tag{7.1}$$

where n is the total number of possible predictors ($n = 12^2\times$ n° of variables); and k is the number of predictors in our MLR model. The reason for the number 12 to be squared is because we also consider as possible predictors accumulated anomalies through different consecutive months. In the final MLR, however, we only allow each month to appear once for the same variable (i.e. we cannot have TN_{3-6} and TN_{5-6} in the same MLR). This initial phase, devoted to predictor choice, involved working with E-OBS data as well as in-flow and volume observations in an in-sample and perfect prognosis scheme. Moreover, to avoid artificial skill due to any existing trends all the predictors and predictands have been linearly de-trended previously (see i.e. Boer *et al.*, 2013; Benestad *et al.*, 2008). As for the ordering of the best predictors to build our models we have considered different verification metrics: *a*) coefficient of determination; *b*) Akaike information criterion; *c*) MAE; *d*) RMSE; and *e*) trend and autocorrelation of the residuals. Predictor combinations with trend and/or autocorrelation in the residuals have been discarded in favour of similar models without these features because they point towards flaws in the model (Wilks, 2006). Thus, at the end of this process we have obtained one MLR model for each month, making a total of 36 models for the study. Tables 7.1, 7.2 and 7.3 summarise the models for each month and variable with the predictors ordered in accordance to their explanatory power.

In the case of in-flow anomalies (see table 7.1) we can see that the best predictors are rainfall anomalies for the same predicted month or the accumulated anomalies for the same month and the previous one. This is not surprising for the river flow in this upper sub-basin has a rather linear response with precipitation at a monthly scale. The influence detected in months far away the same modelled month are probably due to the effect of rainfall anomalies upon soil moisture, groundwater and their impacts on the river flow (i.e. Sear and Dawson, 1999; Sophocleous, 2002). The second most influencing predictor at in-flow level is the maximum temperature, indicative of the effects of evapotranspiration on soil moisture and run-off. In fact, as the soil becomes more (less) saturated, the response of the rainfall threshold to produce run-off becomes smaller (greater) and, consequently, there is an enhancement (decrease) of the linear relationship between rainfall and stream-flow. Finally, minimum temperatures act as a proxy of the snow amount in winter (Bednorz, 2004) and also as a less effective proxy of evapotranspiration.

Table 7.1. Performance of the best predictor combinations for in-flow monthly anomalies in the Boadella reservoir (precipitation, *rr*; maximum temperature, *Tx*; and minimum temperature, *Tn*). The subscript numbers indicate the months for which the anomalies are accumulated. Each row contains the month, the best predictor combination and the reproduced variance, R^2, in LOOCV perfect conditions.

	Best Predictor Combination	$\mathbf{R^2}$
Jan	$\left\{rr_{(12-12)}, rr_{(1-1)}, Tx_{(1-1)}, Tn_{(10-1)}\right\}$	0.81
Feb	$\left\{rr_{(12-12)}, rr_{(1-2)}, Tx_{(6-9)}, Tn_{(11-1)}\right\}$	0.77
Mar	$\left\{rr_{(5-8)}, rr_{(1-2)}, Tx_{(7-7)}\right\}$	0.41
Apr	$\left\{rr_{(4-4)}, rr_{(12-3)}, Tx_{(6-6)}, Tn_{(5-7)}, Tn_{(8-10)}\right\}$	0.66
May	$\left\{rr_{(4-5)}, rr_{(7-7)}, rr_{(10-10)}\right\}$	0.73
Jun	$\left\{rr_{(6-6)}, Tn_{(9-9)}, Tn_{(11-11)}\right\}$	0.63
Jul	$\left\{rr_{(6-7)}, Tx_{(10-1)}, Tn_{(11-12)}\right\}$	0.79
Aug	$\left\{rr_{(7-8)}, Tx_{(1-7)}, Tn_{(10-1)}\right\}$	0.51
Sep	$\left\{rr_{(6-9)}, Tx_{(3-3)}, Tn_{(8-8)}\right\}$	0.31
Oct	$\left\{rr_{(9-10)}, Tx_{(11-3)}, Tn_{(1-2)}\right\}$	0.76
Nov	$\left\{rr_{(11-11)}, rr_{(7-7)}, Tn_{(4-5)}\right\}$	0.57
Dec	$\left\{rr_{(10-12)}, Tn_{(5-7)}, Tn_{(8-9)}\right\}$	0.55

Turning to volume anomalies (see table 7.2), the predictor with most explanatory power is the accumulated in-flow anomaly observed in the preceding months and/or for the same month. This is logical since the main driver of the stored volume in a reservoir is its supplying river flow. This is followed by antecedent accumulated rainfall anomalies (from antecedent months and/or for the modelled month), which are related to the groundwater state (see i.e. Sophocleous, 2002). Finally, we have maximum and minimum temperature predictors, proxies for evapotranspiration in the reservoir and the upper sub-basin. In the case of minimum temperature, it also acts as proxy for winter snow amount.

Finally, moving on to the out-flow models we can see that they are, generally, the ones with the less number of predictors (see table 7.3). That is because in many months its modelling through a MLR has limited success and the inclusion of further predictors does not increase the performance of the model. This is consequence of the high level of human intervention in its evolution, because not only it depends on meteorological anomalies but also on other non-linear factors such as regulation protocols. Thus, if we turn our attention to the best modelled months we will see that they coincide with the maximum and the end of the irrigation season (July-August-September) with volume and temperature anomalies as main predictors. An hypothesis to explain June's decreased predictability would be that this month's outflow could be more influenced by the human decision to set the beginning and intensity of the irrigation season. Conversely, in the center and final months of summer the soil moisture conditions and irrigation needs are more settled and human decisions would be mainly driven by antecedent volume anomalies and evapotranspiration (ACA, 2009b).

Table 7.2. Performance of the best predictor combinations for volume monthly anomalies in the Boadella reservoir (in-flow, *flwin*; precipitation, *rr*; maximum temperature, *Tx*; and minimum temperature, *Tn*). The subscript numbers indicate the months for which the anomalies are accumulated. Each row contains the month, the best predictor combination and the reproduced variance, R^2, in LOOCV perfect conditions.

	Best Predictor Combination	$\mathbf{R^2}$
Jan	$\left\{flwin_{(6-12)}, Tn_{(2-5)}, Tn_{(11-12)}\right\}$	0.74
Feb	$\left\{flwin_{(5-11)}, rr_{(9-9)}, rr_{(11-2)}, Tn_{(4-5)}, Tn_{(11-12)}\right\}$	0.82
Mar	$\left\{flwin_{(4-11)}, flwin_{(2-2)}, rr_{(9-12)}, Tx_{(5-5)}, Tn_{(4-5)}\right\}$	0.72
Apr	$\left\{flwin_{(2-4)}, flwin_{(5-5)}, rr_{(7-7)}, rr_{(9-12)}, Tx_{(1-1)}\right\}$	0.69
May	$\left\{flwin_{(6-2)}, flwin_{(3-4)}, rr_{(11-5)}, Tx_{(1-1)}, Tn_{(9-9)}\right\}$	0.61
Jun	$\left\{flwin_{(11-12)}, flwin_{(3-4)}, rr_{(11-12)}, Tx_{(12-1)}, Tn_{(9-9)}\right\}$	0.62
Jul	$\left\{flwin_{(3-5)}, flwin_{(6-7)}, rr_{(10-11)}, Tx_{(9-4)}, Tn_{(6-7)}\right\}$	0.76
Aug	$\left\{flwin_{(3-5)}, flwin_{(6-7)}, rr_{(9-11)}, Tx_{(9-9)}, Tn_{(6-7)}\right\}$	0.76
Sep	$\left\{flwin_{(3-5)}, flwin_{(6-7)}, Tx_{(7-8)}, Tn_{(6-7)}\right\}$	0.81
Oct	$\left\{flwin_{(3-5)}, flwin_{(6-8)}, rr_{(7-10)}, Tx_{(6-6)}, Tn_{(11-5)}\right\}$	0.85
Nov	$\left\{flwin_{(4-4)}, flwin_{(7-10)}, rr_{(6-11)}, Tx_{(2-2)}, Tn_{(12-2)}\right\}$	0.86
Dec	$\left\{flwin_{(4-11)}, rr_{(8-12)}, Tx_{(2-2)}, Tn_{(1-8)}\right\}$	0.84

Table 7.3. Performance of the best predictor combinations for out-flow monthly anomalies in the Boadella reservoir (volume, *vl*; precipitation, *rr*; maximum temperature, *Tx*; and minimum temperature, *Tn*). The subscript numbers indicate the months for which the anomalies are accumulated. Each row contains the month, the best predictor combination and the reproduced variance, R^2, in LOOCV perfect conditions.

	Best Predictor Combination	$\mathbf{R^2}$
Jan	$\{rr_{(10-1)}, Tx_{(11-1)}\}$	0.60
Feb	$\{rr_{(1-2)}, Tx_{(7-9)}, Tn_{(1-1)}\}$	0.86
Mar	$\{rr_{(7-8)}, vl_{(12-2)}\}$	0.42
Apr	$\{rr_{(10-2)}\}$	0.32
May	$\{rr_{(4-5)}, vl_{(10-4)}\}$	0.55
Jun	$\{vl_{(3-4)}\}$	0.33
Jul	$\{Tx_{(4-4)}, Tn_{(10-11)}, vl_{(3-5)}\}$	0.75
Aug	$\{vl_{(3-5)}, vl_{(6-6)}, vl_{(7-7)}\}$	0.90
Sep	$\{Tx_{(5-6)}, vl_{(7-7)}, vl_{(8-8)}\}$	0.85
Oct	$\{rr_{(1-4)}, rr_{(6-6)}, Tn_{(7-7)}\}$	0.51
Nov	$\{rr_{(1-6)}, Tn_{(4-5)}\}$	0.18
Dec	$\{rr_{(10-12)}, Tx_{(9-11)}, Tn_{(10-12)}\}$	0.18

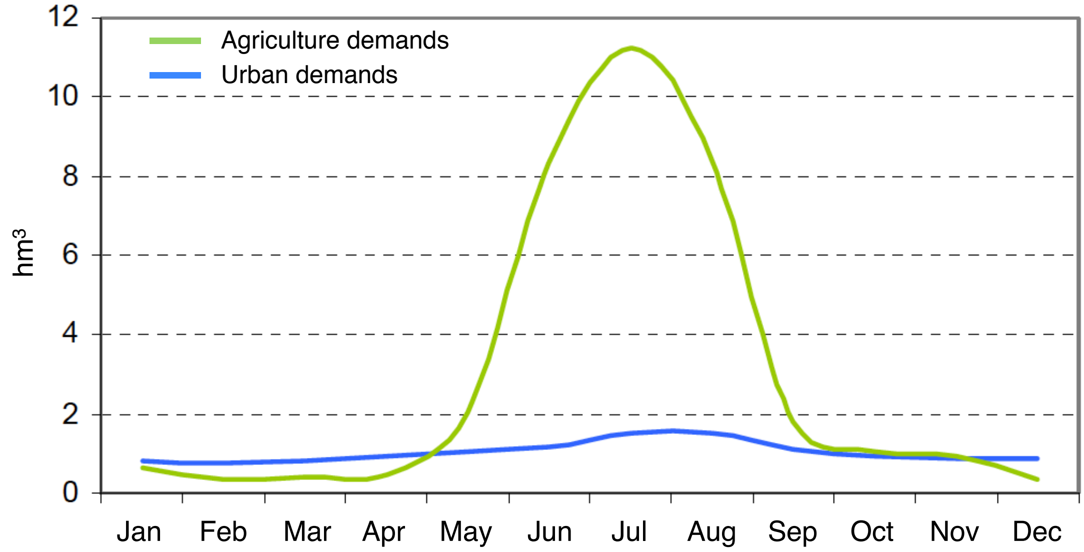


Figure 7.1. Boadella monthly water demands from agriculture and urban areas (in hm^3). Adapted from ACA (2009b).

Concerning the other months, from October to May the main predictors are the accumulated rainfall, temperature and volume anomalies. During these months the out-flow is determined by a combination of human and meteorological factors, without anyone being clearly predominant (see figure 7.1).

Afterwards we have proceeded to forecast the monthly anomalies with those models. To do so we have carried out a LOOCV approach (see section 5.4). The uncertainty on the series has been estimated with a bootstrap method based on a methodology of Calmanti *et al.* (2007, see also 5.4). Subsequently we have tested the aforementioned models with six strategies: *a*) Climatology; *b*) Persistence; *c*) Antecedent observations + climatology (A+Cl); *d*) Antecedent observations + S4 anomalies (A+S4); *e*) Antecedent observations + MOS-analog calibrated S4 anomalies (A+MOS); and *f*) Antecedent observations + LR calibrated S4 anomalies (A+LR). The first three approaches act as controls, and they comprise three approximations that only use observations as input data; the other three combine both observations and S4 calibrated forecasts.

As for the in-flow and volume anomalies acting as predictors, aside from using direct observations, we have also derived the corresponding series for each of the forecast leads from the different model approaches. Finally, we have conducted a verification analysis of each forecast lead (up to seven months), based on a selection of deterministic and probabilistic criteria. In the deterministic approach we have considered the 50th percentile for the S4 and its calibrations, whereas in the probabilistic approximation we have used the information coming from the ensemble to test the performance of the forecast methods in three conditions: dry (lower climatic thirtile), normal (middle climatic thirtile) and humid (middle climatic thirtile). The deterministic metrics applied have been the MAE and the Pearson correlation coefficient. In the MAE comparison of our forecast systems against climatology we have set a threshold of 5% MAE amelioration below which we do not consider that our method enhances climatic results to try to minimize random noise influence. With respect to the probabilistic validation we have chosen the Attributes diagram and the Economic Value curve (EV; see section 5). When dealing with the EV curves it is important to remark that we have established a threshold of 0.10 EV Area above which we consider the method better than climatology. Actually, we have observed that EV Areas below 0.10 imply value for users when no other skill metric offers positive results. We hypothesise that this behaviour might be related to the intrinsic noise of this scale.

7.3 Results

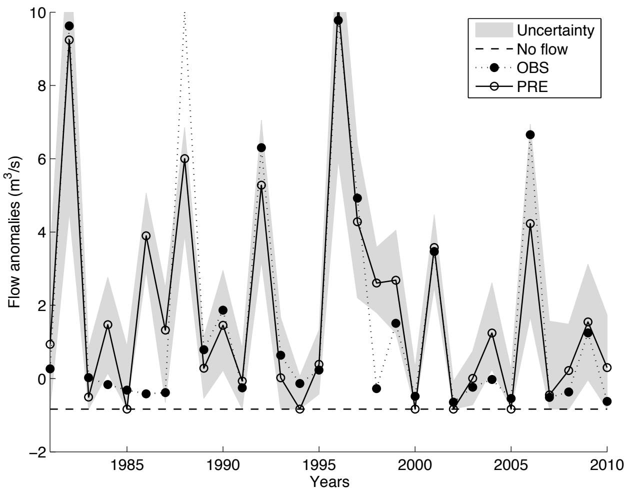
In the previous section we have introduced the framework to obtain the MLR models for each variable. Here we will compare the performance of these models under *perfect prognosis* with the results of *climatology* and *persistence* approaches, and also with the MLR forecasts coming from the combination of antecedent observations with climatology, S4 and its MOS-analog and LR calibrations (see subsection 7.2).

Initially, we have modelled the hindcast anomalies (1981-2010) in a LOOCV *perfect prognosis* approach, that is, knowing the observed values for the modelled month. This initial step provided us with a deterministic benchmark for the study (see tables 7.1, 7.2 and 7.3). It is important to remark that, even we have also considered the perfect prognosis approximation in the probabilistic verification, it is no longer a benchmark in that case. In this way we have found that the inflow perfect prognosis simulation attains an amelioration over MAE climate up to 60%, with January being the best month and September, the worst, with no amelioration whatsoever. Their corresponding Spearman correlation coefficients range from 0.55 in September to 0.90 in January (pval < 0.05). Volume anomalies, on the other hand, improve climatology in all months, from a minimum 35% MAE enhance in October to more than 65% in November and December. In this case the Spearman correlations go from 0.75 to 0.95 (pval < 0.05). The worst month is October and the best, November and December. Finally, out-flow anomalies are generally the most difficult to be modelled through the MLR, with Spearman correlation coefficients from 0.45 to 0.85 (pval < 0.05). The finest results are focused in July, August and September, with potential ameliorations from 30% to 50%. The poorest results comprise April, November and December, with any upgrade beyond climatology. In figure 7.2 we can see the best modelled month for each of the variables.

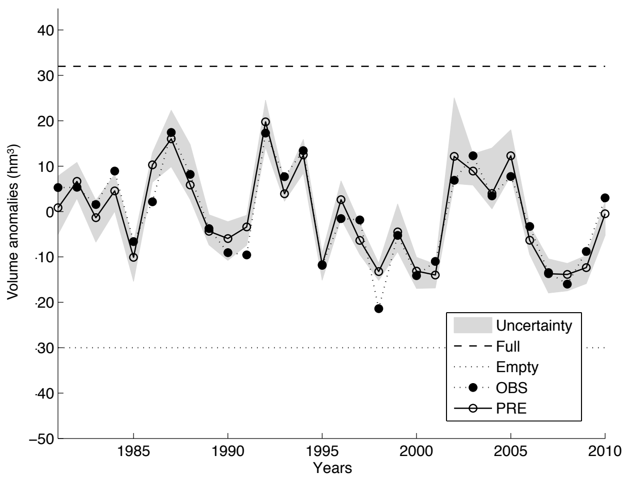
Afterwards we have proceeded to compare these results with the other forecast configurations and controls, taking climatology as the reference forecast (both deterministically and probabilistically). This choice responds to the fact that it is the usual strategy for reservoir's decision making. To ease the readability of this section the results will be presented separately for each variable.

In-Flow

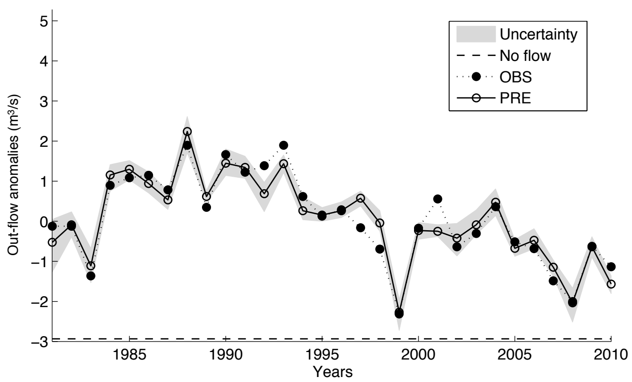
If we analyse the in-flow MAE the best results are obtained at lead one, where in all months except April, June, September and December there is always one or multiple forecast configurations that go beyond the climatology control (see figure 7.3). In general, the most consistent approach at this lead is the use of the MLR models with antecedent anomaly observations, A+Clim. However, in May and October persistence offers better results and, in July, the combination of antecedent observations with the MOS-analog calibrated S4, A+MOS, is the finest. Nevertheless, in all these cases the A+Clim remained as the second option. At lead two we have that A+Clim keeps on showing good performance, surpassing climatology in March, July and August. At further leads only three months can go beyond climatology: February, A+Clim; July, A+MOS; and November, A+S4. At longer leads there is no consistent pattern among the best methods and



(a) January in-flow anomalies



(b) November volume anomalies



(c) August out-flow anomalies

Figure 7.2. Modelled and observed in-flow, out-flow and volume anomaly values of the best MLR model in a LOOCV approximation and perfect prognosis conditions.

in the occasions some skill is shown there is an alternation among persistence, A+Clim, A+MOS and A+S4. The observed predictability in July, though, is an interesting result for this area, attending the potential of seasonal forecasting in summer water management. In figure 7.4 we can see that the percentage of October's in-flow MAE reduction compared to climatology with the six forecast strategies considered is only perceptible at lead one. We show October's performance because it is the month with the highest rainfall in the basin (see figure 4.11).

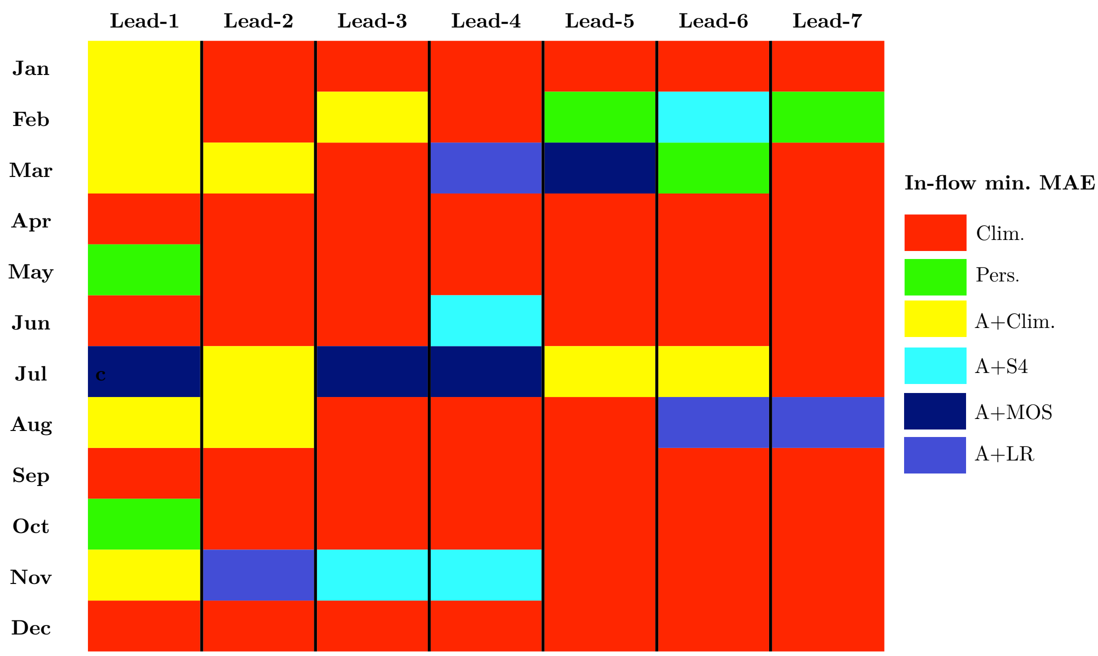


Figure 7.3. In-flow forecast strategy with the lowest MAE for each month and forecast horizon at the Boadella reservoir. The six forecast systems considered are: climatology (Clim), persistence (Pers), antecedent observation combined with climatological values (A+Clim), antecedent observations combined with mean bias corrected S4 (A+S4), antecedent observations combined with MOS-analog calibrated S4 (A+MOS) and antecedent observations combined with S4 calibrated with a linear regression procedure (A+LR). Note that forecast strategies different from climatology only appear in the table if they improve climatology MAE results by a minimum of 5%.

Moving on to the probabilistic performance our attention is centred in the envelope of the Economic Value curve and its enclosed area (EVA). We have to recall that positive values of the EVA mean that it exists a probability threshold

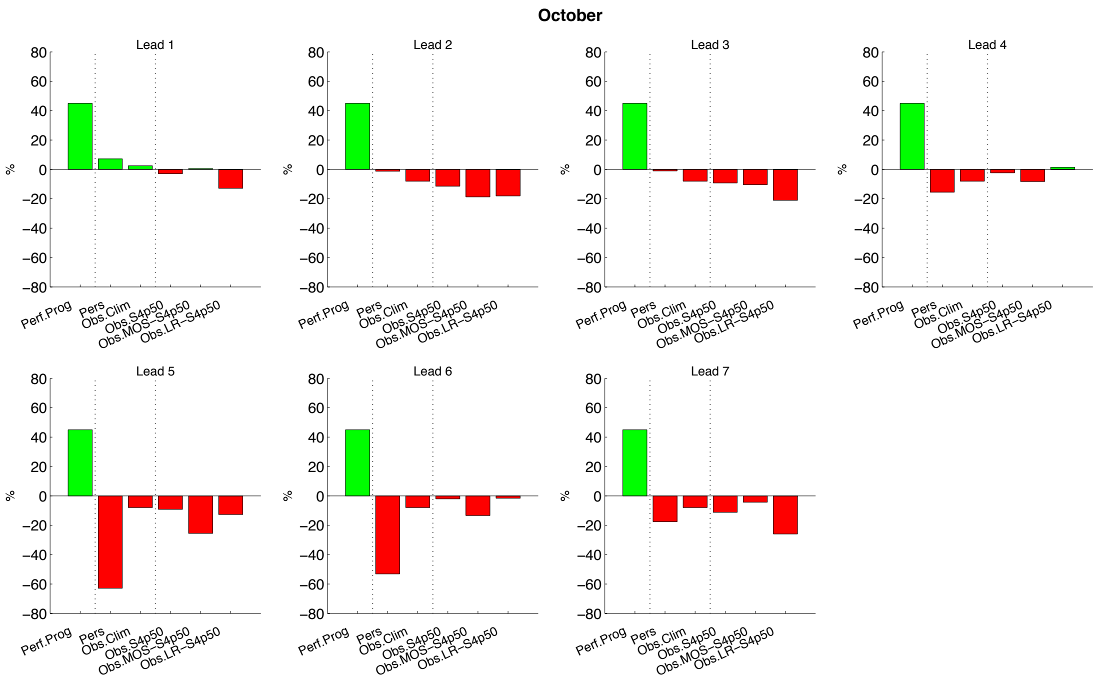


Figure 7.4. Percentage of October's in-flow MAE reduction with respect to climatology. Perf.Prog. refers to the MLR model performance with observed values. The other bars correspond to the five forecast configurations considered: persistence (Pers), antecedent observation combined with climatological values (Obs.Clim), antecedent observations combined with mean bias corrected S4 (Obs.S4.), antecedent observations combined with MOS-analog calibrated S4 (Obs.MOS-S4.) and antecedent observations combined with S4 calibrated with a linear regression procedure (Obs.LR-S4.).

that, for a certain forecast method and a specific range of users, leads to more valuable decisions than using climatology alone (see subsection 5.7.2.7). From the three situations studied it is clear that the upper and lower thirtiles are the best forecast, ameliorating climatology in many leads. The middle thirtile, on the contrary, is the worst forecast with reduced ameliorations (see figure 7.5). As we have said in section 6.2.2 this is probably due to the small number of ensemble members.

Starting with the lower thirtile (corresponding to dry conditions) we have found that the best results are attained at lead one, with an amelioration beyond climatology in all the months except December. Persistence is the option with best results in March, April, May and November. The A+S4 shows the best results in February and July; A+MOS presents the best outcome in June and September; in October the A+LR is the leading option; and in January, A+Clim At other leads there are scattered cases in which we find economic value, mostly centred in the A+MOS, but they do not follow any consistent pattern. If we turn our attention to the upper tercile we find that the EVA is positive in all months except May, June and December. Persistence is the optimal forecast system in January, March, July and September; A+Clim is the best option in February; A+S4 in August and October; and A+MOS in November. At other leads there are scattered positive values that are not systematic and depend on the month and the forecast system. In general there is a dominance of the A+MOS but persistence, A+S4 and A+LR are also good options in particular cases.

Finally, as an example, we show the EV curves for October in-flow forecasts at lead 1, 4 and 7 (see figure 7.6). This is the month with the precipitation maxima in the Muga river basin (see figure 4.11). Note that there are always users that are able to take better decisions in all three leads if they use an appropriate forecast system and that the best results are attained at lead one and for lower and upper thirtiles. Figure 7.7 details the attributes diagram for the lower and upper thirtile of this same forecast focused respectively, in A+LR and A+S4 (the best strategies at this horizon). As we can see the forecast probabilities lie inside and outside the reliability shaded region. This contrasts with the EV results highlighting the importance of facing the verification process from different perspectives.

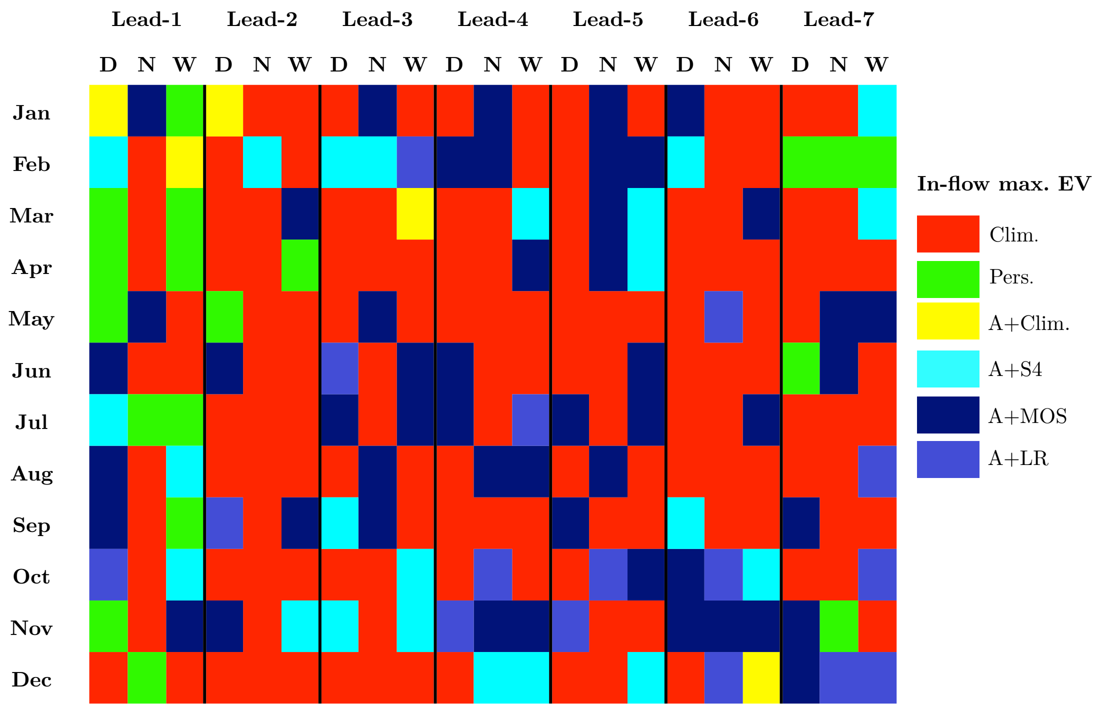


Figure 7.5. In-flow forecast strategy with the highest Economic Value (EV) for each month, forecast horizon and climatic conditions (**D** -lower tercile-, **N** -middle tercile-, and **W** -upper tercile-) at the Boadella reservoir. The six forecast systems considered are: climatology (Clim), persistence (Pers), antecedent observation combined with climatological values (A+Clim), antecedent observations combined with mean bias corrected S4 (A+S4), antecedent observations combined with MOS-analog calibrated S4 (A+MOS) and antecedent observations combined with S4 calibrated with a linear regression procedure (A+LR). Note that forecast strategies different from climatology only appear in the table if they have a minimum EV Area of 0.10.

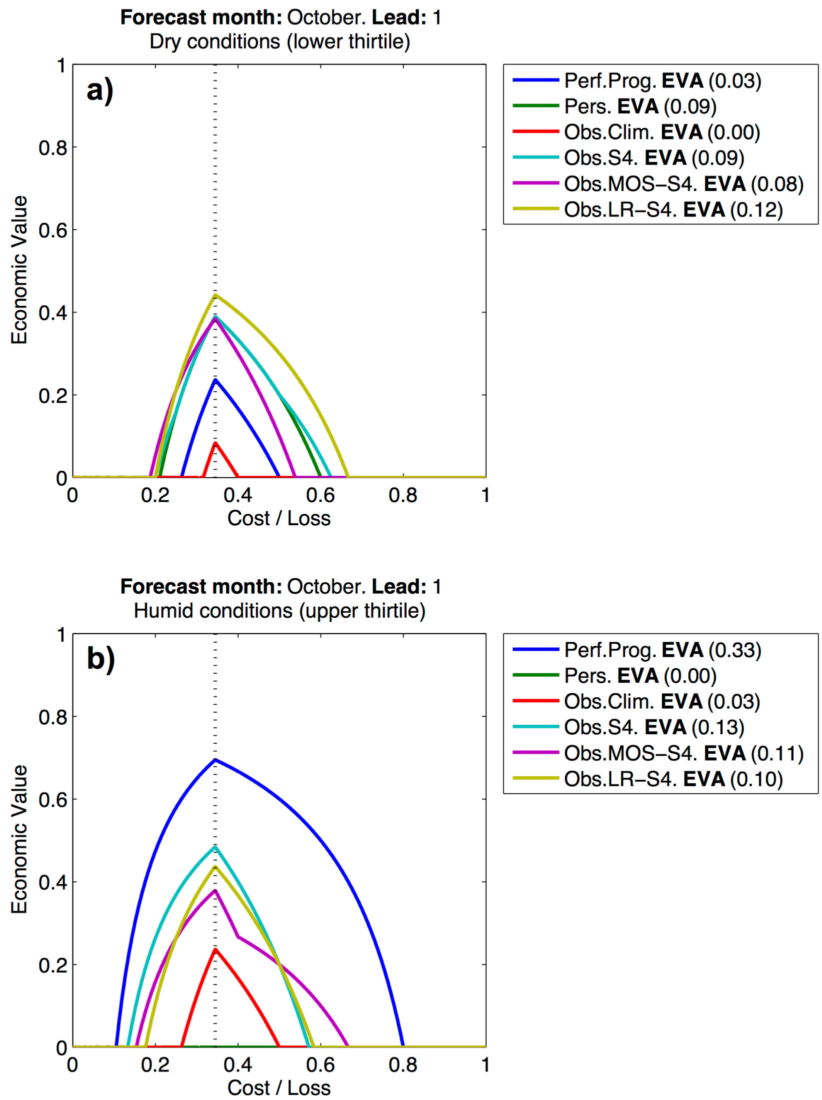


Figure 7.6. EV plot for October's lead one forecast of (a) dry conditions (b) humid conditions. The base line corresponds to climatology. Perf.Prog. refers to the MLR model performance with observed values. The other curves correspond to the five forecast configurations considered: persistence (Pers), antecedent observation combined with climatological values (Obs.Clim), antecedent observations combined with mean bias corrected S4 (Obs.S4.), antecedent observations combined with MOS-analog calibrated S4 (Obs.MOS-S4.) and antecedent observations combined with S4 calibrated with a linear regression procedure (Obs.LR-S4.).

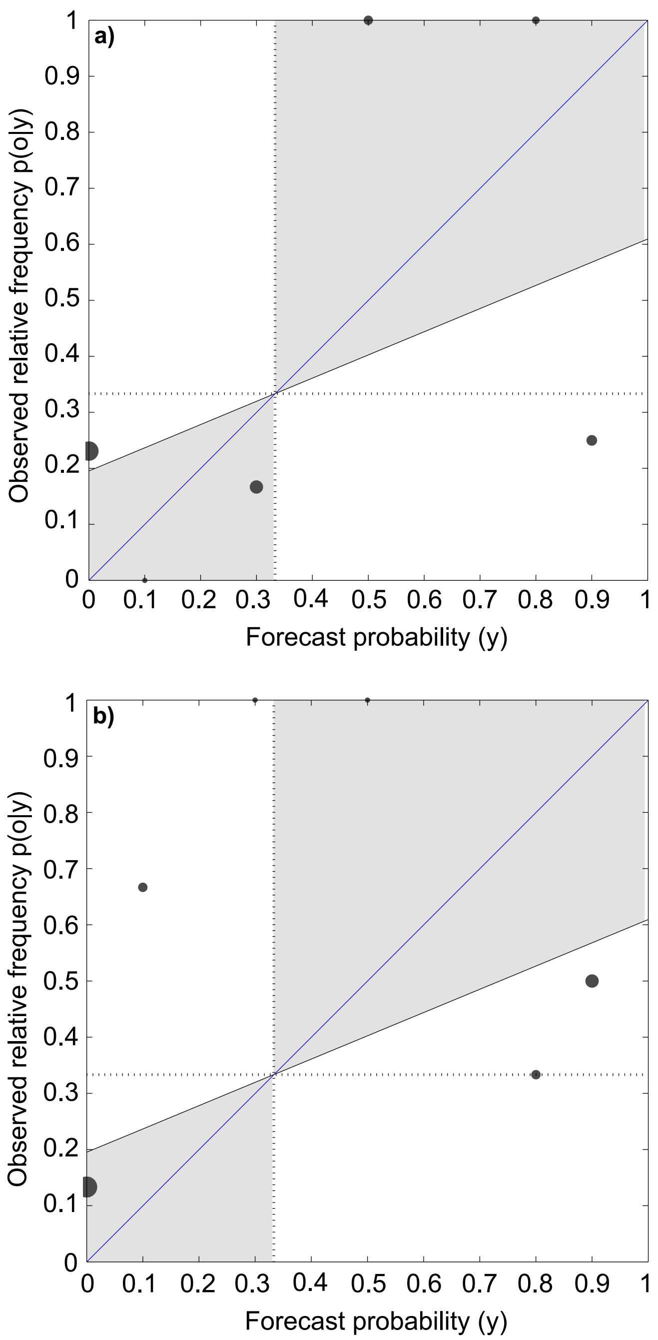


Figure 7.7. Attributes diagram for October's first lead in-flow forecast of (a) dry conditions with A+LR (b) humid conditions with A+S4. Only some of the forecast probabilities lie inside the reliability shaded region.

Volume

Volume anomalies are the best modelled variable through the MLR approximation. When looking at the MAE we see important ameliorations with respect to climate in all months up to lead four (see figure 7.8). Furthermore, in the case of July, November and December the predictability upgrade reaches lead seven; and for January and June, we have positive results up to lead six. Persistence and the A+Clim are the options which give the best results.

At the first lead, persistence shows the best results in all months except from October to January, A+Clim (October, December and January) and A+LR (November) surpass persistence. At lead two persistence keeps on being the most advantageous approach in March to August. At this same horizon the A+Clim ranks first in January and September, whereas the A+MOS should be the choice in February, persistence in October and the A+S4 in November and December. At lead three persistence is the finest forecast system between April and July, followed by the A+Clim in January, August, September and October, and A+S4/MOS/LR (with little differences among them) in February, March, November and December. At lead four A+S4/MOS/LR show better results from November to April; A+Clim should be the preferred option for August, September and October; and persistence is the best approach from May to July. Progressing to lead five, persistence is only useful in January and June; A+Clim in September; A+S4 in March, July and November; A+MOS in December; and A+LR in February. At lead six, A+MOS is the best approach in January and June; A+LR in July and December; and A+S4, November. Finally, at lead seven December and March are better forecast with A+MOS while July and November show better results with A+LR. In figure 7.9 we can see the percentage of July's volume MAE reduction compared to climatology with the six forecast strategies considered. We show July's forecast performance because it is the month with the highest hydrological stress (see figure 4.11) and when there is a peak of the irrigation demands (see figure 7.1).

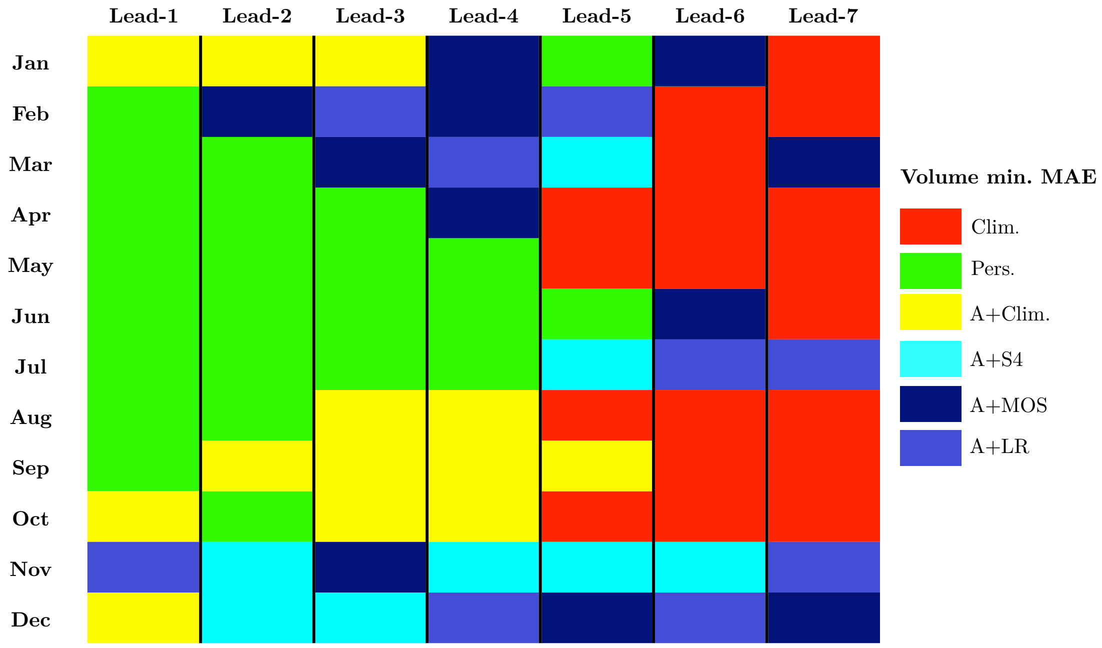


Figure 7.8. Volume forecast strategy with the lowest MAE for each month and forecast horizon at the Boadella reservoir. The six forecast systems considered are: climatology (Clim), persistence (Pers), antecedent observation combined with climatological values (A+Clim), antecedent observations combined with mean bias corrected S4 (A+S4), antecedent observations combined with MOS-analog calibrated S4 (A+MOS) and antecedent observations combined with S4 calibrated with a linear regression procedure (A+LR). Note that forecast strategies different from climatology only appear in the table if they improve climatology MAE results by a minimum of 5%.

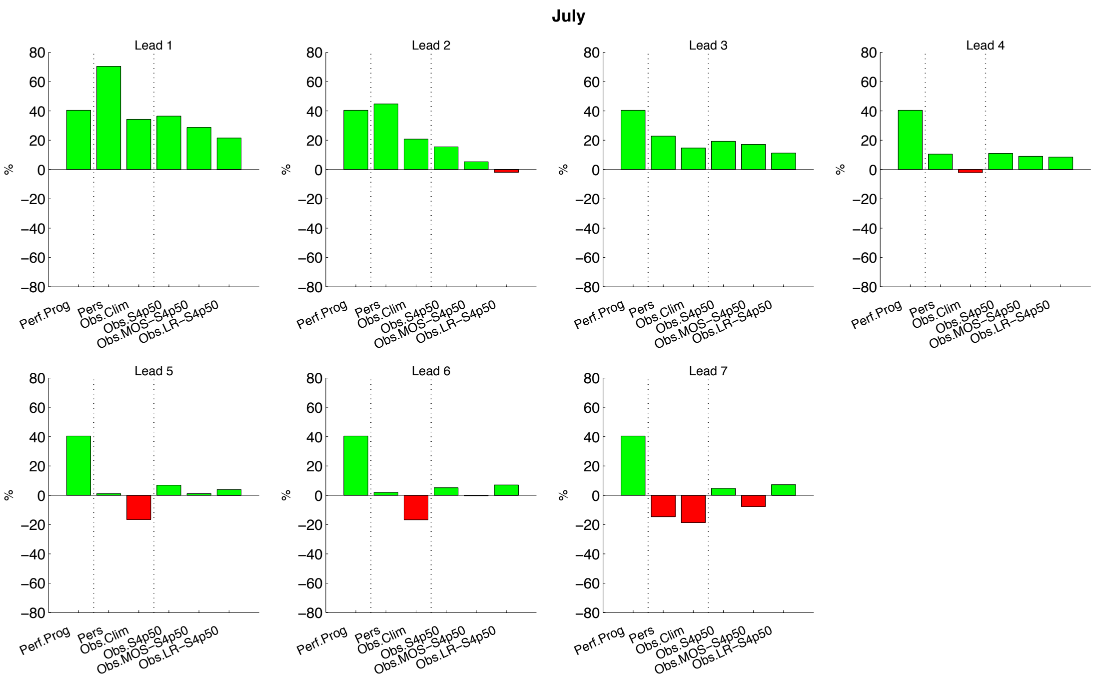


Figure 7.9. Percentage of July's volume MAE reduction with respect to climatology. Perf.Prog. refers to the MLR model performance with observed values. The other bars correspond to the five forecast configurations considered: persistence (Pers), antecedent observation combined with climatological values (Obs.Clim), antecedent observations combined with mean bias corrected S4 (Obs.S4.), antecedent observations combined with MOS-analog calibrated S4 (Obs.MOS-S4.) and antecedent observations combined with S4 calibrated with a linear regression procedure (Obs.LR-S4.).

The probabilistic performance also exhibits better results than in-flow. At first lead persistence is the leading method in all climatic categories except for November and December, where the A+S4 offers the greatest EVA (see figure 7.10). For lead two persistence shows the finest upgrade in all the three terciles in March, April, June, July and September. In February it is the best option in the lower and upper tercile and in August, for the lower and middle thirtiles. At this forecast horizon the A+S4 and A+MOS are the best options in the rest of terciles and months. The middle tercile is not enhanced in January, February and November. At lead three the lower and upper terciles always display a prediction approximation which offers better results than climatology. Yet, we have that the finest is generally persistence, followed by A+S4 and A+MOS. The middle thirtile is only better modelled than climatology in 5 months: January, May, June, August and September. In January, May, June and August the best forecast systems A+S4 and MOS whereas in September persistence is the most advantageous. From August to October there is an upgrade with respect to climatology in the upper and lower thirtiles up to lead seven normally by means of persistence but also with A+S4 and MOS. This pattern can be also seen in May and June, up to lead five and July, up to lead six, but in the latter case being persistence the best option. As for November and December the predictability of the first and third thirtile reaches lead seven, but in this case the dominant forecast system is the A+MOS. This also happens in January and, restricted to the lower thirtile, in February. From lead four to lead seven the middle tercile is normally better forecast by climatology and only in scattered cases there is another forecast approach which gives better results.

In figure 7.11 we can see the EV curves for July's volume forecast issued in March (lead five) with the six forecast strategies considered. We focus our attention in March because it is when the projections of summer water price are normally made (ACA personal communication). Figure 7.12 details the attributes diagram for the lower and upper thirtile of this same forecast focused in persistence (the best strategy at this horizon). In this occasion, all the forecast probabilities lie inside the reliability shaded region in accordance with the EV results.

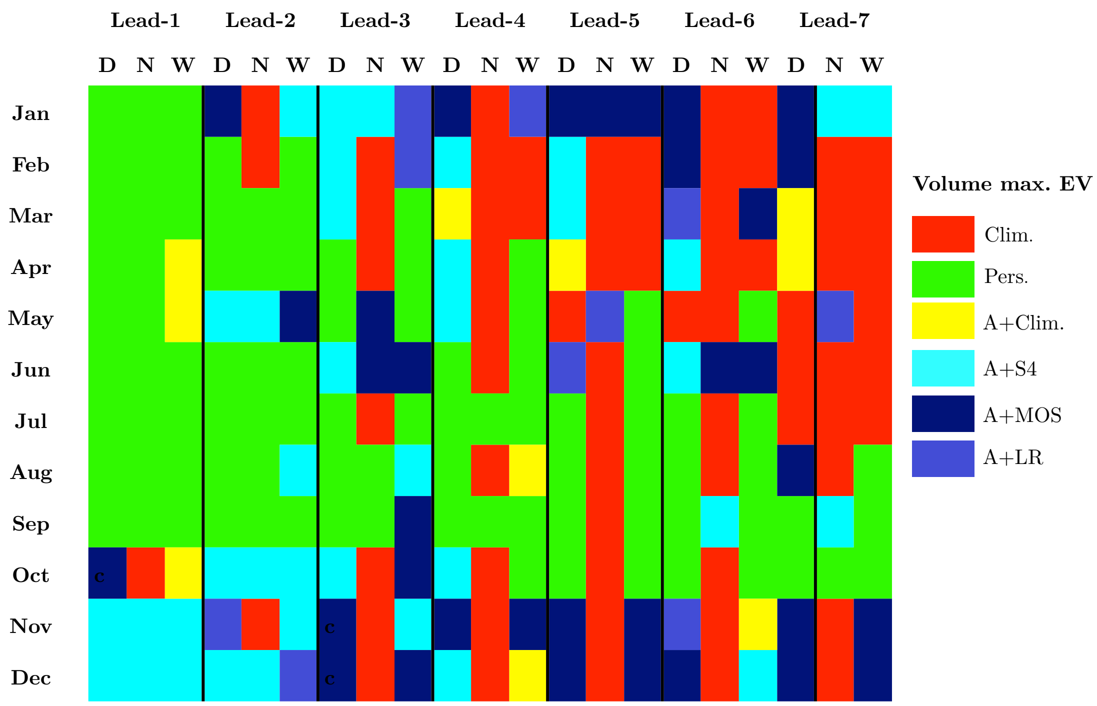


Figure 7.10. Volume forecast strategy with the highest Economic Value (EV) for each month, forecast horizon and climatic conditions (**D** -lower tercile-, **N** -middle tercile-, and **W** -upper tercile-) at the Boadella reservoir. The six forecast systems considered are: climatology (Clim), persistence (Pers), antecedent observation combined with climatological values (A+Clim), antecedent observations combined with mean bias corrected S4 (A+S4), antecedent observations combined with MOS-analog calibrated S4 (A+MOS) and antecedent observations combined with S4 calibrated with a linear regression procedure (A+LR). Note that forecast strategies different from climatology only appear in the table if they have a minimum EV Area of 0.10.

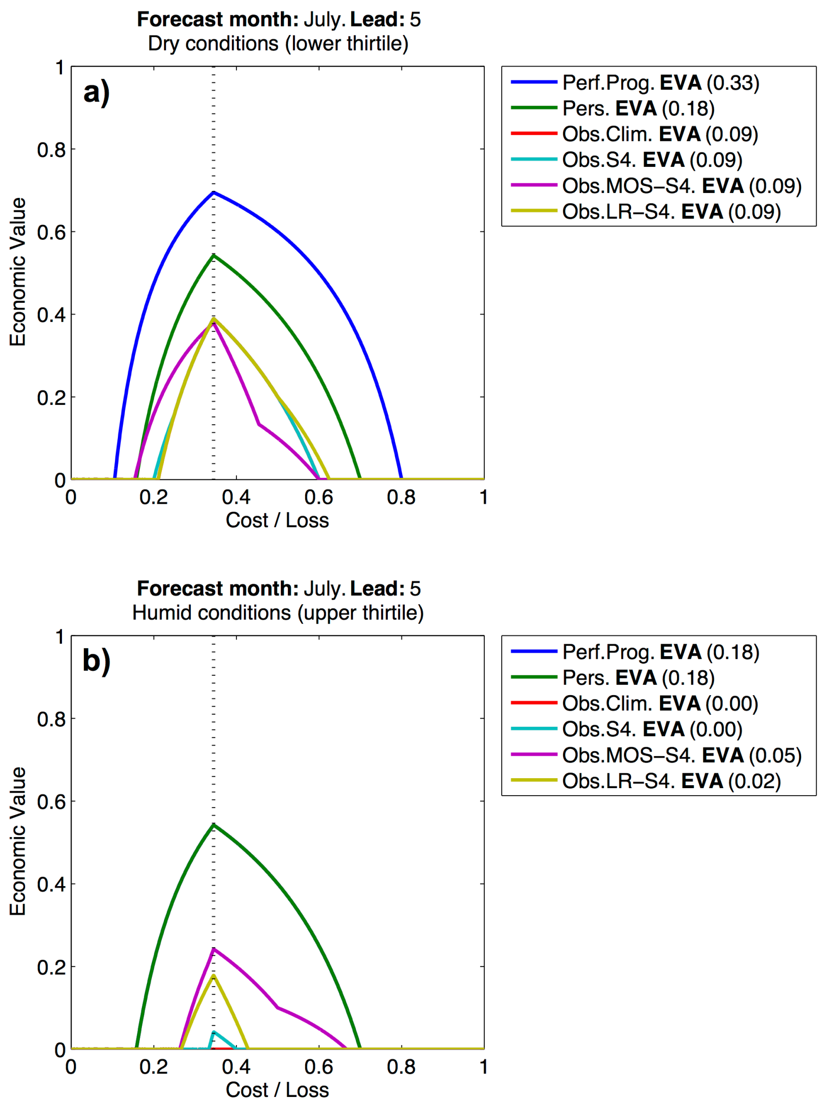


Figure 7.11. EV plot for July's lead five forecast of (a) dry conditions (b) humid conditions. The base line corresponds to climatology. Perf.Prog. refers to the MLR model performance with observed values. The other curves correspond to the five forecast configurations considered: persistence (Pers), antecedent observation combined with climatological values (Obs.Clim), antecedent observations combined with mean bias corrected S4 (Obs.S4.), antecedent observations combined with MOS-analog calibrated S4 (Obs.MOS-S4.) and antecedent observations combined with S4 calibrated with a linear regression procedure (Obs.LR-S4.).

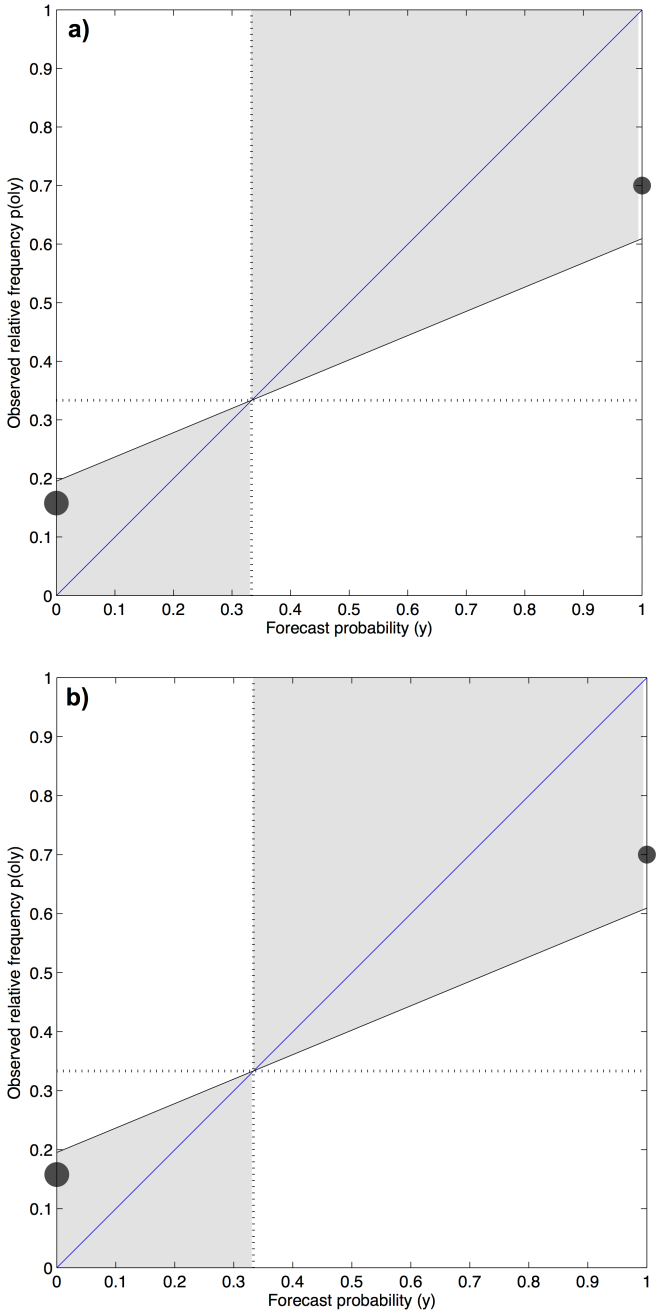


Figure 7.12. Attributes diagram for July's fifth lead volume persistence forecast of (a) dry conditions (b) humid conditions. All forecast probabilities lie in the reliability shaded region.

Out-flow

The out-flow MAE analysis highlights August and September going beyond climatology up to lead seven (see figure 7.13). At lead one there is an upgrade driven by persistence in January, February, April, May and August, and by A+Clim in March, June, July, September and October. At lead two persistence is the leading method in May, whereas A+Clim shows the best results from June to August and also in October. In March the better method is A+MOS and, in September, the A+S4. Lead three presents an upgrade in March and from May to October. A+Clim is the best option in July, August and October. On the other hand, A+MOS is the choice for March and June, while A+S4 is the finest approach in May and September. For lead four February shows an upgrade with persistence, March with A+LR and May with A+S4. From July to September the best MLR configuration is A+Clim. At lead five, July presents an upgrade with A+LR; August, with A+Cl; and September with A+MOS. For leads six to seven August and September improve climatology with A+Clim and with A+MOS at lead seven in September. May also show some enhancement at lead seven. Finally, November and December do not show any MAE amelioration at any forecast horizon. In figure 7.14 we can see the percentage of July's out-flow MAE reduction compared to climatology with the six forecast strategies considered. We show July's forecast performance because it is the month with the highest hydrological stress (see figure 4.11) and when the irrigation demands peak (see figure 7.1).

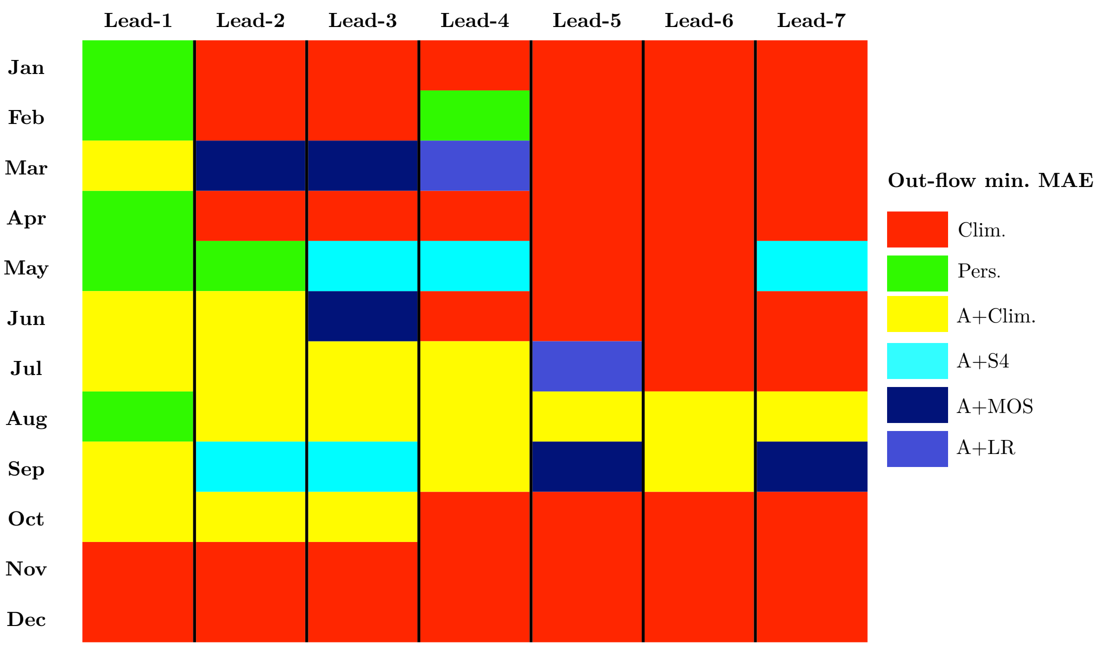


Figure 7.13. Out-flow forecast strategy with the lowest MAE for each month and forecast horizon at the Boadella reservoir. The six forecast systems considered are: climatology (Clim), persistence (Pers), antecedent observation combined with climatological values (A+Clim), antecedent observations combined with mean bias corrected S4 (A+S4), antecedent observations combined with MOS-analog calibrated S4 (A+MOS) and antecedent observations combined with S4 calibrated with a linear regression procedure (A+LR). Note that forecast strategies different from climatology only appear in the table if they improve climatology MAE results by a minimum of 5%.

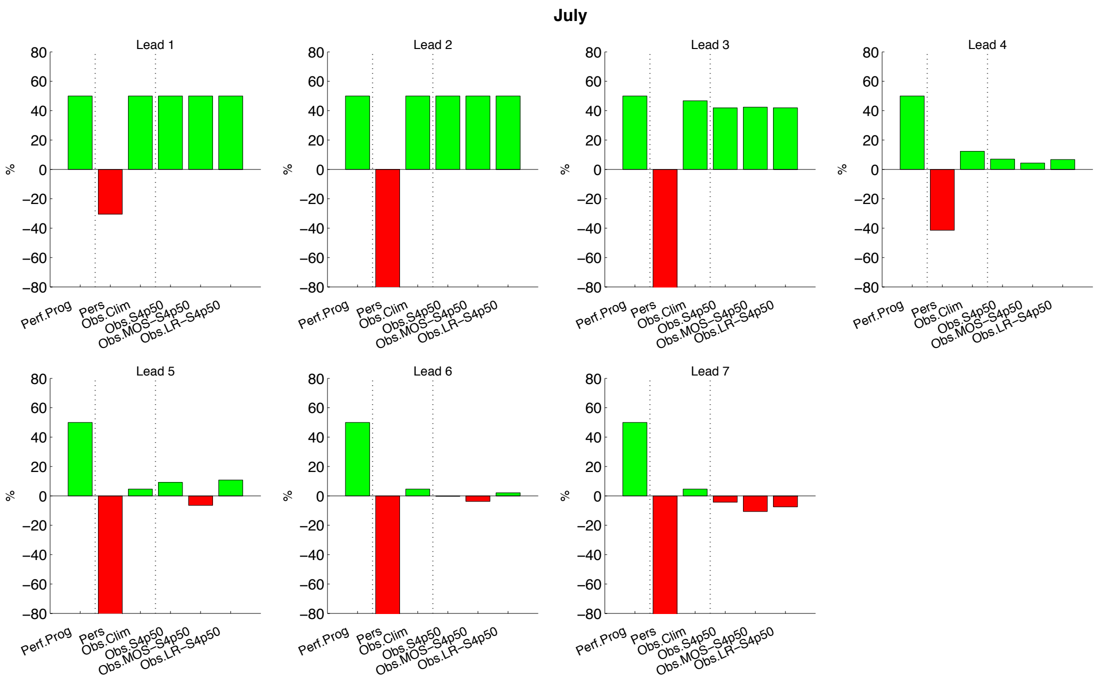


Figure 7.14. Percentage of July's out-flow MAE reduction with respect to climatology. Perf.Prog. refers to the MLR model performance with observed values. The other bars correspond to the five forecast configurations considered: persistence (Pers), antecedent observation combined with climatological values (Obs.Clim), antecedent observations combined with mean bias corrected S4 (Obs.S4.), antecedent observations combined with MOS-analog calibrated S4 (Obs.MOS-S4.) and antecedent observations combined with S4 calibrated with a linear regression procedure (Obs.LR-S4.).

The EVA analysis of the first lead shows that persistence is the best option in the lower and upper terciles from November to February, April, September; the lower thirtile of May and June; and March's upper thirtile (see figure 7.15). In the other months it is the combination A+Clim which shows the greatest amelioration. The middle thirtile does not display any general amelioration beyond climatology at any lead. At lead two persistence is the finest system in the upper thirtile of April, May and the lower thirtile of December. January, February, April and November do not show any enhance beyond climatology for the lower tercile; and January, June and December, neither for the upper. In all the rest is the A+Clim forecast system the one which poses the greatest upgrade with respect to climatology. Progressing to lead three, the lower tercile forecasts only surpasses climatology in March and from May to October. For the upper tercile these amelioration comprise February, March, May, and from July to December. This lead's best forecast strategies are the A+Clim followed by the A+MOS. From lead four onwards we only have scattered positive EVA values distributed between the lower and the upper thirtiles of, specially, the months comprised between May and October. In those cases the predominant forecast approach is the A+MOS, followed by the A+S4 and the A+Clim.

In figure 7.16 we can see the EV curves for July's out-flow forecast issued in May (lead three) with the six forecast strategies considered. We focus our attention in May because it is the farthest lead in which we still have remarkable positive skill in the forecast of July's anomalies (see i.e. figure 7.14). Figure 7.17 details the attributes diagram for the lower and upper thirtile of this same forecast focused in A+Clim (the best strategy at this horizon). As we can see, all the forecast probabilities are inside the reliability shaded region, a result that is in concordance with the EV analysis' outcome.

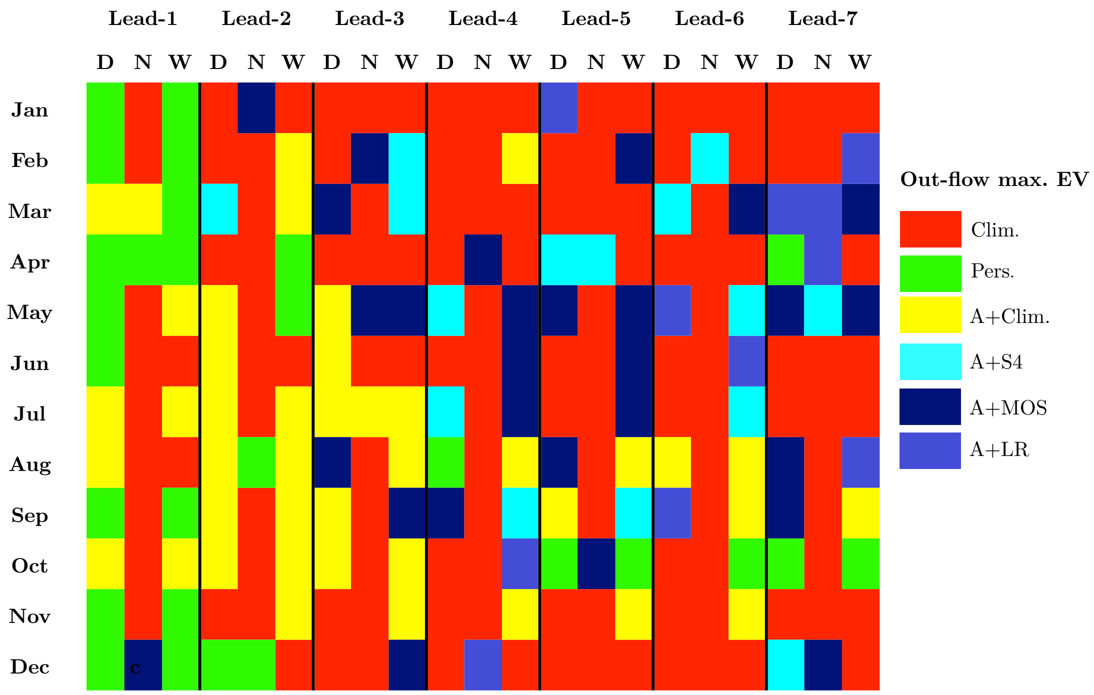


Figure 7.15. Out-flow forecast strategy with the highest Economic Value (EV) for each month, forecast horizon and climatic conditions (**D** -lower tercile-, **N** -middle tercile-, and **W** -upper tercile-) at the Boadella reservoir. The six forecast systems considered are: climatology (Clim), persistence (Pers), antecedent observation combined with climatological values (A+Clim), antecedent observations combined with mean bias corrected S4 (A+S4), antecedent observations combined with MOS-analog calibrated S4 (A+MOS) and antecedent observations combined with S4 calibrated with a linear regression procedure (A+LR). Note that forecast strategies different from climatology only appear in the table if they have a minimum EV Area of 0.10.

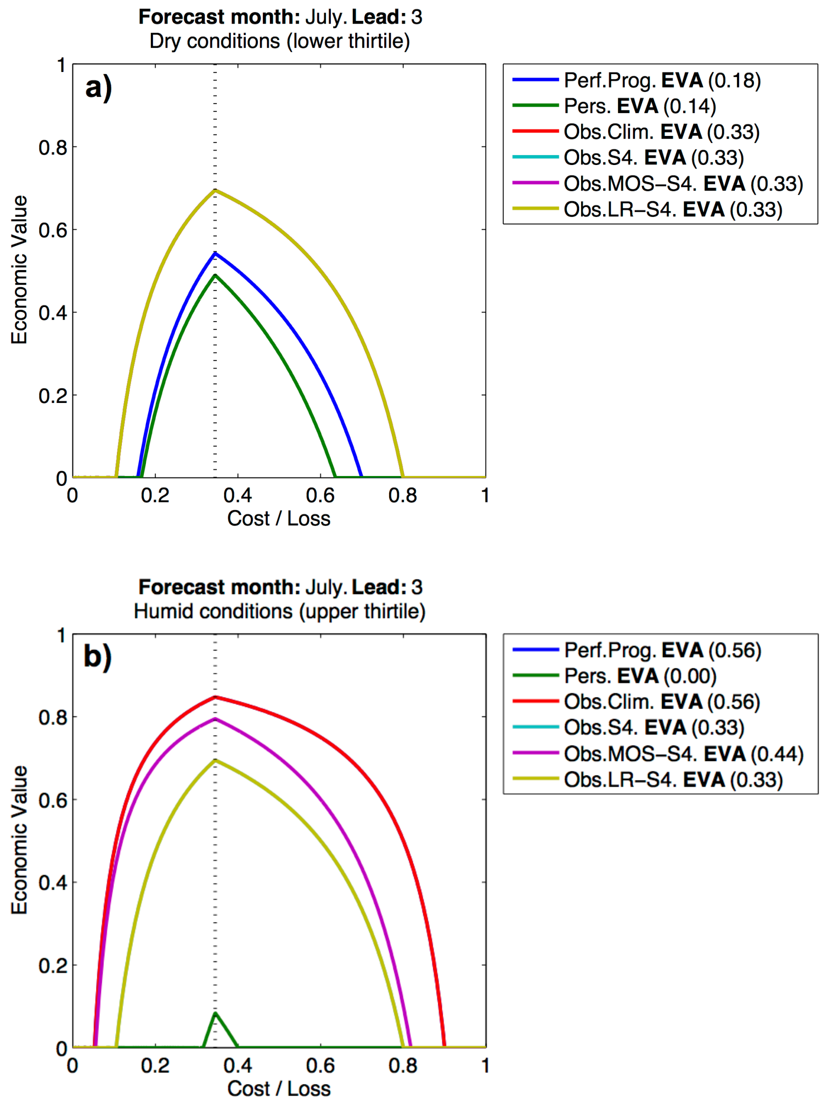
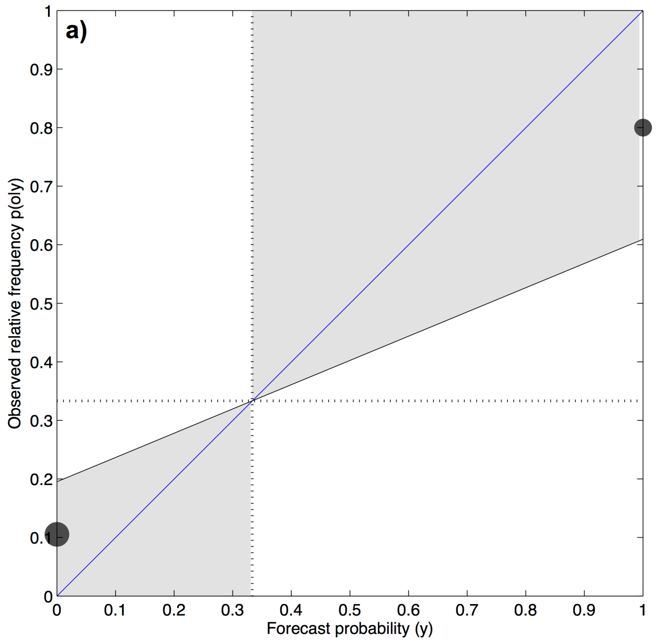


Figure 7.16. EV plot for July's lead three forecast of (a) dry conditions (b) humid conditions. The base line corresponds to climatology. Perf.Prog. refers to the MLR model performance with observed values. The other curves correspond to the five forecast configurations considered: persistence (Pers), antecedent observation combined with climatological values (Obs.Clim), antecedent observations combined with mean bias corrected S4 (Obs.S4.), antecedent observations combined with MOS-analog calibrated S4 (Obs.MOS-S4.) and antecedent observations combined with S4 calibrated with a linear regression procedure (Obs.LR-S4.).



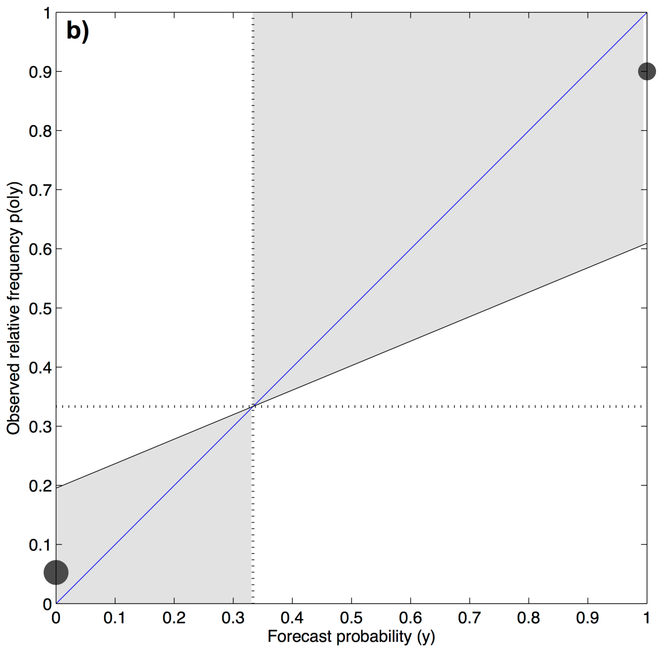


Figure 7.17. Attributes diagram for July's third lead out-flow A+Clim forecast of (a) dry conditions (b) humid conditions. All forecast probabilities lie in the reliability shaded region.

7.4 Discussion & Conclusions

In this chapter we have studied the potential of seasonal forecasting in the prediction of in-flow, out-flow and volume monthly anomalies in the Boadella reservoir (upper Muga river basin). This region has been selected by the Oficina Catalana del Canvi Climàtic (2012) as an area of interest for its particular vulnerability towards water scarcity in a context of climate change. Hence, this case study has been an opportunity to see whether seasonal forecasting can play an active role in the elaboration of future water management strategies. If we recall the first question raised at the beginning of this chapter,

> Can we model monthly anomalies for the in-flow, out-flow and volume stored in a reservoir through a MLR approach?

We have approached the answer through a MLR strategy by searching for the minimum number of predictors with the greatest explanatory power. To do so we have screened the performance of all the possible models with a previously developed algorithm that allows us to control different aspects of the forecast quality (see section 7.2). Although our experiments showed the results obtained are very similar to those derived with a step-wise regression, in the near future we would like to extend our algorithm to the latter because it is more computing efficient.

The LOOCV verification revealed that reservoir volume was the variable best described by the MLR models, followed by in-flow and out-flow anomalies (see section 7.2). In all these cases the meteorological predictors were rainfall, maximum and minimum temperature anomalies. Rainfall can be related to groundwater, river flow and soil-moisture state whereas the temperature predictors are proxies of evapotranspiration with the particularity that, in the case of minimum temperature, it is also a proxy for winter snow amount. Besides, in-flow observations demonstrated to be an important predictor for volume anomalies due to its role in the replenishment of the reservoir; and volume anomalies were also a transcendent driver for out-flow anomalies because they act as proxies for water-manager decisions.

As for the differences in the modelling of the predictands through an MLR, they are probably consequence of three factors: the inherent diversity of dependences of the predictand variables with the predictors, the non-linearitie/linearities introduced by human decisions and the different sources of predictor data. The inherent dependencies of the predictand variables have been already dealt in section 7.2. Out-flow anomalies are affected by the objectives of water demand satisfaction and avenue lamination, which are non-linear in nature for they heavily depend on the environmental conditions. On the other hand, volume anomalies are influenced by the objective of maintaining water supplies which turn this variable very linear. Regarding the third factor, when in-flow and volume anomalies act as predictors the antecedent observations are in-site observations instead of interpolated values from stations far apart.

This means that, in case a linear relation between variables and predictors existed, this will be clearer if the predictors were observations taken in the upper sub-basin rather than data coming from an interpolated gridded dataset. The reason to use E-OBS instead of direct observations was to search for a methodology that could be easily transferred into other regions. Nevertheless, our next step will be to compare the results obtained in this thesis with the outcome got when using observed values of rainfall, maximum and minimum temperatures instead of E-OBS values. Our hypothesis here is that the results found with E-OBS are a lower quality threshold and that the use of observed values can only improve those results. However, this is something that has yet to be proven.

Up to this point, then, it is a safe conclusion to say that the MLR can be an effective way to reproduce volume anomalies, the most part of in-flow anomalies and out-flow anomalies when using E-OBS as meteorological predictors and local in-flow and volume observation anomalies. Consequently, we went on to answer the second question and third questions:

> Can we issue seasonal forecasts of these anomalies? Can these forecasts be more useful than climatology?

To answer this question we have carried out a deterministic and probabilistic verification of the six forecast strategies for every month and each of the forecast leads. The combination of both approximation (mainly through the MAE and

EV) of the exact values and three climatic categories (dry, normal and humid) provided us with a more complete picture to assess the potential utility of this application for end-users. It is worth noting that in some cases there were values of the EVA between 0 and 0.10 of that arose when no other skill metric showed positive values. Since these values did not follow any recognizable scheme nor a specific forecast system we attributed this to noise and established EV Area of 0.10 as a threshold above which we could determine more robust surpassing of climatology. Still, it could be the case that for certain users the application of forecast systems different from climatology might imply some advantages even when the skill demonstrated in other metrics was little to non-existent. However, this is an hypothesis that has to be specifically tested in future studies.

That said, we can answer the question affirmatively. Nevertheless, since each variable has its particular characteristics and dependencies, the performances shown are very different among them. For instance, volume is the best forecast variable with MAE and EVA going beyond climatology up to lead four for the three climatic categories (dry, normal and humid) and even reaching lead seven from August to January. In winter, the best forecast configuration is the A+S4/MOS, an outcome that resembles the results obtained in chapter 6. In the other months, persistence and A+Clim are the best forecast systems up to lead four. In-flow forecast amelioration is almost restricted to the first lead and dry and humid conditions. There are also some enhancements at other scattered leads and months but without following any clear pattern. At first leads the antecedent information plays a determining role, with the A+Clim and persistence as the main sources of predictability. At further leads the combinations of A+S4/MOS/LR are the finest methods. Finally, the out-flow anomalies display some contrasts between the MAE and EVA results, an outcome that confirms the usefulness of looking at data from distinct perspectives. This is specially evident in November and December where there is no MAE amelioration and we can find positive values of EVA for dry and humid conditions at first lead (or even at lead six, in November's reaching upper thirtile). The other months show predictability for dry and humid conditions at the first lead and even at lead two. Persistence and A+Clim, are the better approaches at this horizon. From August to October this is also true almost up to lead seven and, in July, until lead three. In the other months and leads the A+S4 and MOS are the forecast systems which surpass climatology in the scattered occasions this happens. Eventually, it is worth noting that the middle thirtile of the climatology is the worst forecast category, probably due to the small size of the ensemble.

Hence, since most of today's reservoir management and end-user decisions are taken on the basis of climatology, the use of antecedent observations as well as current seasonal forecasts through the development of monthly MLR models constitute a window of opportunity to decision-making optimization. In fact, we have seen that our models already provide added value for each month of the year with a minimum of four months in advance in the case of volume anomalies, but also for in-flow and out-flow first leads, with the latter variable exhibiting enhanced predictability at longer horizons for the months of higher water demands.

The determinant role observed of antecedent conditions in seasonal predictability reinforces the idea of underlying physical mechanisms that connect the chosen predictors with our predictands and, additionally, that the MLR is a good way to uncover them. However, the importance that persistence shows in many occasions points towards the possibility to upgrade our models by including it in the MLR as another possible predictor, and so it is the next step we will work on (starting with volume anomalies).

As for the combination of antecedent observations with the mean bias corrected S4 and its calibrations, the A+S4 along with the A+MOS are the preferred options when they function best than the A+Clim and persistence (with a slight bias towards the A+MOS). The increase of the ensemble size, the expansion of the analog pool, the inclusion of the independent term in the LR and the use of in-site meteorological observations are interesting research lines to follow that might offer better results in the future. It is also our plan to extend our work to other areas such the Tarn basin in France or other basins and reservoirs in Spain and Europe to check the applicability and transferability of our findings.

CHAPTER 8

Seasonal forecasting applications: summer fires

8.1 Overview

In chapter 6 we have seen that *bias corrected* S4 is the best option to tailor seasonal forecast end-user applications that can go beyond *climatology* and *persistence* controls. This result only showed up systematically at lead one, although in all variables and domains there are cases which also exhibited some skill at longer leads. In the previous chapter this result has been applied in the integral study of seasonal forecasting as a tool to foresee reservoir in-flow, out-flow and, specially, volume anomalies with months in advance.

As of now, we turn our gathered learnings to the study of another potential seasonal forecasting application: the study of seasonal predictability of summer burned area. Several works have already analysed the relationships between climate variables and summer fires and found that the inter-annual variability of summer fires in the Mediterranean area is driven by drought and temperature influence upon fuel flammability and structure (see i.e. Pausas, 2004; Meyn *et al.*, 2007; Turco *et al.*, 2013a). This suggests the existence of potential predictability for this kind of fires. In fact, to our knowledge, to date there is no comparative analysis addressing the combination of statistical-empirical approaches and dynamical seasonal forecasting models in this field. To fill this gap we explore the seasonal predictability of summer burned area in Catalonia, a Mediterranean

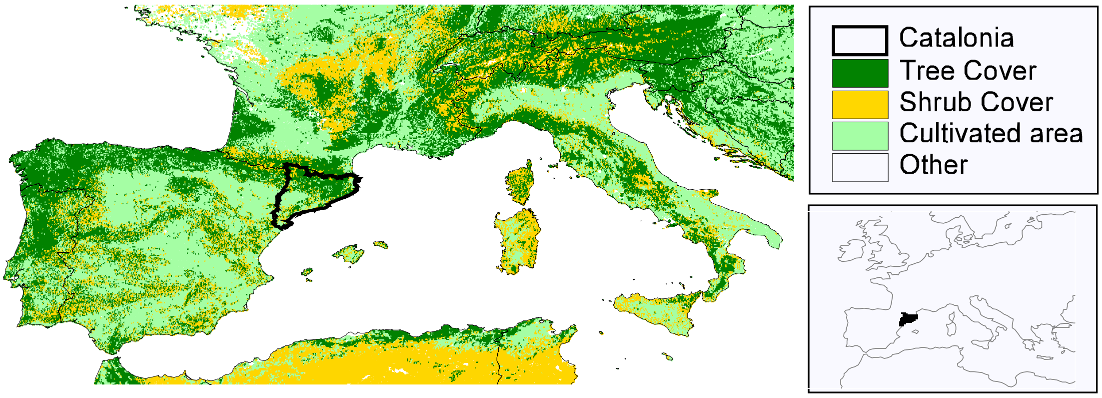


Figure 8.1. Domain of study and dominant land cover from the Global Land Cover dataset GLC2000 (Bartholomé and Belward, 2005). The inset shows a geographical map at larger scale.

environment (see figure 8.1). Approximately 60% of its extent (32000 km2) is covered by shrubland and forest (mostly low-height forest dominated by conifers, such as *Pinus halepensis*, *Pinus sylvestris* and *Pinus nigra*) and is frequently affected by high intensity, stand-replacing fires (Díaz-Delgado *et al.*, 2004; González and Pukkala, 2007). Further information on this region can be found in section 4.3.

Our goal is achieved through three supporting questions:

1. Can we develop a statistical model to link fire activity to current-summer and antecedent drought values?
2. What is the performance of such a model combining the observed antecedent drought with forecasts of the current-summer drought based on the S4, climatology and persistence?
3. Can we draw some general conclusions on the most convenient forecast system to obtain skilful seasonal prediction of fire activity in the region under study?

The answers to these points will help us define the feasibility to establish an operational framework for seasonal fire prevention in a mid-term future. To do so we have evolved our former MLR approach to predict summer burned area anomalies (see section 3.4) with the use of two standard drought indices: the

Standardized Precipitation Index (SPI; Mckee *et al.*, 1993) and the Standard Precipitation and Evaporation index (SPEI; Vicente-Serrano *et al.*, 2010). These new set of variables are then rearranged in two terms: *antecedent* and *current* drought conditions, in which *current* refers to the months of the same summer year for which we issue the forecast, and *antecedent*, which refers to any month lying before.

These SPI and SPEI can be both computed from the three variables that we have already been using throughout this thesis: T_{max}, T_{min} and Pr (see sections 3.1 and 3.2). However, the combination often requires merging of E-OBS and S4 data, so the implementation of a bias correction requires a more complex approach than the addition or subtraction of the *mean bias*. This methodology is known as *Quantile Mapping* (see i.e. Déqué, 2007; Wilcke *et al.*, 2013).

Referring to the validation process, this consists of an out-of-sample verification of the different model approximations against the observational fire dataset. It involves deterministic and probabilistic metrics that match S4 with *climatological* and *persistence* controls to identify which one has the best overall performance.

So, after this brief overview, this chapter is organized as follows: firstly, the *Methodology* section presents and justifies the selection of different data-treatment, modelling and verification strategies; then, *Results* sums up the outcome of the analysis performed; and finally, in *Discussion and conclusions* there is a recap of the answers to the questions raised.

8.2 Methodology

8.2.1 Drought indicators

We considered two standard drought indices: the Standardized Precipitation Index (SPI; Mckee *et al.*, 1993) and the Standard Precipitation and Evaporation index (SPEI; Vicente-Serrano *et al.*, 2010). SPI has been widely used for meteorological drought studies and is recommended by the World Meteorological Organization (WMO, 2012). SPI is a transformation of the accumulated precipitation values over a specific period (usually from 1 to 12 months) into a Gaussian distribution with mean zero and unity standard deviation. Positive values indicate situations of surplus rainfall, while negative values identify dry situations. The SPEI is mathematically similar to SPI, but includes the effects of temperature (Hargreaves, 1994; Vicente-Serrano *et al.*, 2010).

In order to compute SPI (and SPEI) from forecast precipitation (and temperature), we merged the seasonal forecasts of precipitation (and temperature) with the antecedent series of historical records from EOBS, following the methodology from Dutra *et al.* (2013). Since two different datasets are merged, special care has to be taken to ensure that the climatological mean is preserved for both datasets. To this aim, we used a quantile-quantile mapping (QM; Wilcke *et al.*, 2013), a non-parametric technique that can be applied to any type of variable (in this case precipitation and temperature) regardless of its distributional properties. We also upscaled the climatological series to the regional level (Catalonia) prior to their calibration, thus reducing the amount of data to process. To do so, we used the QM implementation provided in the R package downscaleR (Santander Meteorology Group, 2015). After merging the observed reference and the (QM calibrated) series of temperature and precipitation, we aggregated to a monthly scale, and then computed both SPI and SPEI using the implementation of the R package SPEI (Begueria and Serrano, 2015).

8.2.2 Burned area analysis

We have analysed the burned area in Catalonia of the summer months from June to September (BA hereinafter). This is the period with larger fires, which account for about 86% of the annual burned area. For more details on this dataset, the reader is referred to section 3.4.

Since BA follows an approximate log-normal distribution, we normalize the variable by applying a standard *log* transformation. Figure 8.2a shows the total burned area in summer, while figure 8.2b displays the log-transformed BA. These series reveal strong year-to-year oscillations, with two peaks in 1986 and 1994, and a general negative trend. A possible driver of this negative fire trend is improved fire-management strategies resulting in reinforced fire prevention and fighting resources (Turco *et al.*, 2013b, 2014).

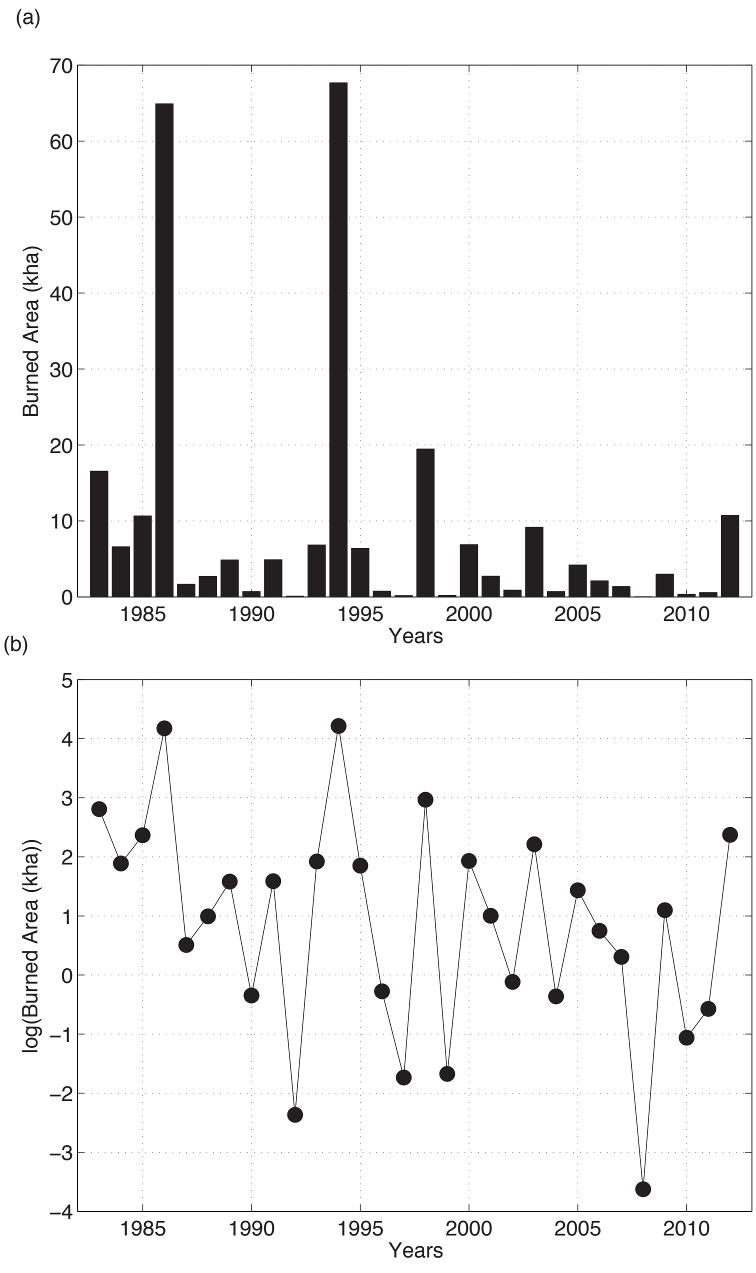


Figure 8.2. Summer Burned Area (BA) in Catalonia (NE Spain) over the 30-year period 1983-2012 (a) and log-transformed Burned Area (b).

8.2.3 Fire-drought model

Recent studies (Turco *et al.*, 2013a, 2014) have shown that summer fires in Catalonia are connected to the temperature and precipitation of both the coincident season and that of two years before. Our approach builds on these studies by exploring the predictive relationship between drought indicators and fires through a statistical model. This method links drought indices to BA through a multiple linear regression model (MLR hereafter) based on the following hypothesis: antecedent droughts influence fuel structure, while current-year drought promotes favourable conditions for ignition and combustion. Essentially, the model relates year-to-year changes in BA with current and antecedent droughts:

$$\log(BA) = a \cdot DIC(\tau_a) + b \cdot DIA(\tau_b) + \varepsilon \tag{8.1}$$

where DIC refers to the Drought Index Current (SPI or SPEI) condition and DIA to the Drought Index Antecedent situation; a and b are coefficients that represent the sensitivities of BA to DIC and DIA, respectively; finally, τ_a and τ_b are the months to which the indexes DIC and DIA refer, respectively. Prior to the analysis, the time series of both fire and drought indexes were linearly detrended to minimize the influence of slowly changing factors.

8.3 Results

In the previous section we have introduced the elements to build our MLR model. After testing several MLR combinations (eq. 8.1) with SPI/SPEI for 3, 6, 9 and 12 accumulation months we have found that the best results show up with SPI/SPEI 6-month accumulation for τ_a and τ_b of 1 and 27 months, respectively. Afterwards, we have computed the parameters a and b by least-square fitting to the data. In eqs. 8.2 and 8.3 we show the resulting models for SPI and SPEI:

$$BA = -0.90 \cdot SPI6(1) + 0.64 \cdot SPI6(27) + \varepsilon \tag{8.2}$$

$$BA = -1.05 \cdot SPEI6(1) + 0.61 \cdot SPEI6(27) + \varepsilon \tag{8.3}$$

In both cases, SPI and SPEI, we identify similar drought variables as efficient predictors from coincident and antecedent years. The Pearson correlations of the models (eqs. 8.2 and 8.3) are respectively 0.68 and 0.70 (pValue < 0.05), indicating that these parsimonious models show skill in reproducing the year-to-year changes in BA. Turning to the model residuals they satisfy the hypothesis of normality, zero autocorrelation and no trend.

To find out which are the relative importances of DIC and DIA in eqs. 8.2 and 8.3 we use the Akaike Information Criterion (AIC). The AIC score (Akaike, 1974) measures the validity of a statistical model based on a trade-off between its accuracy and its complexity (that is, the number of free parameters) being the best models those with the lower AIC. In table 8.1 we can appreciate the added value of including antecedent drought conditions in the analysis.

Table 8.1. AIC values for SPI and SPEI with consideration of only current drought conditions (DIC), or of the combination of current and antecedent drought conditions (DIC +DIA, as in eqs. 8.2 and 8.3)

AIC	DIC	DIC+DIA
SPI	-7.8	-11.5
SPEI	-10.7	-13.4

These results confirm the importance of the year-to-year drought variability in regulating both fuel flammability and fuel structure. A negative SPI or SPEI means a stronger drought whereas a positive value refers to a moisture surplus. In our case it is worth noting that the signs of the coefficients for 6-month accumulation SPI and SPEI for month 1 and 27 are reversed. This means that BA is promoted if a wet period exists at month 27 and it is followed by a drought period in the months just before summer. This is explained because antecedent wet periods increase vegetation growth (fuel structure) whereas coincident droughts increase fuel flammability, so when they are combined (in Catalonia, for months 27 and 1 before summer) they act favouring summer BA.

Given the similarity between the two models, in the following we will focus on the SPEI drought index since it performs slightly better. An important test for the models is assessing their ability to perform out-of-sample predictions: we determine the model parameters on one subset of the data (training set) and validate the prediction on the other subset (testing set). Specifically, we apply a leave-one-out cross-validation (von Storch, 1999) in which we iteratively test one year using the remaining observations as training data.

To avoid artificial skill, the data are linearly detrended in each step of the cross-validation. To estimate the uncertainty of this kind of predictions, we followed the methodology proposed by Calmanti *et al.* (2007). Basically, this consists in calculating the variance, V, of the residuals in the calibration period; then generating 1000 random residual time series with the same variance, V, and finally adding the stochastic residuals to the predicted values to generate an ensemble of 1000 predictions.

Figure 8.3 shows the observed BA data together with the median of the ensemble of 1000 model realizations and its uncertainty bands (defined by the inter-quartile range of the ensemble of the 1000 out-of-sample predictions). This figure reveals valid model skill also in out-of-sample mode, with 0.58 Pearson correlation (pValue < 0.05).

From an operational point of view, it is important to assess whether this model can be used to separate positive and negative anomalies. We thus evaluate whether the MLR model can predict the occurrence of events, defining as events those cases with above-normal fire activity. We consider probabilistic forecast values ranging between 0% and 100% obtained as the percentage of the 1000 different out-of-sample predictions above their mean values. That is, our forecast is not deterministic and users need to take into account the uncertainty of the forecast expressed by these probabilities. For instance, users could decide to take action when a 10% probability of an above-average event is forecast. In this case, the number of missed events may be very low, but this may imply a high number of false alarms.

The Relative Operating Characteristic (ROC) diagram shows the Hit rate (i.e., the relative number of times a forecasted event actually occurred) against the False Alarm Rate (i.e., the relative number of times an event has been forecasted as an *event* but did not actually happen) for different potential decision thresholds (Mason and Graham, 1999). The area under the Roc curve, that is, the ROC Area (RA), is a useful measure to summarize the skill of a model. RA ranges

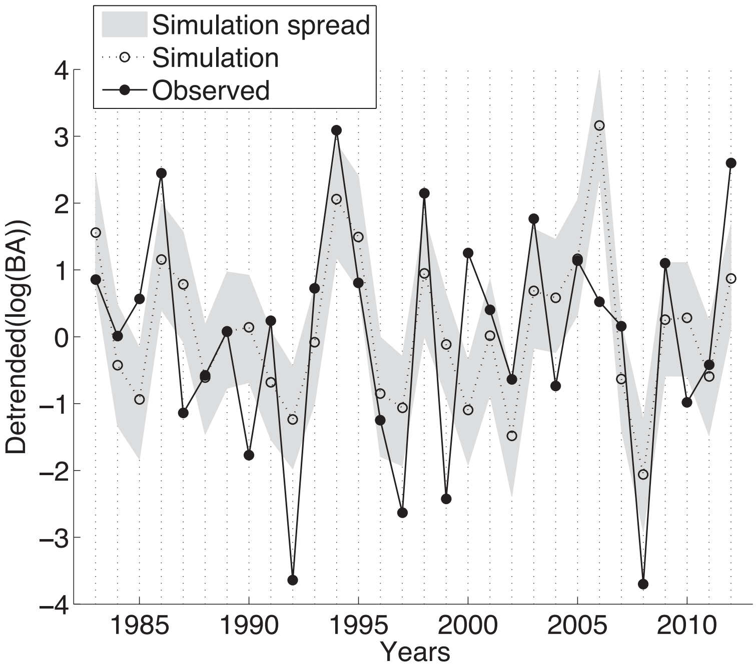


Figure 8.3. Out-of-sample prediction for BA using current-year and antecedent SPEI as predictors (Eq. 8.3). The continuous line with solid circles represents the observed data. The dotted line with empty circles is the median of 1000 different out-of-sample predictions; the dashed bands include the inter-quartile range of the ensemble of out-of-sample predictions. The vertical dotted lines show the edges of the 30 test periods considered.

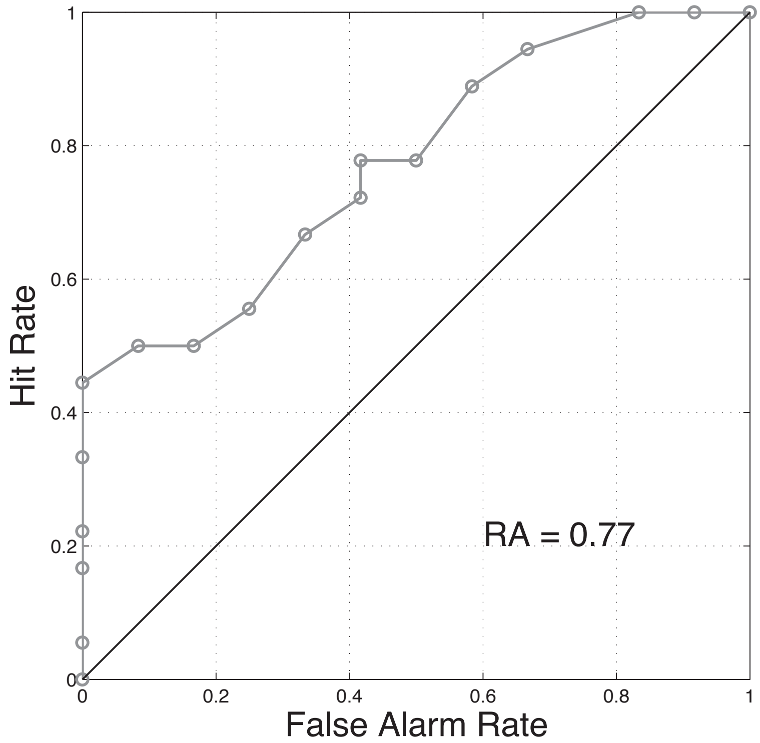


Figure 8.4. ROC diagram for BA using the MLR of eq. 8.3 (i.e. using current year and antecedent SPEI as drought indicator). The grey open dots indicate a set of probability forecasts by stepping a decision threshold with 5% probability through the forecasts. The number inside the plots is the ROC Area (RA).

from zero, for a forecast with no hit and only false alarms, to one, indicating a perfect forecast. Models with ROC areas above 0.5 have more skill than random forecasts.

Figure 8.4 shows that our model (eq. 8.3) has skill: the ROC curve is well above the identity line, with a ROC Area of 0.77. This model is then used to the study of seasonal predictability of summer forest fire BA as discussed below.

Our climate-fire model is based on two terms: antecedent drought, which is already known well before the fire season, and coincident drought, which needs to be forecasted to obtain an estimation of the summer BA. In the following we will drive the MLR model (eq. 8.3) with observed antecedent drought conditions and with different forecast sources for the coincident drought: *a*) seasonal S4 predictions; *b*) climatology; and *c*) persistence. Climatology forecasts are the average condition for the current-summer drought, which is equivalent to setting the SPEI to 0. For persistence forecasts, we consider the SPEI of the preceding 1 to 5 months prior to the first summer month, which is June.

Figure 8.5 shows the observed BA evolution together with the application of the MLR (eq. 8.3) according to the three forecast approaches described above. Figure 8.6 shows the ROC diagrams for these forecast systems. At first glance, the seasonal S4 forecast system seems to follow a similar behaviour to the climatology approach. However, both predictions show some amount of skill, with correlations of 0.36 (pValue = 0.06) and 0.37 (pValue = 0.04) and RA of 0.58. We argue that, in this case, this source of predictability is attributable to antecedent drought variables. The persistence forecast shows the best results considering the drought condition in May, with a correlation of 0.49 (pValue $<$ 0.01) and RA of 0.72.

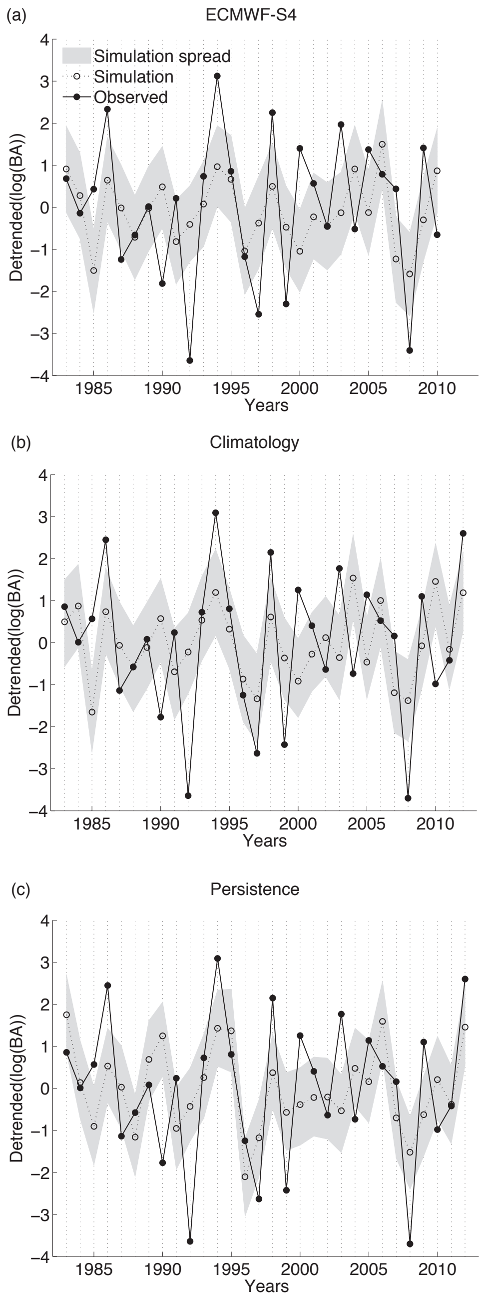


Figure 8.5. MLR model (eq. 8.3) results considering three forecast approaches (a) the seasonal forecast S4 issued in February for the current-summer drought conditions (b) climatology and (c) persistence of May SPEI6 conditions. The continuous line with solid circles represents the observed data. The dotted line with empty circles is the median of 1000 different out-of-sample predictions; the dashed bands include the interquartile range of the members of the ensemble of out-of-sample predictions. The vertical dotted lines show the edges of the 30 test periods considered.

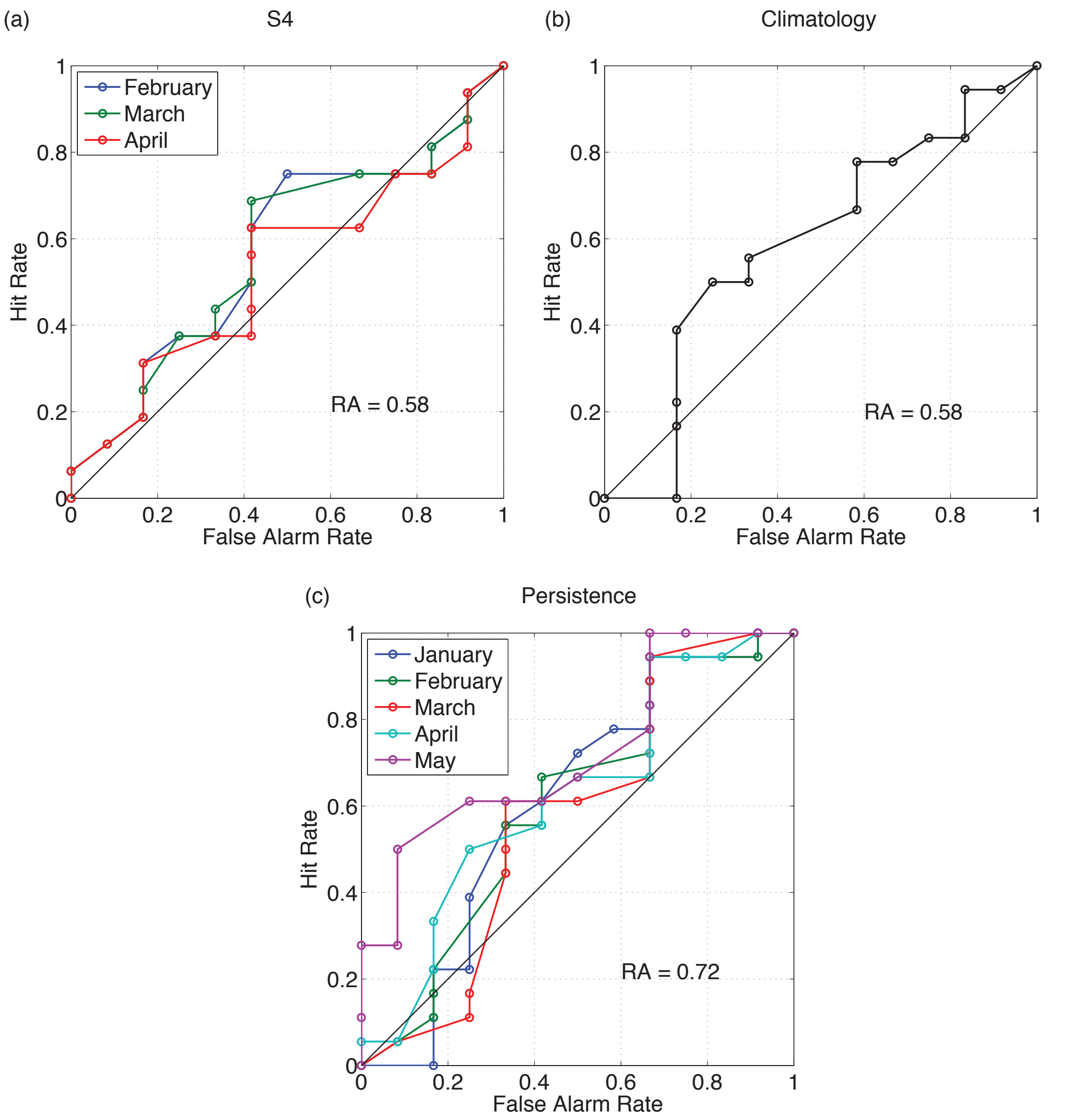


Figure 8.6. ROC diagrams for (a) S4, (b) climatology forecast and (c) persistence forecast. The open dots indicate a set of probability forecasts by stepping a decision threshold with 5% probability through the forecasts. The number inside the plots is the ROC Area (RA).

8.4 Discussion & Conclusions

The goal of this chapter was to evaluate the predictability of summer fires BA in a Mediterranean environment (NE Spain). To this aim, we tested the performance of parsimonious linear models by building on current-summer and antecedent drought variables. So, to the first question,

> Can we develop a statistical model to link fire activity to current-summer and antecedent drought values?

The answer is positive for we have been able to develop a MLR model that can successfully link current-summer and antecedent drought. As a result, one might reasonably think that the addition of accurate seasonal forecasts for the expected summer drought would enable its fire prediction ability. However, we have found a lack of skill of the model when introducing the S4 forecasts, a result that can be directly linked to the uncertainties identified for leads beyond one and the particular misbehaviour of S4 for May maximum temperature at lead one in Catalonia (see section 6.3.3). The importance of lead one is clear when we see that persistence shows the best results when computing May SPEI6. Hence, the answer to the second question,

> What is the performance of such a model combining the observed antecedent drought with forecasts of the current-summer drought based on the S4, climatology and persistence?

Is that when predicting the probability of above/below-normal summer fire BA the use of S4 in this particular approximation only equals climatology, whereas the use of persistence can go far beyond them. Consequently at this point, we have the answer for the third question,

> Can we draw some general conclusions on the most convenient forecast system to obtain skilful seasonal prediction of fire activity in the region under study?

Indeed, persistence for May SPEI6 should be the choice to start building an operational forecast structure to identify positive/negative BA anomalies. However, it cannot be discarded that with another strategy the S4 could offer more valuable results (see chapters 6 and 7) and so will be an objective for future works. In fact, we argue that our persistence model has two main sources of predictability: one linked to the dependency of the fuel structure on antecedent droughts, and the other related to the influence of preceding drought on fuel flammability. Potentially, it is reasonable to expect that the approach presented here could also be applied to other geographical areas with similar characteristics to Catalonia's, such as several Mediterranean regions covered by the so-called *Mediterranean forest, woodland and scrub*. These regions are covered by an abundance of fine fuels that need relatively short periods to dry sufficiently. In addition to these direct effects, climate may influence the abundance, continuity, and type of fuels (fine vs. coarse). In these regions, antecedent climate could favour fuel gaps to be filled within the landscape, resulting in increased abundance and continuity of fuel load.

While many studies have investigated this relationship in North America (Westerling *et al.*, 2002; Preisler *et al.*, 2004; Preisler and Westerling, 2007; Preisler *et al.*, 2008; Roads *et al.*, 2010) less analysis have been done in southern Europe. In spite of that, the existent studies support our hypothesis: fires are related to antecedent climate variables in Mediterranean environments. For example, Pausas (2004) analysed fires and climate in another Mediterranean region, Valencia (close to the area studied in this paper), and found that the year-to-year changes in burned area is (negatively) correlated with concurrent summer rainfall and (positively) correlated with antecedent summer rainfall. Similarly, Koutsias *et al.* (2012) investigated the relationships between forest fire activity and meteorological variables in Greece and found significant correlations with fire-season precipitation and lagged precipitation. Additionally, Gudmundsson *et al.* (2014) explored the relationship between above normal wildfire activity and meteorological drought (SPI) in Mediterranean Europe as well as in the Iberian Peninsula, the South Italy and in Greece sub-regions. In this case their focus was on exploring predictive relationships between fires and meteorological drought by means of fuel moisture. Although they did not specifically analyse the fire link with antecedent climate variables, their results for South Italy and Greece indicated: *significant predictability of regional wildfire activity can be yielded from antecedent meteorological conditions.*

Recently there have been serious warnings related to the impact of climate change and other environmental or socioeconomic changes on forest fire risk (Moriondo *et al.*, 2006; Bowman *et al.*, 2009; Moreira *et al.*, 2011; Pausas and Keeley, 2014). A recent study suggests that without the prevention and fire-fighting efforts undertaken in the last decades, the recent increase in temperature (Giorgi and Lionello, 2008; Turco *et al.*, 2012; IPCC, 2013b) would have significantly aggravated the observed fire impacts in Catalonia (Turco et al. 2014). In addition, future regional projections indicate that climate effects could become stronger and overcome fire prevention efforts (Bedia *et al.*, 2014; Turco *et al.*, 2014). Understanding the link between the interannual variability of drought and fire is therefore important not only to better understand fires and predict their change, but also to support the decision-making process of policy-makers and civil protection agencies for fuel management and resource allocation decisions (Mavsar *et al.*, 2013; San-Miguel-Ayanz *et al.*, 2013a). In fact, the ability to forecast the risk of fire months in advance makes it possible to organize allocation of fire fighting resources (e.g. fire fighters, equipment and aircraft) and target specific burning restrictions. For instance, seasonal forecasts may help to increase preparedness in vulnerable areas such as wildland-urban interfaces and may help to support decisions to reduce fuel load and continuity with prescribed burning and fuel-breaks (San-Miguel-Ayanz *et al.*, 2013b; Moreno *et al.*, 2014).

The empirical drought-fire model proposed here does not require large computational costs and can provide a preliminary estimate of the expected fire conditions for the summer season. Consequently, the relative simplicity and scalability of this approach leaves the door open to the development of an operational predictive framework that could be adopted in other geographical regions. Notwithstanding, there is still room for improvement since there is a fraction of summer fire variance that it is still out of our model. Indeed, although our regression model indicates a strong relationship to drought, there are other factors that also play a major role for determining the burned area, such as soil moisture, pre-existent biomass structure (that might depend on both natural factors as well as on human activities), fire suppression or fire ignition characterization. Hence, a further improvement of the modelling framework presented here may be expected with a better calibration of the predictors, the identification of new physical processes and/or the use of new improved climate data products such as better-skilled seasonal forecast dynamical models. Nevertheless, despite these limitations, our results suggest that by exploiting the relationships between summer BA and preceding drought conditions, the model might allow for a satisfactory long-term prediction of above/below-normal burned area.

CHAPTER 9

Conclusions

In this thesis we have studied the benefits of different calibration approaches on seasonal forecasting and the development of strategies to improve the seasonal prognosis of water resources and forest fires. To do so we have established three sub-objectives with a set of embedded tasks that have been addressed in chapters 6, 7 and 8:

1. Skill assessment

- **a)** Evaluation of the skill of the raw ECMWF System-4 (S4) output in Europe, Spain, Catalonia and the Muga river basin.
- **b)** Impact on the S4 performance of the MOS-analog and linear regression (LR) calibrations in comparison to mean bias correction in Europe, Spain, Catalonia and the Muga river basin.

2. Seasonal forecast of water resources

- **a)** Modelling of the Boadella reservoir in-flow, out-flow and volume anomalies through a Multiple Linear Regression (MLR) procedure.
- **b)** Evaluation of the seasonal predictability of the Boadella reservoir predictand anomalies through several seasonal forecast approaches.
- **c)** Are these seasonal forecast more advantageous than climatology?

3. Seasonal forecast of forest fires

- **a)** MLR modelling of summer (JJAS) burned area in Catalonia taking into account antecedent and current year drought conditions with the Standardized Precipitation Index and the Standardized Precipitation and Evapotranspiration Index (SPI/SPEI).
- **b)** Performance of the MLR model under different seasonal forecast configurations.

Now it is our aim to summarize our central findings and the original contributions to the field.

9.1 Overall

1. Skill assessment

The ECMWF System-4 (S4) model output has been calibrated on four domains: *a*) Europe *b*) Spain *c*) Catalonia and *d*) the Muga river basin, with a MOS-analog, linear regression (LR) and mean bias correction strategies. Afterwards we have compared the original and calibrated S4 forecasts with climatology and persistence. We have studied the S4 hindcast period 1981-2010 for each month of the year and up to lead seven considering an ensemble of 15 members rearranged in the so-called *generalized* form (see section 6.2.2). The verification metrics chosen to depict the results have been the mean absolute error (MAE), Taylor diagrams (taking into account the root mean squared error -RMSE-, the standard deviation and the Pearson correlation coefficient) and the discrete Brier skill score (dBSS).

- **(a)** **Evaluation of the skill of the raw S4 output in Europe, Spain, Catalonia and the Muga river basin.** The original S4 forecasts surpass climatology skill generally at lead one and in winter months. In the European domain this behaviour is observed for maximum and minimum temperatures but not for precipitation. Turning to Spain and Catalonia, precipitation forecasts overcome climatology while forecasts for maximum and minimum temperatures don't. Finally, for the Muga river basin we spot winter skill in the three analysed variables and also in autumn and the early spring. At further leads climatological improvements are scarce.

The observed first-lead winter predictability can be related to the stability of the winter general circulation anomalies. In the occasions when there is no first lead upgrade beyond climatology the bias study seems to confirm that first order model biases are the cause for this reduction in skill. Finally, focusing on persistence, it is surpassed by the S4 in virtually all leads, months and domains.

(b) **Impact on the S4 performance of the MOS-analog and LR in comparison to mean bias correction in Europe, Spain, Catalonia and the Muga river basin.** Our study showed that although the MOS-analog and LR calibrations have improved the original S4, they seldom surpass the performance of mean bias corrected S4. That said, the MOS-analog calibration sometimes gives larger ameliorations than the mean bias correction and the linear regression calibration (though the latter is normally better than the MOS-analog approach). It is worth noting that the mean bias S4 calibration enhancements are rather limited to the first lead but with exceptions in all domains.

Incidentally, it is also important to highlight that in all regions the first lead MAE of the mean bias calibrated S4 forecasts (50th percentile) improves climatology in practically all months. However, this correction seems to be more effective on temperatures and larger domains suggesting that these variables are more affected by first order biases and that these departures tend to arise more consistently on larger regions. When looking to the dBSS of the lower and upper terciles, though, these improvements are more restricted to winter and some scattered months in other seasons.

Going more specifically, and still focusing in the first lead, in the European domain all the variables are best forecast in winter (though

June and July are also good months). In Spain, this is also the case for precipitation and minimum temperature, where winter is the best forecast period and August, July and June also show good performance. For maximum temperatures, on the other hand, February, June and October are the highest upgrade exponents. In Catalonia, February and October are the months which show the best improvements for the three variables. Besides, in the case of minimum temperature, winter is also a good lead one forecast period. Finally, in the Muga river basin we find that for precipitation the February-March and October-November periods show the largest ameliorations. The same months are found for minimum temperature only exchanging November by January. Finally, for maximum temperature, the best months are January-March, August and October-November.

The recurrent predictability of February and October in the Spanish Mediterranean is highly valuable because both are key months regarding water management (ACA personal communication). Occasionally, the results have also shown some ameliorations beyond climatology at other leads. This is mostly perceptible in the smaller domains, Catalonia and the Muga river basin, but it also happens seldom in the larger ones. Yet, this predictability is not as systematic as the corresponding to lead one and, therefore, to discard the possibility of being noise, it has to be independently studied case by case (i. e. lead 6 maximum temperature forecasts for the Muga river basin for March-May).

2. Seasonal forecast of water resources

The application of seasonal forecast to water resources has been based in the construction of monthly multiple linear regression models (MLR) for the inflow, out-flow and volume anomalies in the Boadella reservoir. To build these models we have identified the underlying physical relationships between our predictands and the potential predictors to identify the models with the minimum number of independent variables and the maximum explanatory power. This initial phase, devoted to predictor choice, involved working with E-OBS data as well as in-flow and volume observations in an in-sample and perfect prognosis scheme. Moreover, to avoid artificial skill due to any existing trends all the predictors and predictands have been linearly de-trended previously. As for the ordering of the best predictors to build our models we have considered different verification metrics: *a*) coefficient of determination *b*) Akaike information criterion *c*) MAE *d*) RMSE and *e*) trend and autocorrelation of the residuals.

Afterwards, we have proceeded to forecast the monthly anomalies with those models for the period 1981-2010 (for each month of the year and up to lead seven). To do so we have carried out a leave-one-out cross-validation (LOOCV). Subsequently we have tested the aforementioned models with six strategies: *a*) Climatology *b*) Persistence *c*) Antecedent observations + climatology (A+Cl) *d*) Antecedent observations + S4 anomalies (A+S4) *e*) Antecedent observations + MOS-analog calibrated S4 anomalies (A+MOS) and *f*) Antecedent observations + LR calibrated S4 anomalies (A+LR). The first three approaches act as controls, and they comprise three approximations that only use observations as input data; the other three combine both observations and S4 calibrated forecasts. The verification metrics chosen to depict the results have been the mean absolute error (MAE), the coefficient of determination (R^2), the attributes diagram, the economic value curve (EV) and the economic value area (EVA).

(a) **Modelling of the Boadella reservoir in-flow, out-flow and volume anomalies through a MLR procedure.** The LOOCV revealed that reservoir volume was the variable best described by the MLR models, followed by in-flow and out-flow anomalies. In the case of volume anomalies, the predictor with most explanatory power is the accumulated in-flow anomaly observed in the preceding months and/or for the same month. This is logical since the main driver of the stored volume in a reservoir is its supplying river flow. This is followed by antecedent accumulated rainfall anomalies (from antecedent months and/or for the modelled month), which are related to the groundwater state. Finally, we have maximum and minimum temperature predictors, proxies for evapotranspiration in the reservoir and the upper sub-basin. In the case of minimum temperature, it also acts as proxy for winter snow amount.

Turning to in-flow anomalies the best predictors are rainfall anomalies for the same predicted month or the accumulated anomalies for the same month and the previous one. This is not surprising for the river flow in this upper sub-basin has a rather linear response with precipitation at a monthly scale. The influence detected in months far away the same modelled month are probably due to the effect of rainfall anomalies upon soil moisture, groundwater and their impacts on the river flow. The second most influencing predictor at in-flow level is the maximum temperature, indicative of the effects of evapotranspiration on soil moisture and run-off. In fact, as the soil becomes more (less) saturated, the response of the rainfall threshold to produce run-off becomes smaller (greater) and, consequently, there is an enhancement (decrease) of the linear relationship between rainfall and stream-flow. Finally, minimum temperatures act as a proxy of the snow amount in winter and also as a less effective proxy of evapotranspiration.

Finally, moving on to the out-flow models we can see that they are, generally, the ones with the less number of predictors. That is because in many months its modelling through a MLR has limited success and the inclusion of further predictors does not increase the performance of the model. This is consequence of the high level of human intervention in its evolution, because not only it depends on meteorological anomalies but also on other non-linear factors such as regulation protocols.

If we turn our attention to the best modelled months we will see that they coincide with the maximum and the end of the irrigation season (July-August-September) with volume and temperature anomalies as main predictors. An hypothesis to explain June's decreased predictability would be that this month's outflow could be more influenced by the human decision to set the beginning and intensity of the irrigation season. Conversely, in the center and final months of summer the soil moisture conditions and irrigation needs are more settled and human decisions would be mainly driven by antecedent volume anomalies and evapotranspiration. Concerning the other months, from October to May the main predictors are the accumulated rainfall, temperature and volume anomalies. During these months the out-flow is determined by a combination of human and meteorological factors, without anyone being clearly predominant.

(b) **Evaluation of the seasonal predictability of the Boadella reservoir predictand anomalies through several seasonal forecast approaches.** Volume is the best forecast variable with MAE and EVA going beyond climatology up to lead four for the three climatic categories (dry, normal and humid) and even reaching lead seven from August to January. In winter, the best forecast configuration is the Antecedent+S4/MOS (A+S4/MOS), an outcome that resembles the results obtained in chapter 6. In the other months, persistence and Antecedent+Clim (A+Clim) are the best forecast systems up to lead four.

In-flow forecast amelioration is almost restricted to the first lead and dry and humid conditions. There are also some enhancements at other scattered leads and months but without following any clear pattern. At first leads the antecedent information plays a determining role, with the A+Clim and persistence as the main sources of predictability. At further leads the combinations of A+S4/MOS/LR are the finest methods.

Finally, the out-flow anomalies display some contrasts between the MAE and EVA results, an outcome that confirms the usefulness of looking at data from distinct perspectives. This is specially evident in November and December where there is no MAE amelioration and we can find positive values of EVA for dry and humid conditions at first lead (or even at lead six, in November's reaching upper thirtile). The other months show predictability for dry and humid conditions at the first lead and even at lead two. Persistence and A+Clim, are the better approaches at this horizon. From August to October this is also true almost up to lead seven and, in July, until lead three. In the other months and leads the A+S4 and MOS are the forecast systems which surpass climatology in the scattered occasions this happens. Eventually, it is worth noting that the middle thirtile of the climatology is the worst forecast category, probably due to the small size of the ensemble.

(c) **Performance comparison of the considered seasonal forecast strategies against climatology.** We have seen that our models already provide added value for each month of the year with a minimum of four months in advance in the case of volume anomalies, but also for in-flow and out-flow first leads, with the latter variable exhibiting enhanced predictability at longer horizons for the months of higher water demands (which can be very valuable for water-managers). Therefore, there is the possibility to work on the construction of operational frameworks to replace climatology for these variables and horizons.

Besides, the determinant role of antecedent conditions in seasonal predictability reinforces the idea of underlying physical mechanisms that connect the chosen predictors with our predictands and, additionally, that the MLR is a good way to uncover them. Moreover, the importance that persistence shows in many occasions points towards the possibility to upgrade our models by including it in the MLR as another possible predictor. As for the combination of antecedent observations with the mean bias corrected S4 and its calibrations, the A+S4 along with the A+MOS are the preferred options when they function best than the A+Clim and persistence (with a slight bias towards the A+MOS).

3. **Seasonal forecast of forest fires**

   In this part we have explored the seasonal predictability of summer (JJAS) wildfires in Catalonia developing a multiple linear regression model (MLR) with antecedent and current-summer drought indices (standardised precipitation index, SPI; and standardised precipitation evapotranspiration Index, SPEI). We have tested three forecast systems based on *a*) ECMWF System-4 (S4) *b*) persistence and *c*) climatology. These approaches are evaluated through a leave-one-out cross-validation (LOOCV) over the period 1983–2012 (for each month of the year and up to lead seven). The verification metrics chosen to depict our results have been the Pearson correlation coefficient, the Relative Operating Characteristic curve (ROC) and the ROC area.

   (a) **MLR modelling of summer burned area in Catalonia taking into account antecedent and current year drought conditions with the SPI/SPEI indices.** We have developed a MLR model that can successfully link current-summer (SPEI6) and antecedent drought (SPEI27). This SPEI configuration responds to the particular influence between drought conditions and Catalonia's Mediterranean ecosystem in which there is usually an abundance of fine fuels that need relatively short periods to dry. In fact, in Mediterranean-type environments drought acts influencing burned area by controlling fuel flammability and structure. More specifically, present drought conditions can induce fine fuel drying whereas antecedent climate could favour the filling of fuel gaps within the landscape, resulting in an increased abundance and continuity of fuel load. Thus, it is reasonable to expect that the approach presented in this thesis could also be applied to other geographical areas with similar characteristics to Catalonia's.

   (b) **Performance of the MLR model under different seasonal forecast configurations.** We have observed that the probability prediction of above/below-normal summer fire burned area with the calibrated S4 forecasts has the same skill as climatology. Persistence, on the other hand, can enhance climatological forecasts with May's

SPEI6. Therefore, persistence should be the choice to start building an operational forecast structure to identify positive/negative burned area anomalies. The lack of skill of the model when introducing the S4 forecasts is a result that can be directly linked to the uncertainties identified for leads beyond one and the particular misbehaviour of S4 for May's maximum temperature at lead one in Catalonia (see section 6.3.3). The importance of lead one is clear when we see that persistence shows the best results when computing May's SPEI6. Hence, it cannot be discarded that with another approach the S4 could offer more valuable results.

9.2 Contributions to the field

The accomplishment of the objectives planned has lead to a number of original results and methodological approaches that might contribute to the progression of the seasonal forecasting field. This has been achieved, for instance, by means of the evidences provided on the skill of the S4 on different domains, the impacts of applying linear/non-linear calibrations on the S4 predictions or the utility found in water reservoir and summer forest fire seasonal forecast applications. Besides, the comprehensibility of our methodologies and the accessibility of the datasets used eases the transferability of these strategies virtually to any part of the world. More specifically we can detail these contributions as,

- **Multi-scale S4 skill assessment in Europe from continental to local domains.** Up to this moment few studies have analysed the skill of the S4 forecasts in Europe. Furthermore, to our knowledge there is no work studying the skill at a monthly level for the whole year from a deterministic and probabilistic perspective and ranging from continental to grid-point domains. We have found that the majority of skill is focused in the first lead. However, although the probabilistic assessment showed that the most part of this skill was focused in the winter months, we have seen that deterministic forecasts improved climatology and persistence controls also in the majority of months and for all variables. Besides, this amelioration appears to be more effective on temperatures and bigger domains. This suggests that first order model biases are more focused in temperatures and that they arise more easily in larger regions due to their aggregated nature.

- **Application of the MOS-analog method as a calibration strategy for seasonal predictions.** In the recent years the MOS-analog method has been introduced as a post-processing tool for the calibration and downscaling of climate change projections. However, its use in seasonal forecasting is rather uncommon and we could not find any examples of its application, particularly for Europe. With our implementation we put forward the potential of this non-linear post-processing technique in the correction and improvement of seasonal model output.

- **Evaluation of the linear regression and MOS-analog calibration performance with respect to mean bias correction in downscaling at different scales.** The reviewed studies consider that calibration post-processing techniques add little skill to the original seasonal model output, essentially correcting first order biases. In this thesis we have reached similar conclusions but with an important difference for it seems that with the MOS-analog some results hint the possibility to go beyond mean bias correction if the analog pool is sufficiently increased. Something similar might arise with the linear regression calibration approach if working also with an independent term. Since this behaviour seems to appear fundamentally in the smaller domains this also highlights the importance of conducting multi-scale analysis when evaluating the performance of post-processing methods.

- **Seasonal MLR model of monthly in-flow, out-flow and volume anomalies with MLR procedure.** We have performed a comprehensive study of the possibility to model reservoir water-supply variables through monthly MLR models. Although the MLR strategy has been applied thoroughly in hydrology we have not found any references in which the three main reservoir's water-supply variables (in-flow, volume and out-flow) had been treated from an integrated MLR perspective. That is, in views of using the MLR model output of one variable as predictor information for another predictand MLR forecast. Generally, the perfect prognosis approach revealed that all three variables can be modelled through the MLR though volume has been the most successful.

- **Value of seasonal forecasts to foresee reservoir's in-flow, out-flow and volume monthly anomalies.** We have extensively studied the performance of our MLR model suite for every month of the year under six seasonal forecast configurations considering that decision-makers use climatology as the base predictive model. The results showed that volume anomaly seasonal forecasts can begin the operational switch from climatology to another forecast strategy. This is also true for some months in the out-flow's modelling. For the in-flow case, though, there is still further research needed before reaching that sate but we think there is potential in achieving it somewhere in the short-middle term. Besides, the nature of the MLR relationships obtained makes us think that, specially for volume anomalies, this behaviour might transcend the specificity of the Boadella reservoir.

- **Seasonal MLR model for summer fire burned in Catalonia.** Our results suggest that by exploiting the relationships between summer burned area and preceding drought conditions there is the possibility to develop MLR models that can provide a preliminary seasonal estimate of the expected above/below-normal summer fire burned area in Mediterranean regions such as Catalonia.

- **Value of seasonal fire forecast with respect to climate change impacts on forest fires.** Understanding the link between the interannual variability of drought and fires is important not only to better understand fires and predict their change, but also to support the decision-making process of policy-makers and civil protection agencies for fuel management and resource allocation decisions. In fact, the ability to forecast the risk of fire months in advance makes it possible to organize allocation of fire fighting resources (e.g. fire fighters, equipment and aircraft) and target specific burning restrictions. For instance, seasonal forecasts may help to increase preparedness in vulnerable areas such as wildland-urban interfaces and may help to support decisions to reduce fuel load and continuity with prescribed burning and fuel-breaks.

CHAPTER 10

Prospects & Future work

The development of this thesis has risen some questions that require further analysis and research, the extension of some datasets or that directly lie beyond the original objectives for this work. In this chapter we will list some of these matters, in the hope that we can progress in their resolution somewhere in the future.

- **Increase of the ensemble and hindcast size.** We hope that with the extension of the re-forecast the MOS-analog calibration technique could be substantially improved. The probabilistic verification results would also benefit by an enlargement of the ensemble size, to avoid the underrepresentation of the centre ranks in the analysis. In this sense there are two ways to proceed: either see whether we can increase our collaboration with the ECMWF accessing to the operative or near-operative configuration of the S4, or switch to another forecast system that fulfils our requirements.

- **MOS-analog pool optimization expression.** In the MOS-analog calibration we have found that for each region there is a certain number of analogs that optimize the representativeness of each member of the analog pool. In accordance with literature this means that a certain number of historical analogs may suffice or not depending on the number of possible situations that can arise on the studied region, that is to say, on the inherent variability of the variable field map. This result may be used to derive an expression to relate the inherent variability of a region with the optimal number of analogs needed to find similar monthly situations within an arbitrary error band. Conversely, this could also serve to state the maximum variability for which an analog pool is able to offer useful analogs.

- **Check the impact of adding the independent term to the linear regression calibration.** Although we think that the influence should be small it is our aim to effectively check this perception by comparing the value of adding or not this term to the linear calibration procedure.

- **Assess the skill and variability of the ensemble forecasts from spatial and temporal perspectives through a Taylor diagram representation.** In chapter 6 we have rearranged the original ensemble in accordance to percentiles (the so-called generalized ensemble, GE). When representing it in a Taylor diagram along with the spatial climate correlation we have hypothesised that we can distinguish whether our model can go beyond spatial correlation by adding temporal skill. Besides, we think that this representation can be used to evaluate the ensemble variability with respect to the temporal and spatial variabilities observed. However, since they are only preliminary results it is our goal for the near future to state their theoretical basis and try to establish a clear protocol to apply the approach in the aforementioned analysis.

- **Reconsider the unification approach followed to issue a single forecast from the ensemble.** The unification strategy presented in section 6.2.3 showed a somewhat erratic behaviour, alternating good and bad results. Two reasons may explain this outcome. The first one is concerned with the possibility that no specific percentile of the *pdf* gives better results at each of the grid-points. The second is that the differences between the discriminating parameter (Mean Absolute Error, MAE) among the candidate members are too small and the method cannot efficiently identify the best option. One solution could be, either changing the discriminating parameter or increasing the re-forecast sample. We plan to address this question in a future study.

- **Change the process of predictor selection from total screening to step-wise regression.** In chapter 7 we have approached the best predictor Multiple Linear Regression (MLR) search with a total screening process. Further exploration showed the results obtained are very similar to those coming from a step-wise regression but at a much less computational cost. Thus, it is our aim to switch to this methodology as soon as we can.

- **Evaluate the performance of the reservoir's in-flow, out-flow and volume anomaly MLR models with in-site meteorological predictors.** When in-flow and volume anomalies act as predictors the antecedent observations are in-site observations instead of interpolated values from stations far apart. This means that, in case a linear relation between variables and predictors existed, this will be clearer if the predictors are observations taken in the reservoir rather than data coming from an interpolated gridded dataset. The reason to use E-OBS instead of direct observations was to search for a methodology that could be easily transferred into other regions. Nevertheless, we plan to compare the results obtained in this thesis with the outcome got when using values of rainfall, maximum and minimum temperatures observed at the reservoir instead of E-OBS values. Our hypothesis is that the results found with E-OBS are a lower quality threshold and that the use of observed values can only improve those results. However, this is something that has still to be proven.

- **Analyse the advantages and disadvantages of introducing persistence as a predictor in the MLR models**. The importance of persistence in many occasions points towards the possibility to upgrade our models by including it as another predictor when searching for the best MLR models. This is specially the case for volume anomalies but it will also be checked in the other two variables (in-flow and out-flow).

- **Establish the robustness of setting 5% MAE and 0.10 Economic Value Area (EVA) thresholds to consider enhancement beyond climatology.** In some cases there were values of the EVA between 0 and 0.10 and MAE, between 0 and 5% that arose when no other skill metric showed positive results. Since these values did not follow any recognizable scheme nor a specific forecast system we attributed this to noise and established EVA of 0.10 and a MAE of 5% as thresholds above which we could determine more robust surpassing of climatology. Still, it could be the case that for certain users the application of forecast systems different from climatology might imply some advantages even when the skill demonstrated in other metrics was little to non-existent. However, this is an hypothesis that has to be specifically tested and will be a matter of future studies.

- **Contact the Boadella reservoir's water managers to examine the possibility to tailor specific seasonal forecast solutions from the results obtained**. We have seen that our MLR approach for the modelling and forecast of reservoir's main variable anomalies can have more skill than climatology. Hence, there is the potential to start working with decision-makers to develop better seasonal predictive frameworks ahead than the customary use of climatology. For example, one possible approach could be to look into the *water guarantee* curves just before the beginning of the irrigation season and issue seasonal forecasts in order to provide water-managers with different future projections for the available water amount. Of course, after the beginning of the season this information would be still useful to dynamically re-adapt management decisions, but having the information before its onset would be very valuable for defining aspects such as water pricing or to contract insurances against possible losses derived from the advent of hypothetical drought restrictions.

- **Explore the improvement of the fire-drought model by including other predictors and the use of other calibrations.** The empirical drought-fire model proposed here does not include a fraction of the summer fire variance. Indeed, although our regression model indicates a strong relationship to drought, there are other factors that might play an important role in determining the burned area, such as soil moisture, pre-existent biomass structure (that might depend on both natural factors as well as on human activities), fire suppression or fire ignition characterization. Therefore, it is our aim to inspect whether they can add skill in our model. Furthermore, we have only used S4 information calibrated through a Quantile-Mapping (QM) procedure. In the future we will also explore the possibility to build the Standardized Precipitation and Standardized Precipitation and Evapotranspiration (SPI/SPEI) indices with MOS-analog and linear calibrated S4 forecasts. Therefore, a further improvement of the modelling framework may be expected with the identification of new physical processes and/or the better calibration of the model's predictors. With these enhancements, the prospects of a summer fire seasonal forecast operational structure in cooperation with the fire prevention service of the *Generalitat de Catalunya* would be closer.

- **Extend these methodologies to other models, regions and fields.** The simplicity and scalability of these methods along with the positive results obtained commit ourselves to continue this line of research by trying to expand these studies to other seasonal forecast models, different geographical domains and distinct disciplines. For instance, we have begun a collaboration with EDF-DTG to implement the MLR modelling to study seasonal forecast of the Montbrun's stream-flow, in the Tarn's river basin (France) with preliminary results that are similar to those found in the Boadella reservoir.

References

ACA (2009a). Pla de gestió del districte de conca fluvial de Catalunya. Tech. rep. 2

ACA (2009b). Pla de gestió del districte de conca fluvial de Catalunya (Annex XIII). Tech. rep. 145, 148

Accadia C., Mariani S., Casaioli M., Lavagnini A., and Speranza A. (2003). Sensitivity of Precipitation Forecast Skill Scores to Bilinear Interpolation and a Simple Nearest-Neighbor Average Method on High-Resolution Verification Grids. *Weather and Forecasting* **18**: 918–932. 79

Adamowski J., Fung-Chan H., Prasher S.O., Ozga-Zielinski B., and Sliusarieva A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* **48**(1): 1–14. 141

Adamowski J. and Karapataki C. (2010). Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms. *Journal of Hydrologic Engineering* **15**(10): 729–743. 141

Ajami H., McCabe M.F., Evans J.P., and Stisen S. (2014). Assessing the impact of model spin-up on surface water-groundwater interactions using an integrated hydrologic model. *Water Resources Research* **50**(3): 2636–2656. 142

Akaike H. (1974). A new look at the statistical model identification. *IEEE Transactionson Automatic Control* **19**: 716–723. 58, 183

Altava-Ortiz V., Barrera A., Llasat M.C., and Prat M.A. (2006). Application of the MM5 and the analogous method to heavy rainfall event, the case of 16-18 October 2003 in Catalonia (NE Spain). *Advances in Geosciences* **7**: 313–319. 5

Álvarez J., Potenciano A., and Villaverde J. (2008). Evolución histórica de sequías en España. *Revista de Obras Públicas: Organo profesional de los ingenieros de caminos, canales y puertos* (3486): 53–68. 19

Anscombe F.J. (1973). Graphs in Statistical Analysis. *The American Statistician* **27**: 17–21. 60

Baldwin M.P. and Dunkerton T.J. (2001). Stratospheric harbingers of anomalous weather regimes. *Science* **294**(5542): 581–4. 14

Balmaseda M.A., Ferranti L., Molteni F., and Palmer T.N. (2010). Impact of 2007 and 2008 Arctic ice anomalies on the atmospheric circulation: Implications for long-range predictions. *Quarterly Journal of the Royal Meteorological Society* **136**(652): 1655–1664. 14

Bao Z., Zhang J., Liu J., Fu G., Wang G., He R., Yan X., Jin J., and Liu H. (2012). Comparison of regionalization approaches based on regression and similarity for predictions in ungauged catchments under multiple hydro-climatic conditions. *Journal of Hydrology* **466-467**: 37–46. 141

Barnett T.P. and Preisendorfer R. (1987). Origins and Levels of Monthly and Seasonal Forecast Skill for United States Surface Air Temperatures Determined by Canonical Correlation Analysis. *Monthly Weather Review* **115**(9): 1825–1850. 12

Barnston A.G., Mason S.J., Goddard L., DeWitt D.G., and E. Zebiak S. (2003). Multimodel ensembling in seasonal climate forecasting at IRI. *Bulletin of the American Meteorological Society* (December): 1783–1796. 16

Barnston A.G., van den Dool H.M., Rodenhuis D.R., Ropelewski C.R., Kousky V.E., O'Lenic E.A., Livezey R.E., Zebiak S.E., Cane M.A., Barnett T.P., Graham N.E., Ji M., and Leetmaa A. (1994). Long-Lead Seasonal Forecasts—Where Do We Stand? *Bulletin of the American Meteorological Society* **75**(11): 2097–2114. 9, 16, 53, 60

Barrera A., Altava-Ortiz V., Llasat M.C., and Barnolas M. (2007). Heavy rain prediction using deterministic and probabilistic models – the flash flood cases of 11–13 October 2005 in Catalonia (NE Spain). *Advances in Geosciences* **12**: 121–126. 55

Barrera A., Llasat M.C., and Barriendos M. (2006). Estimation of extreme flash flood evolution in Barcelona County from 1351 to 2005. *Natural Hazards and Earth System Science* **6**: 505–518. 19

Barriendos M., Coeur D., Lang M., Llasat M.C., Naulet R., Lemaitre F., and Barrera A. (2003). Stationarity analysis of historical flood series in France and Spain (14th-20th centuries). *Natural Hazards and Earth System Science* **3**: 583–592. 1, 19

Bartholomé E. and Belward A.S. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing* **26**(9): 1959–1977. 178

Bedia J., Herrera S., Camia A., Moreno J.M., and Gutiérrez J.M. (2013). Forest fire danger projections in the Mediterranean using ENSEMBLES regional climate change scenarios. *Climatic Change* **122**: 185–199. 21

Bedia J., Herrera S., and Gutiérrez J.M. (2014). Assessing the predictability of fire occurrence and area burned across phytoclimatic regions in Spain. *Natural Hazards and Earth System Sciences* **14**: 53–66. 20, 192

Bednorz E. (2004). Snow cover in eastern Europe in relation to temperature, precipitation and circulation. *International Journal of Climatology* **24**(5): 591–601. 143

Begueria S. and Serrano V. (2015). SPEI: Calculation of Standardised Precipitation-Evapotranspiration index. R package version 1.6. 180

Beljaars A.C.M., Viterbo P., Miller M.J., and Betts A.K. (1996). The Anomalous Rainfall over the United States during July 1993: Sensitivity to Land Surface Parameterization and Soil Moisture Anomalies. *Monthly Weather Review* **124**: 362–383. 13

Benestad R.E., Hanssen-Bauer I., and Chen D. (2008). *Empirical-statistical downscaling*. World Scientific Pub Co Inc, Oslo. 16, 52, 54, 55, 58, 134, 143

Bengtsson L., Schlese U., Roeckner E., Latif M., Barnett T.P., and Graham N. (1993). A two-tiered approach to long-range climate forecasting. *Science (New York, N.Y.)* **261**(5124): 1026–1029. 13, 15

Bianucci P., Sordo-Ward A., Moralo J., and Garrote L. (2015). Probabilistic-Multiobjective Comparison of User-Defined Operating Rules. Case Study: Hydropower Dam in Spain. *Water* **7**: 956–974. 19

Block P. (2011). Tailoring seasonal climate forecasts for hydropower operations. *Hydrology and Earth System Sciences* **15**(4): 1355–1368. 18

Blumenthal M.B. (1991). Predictability of a Coupled Ocean–Atmosphere Model. *Journal of Climate* **4**(8): 766–784. 13, 15

Boer G.J., Terray L., Scaife A.A., Branstator G., Ishii M., Alves O., Déqué M., Stockdale T., Cassou C., Doblas-Reyes F.J., Coelho C., Smith D., Douville H., and Dubus L. (2013). International workshop on seasonal to decadal prediction. Tech. rep., WCRP Report No. 23/2013. 2, 10, 12, 14, 134, 143

Bonan G.B. (2008). Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* **320**(5882): 1444–9. 20

Bougeault P., Binder P., Buzzi A., Dirks R., Houze R., Kuettner J., Smith R.B., Steinacker R., and Volkert H. (2001). The MAP Special Observing Period. *Bulletin of the American Meteorological Society* **82**(3): 433–462. 24

Bowman D.M., Balch J.K., Artaxo P., Bond W.J., Carlson J.M., Cochrane M.A., D'Antonio C.M., Defries R.S., Doyle J.C., Harrison S.P., Johnston F.H., Keeley J.E., Krawchuk M.A., Kull C.A., Marston J.B., Moritz M.A., Prentice I.C., Roos C.I., Scott A.C., Swetnam T.W., van der Werf G.R., and Pyne S.J. (2009). Fire in the Earth system. *Science* **324**(5926): 481–484. 20, 192

Brands S. (2013). Skillful seasonal predictions of boreal winter accumulated heating degree-days and relevance for the weather derivative market. *Journal of Applied Meteorology and Climatology* **52**: 1297–1302. 18

Brands S., Manzanas R., Gutiérrez J.M., and Cohen J. (2012). Seasonal Predictability of Wintertime Precipitation in Europe Using the Snow Advance Index. *Journal of Climate* **25**(12): 4023–4028. 2, 14

Branković Č., Palmer T.N., Molteni F., Tibaldi S., and Cubasch U. (1990). Extended-range predictions with ECMWF models: Time-lagged ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society* **116**(494): 867–912. 15

Brewer S., Alleaume S., Guiot J., and Nicault A. (2007). Historical droughts in Mediterranean regions during the last 500 years: a data/model approach. *Climate Of The Past* **3**(2): 355–366. 1, 18, 139

Callado A., Escribà P., García-Moya A.J., Montero J., Santos C., Santos-Muñoz D., and Simarro J. (2013). Ensemble Forecasting. In P. Ray (ed.) *Climate Change and Regional/Local Responses*. InTech, 56. 67

Calmanti S., Motta L., Turco M., and Provenzale A. (2007). Impact of climate variability on Alpine glaciers in northwestern Italy. *International Journal of Climatology* **27**: 2041–2053. 54, 148, 184

Campbell S.D. and Diebold F.X. (2005). Weather Forecasting for Weather Derivatives. *Journal of the American Statistical Association* **100**(469): 6–16. 18

Cane M.A., Zebiak S.E., and Dolan S.C. (1986). Experimental forecasts of El Niño. *Nature* **321**: 827–832. 15

Carroll N., Frijters P., and Shields M.A. (2007). Quantifying the costs of drought: new evidence from life satisfaction data. *Journal of Population Economics* **22**(2): 445–461. 1

Carson D.J. (1998). Seasonal forecasting. *Quarterly Journal of the Royal Meteorological Society* **124**: 1–26. 9, 12, 60

Castro A., Vidal M.I., Calvo A.I., Fernández-Raga M., and Fraile R. (2011). May the NAO index be used to forecast rain in Spain? *Atmósfera* **24**(3): 251–265. 10

Céron J.P.P., Tanguy G., Franchistéguy L., Martin E., Regimbeau F., Vidal J.P.P., Franchistguy L., Martin E., Regimbeau F., and Vidal J.P.P. (2010). Hydrological seasonal forecast over France: feasibility and prospects. *Atmospheric Science Letters* **11**: 78–82. 14, 18, 19

Charney J.G. and Shukla J. (1981). Predictability of Monsoons. In J. Lighthill and R.P. Pearce (eds.) *Monsoon Dymics*. Cambridge University Press, Cambridge, 99–109. 9

Chavez-Jimenez A., Granados A., Garrote L., and Martín-Carrasco F. (2014). Adapting Water Allocation to Irrigation Demands to Constraints in Water Availability Imposed by Climate Change. *Water Resources Management* **29**: 1413–1430. 139

Chen Y., Randerson J.T., Morton D.C., DeFries R.S., Collatz G.J., Kasibhatla P.S., Giglio L., Jin Y., and Marlier M.E. (2011). Forecasting fire season severity in South America using sea surface temperature anomalies. *Science (New York, N.Y.)* **334**(6057): 787–91. 21

Chowdhury S. and Sharma A. (2009). Multisite seasonal forecast of arid river flows using a dynamic model combination approach. *Water Resources Research* **45**. 18

Chu P., Yan W., Fujioka D.F., and Fujioka F. (2002). Fire-climate relationships and long-lead seasonal wildfire prediction for Hawaii. *International Journal of Wildland Fire* **11**: 25–31. 21

Coelho C. and Costa S. (2010). Challenges for integrating seasonal climate forecasts in user applications. *Current Opinion in Environmental Sustainability* **2**(5-6): 317–325. 2, 17

Cohen J., Barlow M., Kushner P.J., and Saito K. (2007). Stratosphere–Troposphere Coupling and Links with Eurasian Land Surface Variability. *Journal of Climate* **20**(21): 5335–5343. 14

Colomer J., Serra T., Casamitjana X., and Baserba C. (2004). L'embassament de Boadella. *Revista de Girona* : 52–54. 44, 45

Dai A. (2011). Drought under global warming: A review. *Wiley Interdisciplinary Reviews: Climate Change* **2**(1): 45–65. 1, 19

Dai A., Trenberth K.E., and Karl T.R. (1999). Effects of Clouds, Soil Moisture, Precipitation, and Water Vapor on Diurnal Temperature Range. *Journal of Climate* **12**(8): 2451–2473. 13

Damrongkulkamjorn P. and Churueang P. (2005). Monthly energy forecasting using decomposition method with application of seasonal ARIMA. In *Conference, 2005. IPEC*. 18

Danielson J.J. and Gesch D.B. (2011). Global Multi-resolution Terrain Elevation Data 2010. Tech. rep., USGS. 34, 38, 42, 43, 45

Davey M.K. and Brookshaw A. (2011). Long-range meteorological forecasting and links to agricultural applications. *Food Policy* **36**: S88–S93. 18

de Groot W.J. and Flannigan M.D. (2014). Climate change and early warning ystems for wildland fire. In A. Singh and Z. Zommers (eds.) *Reducing Disaster: Early Warning Systems and Climate Change*, chap. 7. Springer Netherlands, Dordrecht, 127–251. 21

Dee D.P., Uppala S.M., Simmons A.J., Berrisford P., Poli P., Kobayashi S., Andrae U., Balmaseda M.A., Balsamo G., Bauer P., Bechtold P., Beljaars A.C.M., van de Berg L., Bidlot J., Bormann N., Delsol C., Dragani R., Fuentes M., Geer a.J., Haimberger L., Healy S.B., Hersbach H., Hólm E.V., Isaksen L., Kållberg

P., Köhler M., Matricardi M., Mcnally a.P., Monge-Sanz B.M., Morcrette J.J., Park B.K., Peubey C., de Rosnay P., Tavolato C., Thépaut J.N., and Vitart F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society* **137**: 553–597. 27

DelSole T. and Shukla J. (2010). Model Fidelity versus Skill in Seasonal Forecasting. *Journal of Climate* **23**: 4794–4806. 15

DelSole T. and Shukla S. (2009). Artificial skill due to predictor selection. *Journal of Climate* **22**: 331–345. 53

Déqué M. (2007). Frequency of precipitation and temperature extremes over France in an anthropogenic. *Global and Planetary Change* **57**: 16–26. 50, 179

Díaz-Delgado R., Lloret F., and Pons X. (2004). Spatial patterns of fire occurrence in Catalonia, NE, Spain. *Landscape Ecology* **19**(7): 731–745. 178

Diez E., Primo C., García-Moya J.A., Gutiérrez J.M., and Orfila B. (2005). Statistical and dynamical downscaling of precipitation over Spain from DEMETER seasonal forecasts. *Tellus A* **57**(3): 409–423. 16

Diffenbaugh N.S. and Giorgi F. (2012). Climate change hotspots in the CMIP5 global climate model ensemble. *Climatic Change* **114**(3-4): 813–822. 1

Diffenbaugh N.S., Pal J.S., Giorgi F., and Gao X. (2007). Heat stress intensification in the Mediterranean climate change hotspot. *Geophysical Research Letters* **34**(11): L11706. 1

Diomede T., Nerozzi F., Paccagnella T., and Todini E. (2006). The use of meteorological analogues to account for LAM QPF uncertainty. *Hydrology and Earth System Sciences Discussions* **3**(5): 3061–3097. 55

Doblas-Reyes F.J., Coelho C.A.S., and Stephenson D.B. (2008). How much does simplification of probability forecasts reduce forecast quality? *Meteorological Applications* **15**: 155–162. 60

Doblas-Reyes F.J., Déqué M., and Piedelievre J.P. (2000). Multi model spread and probabilistic seasonal forecasts in PROVOST. *Quarterly Journal of the Royal Meteorological Society* **126**: 2069–2087. 10

Doblas-Reyes F.J., García-Serrano J., Lienert F., Biescas A.P., and Rodrigues L.R. (2013). Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdisciplinary Reviews: Climate Change* **4**: 245–268. 2, 3, 11, 12, 15

Doblas-Reyes F.J., Hagedorn R., and Palmer T.N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination. *Tellus A* **57**(3): 234–252. 16, 80

Doblas-Reyes F.J., Weisheimer A., Déqué M., Keenlyside N., McVean M., Murphy J.M., Rogel P., Smith D., and Palmer T.N. (2009). Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society* **135**(643): 1538–1559. 16, 17

Duband D. (1981). Prévision spatiale des hauteurs de précipitations journalières. *La Houille Blanche* **7**(8): 497–512. 54

Dube O.P. (2009). Linking fire and climate: interactions with land use, vegetation, and soil. *Current Opinion in Environmental Sustainability* **1**(2): 161–169. 20

Ducré-Robitaille J.F., Vincent L.A., and Boulet G. (2003). Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23**(9): 1087–1101. 25

Dutra E., Di Giuseppe F., Wetterhall F., and Pappenberger F. (2013). Seasonal forecasts of droughts in African basins using the Standardized Precipitation Index. *Hydrology and Earth System Sciences* **17**: 2359–2373. 18, 180

Dutra E., Pozzi W., Wetterhall F., Di Giuseppe F., Magnusson L., Naumann G., Barbosa P., Vogt J., and Pappenberger F. (2014). Global meteorological drought – Part 2: Seasonal forecasts. *Hydrology and Earth System Sciences* **18**(7): 2669–2678. 2

Dutton J.A., James R.P., and Ross J.D. (2013). Calibration and combination of dynamical seasonal forecasts to enhance the value of predicted probabilities for managing risk. *Climate Dynamics* **40**(11-12): 3089–3105. 18

Dwyer E., Grégoire J.M., and Pereira J.M.C. (2000). Climate and vegetation as driving factors in global fire activity. In *Biomass Burning and Its Inter-Relationships with the Climate System*. Springer Netherlands, 171–191. 20

el Kenawy A., López-Moreno J.I., and Vicente-Serrano S.M. (2012). Trend and variability of surface air temperature in northeastern Spain (1920–2006): Linkage to atmospheric circulation. *Atmospheric Research* **106**: 159–180. 10

Falloon P., Challinor A., Dessai S., Hoang L., Johnson J., and Koehler A.K. (2014). Ensembles and uncertainty in climate change impacts. *Frontiers in Environmental Science* **2**. 10

Fan Y.R., Huang W., Huang G.H., Li Z., Li Y.P., Wang X.Q., Cheng G.H., and Jin L. (2015). A stepwise-cluster forecasting approach for monthly streamflows based on climate teleconnections. *Stochastic Environmental Research and Risk Assessment* **29**(6): 1557–1569. 141

Fil C. and Dubus L. (2005). Winter climate regimes over the North Atlantic and European region in ERA40 reanalysis and DEMETER seasonal hindcasts. *Tellus A* **57**(3): 290–307. 12

Flannigan M.D., Krawchuk M., de Groot W.J., Wotton B., and Gowman L. (2009). Implications of changing climate for global wildland fire. *International Journal of Wildland Fire* **18**: 483–507. 1, 20

Flannigan M.D., Stocks B.J., and Wotton B.M. (2000). Climate change and forest fires. *The Science of the total environment* **262**(3): 221–229. 20

Folland C.K., Scaife A.A., Lindesay J., and Stephenson D.B. (2012). How potentially predictable is northern European winter climate a season ahead? *International Journal of Climatology* **32**(6): 801–818. 12

Fraisse C.W., Breuer N.E., Zierden D., Bellow J.G., Paz J., Cabrera V.E., Garcia y Garcia A., Ingram K.T., Hatch U., Hoogenboom G., Jones J.W., and O'Brien J.J. (2006). AgClimate: A climate forecast information system for agricultural risk management in the southeastern USA. *Computers and Electronics in Agriculture* **53**: 13–27. 18

Froude L.S.R., Bengtsson L., and Hodges K.I. (2013). Atmospheric predictability revisited. *Tellus A* **65**: 1–13. 11

Fukutome S., Prim C., and Schär C. (2001). The role of soil states in medium-range weather predictability. 14

Ganteaume A., Camia A., Jappiot M., San-Miguel-Ayanz J., Long-Fournel M., and Lampin C. (2013). A review of the main driving factors of forest fire ignition over Europe. *Environmental management* **51**(3): 651–62. 20

Garbrecht J.D. and Schneider J.M. (2007). Climate forecast and prediction product dissemination for agriculture in the United States. *Australian Journal of Agricultural Research* **58**(10): 966–974. 2, 18

García-Morales M.B. and Dubus L. (2007). Forecasting precipitation for hydroelectric power management: how to exploit GCM's seasonal ensemble forecasts. *International Journal of Climatology* **27**(12): 1691–1705. 18

Garrote de Marcos L. and Cubillo F. (2008). Drought risk and vulnerability in water supply systems. *Options Mediterraneennes Series A* **80**: 21–27. 139

Gaushell D. (2002). Forecasting loads and prices for New England market. In *IEEE Power Engineering Society Summer Meeting*, vol. 3. 18

Gibergans-Báguena J. and Llasat M.C. (2007). Improvement of the analog forecasting method by using local thermodynamic data. Application to autumn precipitation in Catalonia. *Atmospheric Research* **86**(3-4): 173–193. 4

Giorgi F. and Lionello P. (2008). Climate change projections for the Mediterranean region. *Global and Planetary Change* **63**(2-3): 90–104. 192

Gneiting T. and Raftery A.E. (2005). Atmospheric science. Weather forecasting with ensemble methods. *Science (New York, N.Y.)* **310**(5746): 248–249. 15

Gobena A.K. and Gan T.Y. (2010). Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology* **385**: 336–352. 19

Goddard L., Aitchellouche Y., Baethgen W., Dettinger M., Graham R., Hayman P., Kadi M., Martínez R., and Meinke H. (2010). Providing Seasonal-to-Interannual Climate Information for Risk Management and Decision-making. *Procedia Environmental Sciences* **1**: 81–101. 18

González J.R. and Pukkala T. (2007). Characterization of forest fires in Catalonia (north-east Spain). *European Journal of Forest Research* **126**(3): 421–429. 178

Graham R.J., Yun W.T., Kim J., Kumar A., Jones D., Bettio L., Gagnon N., Kolli R.K., and Smith D. (2011). Long-range forecasting and the Global Framework for Climate Services. *Climate Research* **47**: 47–55. 17

Gudmundsson L., Rego F.C., Rocha M., and Seneviratne S.I. (2014). Predicting above normal wildfire activity in southern Europe as a function of meteorological drought. *Environmental Research Letters* **9**: 084008. 20, 21, 191

Guéremy J.F., Laanaia N., and Céron J.P. (2012). Seasonal Forecast of French Mediterranean heavy precipitating events linked to weather regimes. *Natural Hazards and Earth System Science* **12**(7): 2389–2398. 12

Gutiérrez J.M., Cano R., Cofiño A.S., and Sordo C.M. (2004). *Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas*. Centro de Publicaciones. Secretaría Técnica. Ministerio de Medio Ambiente, Santander, series mon edn. 16, 52, 71, 72, 73, 85

Gutiérrez J.M., San-Martín D., Brands S., Manzanas R., and Herrera S. (2013). Reassessing Statistical Downscaling Techniques for Their Robust Application under Climate Change Conditions. *Journal of Climate* **26**: 171–178. 17

Hagedorn R., Doblas-Reyes F.J., and Palmer T.N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. *Tellus A* **57**(3): 219–233. 16

Haren R., Oldenborgh G.J., Lenderink G., Collins M., and Hazeleger W. (2012). SST and circulation trend biases cause an underestimation of European precipitation trends. *Climate Dynamics* **40**(1-2): 1–20. 12

Hargreaves G.H. (1994). Defining and Using Reference Evapotranspiration. *Journal of Irrigation and Drainage Engineering* **120**: 1132–1139. 180

Harris S., Nicholls N., and Tapper N. (2014). Forecasting fire activity in Victoria, Australia, using antecedent climate variables and ENSO indices. *International Journal of Wildland Fire* **23**(2): 173. 18, 21

Hastenrath S., Sun L., and Moura A.D. (2009). Climate prediction for Brazil's Nordeste by empirical and numerical modeling methods. *International Journal of Climatology* **29**(6): 921–926. 16

Haylock M.R., Hofstra N., Klein-Tank A., Klok E.J., Jones P.D., and New M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research* **113**(D20119): 1–12. 24, 25

Herrera S., Gutiérrez J.M., Ancell R., Pons M.R., Frías M.D., and Fernández J. (2012). Development and analysis of a 50-year high-resolution daily gridded precipitation dataset over Spain (Spain02). *International Journal of Climatology* **32**(1): 74–85. 24

Herrera S., Pazó D., Fernández J., and Rodríguez M.M. (2011). The role of largescale spatial patterns in the chaotic amplification of perturbations in a Lorenz'96 model. *Tellus A* : 1–11. 5

Hessl A.E. (2011). Pathways for climate change effects on fire: Models, data, and uncertainties. *Progress in Physical Geography* **35**(3): 393–407. 20

Hewitt C., Mason S., and Walland D. (2012). The Global Framework for Climate Services. *Nature Climate Change* **2**(12): 831–832. 17

Hill J., Stellmes M., Udelhoven T., Röder A., and Sommer S. (2008). Mediterranean desertification and land degradation. *Global and Planetary Change* **64**(3-4): 146–157. 19

Hirschi M., Seneviratne S.I., Alexandrov V., Boberg F., Boroneant C., Christensen O.B., Formayer H., Orlowsky B., and Stepanek P. (2010). Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nature Geoscience* **4**(1): 17–21. 13

Hobbins M.T., Dai A., Roderick M.L., and Farquhar G.D. (2008). Revisiting the parameterization of potential evaporation as a driver of long-term water balance trends. *Geophysical Research Letters* **35**: L12403. 141

Hofstra N., Haylock M., New M., Jones P., and Frei C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres* **113**(D21): D21110. 25

Huth R. (2002). Statistical Downscaling of Daily Temperature in Central Europe. *Journal of Climate* **15**(13): 1731–1742. 58

Iglesias A. and Garrote L. (2014). Drought in the Light of Climate Change in the Mediterranean Area. In P. Quevauviller (ed.) *Hydrometeorological Hazards: Interfacing Science and Policy*. John Wiley & Sons, Ltd., Chichester, 203–225. 2, 19

Iglesias A., Garrote L., Cancelliere A., Cubillo F., and Wilhite D. (2009). Coping with drought risk in agriculture and water supply systems: Drought management and policy development in the mediterranean. In *Advances in natural and technological hazards research*, vol. 26. 320. 19

Iglesias A., Garrote L., Flores F., and Moneo M. (2006). Challenges to Manage the Risk of Water Scarcity and Climate Change in the Mediterranean. *Water Resources Management* **21**(5): 775–788. 139

Ineson S. and Scaife A.A. (2008). The role of the stratosphere in the European climate response to El Niño. *Nature Geoscience* **2**(1): 32–36. 14

IPCC (2013a). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, vol. AR5. Cambridge University Press, Cambridge. 1

IPCC (2013b). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Intergovernmental Panel on Climate Change, Geneva. 192

IPCC (2014). *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Parts A and B: Global and Sectorial and Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Contribution of.* Cambridge University Press, Cambridge. 1

Ji M., Kumar A., and Leetmaa A. (1994). A Multiseason Climate Forecast System at the National Meteorological Center. *Bulletin of the American Meteorological Society* **75**: 569–577. 15

Johnson C. and Bowler N. (2009). On the Reliability and Calibration of Ensemble Forecasts. *Monthly Weather Review* **137**: 1717–1720. 2, 80

Johnston P.A., Archer E., Vogel C.H., Bezuidenhout C.N., Tennant W.J., and Kuschke R. (2004). Review of seasonal forecasting in South Africa: producer to end-user. *Climate Research* **28**: 67–82. 2

Journel A.G. and Huijbregts C. (1978). *Mining geostatistics.* Academic Press. 25

Kanamitsu M., Lu C.H., Schemm J., and Ebisuzaki W. (2003). The Predictability of Soil Moisture and Near-Surface Temperature in Hindcasts of the NCEP Seasonal Forecast Model. *Journal of Climate* **16**(1984): 510–521. 13

Kim H., Webster P.J., and Curry J.A. (2012). Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dynamics* **39**(12): 2957–2973. 24

Kim H.M. and Webster P.J. (2010). Extended-range seasonal hurricane forecasts for the North Atlantic with a hybrid dynamical-statistical model. *Geophysical Research Letters* **37**(21): L21705. 18

Klok E.J. and Klein-Tank A. (2009). Updated and extended European dataset of daily climate observations. *International Journal of Climatology* **29**: 1182–1191. 24

Koenigk T. and Mikolajewicz U. (2009). Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Climate Dynamics* **32**(6): 783–798. 3

Koutsias N., Arianoutsou M., Kallimanis A.S., Mallinis G., Halley J.M., and Dimopoulos P. (2012). Where did the fires burn in Peloponnisos, Greece the summer of 2007? Evidence for a synergy of fuel and weather. *Agricultural and Forest Meteorology* **156**: 41–53. 191

Koutsias N., Xanthopoulos G., Founda D., Xystrakis F., Nioti F., Pleniou M., Mallinis G., and Arianoutsou M. (2013). On the relationships between forest fires and weather conditions in Greece from long-term national observations (1894-2010). *International Journal of Wildland Fire* **22**: 493–507. 20

Kug J.S., Kang I.S., and Choi D.H. (2008). Seasonal climate predictability with Tier-one and Tier-two prediction systems. *Climate Dynamics* **31**(4): 403–416. 13, 15, 27

Kumar A. (2009). On the assessment of the value of the seasonal forecast information. *Meteorological Applications* **392**(December 2009): 385–392. 2, 16

LaRow T.E., Stefanova L., Shin D.W., and Cocke S. (2010). Seasonal Atlantic tropical cyclone hindcasting/forecasting using two sea surface temperature datasets. *Geophysical Research Letters* **37**(2): L02804. 18

Leckebusch G.C., Befort D.J., Weisheimer A., Knight J., Thornton H., Roberts J., and Hermanson L. (2015). Extra-tropical Cyclones and Windstorms in Seasonal Forecasts. *EGU General Assembly Conference Abstracts* **17**: 11120. 18

Llasat M.C. and Puigcerver M. (1997). Total rainfall and convective rainfall in Catalonia, Spain. *International Journal of Climatology* **17**(15): 1683–1695. 4

Llasat M.C., Zaragoza A., Cabot J., and Aznar B. (2010). Une approximation de la prévision saisonnière des étiages et sécheresses en Catalogne. *La Houille Blanche* **4**: 60–66. 2, 4

Lloyd-Hughes B. and Saunders M.A. (2002). A drought climatology for Europe. *International Journal of Climatology* **22**(13): 1571–1592. 18

Lorenz E.N. (1969a). Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *Journal of the Atmospheric Sciences* **26**: 636–646. 54, 56

Lorenz E.N. (1969b). Three approaches to atmospheric predicitability. *Bulletin of the American Meteorological Society* **50**: 345. 9

Macias Fauria M., Michaletz S.T., and Johnson E.A. (2011). Predicting climate change effects on wildfires requires linking processes across scales. *Wiley Interdisciplinary Reviews Climate Change* **2**(1): 99–112. 20

Manzanas R., Bedia J., Magarino M.E., Fernández J., Cofiño A.S., and Gutiérrez J.M. (2013). Preliminary Validation of the Seasonal Forecasting Systems included in the SPECS-EUPORIAS Data Portal: The ECMWF System4 Hindcast. Tech. rep. 10

Maraun D., Wetterhall F., Ireson A.M., Chandler R.E., Kendon E.J., Widmann M., Brienen S., Rust H.W., Sauter T., Themessl M., Venema V.K.C., Chun K.P., Goodess C.M., Jones R.G., Onof C., Vrac M., and Thiele-Eich I. (2010). Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics* **48**: 1–38. 17, 55

Marshall A.G. and Scaife A.A. (2009). Impact of the QBO on surface winter climate. *Journal of Geophysical Research* **114**(D18): D18110. 14

Martín-Vide J. and Barriendos M. (1995). The use of rogation ceremony records in climatic reconstruction: a case study from Catalonia (Spain). *Climatic Change* **30**(2): 201–221. 18

Mason S.J., Goddard L., Graham N.E., Yulaleva E., Sun L., and Arkin P.A. (1999). The IRI seasonal climate prediction system and the 1997/1998 El Niño event. *Bulletin of the American Meteorological Society* **80**: 1853–1873. 88

Mason S.J. and Graham N.E. (1999). Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels. *Weather and Forecasting* **14**: 713–725. 184

Matulla C., Schmutz S., Melcher A., Gerersdorfer T., and Haas P. (2007). Impact of a downscaled climate change simulation on the fish fauna in an Inner-Alpine river. *International Journal of Biometeorology* **52**(2): 127–137. 55

Mavsar R., González Cabán A., and Varela E. (2013). The state of development of fire management decision support systems in America and Europe. *Forest Policy and Economics* **29**: 45–55. 2, 20, 192

Mckee T.B., Doesken N.J., and Kleist J. (1993). The relationship of drought frequency and duration to time scales. In *Proceedings of the Ninth Conference on Applied Climatology*. American Meteorological Society, Boston, 179–184. 179

MedCLIVAR (2012). *The climate of the Mediterranean region: understanding its evolution and effects on environment and societies Organizing Committee*. September. 19

Meyn A., White P.S., Buhk C., and Jentsch A. (2007). Environmental drivers of large, infrequent wildfires: the emerging conceptual model. *Progress in Physical Geography* **31**(3): 287–312. 20, 177

Michaelsen J. (1987). Cross-Validation in Statistical Climate Forecast Models. *Journal of Climate and Applied Meteorology* **26**: 1589–1600. 52

MIMAM (2007). Planes especiales de sequía. Ministerio de medio ambiente. Secretaría de Estado de Aguas y Costas. Dirección general de Obras Hidráulicas y Calidad de las Aguas. Tech. rep. 2, 19

Moberg A. and Jones P.D. (2005). Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901-99. *International Journal of Climatology* **25**(9): 1149–1171. 81

Mogensen K., Balmaseda M.A., and Weaver A. (2012). The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4. Tech. rep., ECMWF. 27

Molteni F., Cubasch U., and Tibaldi S. (1987). 30 and 60-day forecast experiments with the ECMWF spectral models. In *Proceedings of the first ECMWF Workshop on Predictability in the medium and extended range*. ECMWF, Reading, 51–108. 15

Molteni F., Stockdale T., Balmaseda M.A., Balsamo G., Buizza R., Ferranti L., Magnusson L., Mogensen K., and Palmer T. (2011). The new ECMWF seasonal forecast system (System 4). *ECMWF Technical Memorandum* **656**(November): 1–51. 2, 26, 27, 29

Moreira F., Viedma O., Arianoutsou M., Curt T., Koutsias N., Rigolot E., Barbati A., Corona P., Vaz P., Xanthopoulos G., Mouillot F., and Bilgili E. (2011). Landscape–wildfire interactions in southern Europe: implications for landscape management. *Journal of environmental management* **92**: 2389–2402. 1, 20, 192

Moreno M.V., Conedera M., Chuvieco E., and Pezzatti G.B. (2014). Fire regime changes and major driving forces in Spain from 1968 to 2010. *Environmental Science & Policy* **37**: 11–22. 192

Moriondo M., Good P., Durao R., Bindi M., Giannakopoulos C., and Corte-Real J. (2006). Potential impact of climate change on fire risk in the Mediterranean area. *Climate Research* **31**: 85–95. 192

Moritz M.A., Parisien M.A., Batllori E., Krawchuk M.A., Van Dorn J., Ganz D.J., and Hayhoe K. (2012). Climate change and disruptions to global fire activity. *Ecosphere* **3**(6): 1–22. 20

Murphy S.J., Washington R., Downing T.E., Martin R.V., Ziervogel G., Preston A., Todd M., Butterfield R., and Briden J. (2001). Seasonal forecasting for climate hazards: prospects and responses. *Natural Hazards* **23**: 171–196. 9, 16

Nepstad D., Lefebvre P., Lopes da Silva U., Tomasella J., Schlesinger P., Solorzano L., Moutinho P., Ray D., and Guerreira-Benito J. (2004). Amazon drought and its implications for forest flammability and tree growth: a basin-wide analysis. *Global Change Biology* **10**: 704–717. 13

Nola P., Carrer M., Brewer S., Urbinati C., Edouard J.L., Alleaume S., Guiot J., Nicault A., and Guttierez E. (2008). Mediterranean drought fluctuation during the last 500years based on tree-ring data. 1, 18, 139

Oficina Catalana del Canvi Climàtic (2012). Estratègia Catalana d'Adaptació al Canvi Climàtic. Tech. rep. 172

O'Lenic E.A., Unger D.A., Halpert M.S., and Pelman K.S. (2008). Developments in operational long-range climate prediction at CPC. *Weather and Forecasting* **23**: 496–515. 88

Orsolini Y.J., Kindem I.T., and Kvamstø N.G. (2009). On the potential impact of the stratosphere upon seasonal dynamical hindcasts of the North Atlantic Oscillation: a pilot study. *Climate Dynamics* **36**(3-4): 579–588. 14

Orsolini Y.J., Senan R., Benestad R.E., and Melsom A. (2011). Autumn atmospheric response to the 2007 low Arctic sea ice extent in coupled ocean–atmosphere hindcasts. *Climate Dynamics* **38**(11-12): 2437–2448. 14

Orth R. and Seneviratne S.I. (2012). Analysis of soil moisture memory from observations in Europe. 13

Palmer T. and Anderson D. (1994). The prospects for seasonal forecasting—A review paper. *Quarterly Journal of the Royal Meteorological Society* **120**(518): 755–793. 2, 9, 11, 12, 13

Palmer T.N. (1999). Predicting uncertainty in forecasts of weather and climate. Tech. Rep. 294. 15

Palmer T.N., Alessandri A., Andersen U., Cantelaube P., Davey M., Délécluse P., Déqué M., Díez E., Doblas-Reyes F.J., Feddersen H., Graham R., Gualdi S., Guérémy J.F., Hagedorn R., Hoshen M., Keenlyside N., Latif M., Lazar A., Maisonnave E., Marletto V., Morse A.P., Orfila B., Rogel P., Terres J.M., and Thomson M.C. (2004). Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *Bulletin of the American Meteorological Society* **85**(6): 853–872. 10, 12, 28

Pascual M. and Dobson A. (2005). Seasonal patterns of infectious diseases. *PLoS Medicine* **2**(1): 0018–0020. 18

Pausas J.G. (2004). Changes in Fire and Climate in the Eastern Iberian Peninsula (Mediterranean Basin). *Climatic Change* **63**(3): 337–350. 20, 177, 191

Pausas J.G. and Keeley J.E. (2014). Evolutionary ecology of resprouting and seeding in fire-prone ecosystems. *New Phytologist* **204**: 55–65. 192

Pavón D. (2001a). L'embassament de Boadella: antecedents d'una obra, transformacions d'un territori. *Revista de Girona* : 38–45. 44

Pavón D. (2001b). L'embassament de Boadella: antecedents, execució i repercussions locals d'una infraestructura hidràulica. *AIEE* **34**: 179–217. 44

Pechony O. and Shindell D.T. (2010). Driving forces of global wildfires over the past millennium and the forthcoming century. *Proceedings of the National Academy of Sciences* **107**(45): 19167–19170. 20

Peng Z., Wang Q.J., Bennett J.C., Schepen A., Pappenberger F., Pokhrel P., and Wang Z. (2014). Statistical calibration and bridging of ECMWF System4 outputs for forecasting seasonal precipitation over China. *Journal of Geophysical Research: Atmospheres* **119**(12): 7116–7135. 17

Pereira M.G., da Camara C.C., Trigo R.M., Pereira J.M.C., and Leite S.M. (2005). Synoptic patterns associated with large summer forest fires in Portugal. 20

Peterson T.C., Easterling D.R., Karl T.R., Groisman P., Nicholls N., Plummer N., Torok S., Auer I., Boehm R., Gullett D., Vincent L., Heino R., Tuomenvirta H., Mestre O., Szentimrey T., Salinger J., Førland E.J., Hanssen-Bauer I., Alexandersson H., Jones P., and Parker D. (1998). Homogeneity adjustments ofin situ atmospheric climate data: a review. *International Journal of Climatology* **18**(13): 1493–1517. 25

Piani C., Haerter J.O., and Coppola E. (2010). Statistical bias correction for daily precipitation in regional climate models over Europe. *Theoretical and Applied Climatology* **99**(1-2): 187–192. 17

Plana E. (2011). Integració del risc d'incendis en la planificació forestal estratègica i l'ordenació del territori. *Treball de la societat catalana de geografia* **71**: 69–91. 2

Preisler H.K., Brillinger D.R., Burgan R.E., and Benoit J.W. (2004). Probability based models for estimation of wildfire risk. *International Journal of Wildland Fire* **13**: 133–142. 21, 191

Preisler H.K., Chen S.C., Fujioka F., Benoit J.W., and Westerling A.L. (2008). Wildland fire probabilities estimated from weather model-deduced monthly mean fire danger indices. *International Journal of Wildland Fire* **17**(3): 305. 21, 191

Preisler H.K. and Westerling A.L. (2007). Statistical Model for Forecasting Monthly Large Wildfire Events in Western United States. *Journal of Applied Meteorology and Climatology* **46**: 1020–1030. 21, 191

Quan X., Hoerling M., Whitaker J., Bates G., and Xu T. (2006). Diagnosing Sources of U.S. Seasonal Forecast Skill. *Journal of Climate* **19**: 3279–3293. 12

Quintana-Seguí P., Habets F., and Martin E. (2011). Comparison of past and future Mediterranean high and low extremes of precipitation and river flow projected using different statistical downscaling methods. *Natural Hazards and Earth System Science* **11**(5): 1411–1432. 139

Quintana-Seguí P., Le Moigne P., Durand Y., Martin E., Habets F., Baillon M., Canellas C., Franchistéguy L., and Morel S. (2008). Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. *Journal of Applied Meteorology and Climatology* **47**(1): 92–107. 4

Quintana-Seguí P., Martin E., Habets F., and Noilhan J. (2009). Improvement, calibration and validation of a distributed hydrological model over France. *Hydrol. Earth Syst. Sci.* **13**(2): 163–181. 4

Quintana-Seguí P., Ribes A., Martin E., Habets F., and Boé J. (2010). Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins. *Journal of Hydrology* **383**(1-2): 111–124. 4

Rahman M. and Lu M. (2015). Model Spin-Up Behavior for Wet and Dry Basins: A Case Study Using the Xinanjiang Model. *Water* **7**(8): 4256–4273. 142

Reeves J., Chen J., Wang X.L., Lund R., and Lu Q.Q. (2007). A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology* **46**(6): 900–915. 25

Roads J., Fujioka F., Chen S., and Burgan R. (2005). Seasonal fire danger forecasts for the USA. *International Journal of Wildland Fire* **14**: 1–18. 21

Roads J., Tripp P., Juang H., Wang J., Fujioka F., and Chen S. (2010). NCEP-ECPC monthly to seasonal US fire danger forecasts. *International Journal of Wildland Fire* **19**: 399–414. 21, 191

Robertson D.E., Pokhrel P., and Wang Q.J. (2012). Improving statistical forecasts of seasonal streamflows using hydrological model output. *Hydrology and Earth System Sciences Discussions* **9**(7): 8701–8736. 18

Rodrigues L.R., García-Serrano J., and Doblas-Reyes F. (2014). Seasonal forecast quality of the West African monsoon rainfall regimes by multiple forecast systems. *Journal of Geophysical Research: Atmospheres* **119**(13): 7908–7930. 27

Rodwell M.J. and Doblas-Reyes F.J. (2006). Medium-range, monthly, and seasonal prediction for Europe and the use of forecast information. *Journal of Climate* **19**: 6025–6046. 11, 13, 19

Roncoli C., Breuer N., Zierden D., and Fraisse C. (2012). The art of the science: climate forecasts for wildfire management in the southeastern United States. *Climatic Change* : 1113–1121. 18

Ropelewski C.F. and Halpert M.S. (1986). North American Precipitation and Temperature Patterns Associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review* **114**(12): 2352–2362. 10

Rossi G., Caporali E., Garrote L., and Federici G.V. (2012). Defining operating rules for mitigation of drought effects on water supply systems. *EGU General Assembly Conference Abstracts* **14**: 12127. 139

San-Miguel-Ayanz J., Moreno J.M., and Camia A. (2013a). Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives. *Forest Ecology and Management* **294**: 11–22. 192

San-Miguel-Ayanz J., Schulte E., Schmuck G., and Camia A. (2013b). The European Forest Fire Information System in the context of environmental policies of the European Union. *Forest Policy and Economics* **29**: 19–25. 1, 20, 192

Santander Meteorology Group (2015). downscaleR: Climate data manipulation and statistical downscaling. R package version 0.5-2. 180

Scaife A.A., Arribas A., Blockley E., Brookshaw A., Clark R.T., Dunstone N., Eade R., Fereday D., Folland C.K., Gordon M., Hermanson L., Knight J.R., Lea D.J., MacLachlan C., Maidens A., Martin M., Peterson A.K., Smith D., Vellinga M., Wallace E., Waters J., and Williams A. (2014). Skillful long-range prediction of European and North American winters. *Geophyisical Research Letters* **41**: 1–6. 12

Schepen A., Wang Q.J., and Robertson D.E. (2012). Combining the strengths of statistical and dynamical modeling approaches for forecasting Australian seasonal rainfall. *Journal of Geophysical Research: Atmospheres* **117**: D20107. 16

Schneider J.M. and Garbrecht J.D. (2006). Dependability and Effectiveness of Seasonal Forecasts for Agricultural Applications. *Transactions fo the American Society of Agricultural and Biological Engineers* **49**(6): 1737–1753. 2, 18

Schwierz C., Appenzeller C., and Davies H.C. (2006). Challenges posed by and approaches to the study of seasonal-to-decadal climate variability. *Climatic Change* : 1–54. 3

Sear D.A. and Dawson F.H. (1999). Groundwater dominated rivers. *Hydrological Processes* **13**(3): 255–276. 143

Seck A., Welty C., and Maxwell R.M. (2015). Spin-up behavior and effects of initial conditions for an integrated hydrologic model. *Water Resources Research* **51**: 2188–2210. 142

Shabbar A., Skinner W., and Flannigan M.D. (2011). Prediction of seasonal forest fire severity in Canada from large-scale climate patterns. *Journal of Applied Meteorology and Climatology* **50**: 785–799. 21

Shafiee-Jood M., Cai X., Chen L., Liang X., and Kumar P. (2012). Assessing the value of seasonal climate forecast information through an end-to-end forecasting framework: Application to U.S. 2012 drought in central Illinois. *Water Resources Research* : 1–55. 19

Shongwe M.E., Ferro C.A.T., Coelho C., van Oldenborgh G.J., Bilt D., and Paulista C. (2007). Predictability of Cold Spring Seasons in Europe. *Monthly Weather Review* **135**(12): 4185–4201. 14

Shukla S. and Lettenmaier D.P. (2011). Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill. *Hydrology and Earth System Sciences* **15**: 3529–3538. 19

Sigmond M., Fyfe J.C., Flato G.M., Kharin V.V., and Merryfield W.J. (2013a). Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. *Geophysical Research Letters* **40**(3): 529–534. 14

Sigmond M., Scinocca J.F., Kharin V.V., and Shepherd T.G. (2013b). Enhanced seasonal forecast skill following stratospheric sudden warmings. *Nature Geoscience* **6**: 98–102. 14

Singla S. (2012). *Prévisibilité des ressources en eau à l'échelle saisonnière en France*. Ph.D. thesis. 19

Sirdas S., Ross R.S., Krishnamurti T.N., and Chakraborty A. (2007). Evaluation of the FSU synthetic superensemble performance for seasonal forecasts over the Euro-Mediterranean region. *Tellus A* **59**(1): 50–70. 16

Sophocleous M. (2002). Interactions between groundwater and surface water: the state of the science. *Hydrogeology Journal* **10**(1): 52–67. 143, 145

Soubeyroux J.M., Vidal J.P., Baillon M., Blanchard M., Céron J.P., Franchistéguy L., Régimbeau F., Martin E., and Vincendon J.C. (2010). Caractérisation et prévision des sécheresses et étiages en France à partir de la chaîne hydrométéorologique Safran-Isba-Modcou. *La Houille Blanche* : 30–39. 19

Spessa A.C., Field R.D., Pappenberger F., Langner A., and Englhart S. (2014). Seasonal forecasting of fire over Kalimantan, Indonesia. *Natural Hazards and Earth System Sciences Discussions* **2**: 5079–5111. 18, 21

Steinemann A.C. (2006). Using Climate Forecasts for Drought Management. *Journal of Applied Meteorology and Climatology* **45**: 1353–1361. 20

Stockdale T.N. (2000). An overview of techniques for seasonal forecasting. *Stochastic Environmental Research and Risk Assessment* **14**(4): 0305–0318. 11, 13, 19

Stockdale T.N., Anderson D.L.T., Balmaseda M.A., Doblas-Reyes F., Ferranti L., Mogensen K., Palmer T.N., Molteni F., and Vitart F. (2011). ECMWF seasonal forecast system 3 and its prediction of sea surface temperature. *Climate Dynamics* **37**(3-4): 455–471. 27

Tall A., Mason S.J., Van Aalst M., Suarez P., Ait-Chellouche Y., Diallo A.A., and Braman L. (2012). Using seasonal climate forecasts to guide disaster management: The Red Cross experience during the 2008 West Africa floods. *International Journal of Geophysics* **2012**: 1–12. 18

Taylor K.E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* **106**(D7): 7183–7192. 63, 64, 87

Thomson M.C., Doblas-Reyes F.J., Mason S.J., Hagedorn R., Connor S.J., Phindela T., Morse A.P., and Palmer T.N. (2006). Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature* **439**: 576–579. 18

Thornthwaite C.W. (1948). An Approach toward a Rational Classification of Climate. *Geographical Review* **38**(1): 55–94. 141

Tompkins A.M. and Di Giuseppe F. (2014). Potential predictability of malaria in Africa using ECMWF monthly and seasonal climate forecasts. *Journal of Applied Meteorology and Climatology* (2002): 141230123742001. 18

Trenberth K.E. (2005). Relationships between precipitation and surface temperature. *Geophysical Research Letters* **32**(L14703): 1–4. 12

Trenberth K.E., Dai A., Rasmussen R.M., and Parsons D.B. (2003). The Changing Character of Precipitation. *Bulletin of the American Meteorological Society* **84**(9): 1205–1217. 55

Turco M. and Llasat M.C. (2011). Trends in indices of daily precipitation extremes in Catalonia (NE Spain), 1951-2003. *Natural Hazards and Earth System Science* **11**: 3213–3226. 139

Turco M., Llasat M.C., Hardenberg J., and Provenzale A. (2013a). Impact of climate variability on summer fires in a Mediterranean environment (northeastern Iberian Peninsula). *Climatic Change* **116**: 665–678. 5, 13, 18, 20, 30, 177, 182

Turco M., Llasat M.C., Tudela A., Castro X., and Provenzale A. (2013b). Decreasing fires in a Mediterranean region (1970 – 2010 , NE Spain). *Natural Hazards and Earth System Science* **13**: 649–652. 2, 5, 31, 180

Turco M., Llasat M.C., von Hardenberg J., and Provenzale A. (2014). Climate change impacts on wildfires in a Mediterranean environment. 1, 5, 30, 180, 182, 192

Turco M., Marcos R., Quintana-Seguí P., and Llasat M.C. (2012). Testing instrumental and downscaled reanalysis time series for temperature trends in NE of Spain in the last century. *Regional Environmental Change* **14**: 1811–1823. 25, 53, 192

Turco M., Quintana-Seguí P., Llasat M.C., Herrera S., and Gutiérrez J.M. (2011). Testing MOS precipitation downscaling for ENSEMBLES regional climate models over Spain. *Journal of Geophysical Research: Atmospheres* **116**(D18): 1–14. 5, 17, 83

van Den Dool H.M. (1989). A new look at weather forecasting through analogues. *Monthly Weather Review* **117**(10): 2230–2247. 54

van den Dool H.M. (1994). Searching for analogues, how long must we wait. *Tellus A* **46**: 314–324. 55, 133

van Wagner C.E. and Pickett T.L. (1985). Equations and FORTRAN program for the Canadian Forest Fire Weather Index System. 21

Vicente-Serrano S.M., Beguería S., and López-Moreno J.I. (2010). A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate* **23**: 1696–1718. 179, 180

von Storch H. (1999). On the Use of "Inflation" in Statistical Downscaling. *Journal of Climate* **12**(12): 3505–3506. 184

Vrac M., Drobinski P., Merlo a., Herrmann M., Lavaysse C., Li L., and Somot S. (2012). Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. *Natural Hazards and Earth System Science* **12**(9): 2769–2784. 17

Wang H., Schemm J.K.E., Kumar A., Wang W., Long L., Chelliah M., Bell G.D., and Peng P. (2009). A Statistical Forecast Model for Atlantic Seasonal Hurricane Activity Based on the NCEP Dynamical Seasonal Forecast. *Journal of Climate* **22**(17): 4481–4500. 18

Wang H.J., Zhang R.H., Cole J., and Chavez F. (1999). El Nino and the related phenomenon Southern Oscillation (ENSO): The largest signal in interannual climate variation. *Proceedings of the National Academy of Sciences* **96**(20): 11071–11072. 10, 12

Wang Q.J. and Robertson D.E. (2011). Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research* **47**(2): 1–19. 18

Watkins D.W. and Wei W. (2008). The value of seasonal climate forecasts and why water managers don't use them. *World Environmental and Water Resources Congress 2008* **316**: 1–5. 19

Wedgbrow C.S., Wilby R.L., Fox H.R., and O'Hare G. (2002). Prospects for seasonal forecasting of summer drought and low river flow anomalies in England and Wales. *International Journal of Climatology* **22**(2): 219–236. 19

Weigel A.P., Liniger M.A., and Appenzeller C. (2007). The Discrete Brier and Ranked Probability Skill Scores. *Monthly Weather Review* **135**: 118–124. 68, 71

Weigel A.P., Liniger M.A., and Appenzeller C. (2009). Seasonal Ensemble Forecasts: Are Recalibrated Single Models Better than Multimodels? *Monthly Weather Review* **137**(4): 1460–1479. 17

Weisheimer A., Corti S., Palmer T., and Vitart F. (2014). Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* **372**: 1–12. 17

Weisheimer A., Doblas-Reyes F.J., Palmer T.N., Alessandri A., Arribas A., Déqué M., Keenlyside N., MacVean M., Navarra A., and Rogel P. (2009). ENSEM-

BLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters* **36**(21): 1–6. 10, 28

Weisheimer A. and Palmer T.N. (2014). On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface* **11**: 1–10. 3, 69

Weisheimer A., Palmer T.N., and Doblas-Reyes F.J. (2011). Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophysical Research Letters* **38**: L16703. 17

Westerling A.L., Brown T.J., Gershunov A., Cayan D.R., and Dettinger M.D. (2003). Climate and Wildfire in the Western United States. *Bulletin of the American Meteorological Society* **84**: 595–604. 21

Westerling A.L., Gershunov A., Cayan D.R., and Barnett T.P. (2002). Long lead statistical forecasts of area burned in western U.S. wildfires by ecosystem province. *International Journal of Wildland Fire* **11**: 257–266. 21, 191

Westerling A.L., Hidalgo H.G., Cayan D.R., and Swetnam T.W. (2006). Warming and earlier spring increase western U.S. forest wildfire activity. *Science* **313**(5789): 940–3. 21

Wielicki B.A., Wong T., Allan R.P., Slingo A., Kiehl J.T., Soden B.J., Gordon C.T., Miller A.J., Yang S.K., Randall D.A., Robertson F., Susskind J., and Jacobowitz H. (2002). Evidence for large decadal variability in the tropical mean radiative energy budget. *Science* **295**: 841–844. 14

Wilby R.L., Charles S.P., Zorita E., Timbal B., Whetton P., and Mearns L.O. (2004). Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods. *Analysis* **27**(August): 1–27. 55

Wilcke R.A.I., Mendlik T., and Gobiet A. (2013). Multi-variable error correction of regional climate models. *Climatic Change* **120**: 871–887. 50, 179, 180

Wilks D.S. (2006). *Statistical methods in the atmospheric sciences*. Academic Press, London. 16, 49, 52, 57, 58, 60, 61, 62, 64, 65, 68, 69, 71, 74, 82, 141, 143

Wilks D.S. (2008). Improved statistical seasonal forecasts using extended training data. *International Journal of Climatology* **28**: 1589–1598. 13, 16

WMO (2012). Standardized Precipitation Index User Guide **1090**: 24. 179

Xu C.Y. and Singh V.P. (2001). Evaluation and generalization of temperature-based methods for calculating evaporation. *Hydrological Processes* **15**(2): 305–319. 141

Yang S. and Lau K.M. (1998). Influences of Sea Surface Temperature and Ground Wetness on Asian Summer Monsoon. *Journal of Climate* **11**: 3230–3246. 12, 14

Ye Z. and Hsieh W.W. (2008). Enhancing predictability by increasing nonlinearity in ENSO and Lorenz systems. *Nonlinear Processes in Geophysics* **15**: 793–801. 12

Yoon J.H., Ruby Leung L., and Correia J. (2012). Comparison of dynamically and statistically downscaled seasonal climate forecasts for the cold season over the United States. *Journal of Geophysical Research: Atmospheres* **117**(21): 1–17. 16

Zalachori I., Ramos M.H., Garçon R., Mathevet T., and Gailhard J. (2012). Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science and Research* **8**: 135–141. 19

Zorita E. and von Storch H. (1999). The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods. *Journal of Climate* **12**(8): 2474–2489. 54, 55