Drought impact links to meteorological drought indicators and predictability in Spain

Herminia Torelló-Sentelles[1] and Christian L. E. Franzke[2,3]

[1]School of Integrated Climate System Science, Universität Hamburg, Hamburg, 20148, Germany
[2]Center for Climate Physics, Institute for Basic Science (IBS), Busan, 46241, Republic of Korea
[3]Pusan National University, Busan, 46241, Republic of Korea

Correspondence: Christian L. E. Franzke (christian.franzke@pusan.ac.kr)

Abstract. Drought affects many regions worldwide and future climate projections imply that drought severity and frequency will increase. Hence, the impacts of drought on the environment and society will also increase considerably. Monitoring and early warning systems for drought rely on several indicators; however, assessments on how these indicators are linked to impacts are still lacking. Here, we explore the links between different drought indicators and drought impacts within six sub-regions in Spain. We used impact data from the European Drought Impact Report Inventory database, and provide a new case study to evaluate these links. We provide evidence that a region with a small sample size of impact data can still provide useful insights regarding indicator-impact links. As meteorological drought indicators, we use the Standardised Precipitation Index and the Standardised Precipitation-Evapotranspiration Index, as agricultural and hydrological drought indicators we use a Standardised Soil Water Index and, a Standardised Streamflow Index and a Standardised Reservoir Storage Index. We also explore the links between drought impacts and teleconnection patterns and surface temperature by conducting a correlation analysis and then test the predictability of drought impacts using a Random Forest model. Our results show meteorological indices are best linked to impact occurrences overall, and at long time scales between 15 and 33 months. However, we also find robust links for agricultural and hydrological drought indices, depending on the sub-region. The Arctic Oscillation, Western Mediterranean Oscillation and the North Atlantic Oscillation at long accumulation periods (15 to 48 months), are top predictors of impacts in the northwest and northeast regions, the Community of Madrid, and the south regions of Spain respectively. We also find links between temperature and drought impacts. The Random Forest model produces skilful models for most sub-regions. When assessed using a cross-validation analysis, the models in all regions show precision, recall, or R^2 values higher than 0.97, 0.62 and 0.68 respectively. Since we find the models to be skilful, we encourage other types of impact data to be used to investigate these links and to predict drought impacts.

1 Introduction

Drought, as defined by Wilhite and Glantz (1985), "is a condition relative to some long-term average condition of balance between rainfall and evapotranspiration in a particular area, a condition often perceived as 'normal'". The prediction of drought onset or end is a complex task. Drought severity is also difficult to measure or quantify. This is because drought depends on several factors, for instance, the duration, intensity and the geographical extent of the event. Additional factors specific to each

region also play a large role, such as the water demand with respect to water supply. All of these characteristics make drought difficult to identify and quantify. In addition, drought has far-reaching impacts on society and the environment that may last for long time periods. These are highly dependent on a region's vulnerability to drought at a particular point in time (Wilhite and Glantz, 1985).

There is not a common and straightforward definition of drought, however, all types of drought originate from a lack of precipitation. Many different definitions of drought have been developed by different disciplines. There are four main types of disciplinary definitions: meteorological, agricultural, hydrological, and socioeconomic. Meteorological drought usually uses precipitation as its atmospheric parameter. Agricultural drought usually considers links between meteorological drought and its impacts on agriculture. Hydrological drought usually accounts for the repercussions of dry periods on surface and subsurface hydrology, for instance, on streamflow, groundwater and reservoirs. Finally, socioeconomic drought considers the effects that drought has on the supply and demand of economic goods (Wilhite and Glantz, 1985).

There is already evidence that climate change, as a result of anthropogenic actions, has increased the risk of meteorological drought in Southern Europe (Gudmundsson and Seneviratne, 2016). Similarly, warmer temperatures have increased atmospheric evaporative demand, which have in turn increased drought severity over the past 50 years (Vicente-Serrano et al., 2014b). Climate change projections point towards a reduction in the water resources in Spain. Hydraulic infrastructures have been designed with safety margins; however, these might be surpassed due the effects of climate change. Increased evapotranspiration, as a result of increased temperatures, together with a possible increase in the length of the irrigation period, might increase the water demand for irrigation and for agricultural use, which currently account for more than 70% of total water demand. In addition, the energy sector is also dependant on the availability of water, which also makes it vulnerable to this increased drought risk (Ministerio para la Transición Ecológica y el Reto Demográfico, 2020). Vulnerability to water scarcity and drought is also likely to increase due to challenges such as a growing population, population migration to more arid regions, urbanisation, increasing tourism and pollution (Rossi and Cancelliere, 2013). As a result, the consequences of drought on the environment and society are becoming more important. Spain is a country that already has an intense use of water resources; hence, it is crucial to address the water use priorities and reinforce water management to provide future water security.

In order to lessen the impacts of drought on society and the environment, efficient mitigation measures are necessary. This means that effective drought monitoring and early warning systems (DEWS) are essential. DEWS reduce societal vulnerability to drought by maximising the lead time of early warnings to allow more time for the implementation of mitigation measures (Pozzi et al., 2013). These systems usually rely on different drought indicators that represent different parts of the water cycle. Drought indicators describe drought conditions, and examples of commonly used variables are: precipitation, temperature, streamflow, groundwater and reservoir levels, soil moisture and snowpack (Svoboda et al., 2016). DEWS use a large variety of drought indicators, and the most commonly used one is the Standardised Precipitation Index (SPI) (McKee et al., 1993). DEWS usually use indicators based on variables that can be measured with ease and that are readily obtainable in time (Bachmair et al., 2016a). However, Bachmair et al. (2016a) revealed that although there has been increasing efforts on the research and practice on drought indicators, DEWS are still not well linked with assessments on how drought impacts the environment and society. This is because links between drought indicators and impacts have not been sufficiently studied. This means impact data are

not being used to determine whether indicators are linked to the impacts of drought. The authors call for drought to be framed as a coupled dynamical system of the environment and society to fully understand drought impacts.

Drought-related impacts are complex to study and document because many sectors depend on water availability to produce goods and provide services. Bachmair et al. (2016a) also revealed that there are very few systematic approaches for the collection of impact data, except for agricultural drought. A good example of efforts to improve such documentation exists in Europe, where a drought impact inventory has been created: the European Drought Impact Report Inventory (EDII) (Stahl et al., 2016). This database collects reported drought impacts for different European countries. Impacts are classified into major impact categories (e.g., agriculture and livestock farming, wildfires, public water supply, forestry…), and each category has several sub-types. Also, each drought impact event has at least information on; the source of information, location, duration and impact category and has a description.

Here, we investigate the links between different drought indicators and reported impacts from the EDII database for Spain. The time period studied is from August 1975 to May 2013. We aim to assess two meteorological indicators, the SPI and the Standardised Precipitation-Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010), two hydrological indicators, from streamflow and reservoir storage levels, and an agricultural indicator, from soil water content. The main motivation for this study is to provide a new case study for the evaluation of these links, and to test the usefulness of impact data in this region. Impact data from the EDII has already shown linkages with drought indicators in similar studies (Bachmair et al., 2015, 2016b; Stagge et al., 2015). Our first research question is:

1. How strong and robust is the link between drought indicators and drought impacts?

In order to further examine the indicator-impact links we also use a machine learning model, a Random Forest (RF) model, to model and predict drought impact occurrences. The aim of this analysis is to test the potential of such a method, with the available data, to predict future impacts. Thus, our second research question is:

2. Can drought impacts be skilfully predicted using a RF model in this region?

To our knowledge, links between teleconnection patterns or temperature and drought impacts using the EDII database have not been studied before. We therefore investigate five teleconnection patterns (Feldstein and Franzke, 2017), the El Niño Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), the East Atlantic (EA) pattern, the Western Mediterranean Oscillation (WeMO) and the Arctic Oscillation (AO), as well as a surface temperature index, as possible predictors of drought impact occurrences. We aim to investigate whether these climate indices show links and are better predictors of impacts than the previously presented drought indicators. We chose these teleconnection patterns because they have shown correlations with precipitation in Spain (e.g. Rodó et al., 1997; Martinez-Artigas et al., 2021; Ríos-Cornejo et al., 2015), and the NAO, EA, AO and WeMO have also shown links with the drought index, SPEI (Manzano et al., 2019). Furthermore, empirical links between drought impacts and ENSO and NAO in Spain have already been established. For instance, Gimeno et al. (2002) looked at the influence of ENSO and the NAO on the most important Spanish crops. They detected significant effects on yield for most of these crops. They found low yields during La Niña years and higher yields during positive NAO phases. Therefore, the third research question arises:

3. What type of indicators are the best predictors of impacts in each sub-region?

The incompleteness of the impact reports from the EDII database challenges the quantification of impact occurrences, hence we also investigate different ways of counting incomplete reports. As a final research question, we ask:

4. Do different impact quantification methods change the results in a significant way? If so, how?

Our manuscript is structured as follows: in Sect. 2 we introduce the study area, used data and indicators. In Sect. 3 we present the results of our correlation and RF analyses. In Sect. 4 we provide a discussion of our results and conclude.

2 Methods and data

2.1 Study area

Spain is located in a geographical area with a high recurrence of drought events due to it being in a transition zone between polar and subtropical atmospheric circulation influences (Sivakumar et al., 2011). Its precipitation and runoff patterns are highly diverse and complex, which is characteristic of the Mediterranean area. As described by Vide (1994), the characteristic climate of Spain has:

- modest rainfall overall,
- high interannual variability with a high concentration of rainfall in relatively few days,
- long rainless periods,
- an "aridity problem", meaning the amount of potential evapotranspiration is greater than rainfall,
- large regional variations in seasonal rainfall patterns, and
- anomalies that may be related to atmospheric teleconnection patterns.

Because of the chaotic patterns of precipitation, Spaniards have been attempting to increase water availability for at least the last 2000 years (del Moral and Saurí, 1999). Water scarcity and frequent droughts are recurrent problems that Spain suffers, and this is mainly because the spatial and time distribution of its water resources is irregular. Furthermore, because water demands are highest in the more water scarce areas and during the seasons when precipitation is the lowest and evapotranspiration the highest. Climate change will most likely exacerbate the existing problems with water resources (Estrela and Vargas, 2012; Ministerio de Medio Ambiente, 2005).

Droughts cause extensive impacts in Spain, for instance, during a very intense drought event in 1991-1995, water supply was restricted for more than 25% of the total population (12 million people). In the most affected regions, evacuations plans were activated. Agricultural production was also severely affected (del Moral and Hernandez-Mora, 2015). The following major drought event in 2004-2005 led to social unrest and created disputes over future water infrastructure (Iglesias et al.,

2009). Drought periods can impose significant costs on farmers and affect crop productivity (e.g. Iglesias et al., 2003; Austin et al., 1998; Páscoa et al., 2017; Peña-Gallardo et al., 2019). Moreover, vegetation activity has been shown to be linked to the interannual variability of drought (Vicente-Serrano et al., 2019), and drought has been related to burned areas from wildfires (Russo et al., 2017). Drought events have also shown links with daily mortality across Spain (Salvador et al., 2020). Overall, the economic damages of drought are severe: according to the international Emergency Events Database (EM-DAT) (Guha-Sapir et al., 2016), it ranked fourth worldwide and first in Europe on total economic damages resulting from drought events from 1990 to 2018 (US$7.7 billion).

We chose to perform this study in the specific region of Spain because this region is severely impacted by droughts, hence, a better understanding of indicator-impact links here is urgently needed. Such links have been already been successfully investigated in five other European regions (Bachmair et al., 2015, 2016b; Stagge et al., 2015). We focus on studying these links in six sub-regions within Spain because studying a small region means that indicator-impact links can be better detected and quantified (Blauhut et al., 2016).

We also chose this region to investigate indicator-impact links in a region that has a very dense network of reservoirs and to investigate the time scales at which drought conditions lead to drought impacts. As described by González-Hidalgo et al. (2018), Spain has a high density of hydraulic reservoirs. Spain has, after China, the second largest number of dams in the world. This is because Spain's climate is characterised by dry summers and high interannual variability. Such a dense network increases Spain's resilience to short-term droughts by guaranteeing water supply during these. However, at longer time scales, drought conditions still produce severe drought impacts to water supply; conditions that last more than two or three years have been shown to limit the capacity of these infrastructures (González-Hidalgo et al., 2018).

2.2 Indicators and data sets

As drought indicators we considered the SPI, the SPEI and a streamflow index because these are commonly used in DEWS (Bachmair et al., 2016a). The SPI and especially SPEI, have also shown higher correlations than other drought indices with crop yields in Spain (Peña-Gallardo et al., 2019). Furthermore, we included an additional hydrological indicator (using reservoir storage data), and an agricultural drought index (using soil water content data) to investigate their performance against the other types. Soil moisture has been found to be an important factor when studying drought impacts on the productivity of some agricultural crops in Spain (Sainz de la Maza and Del Jesus, 2020). Furthermore, the multi-scalar nature of all these drought indicators is useful when assessing the time scales of drought impacts. We also explored the use of several teleconnection patterns (Feldstein and Franzke, 2017) as predictors of drought impacts.

The SPI is a commonly used drought index that is simple to compute. It can be used to compare droughts in different regions and it can be temporally aggregated over different time scales (Guttman, 1999). It calculates "the precipitation deviation for a normally distributed probability density with a mean of zero and standard deviation of unity" (McKee et al., 1993). It is computed by fitting precipitation data to a distribution and then transforming it to a normal distribution (McKee et al., 1993). In this study, we calculated the SPI using the "SPEI" R package (Vicente-Serrano et al., 2010; Beguería et al., 2014). We used a Gamma probability distribution to model the observed precipitation values. The SPEI is similar to the SPI but is also based

on water balance, which depends on temperature data. This means that the effects of temperature variability on drought are included here. The advantages of this index, especially under global warming conditions, are that it identifies increased drought severity when the water demand is higher as a result of increased evapotranspiration. In addition, its multi-scalar nature allows its use for drought analysis and monitoring (Vicente-Serrano et al., 2010). We calculated the SPEI also with the "SPEI" R package. To compute it, a calculation of a simple climatic water balance (Thornthwaite, 1948) is required. This is computed using the monthly difference between precipitation and potential evapotranspiration (PET) at different time scales. To obtain the final index, the same procedure as for the SPI was followed, however, a log-logistic probability distribution was used to model the precipitation–PET values.

The volumetric soil water variable is the volume of water (m^3) in a soil layer (m^3). It is measured at four layers, and the quantity is dimensionless (Muñoz Sabater, 2019). To obtain an agricultural drought indicator from volumetric soil water content data, we created a Standardised Soil Water Index (SSWI) using the Standardised Drought Analysis Toolbox (Hao and AghaKouchak, 2014; Farahmand and AghaKouchak, 2015). This toolbox provides a generalised framework for deriving nonparametric univariate indices that can be interpreted similarly to the rest of the indices used in this study. This index was calculated at each layer, hereafter referred to as SSWI1, SSWI2, SSWI3 and SSWI4. To create a Standardised Streamflow Index (SSFI) and a Standardised Reservoir Storage Index (SRSI), we standarardised streamflow and reservoir values using the same methodology as for SPI. We also computed a Standardised Temperature Index (STI) using the same methodology, with temperature data only.

We used the Iberia01 daily precipitation and temperature observational gridded data set to calculate the SPI and SPEI (available at: http://hdl.handle.net/10261/183071) (Gutiérrez et al., 2019; Herrera et al., 2019). This is a high-resolution data set produced using a dense network of stations over the Iberian Peninsula: 3481 and 276 stations for precipitation and temperature, respectively. Gridded values are provided at a spatial resolution of 0.1º, and they cover the entire time period studied here. This data set has been shown to produce more realistic patterns in the case of precipitation than other frequently used data sets. We used data from individual streamflow and reservoir level monitoring stations from Ministerio para la Transición Ecológica y el Reto Demográfico to calculate the SSFI and SRSI. There were a total of 1447 and 367 streamflow and reservoir storage monitoring stations respectively, however, after removing stations with more than 20% missing data, data from 786 and 322 stations remained. We obtained volumetric soil water content data from the ERA5-Land data set (Muñoz Sabater, 2019). This is a reanalysis data set that provides estimates for land variables. It has a horizontal spatial resolution of 0.1º x0.1º and has a vertical resolution that consists of four levels of surface: layer 1: 0-7cm, layer 2: 7-28cm, layer 3: 28-100cm and layer 4: 100-289cm. We obtained data for the NAO, EA, AO and ENSO from the NOAA Climate Prediction Center (available at: https://psl.noaa.gov/data/climateindices/list/ and https://www.cpc.ncep.noaa.gov/data/teledoc/ea.shtml) and data for the WeMO (Martin-Vide and Lopez-Bustins, 2006) (available at: http://www.ub.edu/gc/wemo/). We used data from these datasets for the period 1975-2013, except for volumetric soil water content data, where we used data from 1981-2013, since data were only available starting from 1981. We also aggregated all the indices over a range of time scales, these ranges differed depending on up to what timescales the indicator-impact correlation strengths were greatest. These were; 1-33 months for the SPI, SPEI and SRSI, and 1-48 months for the SSWI, STI and teleconnection patterns.

The Nomenclature of Territorial Units for Statistics (NUTS) classification divides economic territories of the European Union (Eurostat). NUTS-1 regions represent major socioeconomic regions, and these were the sub-regions we considered in this study. These were: the Northwest (NW), Northeast (NE), Community of Madrid (MA), Centre (CE), East (E) and South (S). The Canary Islands were excluded due to a lack of impact data in this region. We aggregated all of the indicators studied over each NUTS-1 region and produced a mean monthly time series for each sub-region using the R package 'panas' (De Felice, 2020).

2.3 Drought impact data

We retrieved drought impact information from the EDII. This database had 388 impact report entries for Spain, which covered the time period from August 1975 to May 2013. Each reported impact has three spatial references which correspond to the three levels of the NUTS regions. We aggregated impact information by NUTS-1 region and did not differentiate impact types from one another, hence, treated all impacts as equal and of a general type. The impact categories considered in the EDII were:

- Agriculture and Livestock Farming
- Forestry
- Freshwater Aquaculture and Fisheries
- Energy and Industry
- Waterborne transportation
- Tourism and Recreation
- Public Water Supply
- Water Quality
- Freshwater Ecosystem: Habitats, Plants and Wildlife
- Terrestrial Ecosystem: Habitats, Plants and Wildlife
- Soil System
- Wildfires
- Air Quality
- Human Health and Public Safety
- Conflicts

In order to evaluate links between indicators and impacts we mainly followed the methodology by Bachmair et al. (2015, 2016b), who assessed links between hydro-meteorological indicators and impacts for Germany and the UK. We first explored correlations between indicators and impacts and we then used a RF model to evaluate the predictive potential and predictor importance of the different indicators. Before conducting the analysis, we first converted impact reports into a monthly time series of number of drought impact occurrences for each sub-region. To do this, we imposed criteria to convert a single 'drought impact report' (an entry from the EDII) into a 'drought impact occurrence', which will be referred to as 'DIO'. We converted impact reports into a monthly time series by creating a DIO for every month in between the start and end date. However, a large proportion of the reports were incomplete. The data had the following five main problems:

1. the specific sub-region affected was not indicated,
2. only the start and the end year were indicated,
3. only the start year was indicated,
4. only the start month was indicated, and
5. only the start month and the end year were indicated.

24% of the reports had problem 1, 33% had problem 2, 37% had problem 3, 9% had problem 4 and 3% had problem 5. Therefore, to overcome and estimate the uncertainty in our analysis as a result of the incompleteness of the data, we developed different 'counting methods'. This meant that we tested the effects of including or excluding reports with these problems.

The most censoring method, counting method 1 (CM1), did not include reports with problems 1, 2 and 3. If reports with problem 5 started and ended in the same year, one DIO was created at the start month only, otherwise, DIOs were created from the start month until December of the previous-to-last year. Next, counting method 2 (CM2), was identical to CM1 except, if reports with problem 5 started and ended in a different year, DIOs were created from the start month until December of the end year. Third, was counting method 3 (CM3), which included reports with problems 2-5. Problem 2 was addressed by creating DIOs for every month from the start month of the start year until December of the end year. Problem 3 was addressed by creating a DIO for every month of the start year, and problem 5 by creating DIOs from the start month until December of the end year. Finally, counting method 4 (CM4) was the least censoring method and included reports with all problems. It was identical to CM3, except, if reports did not indicate the specific region affected (problem 1), it assumed that the whole country was affected, and DIOs were created for all the NUTS-1 regions. All counting methods addressed problem 4 by creating one DIO for the specified month only.

Figure 1 shows the distribution of impacts types for each NUTS-1 region and for the whole of Spain using the most and least censoring counting methods. It shows that most of the impacts recorded in the EDII for Spain were on agriculture and livestock farming, public water supply and freshwater ecosystems. Also, depending on the censoring criteria, the distribution of impact types varied slightly. For instance, the most censoring counting methods showed a larger proportion of impacts on terrestrial ecosystems in the S and for the whole of Spain. The least censoring methods also had a larger variety of impact types.

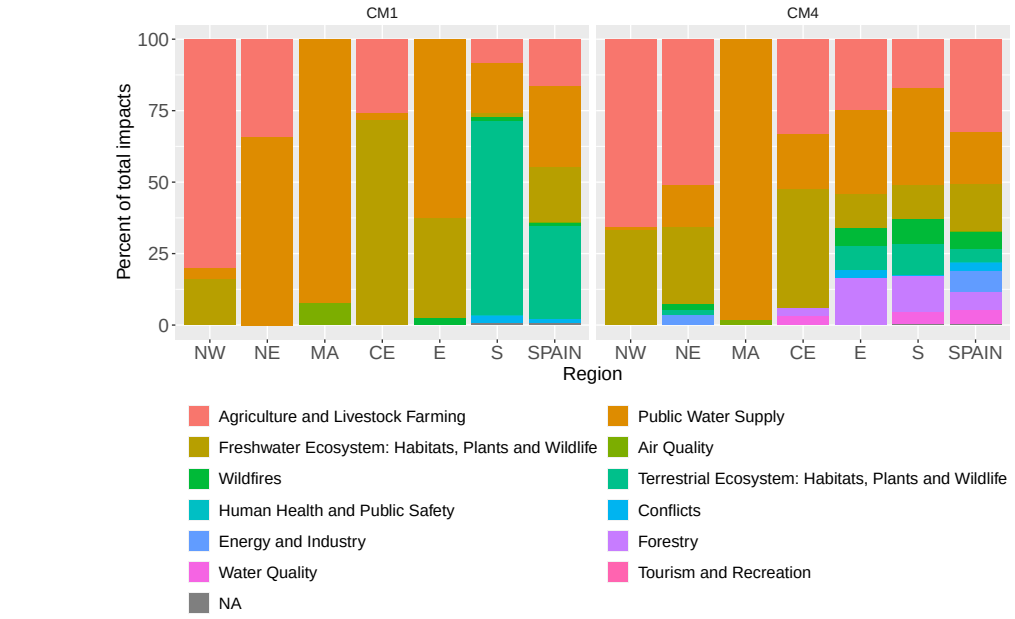

Figure 1. Distribution of impact types for Spain and the sub-regions studied. Results for the most and least censoring counting methods (CM1 and CM4) are shown.

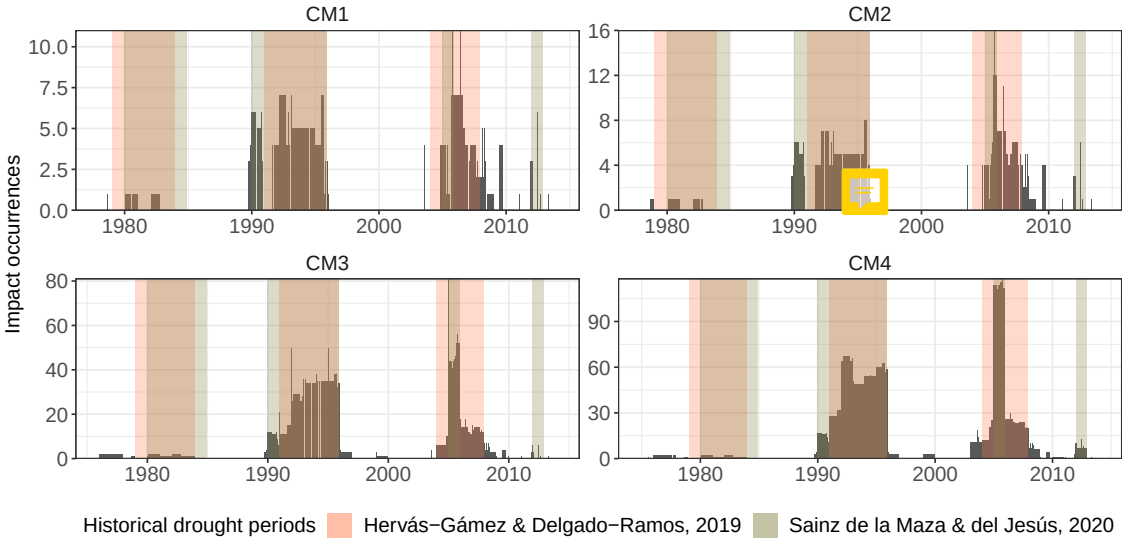

Figure 2. Total monthly DIOs in Spain from 1975 to 2013 using different counting methods (CM1-CM4). Important historical drought periods identified in other studies are highlighted.

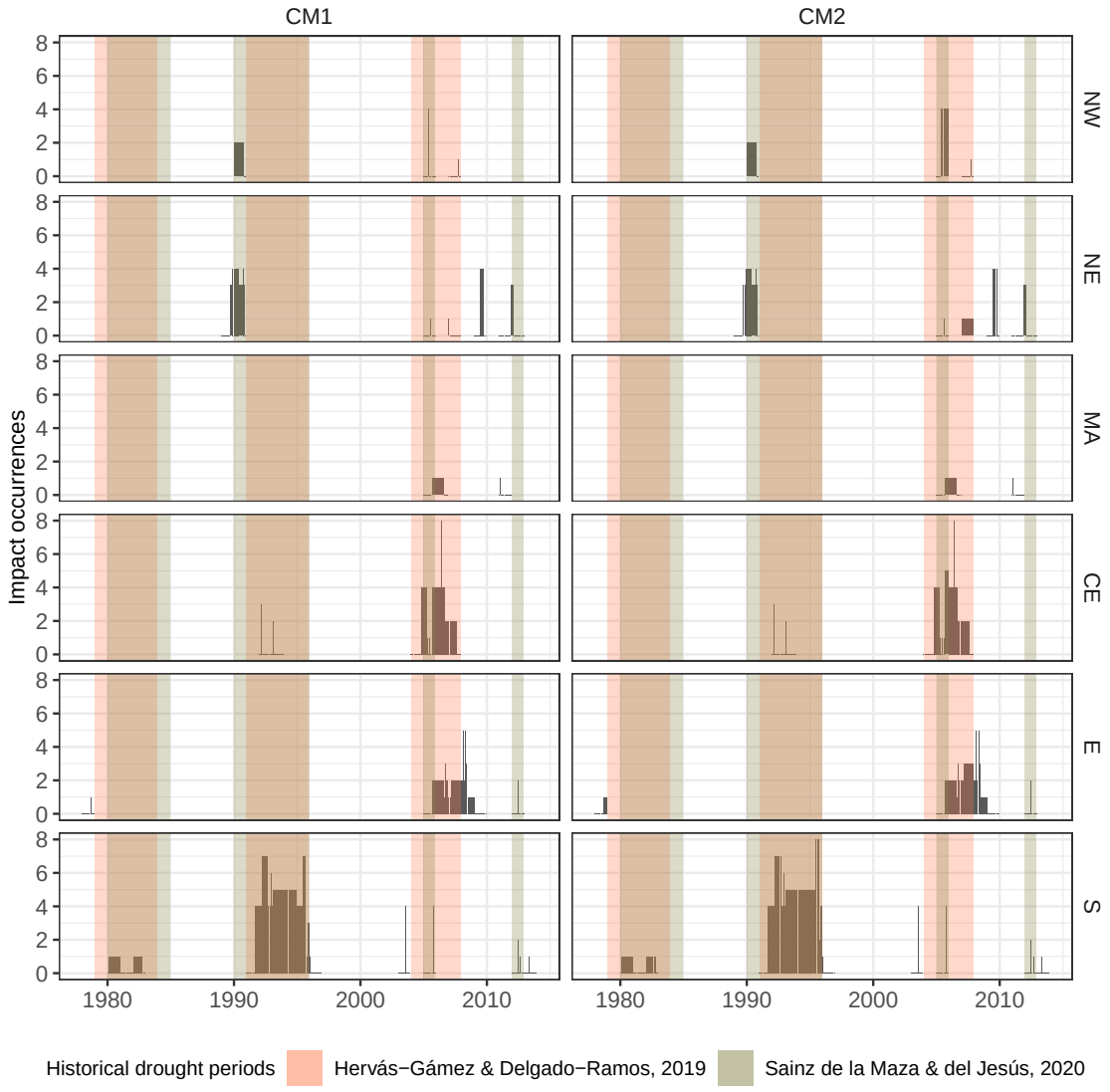

Figure 3. Monthly DIOs in each sub-region from 1975 to 2013 using the two most censoring counting methods (CM1 and CM2). Important historical drought periods identified in other studies are highlighted.

Figure 2 shows the time series of total DIOs for Spain and also shows identified precipitation deficit episodes where Spain has suffered major impacts due to severe drought and water scarcity events (Hervás-Gámez and Delgado-Ramos, 2019; Ministerio de Agricultura, Pesca y Alimentación). Historical drought periods identified by Sainz de la Maza and Del Jesus (2020), determined from economic impacts of past droughts are also shown. The latter authors identified these by using data from the EM-DAT and from another study (Ollero Lara et al., 2018) that used insurance data by the Entidad Estatal de Seguros Agrarios. Figure 2 shows that most of the DIOs occurred during the identified historical drought periods by the authors mentioned. A

Table 1. Information on DIOs for the different sub-regions and the length of the time series for analysis for the most censoring counting method.

NUTS-1 region	Number of DIOs	Length of censored time series (months)
NW	25	36
NE	76	84
MA	13	36
CE	109	72
E	83	84
S	289	156

slight disagreement occurs in 2008 until late 2009, where DIOs continue to occur even though the reported drought episodes end in 2007. Moreover, Fig. 3 shows the time series of DIOs for all sub-regions. We observe that during each drought episode, DIOs do not always occur in all sub-regions and that the amounts and patterns of DIOs do change between counting methods, depending on the harshness of the censoring criteria.

2.4 Correlation analysis

For each NUTS-1 region, we selected a subset of years from August 1975 to December 2013 for the analysis. This selection excluded years where there were no impact occurrences reported. We then included each month of each selected year in a censored time series. We did this to exclude years where regions may have experienced drought impacts but may not have been recorded in the EDII. The lengths of the censored time series are shown in Table 1. To determine the relationship between drought indicators and drought impacts, we first conducted a cross-correlation analysis. We calculated Spearman Rank correlation coefficients (Spearman, 1961) and significance levels for the time series of different indicator versus the time series of DIOs for each NUTS-1 region. For each indicator, we spatially aggregated the indicators over each NUTS-1 region using their mean.

2.5 Random Forest analysis

A RF (Breiman, 1999) is a machine learning approach that uses ensemble trees. This approach has already been used to link drought indicators to impacts (Bachmair et al., 2016b, 2017) and even to forecast drought impacts (Sutanto et al., 2019). A RF is a tree-based ensemble; each tree depends on a random sample of predictor variables and a random response variable. The model then creates a prediction function to predict the response variable by constructing ensembles of trees. The predictions over all trees are combined by being averaged (in regression models) or by selecting the class that is more frequently predicted (in classification models) (Cutler et al., 2012). RFs are appealing because the predictor and response variables can either be continuous or categorical variables. It is a fast model to run computationally, and very few tuning parameters are required.

Apart from training and making predictions, they can also provide variable importance measures. RFs also require minimal human supervision (Cutler et al., 2012).

In this study, we used RF models to model drought impacts, using the R package "randomForest" (Liaw and Wiener, 2002). We trained two RF models for each NUTS-1 region: a regression and a classification model. For the regression models, we used a time series of normalised DIOs as the response variable and the monthly time series of drought indicators as predictors. The number of DIOs was normalised by dividing them by the total number of DIOs for each region. This allowed for a fair comparison of the model errors between regions and counting methods. For the classification models, we used a binary time series of impacts, which was constructed by categorising the response variable by setting DIO = 0 to "no impact" and DIO > 0 to "impact". Unlike for the correlation analysis, the impact occurrence time series were not censored here. As predictors, we included all the indicators mentioned earlier aggregated over the different time scales. We identified the "best" predictors for each region using the variable importance feature. The algorithm estimates variable importance by examining by how much the prediction error increases when that variable is excluded from the model (Liaw and Wiener, 2002).

The main input parameters in a RF model are the number of trees (ntree) and the number of variables randomly sampled at each split (mtry). As Breiman (1999) mentions, for a large number of trees, and as the number of trees increases, the generalisation error converges to its limiting value. We set ntree = 1000. In order to select the optimal mtry parameter for each model, we used the R package "caret" (Kuhn, 2008) for tuning the models. We also used this package to perform a cross-validation analysis.

To assess the predictive potential of the RF models, we first conducted a 10-fold cross-validation analysis using all of the available data (June 1983 to December 2013), repeated five times. We used the root mean squared error (RMSE) and R^2 performance metrics to evaluate the regression models. To assess the classification models, we used precision, recall and F-score metrics. Precision is the number of impact occurrences correctly predicted as a proportion of the total impact occurrence predictions made, and recall is the proportion of impact occurrences correctly predicted (Davis and Goadrich, 2006). The F-score is a combination of precision and recall as their harmonic mean (Hripcsak and Rothschild, 2005). These metrics were chosen because we found large imbalances in the event classes (a large number of "no impact" events).

In order to further evaluate the performance of the RF models, we randomly partitioned the data into a training and testing part, with a 75:25 split. We tuned and trained the RF models (using a 10-fold repeated cross-validation) to then predict the testing set. In this analysis, we also compared the two types of RF models: classification and regression. We did this by converting the outputs of the regression model into binary classes. The threshold to classify the outputs was set to 0.5, 1, 1.25 and 1.5; outputs below each threshold were classified as "no impact", and outputs above each threshold, as "impact".

3 Results

3.1 Correlation analysis

Figure 4 shows significant correlations for the indicators in most regions. This indicates there are clear links between drought indicators and impact occurrences for most regions. Overall, the results from this analysis show that drought impact occurrences

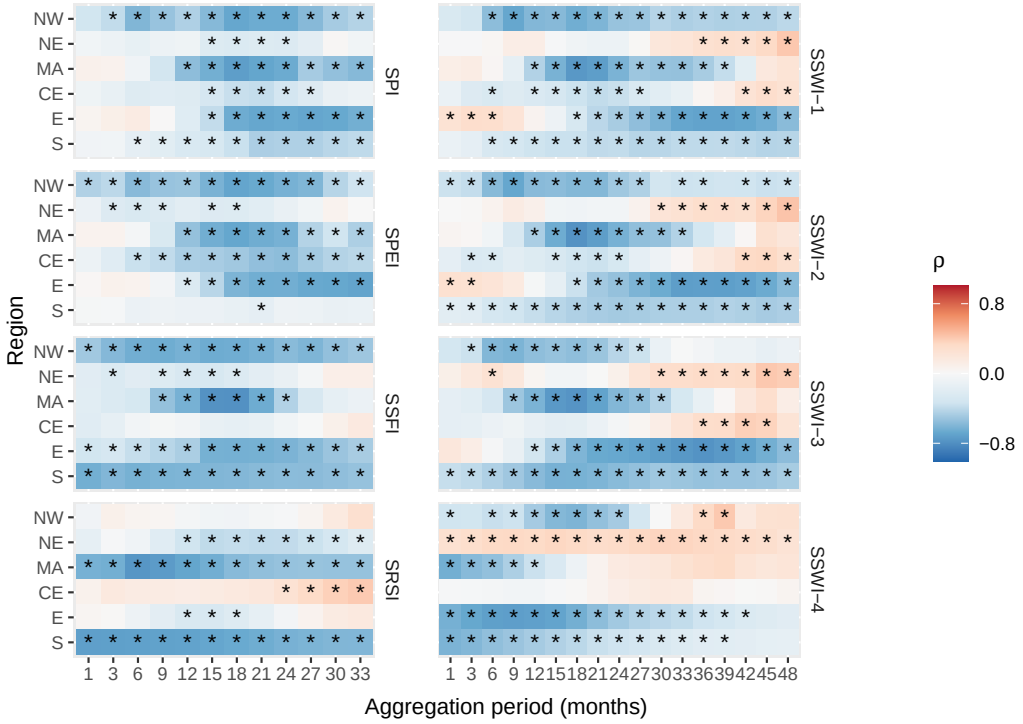

Figure 4. Correlation coefficients (ρ) between time series of drought indicators and impact occurrences for each sub-region using the most censoring counting method. Stars indicate significance ($p < 0.05$).

are negatively correlated to the drought indicators studied and when not, the correlation values were usually weak or not significant. This means that when the drought indicator severity increases, impact occurrences tend to increase, and vice versa. The NE and CE regions show the lowest correlations for all indicators, which indicates that these regions show the weakest links with impact occurrences. In this analysis, the total number of impact occurrences or the length of the censored time series (see Table 1) did not seem to be related with correlation strength. For instance, these two regions were the regions with the fourth-most and second-most number of impact occurrences respectively, but they showed the weakest links.

The SPI and SPEI showed a similar performance to one another, with the exception of the S region, where the SPI showed a larger number of significant and strong correlations. Aggregations over time scales of 18-21 months showed the highest correlations for both these indicators. Moreover, the agricultural indicator, SSWI3 showed the greatest number of significant and strong correlations out of the remaining soil layers. SSWI4 outperformed SPI and SPEI in the S region at a time scale of 18 months; however, its overall performance was lower, especially in the CE and NE regions. The hydrological indicator, SSFI, showed strong and significant correlations in most regions but underperformed in the CE region, when compared to the SPI and SPEI. It also showed very similar patterns to the SSWI but with a slightly better performance in the NE region. SRSI showed strong significant correlations in the MA and S regions, and of slightly lower strength in the NE and E regions.

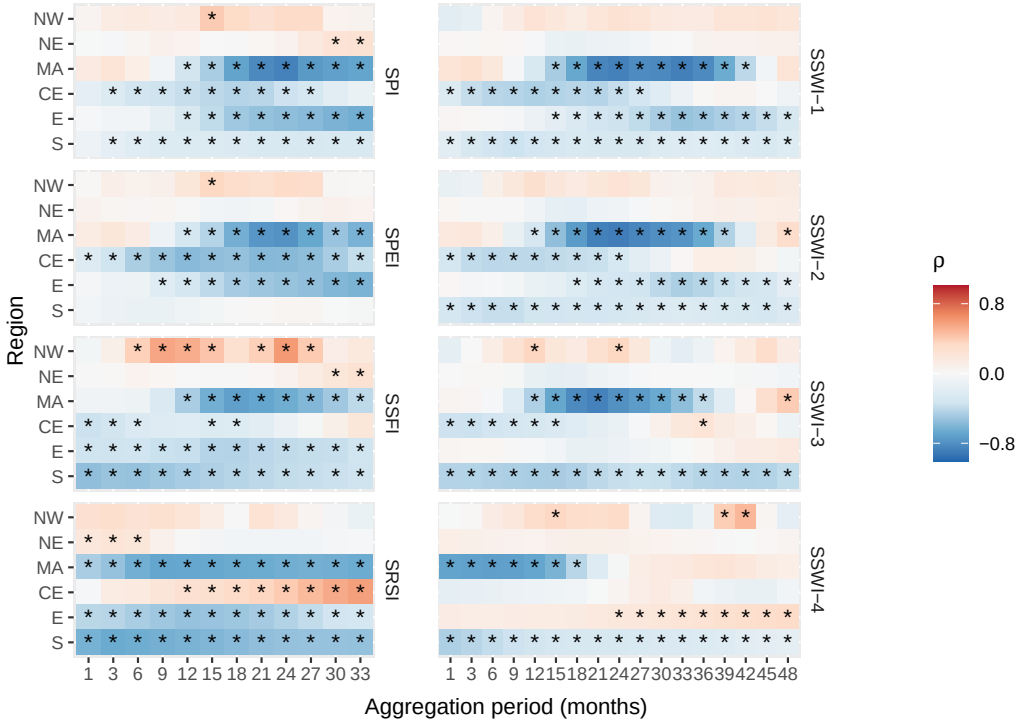

Figure 5. Correlation coefficients (ρ) between time series of drought indicators and impact occurrences for each sub-region using the third-most censoring counting method. Stars indicate significance ($p < 0.05$).

When comparing the correlation patterns across different counting methods we found that overall, the two most censoring counting methods had the highest average correlation coefficients and smallest average p-values over all regions, indicators and aggregation time scales. Correlation patterns remained very similar for the two most censoring methods, except for the SRSI, that in one method showed significant correlations in one region (NW) and did not show this when using the other method. The two least censoring methods showed similar patterns to the two other methods in four sub-regions (MA, CE, E and S), however, most significant correlations disappeared in the remaining two regions (NW and NE) (see Fig. 5). We found that generally, the less censoring the method, the lower the correlation strengths. These results indicate that counting methods do, to some extent, affect correlation patterns and strengths between indicators and impact occurrences. The results obtained here indicate that it is important to investigate different counting methods when working with incomplete impact data. In the following, we will use the two most censoring counting methods to determine the links between indicators and impacts, since the results were most consistent using these two counting methods. The predictors that showed the highest correlation strengths using these counting methods are displayed in Table 2a and 2b.

Overall, the teleconnection patterns and STI (Fig. 6) showed strong and significant correlations with impact occurrences for many sub-regions and time scale aggregations, when using the two most censoring methods. However, the main difference when compared to the correlation patterns with the drought indicators in Fig. 4 and 5, is that the correlation directions varied

Table 2a. Correlation coefficients (ρ) between drought indicators and climate indices, and impact occurrences for the two most censoring counting methods. The indices are ordered by decreasing correlation strength and the time scale at which the indices are aggregated (sc) is shown.

	CM1			CM2			CM1			CM2		
Region	Indicator	sc	ρ	Indicator	sc	ρ	Indicator	sc	ρ	Indicator	sc	ρ
NW	SPEI	18	-0.684	SSWI4	39	0.826	ENSO	27	-0.680	EA	24	-0.669
	SPI	18	-0.674	SSWI4	3	-0.687	NAO	12	0.679	EA	27	-0.576
	SSWI1	9	-0.674	SSWI4	6	-0.676	EA	39	-0.676	WeMO	12	-0.536
	SSWI2	9	-0.674	SPI	21	-0.653	NAO	18	0.660	STI	3	0.534
	SPEI	21	-0.662	SSWI1	9	-0.647	NAO	48	0.657	WeMO	9	-0.533
	SPI	24	-0.651	SSWI1	6	-0.644	NAO	15	0.650	WeMO	6	-0.494
	SPI	21	-0.651	SSWI4	1	-0.638	ENSO	18	-0.642	STI	30	-0.491
	SSFI	15	-0.647	SRSI	18	-0.630	NAO	9	0.635	EA	39	-0.482
	SSFI	9	-0.644	SSWI4	9	-0.627	NAO	30	0.633	EA	30	-0.453
	SSFI	18	-0.641	SPEI	9	-0.622	AO	15	0.625	NAO	48	0.438
NE	SRSI	54	-0.456	SRSI	27	-0.519	STI	6	0.555	AO	39	0.550
	SSWI2	48	0.428	SRSI	24	-0.509	AO	42	0.546	AO	42	0.550
	SSWI1	48	0.410	SRSI	30	-0.507	AO	39	0.540	STI	12	0.541
	SSWI3	45	0.410	SRSI	39	-0.476	AO	48	0.538	STI	9	0.528
	SSWI3	48	0.391	SRSI	33	-0.475	AO	45	0.527	AO	45	0.508
	SRSI	15	-0.390	SRSI	36	-0.458	STI	9	0.498	AO	33	0.508
	SRSI	21	-0.388	SRSI	21	-0.453	AO	33	0.492	AO	48	0.507
	SRSI	24	-0.385	SRSI	42	-0.414	NAO	12	0.491	AO	12	0.506
	SRSI	51	-0.380	SRSI	18	-0.384	AO	36	0.485	AO	36	0.498
	SRSI	18	-0.376	SSWI3	6	0.361	AO	30	0.483	AO	30	0.492
MA	SSFI	18	-0.782	SSFI	18	-0.782	EA	27	-0.732	EA	27	-0.732
	SSFI	15	-0.777	SSFI	15	-0.777	AO	9	-0.688	AO	9	-0.688
	SSWI2	18	-0.777	SSWI2	18	-0.777	STI	27	-0.649	STI	27	-0.649
	SSWI3	18	-0.765	SSWI3	18	-0.765	EA	30	-0.610	EA	30	-0.610
	SRSI	6	-0.760	SRSI	6	-0.760	EA	42	-0.571	EA	42	-0.571
	SSWI1	18	-0.749	SSWI1	18	-0.749	STI	30	-0.565	STI	30	-0.565
	SSWI3	15	-0.743	SSWI3	15	-0.743	EA	33	-0.515	EA	33	-0.515
	SRSI	9	-0.732	SRSI	9	-0.732	EA	24	-0.482	EA	24	-0.482
	SPI	18	-0.721	SPI	18	-0.721	STI	24	-0.470	STI	24	-0.470
	SSWI1	21	-0.721	SSWI1	21	-0.721	NAO	36	0.465	NAO	36	0.465

Table 2b. Correlation coefficients (ρ) between drought indicators and climate indices, and impact occurrences for the two most censoring counting methods. The indices are ordered by decreasing correlation strength and the time scale at which the indices are aggregated (sc) is shown.

	CM1			CM2			CM1			CM2		
Region	Indicator	sc	ρ	Indicator	sc	ρ	Indicator	sc	ρ	Indicator	sc	ρ
CE	SPEI	24	-0.554	SPEI	24	-0.567	WeMO	15	-0.701	WeMO	15	-0.711
	SPEI	21	-0.540	SPEI	21	-0.567	AO	30	-0.687	AO	30	-0.684
	SPEI	18	-0.494	SPEI	18	-0.523	AO	33	-0.645	WeMO	18	-0.649
	SPEI	27	-0.482	SPEI	15	-0.508	WeMO	18	-0.641	AO	39	-0.648
	SPEI	15	-0.470	SPEI	12	-0.500	WeMO	24	-0.633	AO	33	-0.647
	SPEI	12	-0.458	SPEI	27	-0.496	WeMO	12	-0.628	WeMO	12	-0.645
	SRSI	36	0.431	SPEI	30	-0.461	WeMO	21	-0.622	WeMO	24	-0.637
	SPEI	30	-0.430	SRSI	36	0.457	WeMO	27	-0.614	AO	36	-0.630
	SPEI	33	-0.412	SPEI	33	-0.431	AO	39	-0.610	WeMO	21	-0.630
	SRSI	33	0.402	SPEI	9	-0.427	AO	36	-0.605	WeMO	9	-0.625
E	SSWI3	36	-0.772	SSWI3	36	-0.847	WeMO	33	-0.653	WeMO	33	-0.635
	SSWI3	33	-0.751	SSWI3	33	-0.839	WeMO	36	-0.628	WeMO	30	-0.613
	SSWI3	39	-0.743	SSWI3	39	-0.803	WeMO	30	-0.615	WeMO	36	-0.609
	SSWI3	30	-0.720	SSWI2	36	-0.797	WeMO	27	-0.612	WeMO	27	-0.607
	SSWI2	33	-0.720	SSWI2	33	-0.794	WeMO	12	-0.593	WeMO	24	-0.568
	SSWI4	9	-0.718	SSWI3	30	-0.790	WeMO	24	-0.584	EA	21	0.559
	SSWI2	39	-0.716	SSWI2	39	-0.790	WeMO	9	-0.580	EA	24	0.551
	SSWI2	36	-0.713	SSWI1	36	-0.777	WeMO	39	-0.569	WeMO	39	-0.545
	SSWI4	6	-0.712	SSWI4	9	-0.773	WeMO	15	-0.540	EA	12	0.538
	SSWI4	12	-0.709	SSWI1	39	-0.769	WeMO	6	-0.525	EA	18	0.531
S	SRSI	1	-0.714	SRSI	1	-0.732	ENSO	33	0.685	ENSO	33	0.683
	SRSI	3	-0.712	SRSI	3	-0.730	ENSO	36	0.682	ENSO	36	0.683
	SRSI	6	-0.707	SRSI	6	-0.726	ENSO	30	0.676	ENSO	30	0.676
	SRSI	9	-0.700	SRSI	9	-0.720	STI	27	-0.673	NAO	48	0.673
	SRSI	12	-0.697	SRSI	12	-0.717	STI	30	-0.670	STI	30	-0.664
	SRSI	15	-0.680	SRSI	15	-0.700	NAO	48	0.667	STI	27	-0.663
	SRSI	18	-0.660	SRSI	18	-0.681	STI	33	-0.659	ENSO	27	0.658
	SRSI	21	-0.644	SRSI	21	-0.666	ENSO	27	0.656	STI	33	-0.657
	SSFI	1	-0.635	SSWI3	15	-0.651	STI	24	-0.653	ENSO	39	0.655
	SSFI	6	-0.623	SSWI3	18	-0.651	ENSO	39	0.644	ENSO	24	0.641

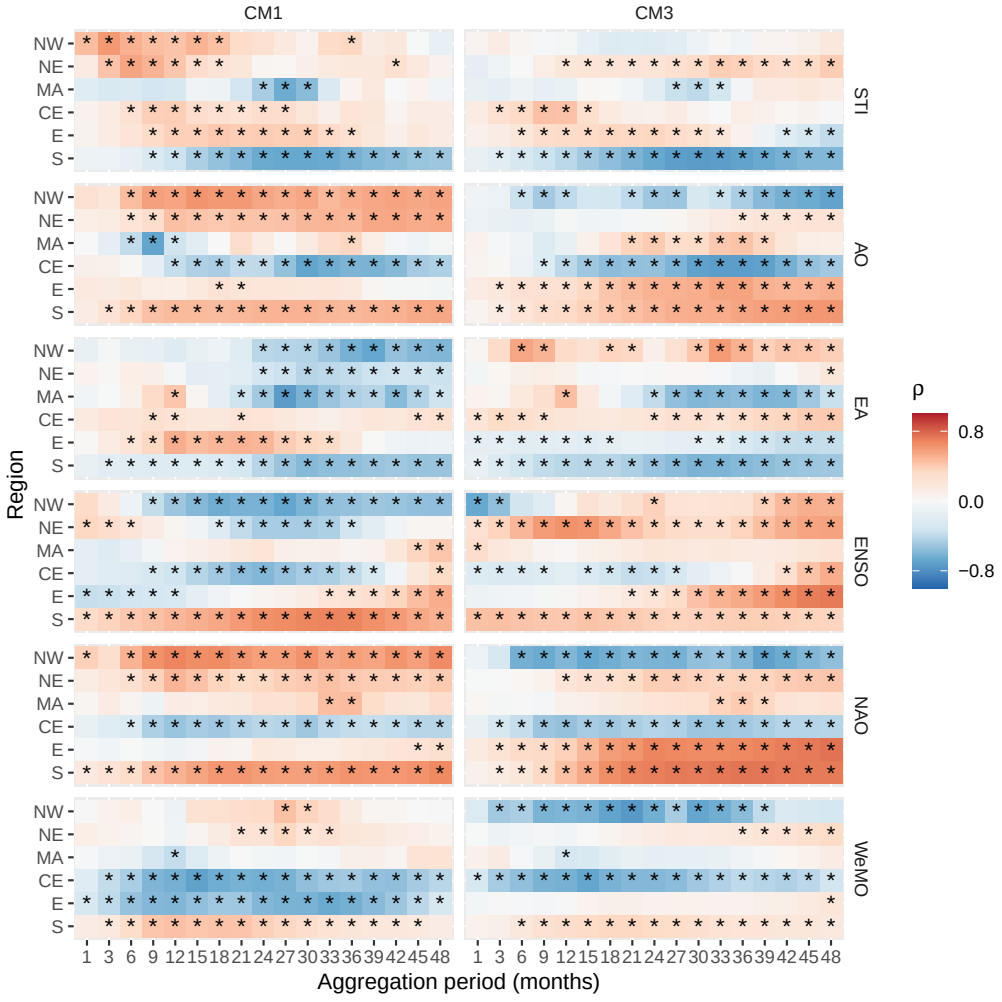

Figure 6. Correlation coefficients (ρ) between time series of climate indices and impact occurrences for each sub-region using the most and third-most censoring counting methods. Stars indicate significance ($p < 0.05$).

across the different sub-regions and counting methods here. This behaviour can be due to there being a negative linear relationship between the two variables or due to a lag between indices and impact occurrences. For example, although STI correlated negatively with impacts in the MA region, when we compared both time series (not shown) we saw that impact occurrences appeared during an abnormally hot period but appeared at a time where there was a short period of decreasing temperatures. This suggests that there is lag between elevated temperatures and impact occurrences in this region. Excluding the MA and S regions, STI and impact occurrences showed positive correlations. This suggests that there is a positive-linear relationship between temperature anomalies and drought impact occurrences.

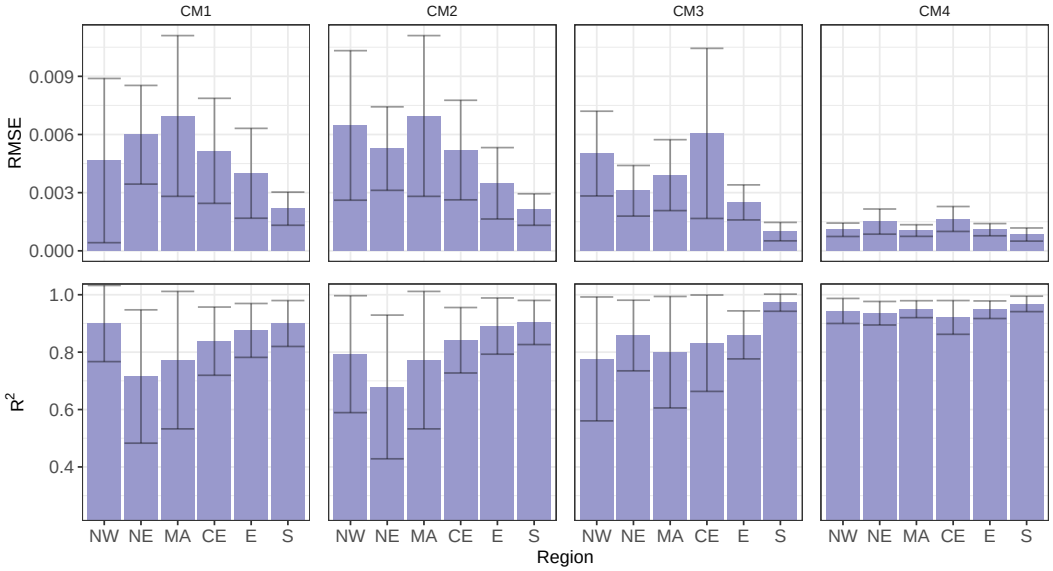

Figure 7. Performance metrics (RMSE and R^2) of the regression RF model when performing a repeated 10-fold cross-validation using all the data. All counting methods are shown. The error bars show the standard deviation of these metrics.

The correlation patterns using the two most censoring methods (Fig. 6) showed that EA at 24-39 months has the strongest correlations in the NW region, AO at 39-48 months and STI at 6-12 months in the NE region, EA at 27-39 months, AO at 9 months and STI at 27-30 months in the MA region, WeMO at 12-24 months and AO at 30-39 months in the CE region, WeMO at 27-36 months in the E region, and ENSO at 30-36 months, STI at 27-30 months and NAO at 48 months in the S region. These results are also displayed in Table 2a and 2b. Out of all the teleconnection patterns, the NAO and AO showed very similar correlation patterns to each other.

The third most censoring method showed similar patterns to the two most censoring methods, except for two sub-regions (NW and NE), where most significant correlation patterns disappeared or changed direction. The least censoring method overall still showed significant correlations but with decreased strength. Therefore, we again find that counting methods sometimes affected correlation patterns. We must also take into account that a correlation analysis only assesses links between two variables; however, the pathways by which these patterns affect drought conditions and the propagation to impacts are usually complex. Possible interactions between different teleconnection patterns or other atmospheric phenomena were not modelled in this analysis due to its bivariate nature.

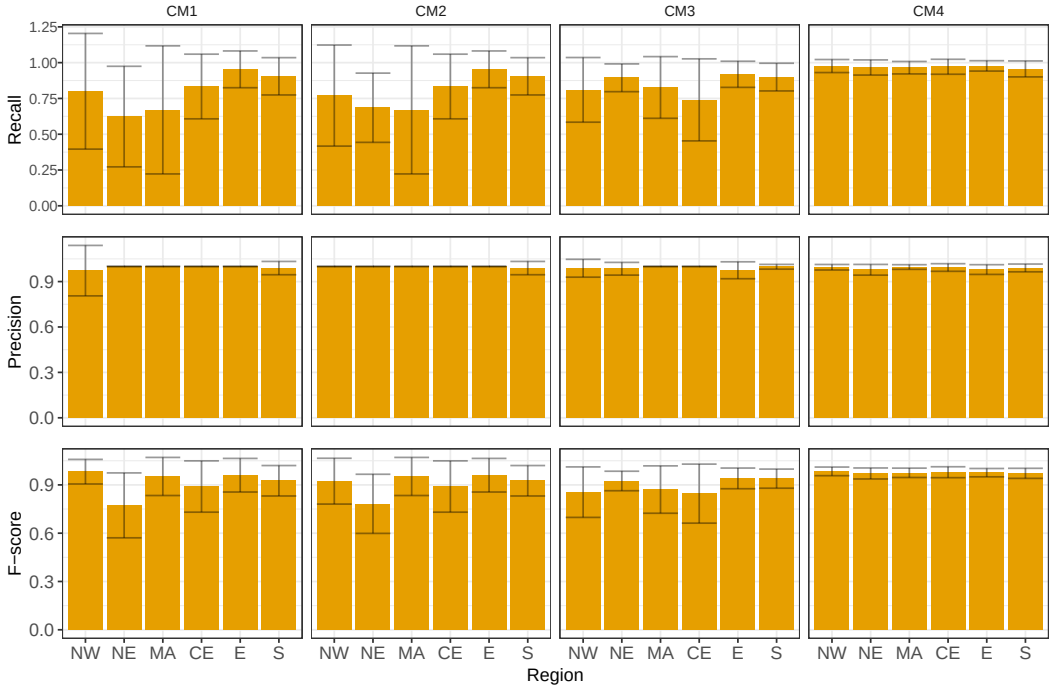

Figure 8. Performance metrics (recall, precision and F-score) of the classification RF model when performing a repeated 10-fold cross-validation using all the data. All counting methods are shown. The error bars show the standard deviation of these metrics.

3.2 Random Forest analysis

3.2.1 Cross-validation analysis

The performance of the regression RF models is shown in Fig. 7. RMSE values ranged from 0.0008 to 0.007 across all counting methods. Since impact occurrences were normalised in this analysis, these values should be interpreted as a fraction of the total number of the impact occurrences for each model. Overall, R^2 values ranged from 0.68 to 0.97; this meant the models explained the variance observed relatively well.

In Fig. 8 we assess the performance of the classification RF. The classification model outputs predictions as either probabilities for each class or directly outputs the class. For this reason, the performance measures are different than in Fig. 7. We see that generally; the precision of all models was higher than the recall. This means that the models predicted fewer than actual impact occurrences (low/moderate recall), but the predictions of impact occurrences were usually correct (high precision). In other words, the model had a low false positive rate and a slightly higher false negative rate. Moreover, recall generally appeared to be highest for the models with more balanced data sets; this was not the case for precision.

Precision values were very high for all the models; when the models predicted an impact occurrence, the predictions were correct 97-100% of the time. Recall values varied more than precision between regions and the standard errors were higher.

However, all models predicted at least 62% and up to 98% of the impact occurrences. Since the F-score metric combines both precision and recall, we used this measure to conclude that the best performing counting method was the one with the least censoring criteria. However, the rest of the methods had very similar average F-scores and their censoring level did not notably affect their F-scores. Furthermore, the results from the two most censoring methods showed that the three regions with the largest number of impact occurrences showed higher recall values than the rest of the regions. The precision across the different sub-regions did not show much variation. These regions also showed the best performance overall when using the regression models (Fig. 7). This suggests that in order to have models with better skill and especially better recall, we need datasets with a greater number of impact occurrences, which means longer impact time series.

When comparing censoring methods, we also found that a small sample size (low numbers of impact occurrences) limits the model's performance. We found that generally, the less strict the censoring criteria, the better the model performance, this is displayed in Fig. 7 and 8. When using the classification models, the balance between the two class types affected model performance; the more balanced, the better the performance. However, although the least strict censoring method shows the best performance, this method could be excluding specific indicator-impact links at the NUTS-1 level, since this method counts impacts that affected the whole country as impacts that occurred independently in all sub-regions. This shows that the choice of a counting method is important when modelling impacts and can significantly affect model performance.

3.2.2 Comparison of regression and classification models using a train-test analysis

Figure 9 shows the performance of the regression and classification RF models after being trained and tested on 75% and 25% of the data respectively. To do this, the regression model's output was converted to the categorical classes: "impact" or "no impact". We tested different thresholds to convert the model outputs and found that when the threshold was lowered, the recall increased, and the precision decreased. The opposite happened when the threshold was increased. We tested thresholds of 0.5, 1, 1.25 and 1.5. Model performance in this analysis was found to be slightly worse than in the cross-validation analysis. It is important to note that model performance for a region can vary depending on how the data are partitioned for testing and training the model. Hence, model performance does not only depend on the strengths between the predictor and the response variable but also depends on the particular splitting. For instance, if an important event or pattern remains in the testing set, it will not be used in the training of the model; hence it will decrease the model's predictive ability.

Although classification models appear to be better at predicting impact events in this analysis, these models did not contain any information on impact severity, whereas regression models did. Since both the model-types performed well, we suggest further evaluation of these to predict impacts. For example, using a classification model with more classes (to represent different impact severities) or using a different threshold to divide the two classes in this study. Also, similar to the previous cross-validation analysis (Sect. 3.1), here we see more clearly that overall, regions with a higher number of total impact occurrences performed best here; these are the S, CE and E regions.

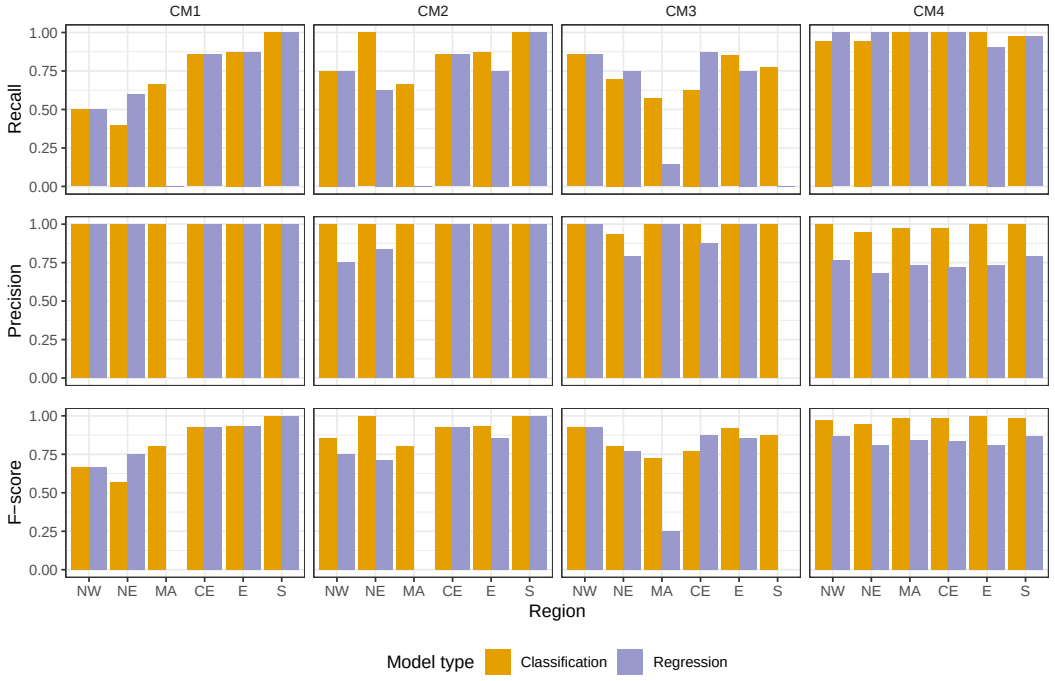

Figure 9. Performance metrics (recall, precision and F-score) of the RF models when training and validating the models on 75% and 25% of the data. All counting methods are shown. The regression model test predictions were converted into binary outcomes using a threshold of 1.25 DIOs.

3.3 Predictor importance

Figure 10 shows the predictor importance using the two most censoring counting methods and regression RF models. The overall patterns of predictor importance did not change significantly when we compared them to the classification RF models (not shown). The top predictors for each region are summarised in Fig. 11. Overall, the meteorological indicator, SPEI, was a top predictor in the CE and MA, at time scales between 24-33 and 15-18 months respectively. In the E and in MA, agricultural indices were top predictors at time scales of 12-21 and 18 months respectively. The hydrological indicator, SSFI, was a top predictor in the NW at a time scale of 21 months. The other hydrological indicator, SRSI, was a top predictor in the S, at timescales of 1-6 months. Out of all teleconnection patterns, AO, NAO and WeMO were the top predictors. The AO in the NW and NE at timescales between 15-21 months, the NAO in the S between 36-48 months, and the WeMO in MA, aggregated at 12 months.

When comparing the most important predictors from both analyses (correlation strength and RF variable importance), we found a general agreement for most sub-regions. When assessing the best drought indicators; MA, CE, E and S regions showed similar results. When assessing the best climate indices; the NE, followed by the S showed the most agreement. These results most probably indicate that for these top predictors, the indicator-impact relationship is linear.

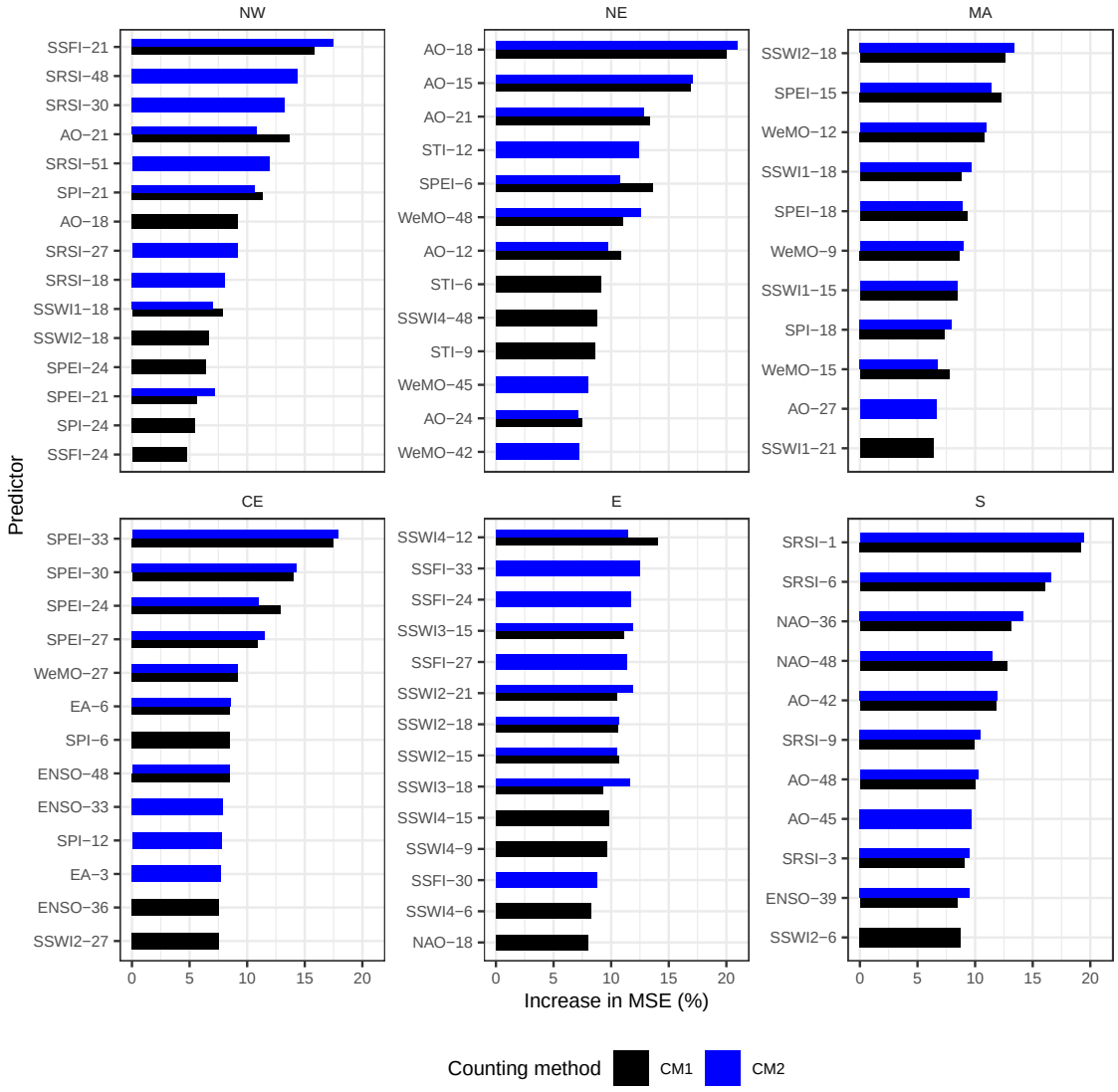

Figure 10. Predictor importance when using the regression RF models and the two most censoring counting methods. The top 10 predictors for each counting method and sub-region are shown.

4 Discussion and conclusion

We found links between the drought indicators and climate indices studied, and drought impact reports from the EDII database. We assessed these links by firstly using a correlation analysis and secondly by modelling drought impacts using drought indicators and teleconnection patterns as predictors in a RF model. RF models were skilful in predicting drought impact occurrences but seemed to be limited by the amount of impact occurrence data. However, we have shown that using drought

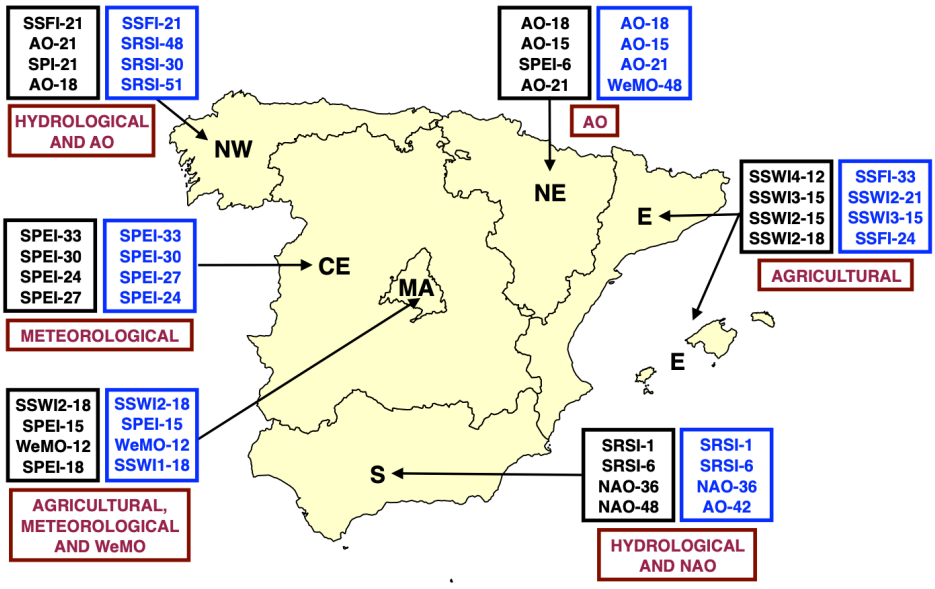

Figure 11. Map with the top four predictors for each sub-region when using the regression RF models and the two most censoring methods, in blue and black respectively. The "best" type of predictors for each sub-region are in red.

impact reports from the EDII with a RF model for Spain, a region with a reduced number of impact report entries, already provides good predictability of impacts for several sub-regions in Spain. Therefore, we encourage drought impact information to continue to be collected and given importance in future drought impact studies, as well as in drought management and early warning systems.

When we searched for the best predictors of impacts for different sub-regions in Spain, our results showed that overall, all the indicators studied showed strong and significant correlations in several regions and time scale aggregations. The indicators that showed strong and significant correlations over all regions were SPI, followed by SPEI. We recommend the use of these indicators if only one indicator is to be used for predictive purposes. However, correlation patterns showed spatial differences. To find the top predictors of impacts, we ranked the different drought indicators by correlation strength and used the results from the two most censoring counting methods. We found the EA and SPI were top five predictors in the NW, STI and AO were top three predictors in the NE, SSFI and SSWI2-3 were the top first and four predictors respectively in MA, WeMO and AO were the top first and second predictors in the CE, SSWI3 was the top predictor in the E, and SRSI was the top predictor in the S (see Table 2a and 2b). Moreover, links between teleconnection patterns or STI, and drought impacts using the EDII database have not been investigated before. Here, we found that in half of the sub-regions, these showed greater correlation strength than the drought indicators.

In the RF analysis, the top predictors for each region were: the SSFI and AO in the NW, AO in the NE, SSWI1-2, SPEI and WeMO in MA, SPEI in the CE, SSWI2-4 in the E and SRSI and NAO in the S (see Fig. 11). As described, both the correlation

and RF analyses showed slightly different predictor importance patterns. We used the two most censoring counting methods to identify the top predictors for each region because these methods showed the most consistent correlation patterns across all the sub-regions. Even though these two methods showed the lowest predictive skill in the RF models, we assume this is because of the reduced number of impact occurrences. Furthermore, we believe that when searching for predictor importance for each sub-region, it is best to exclude the least censoring method, which included in each subregion impact occurrences that affected the whole country or possibly only one specific region but labelled incorrectly. Instead, this method most likely represents the indicator-impact links for the whole region of Spain. Here, the top predictors were: ENSO, SRSI and the year. ENSO was a top four predictor for all sub-regions, SRSI for three sub-regions and the year for two sub-regions. The latter may suggest that impact occurrences have been monotonically increasing over the years. We suggest that to further explore these results, links between these predictors and impacts should be assessed at a country-level scale first. These teleconnection patterns might be the best predictors for impacts at a country level, according to the results obtained using this counting method.

By including the STI we investigated the links between temperature and impact occurrences. The correlation results showed mainly positive and significant correlations, which suggest a relationship between these two variables. However, the STI did not show the strongest correlations nor greatest variable importance (in the RF analysis) when compared to other drought indicators or teleconnection patterns, except for the NE region, where it showed higher correlation strengths than the rest of the indicators. Although we do not recommend the use of this index as a single drought predictor, we believe that its observed connection to drought impacts is important and might become more important as temperatures in Spain continue to increase. Especially, since there already is evidence on increasing trends in evapotranspiration in most meteorological stations in Spain due to decreased relative humidity and increased maximum temperature since the 1960s (Vicente-Serrano et al., 2014a). Moreover, González-Hidalgo et al. (2018) pointed out that since 1990, the role of atmospheric evaporative demand has been playing a large role in drought development. They state that drought is being driven by temperature conditions that affect atmospheric evaporative demand, independently of precipitation evolution. In our study, the SPEI, which includes the effects that temperature has on evapotranspiration, showed higher correlations than the SPI in four out of six regions (NW, NE, CE and E), which again suggests that including the effects of temperature when investigating drought and its impacts is important.

Adequate drought management requires knowledge on the time that different drought types take to propagate through different water resource systems. Both our analyses mostly agreed on the time scales over which different types of drought started to cause impacts. The time scales that showed the strongest links with impact occurrences depended on the sub-region and the method for its analysis. However, using both analysis we overall found, that the strongest links were found at timescales between 15-33 months for the meteorological indices, between 6-33 months for the hydrological indicator SSFI, between 1-18 months for the hydrological indicator SRSI, and between 6-21 months for layers 1-3 of the agricultural index. For the deepest soil layer, the correlation analysis showed strongest correlations at shorter timescales, from 1-9 months. The time scales at which the meteorological indices showed the strongest links were usually longer than those found in Germany and similar to the UK (Bachmair et al., 2016b). In these regions, SPI and SPEI showed the best links with impact occurrences at accumulation periods of 12–24 months for the UK, and at accumulation periods of 2–4 months for Germany. Stagge et al. (2015) found that

Norway, Bulgaria, and Slovenia responded even more rapidly to meteorological drought than Germany and the UK, which shows that Spain has the longest impact response out of these countries.

Furthermore, our results show that systems that respond to precipitation anomalies at the shortest time scales take longer to propagate to impacts. For instance, we have shown that the meteorological indices correlated with impacts at long time scales. The agricultural index in the shallowest layers (1-3) showed correlations at the longest time scales, and the deepest layer showed correlations at shorter time scales. An explanation for this is that soil moisture anomalies take longer to propagate to the deeper layers. Differently, the hydrological index showed strong correlations overall at the earliest time scales. In our analysis, this indicates that drought impacts respond to hydrological droughts and deep soil moisture droughts faster than meteorological and shallow layer soil moisture droughts.

Spain's resilience to short-term droughts, due to its extensive network of hydraulic reservoirs, could explain why we found most indicator-impact links at long time scales (especially meteorological indicators and teleconnection patterns). We found that most of the links between meteorological indicators and teleconnection patterns, and impact occurrences were strongest at time scales between 1-3 years and 1-4 years respectively, depending on the specific indicator and sub-region. As mentioned earlier, drought conditions that last more than two or three years have been shown to limit the capacity of Spain's hydraulic infrastructures (González-Hidalgo et al., 2018). In addition to reservoir systems, groundwater storage also provides resilience (water supply to satisfy demands) during periods of drought. Therefore, groundwater droughts may play a role and be an additional factor that contributes to these long accumulation periods. Especially since 15-20% of all water used in Spain is provided by groundwater (Hernández-Mora et al., 2003).

Moreover, we found that the most frequent type of impacts over all regions was on agriculture and livestock farming, and public water supply (Fig. 1). Both of these sectors depend on reservoir systems for storing water, since irrigation and public water supply are the two sectors that consume most of the stored water from reservoirs. Our results show that SRSI is the best predictor of impacts, outperforming all other indicators, in the S. The correlation analysis also showed strong and significant correlations between SRSI and impacts for several regions. This suggests that drought impacts in Spain depend on reservoir resilience and this could explain why it takes a long time for precipitation anomalies to propagate to impacts (and the response to less frequent but longer drought periods). Reservoir storage has been shown to respond to anomalies in SPI and SPEI at long timescales in some Spanish regions (Vicente-Serrano and López-Moreno, 2005; Lorenzo-Lacruz et al., 2010). This further demonstrates that to understand drought impacts at local scales, we need to consider the effects of local reservoirs systems, in addition to studying the other water resource systems.

The accuracy of our results is dependent on the accuracy of the impact data used, specifically, the method of quantification, the completeness of the data and potential sources of error. Since many impact reports were incomplete and their quantification is subjective, we tested four different versions of counting methods and investigated whether they had an effect on the results. We mainly tested; (1) whether to count an impact that affects the entire country equally as if it only occurred in one sub-region and (2) whether and how to count impacts that lacked information on the start or end date of the impact report. In the correlation analysis, different counting methods mainly produced differences in the strength of the correlations. The least censoring counting methods showed weaker correlations overall, and significant correlations disappeared in one-third of the

regions. However, in the RF analysis, the least censoring counting methods produced models with higher predictive skill than the more censoring methods. Regions with the most impact data also performed best. We infer this is because the performance of a RF model highly depends on the quantity of data used for its training. We therefore conclude that when working with impact data, it is important to compare counting methods and to investigate their effect on the results to overcome potential biases due to subjectivity.

Because Spain is a region that has a reduced sample size of impact reports, we did not conduct an analysis that investigated the links between indicators and sector-specific impacts. However, we believe such an analysis could give an insight into which type of indicators and what times of propagation are better linked to sector-specific impacts. Such a sector-specific analysis has already been conducted in other European regions (Blauhut et al., 2015, 2016; Stagge et al., 2015; Bachmair et al., 2016b). Since drought impact data availability was a limitation in this region, we encourage other types of impact data to be used to investigate these links, for instance, agricultural and economic data.

Code and data availability. The R package "randomForest" (Liaw and Wiener, 2002) is available at: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf. The R package "caret" (Kuhn, 2008) is available at: https://github.com/topepo/caret/. The R package "panas" (De Felice, 2020) is available at: https://github.com/matteodefelice/panas/. The R package "SPEI" (Vicente-Serrano et al., 2010; Beguería et al., 2014) is available at: http://sac.csic.es/spei. The Standardized Drought Analysis Toolbox (SDAT) (Hao et al., 2014; Farahmand and AghaKouchak, 2015) is available at: http://amir.eng.uci.edu/software.php. The research data used in this study are all publicly accessible. The precipitation and temperature data sets used are available at: http://hdl.handle.net/10261/183071) (Gutiérrez et al., 2019; Herrera et al., 2019). Data for streamflow and reservoir levels are available at: https://sig.mapama.gob.es/redes-seguimiento/index.html?herramienta=Aforos (Ministerio para la Transición Ecológica y el Reto Demográfico). Volumetric soil water content data from the ERA5-Land data set (Muñoz Sabater, 2019) are available at: https://doi.org/10.24381/cds.68d2bb30. Data for the climate indices: NAO, EA, AO and ENSO, from the NOAA Climate Prediction Center are available at: https://psl.noaa.gov/data/climateindices/list/ and https://www.cpc.ncep.noaa.gov/data/teledoc/ea.shtml, and data for the WeMO are available at: http://www.ub.edu/gc/wemo/ (Martin-Vide and Lopez-Bustins, 2006).

Author contributions. Both authors contributed to the design of the study, the interpretation of the results and the writing of the manuscript. HTS performed the analysis and prepared the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank the data providers: Ministerio para la Transición Ecológica y el Reto Demográfico, NOAA Climate Prediction Center, Copernicus Climate Change Service (C3S) Climate Data Store, DIGITAL.CSIC and EM-DAT: The Emergency Events Database -

Université Catholique de Louvain (UCL) - CRED, D. Guha-Sapir - www.emdat.be, Brussels, Belgium, for providing us with the data. We also thank Dr Sophie Bachmair for her initial technical help and suggestions. This work was supported by the ClimXtreme project funded by the German Federal Ministry for Education and Research (BMBF) and by the Institute for Basic Science (IBS), Republic of Korea, under IBS-R028-D1.

References

Austin, R., Cantero-Martínez, C., Arrúe, J., Playán, E., and Cano-Marcellán, P.: Yield–rainfall relationships in cereal cropping systems in the Ebro river valley of Spain, Eur. J. Agron., 8, 239–248, https://doi.org/10.1016/S1161-0301(97)00063-4, 1998.

Bachmair, S., Kohn, I., and Stahl, K.: Exploring the link between drought indicators and impacts, Nat. Hazards Earth Syst. Sci., 15, 1381–1397, https://doi.org/10.5194/nhess-15-1381-2015, 2015.

Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., Knutson, C., Smith, K. H., Wall, N., Fuchs, B., Crossman, N. D., and Overton, I. C.: Drought indicators revisited: the need for a wider consideration of environment and society: Drought indicators revisited, Wiley Interdiscip. Rev.: Water, 3, 516–536, https://doi.org/10.1002/wat2.1154, 2016a.

Bachmair, S., Svensson, C., Hannaford, J., Barker, L. J., and Stahl, K.: A quantitative analysis to objectively appraise drought indicators and model drought impacts, Hydrol. Earth Syst. Sci., 20, 2589–2609, https://doi.org/10.5194/hess-20-2589-2016, 2016b.

Bachmair, S., Svensson, C., Prosdocimi, I., Hannaford, J., and Stahl, K.: Developing drought impact functions for drought risk management, Nat. Hazards Earth Syst. Sci., 17, 1947–1960, https://doi.org/10.5194/nhess-17-1947-2017, 2017.

Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, Int. J. Climatol., 34, 3001–3023, https://doi.org/10.1002/joc.3887, 2014.

Blauhut, V., Gudmundsson, L., and Stahl, K.: Towards pan-European drought risk maps: quantifying the link between drought indices and reported drought impacts, Environ. Res. Lett., 10, 014 008, https://doi.org/10.1088/1748-9326/10/1/014008, 2015.

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., De Stefano, L., and Vogt, J.: Estimating drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors, Hydrol. Earth Syst. Sci., 20, 2779–2800, https://doi.org/10.5194/hess-20-2779-2016, 2016.

Breiman, L.: Random forests, UC Berkeley TR567, 1999.

Cutler, A., Cutler, D. R., and Stevens, J. R.: Random Forests, in: Ensemble Machine Learning: Methods and Applications, edited by Zhang, C. and Ma, Y., pp. 157–175, Springer US, Boston, MA, https://doi.org/10.1007/978-1-4419-9326-7_5, 2012.

Davis, J. and Goadrich, M.: The relationship between Precision-Recall and ROC curves, in: Proceedings of the 23rd international conference on Machine learning, pp. 233–240, Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, https://doi.org/10.1145/1143844.1143874, type: 10.1145/1143844.1143874, 2006.

De Felice, M.: Reduce the friction when working with maps and time-series, GitHub [code], https://rdrr.io/github/matteodefelice/panas/, 2020.

del Moral, L. and Hernandez-Mora, N.: La experiencia de sequías en España: inercias del pasado y nuevas tendencias en la gestión de riesgos, in: 5° Water Governance International Meeting, Water Governance Practices under Water Scarcity, Universidade de São Paulo, São Paulo, Brazil, 2015.

del Moral, L. and Saurí, D.: Changing Course: Water Policy in Spain, Environ.: Sci. Policy Sustainable Dev., 41, 12–15, https://doi.org/10.1080/00139159909604640, 1999.

Estrela, T. and Vargas, E.: Drought Management Plans in the European Union. The Case of Spain, Water Resour. Manage., 26, 1537–1553, https://doi.org/10.1007/s11269-011-9971-2, 2012.

Eurostat: https://ec.europa.eu/eurostat/web/nuts/background, last access: 12 February 2020.

Farahmand, A. and AghaKouchak, A.: A generalized framework for deriving nonparametric standardized drought indicators, Adv. Water Resour., 76, 140–145, https://doi.org/10.1016/j.advwatres.2014.11.012, 2015.

Feldstein, S. B. and Franzke, C. L. E.: Atmospheric Teleconnection Patterns, in: Nonlinear and Stochastic Climate Dynamics, edited by Franzke, C. L. E. and O'Kane, T. J., pp. 54–104, Cambridge University Press, Cambridge, https://doi.org/10.1017/9781316339251.004, 2017.

Gimeno, L., Ribera, P., Iglesias, R., de la Torre, L., García, R., and Hernández, E.: Identification of empirical relationships between indices of ENSO and NAO and agricultural yields in Spain, Clim Res, 21, 165–172, https://doi.org/10.3354/cr021165, 2002.

González-Hidalgo, J. C., Vicente-Serrano, S. M., Peña-Angulo, D., Salinas, C., Tomas-Burguera, M., and Beguería, S.: High-resolution spatio-temporal analyses of drought episodes in the western Mediterranean basin (Spanish mainland, Iberian Peninsula), Acta Geophys, 66, 381–392, https://doi.org/10.1007/s11600-018-0138-x, 2018.

Gudmundsson, L. and Seneviratne, S. I.: Anthropogenic climate change affects meteorological drought risk in Europe, Environ. Res. Lett., 11, 044 005, https://doi.org/10.1088/1748-9326/11/4/044005, 2016.

Guha-Sapir, D., Below, R., and Hoyois, P.: EM-DAT: the CRED/OFDA international disaster database, Université Catholique de Louvain [data set], www.emdat.be, 2016.

Gutiérrez, J. M., Herrera, S., Cardoso, R. M., Soares, P. M. M., Espírito-Santo, F., and Viterbo, P.: Iberia01: Daily gridded (0.1º resolution) dataset of precipitation and temperatures over the Iberian Peninsula, DIGITAL.CSIC [data set], https://doi.org/10.20350/DIGITALCSIC/8641, 2019.

Guttman, N. B.: Accepting the Standardized Precipitation Index: a calculation algorithm, JAWRA J. Am. Water Resour. Assoc., 35, 311–322, https://doi.org/10.1111/j.1752-1688.1999.tb03592.x, 1999.

Hao, Z. and AghaKouchak, A.: A Nonparametric Multivariate Multi-Index Drought Monitoring Framework, J. Hydrometeorol., 15, 89–101, https://doi.org/10.1175/JHM-D-12-0160.1, 2014.

Hao, Z., AghaKouchak, A., Nakhjiri, N., and Farahmand, A.: Global integrated drought monitoring and prediction system, Sci. Data, 1, 140 001, https://doi.org/10.1038/sdata.2014.1, 2014.

Hernández-Mora, N., Martínez Cortina, L., and Fornés, J.: Intensive groundwater use in Spain, in: Intensive Use of Groundwater: Challenges and Opportunities, edited by Llamas, M. and Custodio, E., pp. 387–414, Swets & Zeitlinger BV, The Netherlands, 2003.

Herrera, S., Cardoso, R. M., Soares, P. M., Espírito-Santo, F., Viterbo, P., and Gutiérrez, J. M.: Iberia01: a new gridded dataset of daily precipitation and temperatures over Iberia, DIGITAL.CSIC [data set], https://doi.org/10.5194/essd-11-1947-2019, 2019.

Hervás-Gámez, C. and Delgado-Ramos, F.: Drought Management Planning Policy: From Europe to Spain, Sustainability, 11, 1862, https://doi.org/10.3390/su11071862, 2019.

Hripcsak, G. and Rothschild, A. S.: Agreement, the F-Measure, and Reliability in Information Retrieval, J. Am. Med. Inf. Assoc., 12, 296–298, https://doi.org/10.1197/jamia.M1733, 2005.

Iglesias, A., Moneo, M., Garrote, R., and Flores, F.: Drought and climate risks, in: Water policy in Spain, edited by Garrido, A. and Llamas, R. M., pp. 63–75, CRC Press, Cambridge, 2009.

Iglesias, E., Garrido, A., and Gomez-Ramos, A.: Evaluation of drought management in irrigated areas, Agric. Econ., 29, 211–229, https://doi.org/10.1111/j.1574-0862.2003.tb00158.x, 2003.

Kuhn, M.: Building Predictive Models in R using the caret Package, CRAN [code], https://doi.org/10.18637/jss.v028.i05, 2008.

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R news, 2, 18–22, 2002.

Lorenzo-Lacruz, J., Vicente-Serrano, S., López-Moreno, J., Beguería, S., García-Ruiz, J., and Cuadrat, J.: The impact of droughts and water management on various hydrological systems in the headwaters of the Tagus River (central Spain), J. Hydrol., 386, 13–26, https://doi.org/10.1016/j.jhydrol.2010.01.001, 2010.

Manzano, A., Clemente, M. A., Morata, A., Luna, M. Y., Beguería, S., Vicente-Serrano, S. M., and Martín, M. L.: Analysis of the atmospheric circulation pattern effects over SPEI drought index in Spain, Atmos. Res., 230, 104 630, https://doi.org/10.1016/j.atmosres.2019.104630, 2019.

Martin-Vide, J. and Lopez-Bustins, J.-A.: The Western Mediterranean Oscillation and rainfall in the Iberian Peninsula, Int. J. Climatol., 26, 1455–1475, https://doi.org/10.1002/joc.1388, 2006.

Martinez-Artigas, J., Lemus-Canovas, M., and Lopez-Bustins, J. A.: Precipitation in peninsular Spain: Influence of teleconnection indices and spatial regionalisation, Int. J. Climatol., 41, https://doi.org/10.1002/joc.6770, 2021.

McKee, T. B., Doesken, N. J., and Kleist, J.: The relationship of drought frequency and duration to time scales, Proceedings of the 8th Conference on Applied Climatology, 17, 179–183, 1993.

Ministerio de Agricultura, Pesca y Alimentación: Recurrencia y Efectos de las Sequías, https://www.mapa.gob.es/es/desarrollo-rural/temas/politica-forestal/rec_ef_sequia_tcm30-152609.pdf.

Ministerio de Medio Ambiente: A Preliminary Assessment of the Impacts in Spain due to the Effects of Climate Change, https://www.miteco.gob.es/es/cambio-climatico/temas/impactos-vulnerabilidad-y-adaptacion/Full%20report_tcm30-178514.pdf, 2005.

Ministerio para la Transición Ecológica y el Reto Demográfico: Redes de Seguimiento del Estado e Información Hidrológica, https://sig.mapama.gob.es/redes-seguimiento/index.html?herramienta=Aforos, last access: 12 February 2020.

Ministerio para la Transición Ecológica y el Reto Demográfico: Plan Nacional de Adaptación al Cambio Climático 2021-2030, https://www.miteco.gob.es/es/cambio-climatico/temas/impactos-vulnerabilidad-y-adaptacion/plan-nacional-adaptacion-cambio-climatico/default.aspx, 2020.

Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1981 to present, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.68d2bb30, 2019.

Ollero Lara, A., Crespo Vergara, S. I., and Pérez Cimas, M.: Las sequías y España. La respuesta del seguro agrario a un problema intermitente, 2018.

Páscoa, P., Gouveia, C. M., Russo, A., and Trigo, R. M.: The role of drought on wheat yield interannual variability in the Iberian Peninsula from 1929 to 2012, Int. J. Biometeorol., 61, 439–451, https://doi.org/10.1007/s00484-016-1224-x, 2017.

Peña-Gallardo, M., Vicente-Serrano, S. M., Domínguez-Castro, F., and Beguería, S.: The impact of drought on the productivity of two rainfed crops in Spain, Nat. Hazards Earth Syst. Sci., 19, 1215–1234, https://doi.org/10.5194/nhess-19-1215-2019, 2019.

Ríos-Cornejo, D., Penas, Á., Álvarez-Esteban, R., and del Río, S.: Links between teleconnection patterns and precipitation in Spain, Atmos. Res., 156, 14–28, https://doi.org/10.1016/j.atmosres.2014.12.012, 2015.

Rodó, X., Baert, E., and Comín, F. A.: Variations in seasonal rainfall in Southern Europe during the present century: relationships with the North Atlantic Oscillation and the El Niño-Southern Oscillation, Clim. Dyn., 13, 275–284, https://doi.org/10.1007/s003820050165, 1997.

Rossi, G. and Cancelliere, A.: Managing drought risk in water supply systems in Europe: a review, Int. J. Water Resour. Dev., 29, 272–289, https://doi.org/10.1080/07900627.2012.713848, 2013.

Russo, A., Gouveia, C. M., Páscoa, P., DaCamara, C. C., Sousa, P. M., and Trigo, R. M.: Assessing the role of drought events on wildfires in the Iberian Peninsula, Agric. For. Meteorol., 237-238, 50–59, https://doi.org/10.1016/j.agrformet.2017.01.021, 2017.

Sainz de la Maza, M. and Del Jesus, M.: Análisis de sequías históricas a través de los impactos derivados, Ingeniería del agua, 24, 141, https://doi.org/10.4995/ia.2020.12182, 2020.

Salvador, C., Nieto, R., Linares, C., Díaz, J., and Gimeno, L.: Short-term effects of drought on daily mortality in Spain from 2000 to 2009, Environmental Research, 183, 109 200, https://doi.org/https://doi.org/10.1016/j.envres.2020.109200, 2020.

Sivakumar, M. V., Motha, R., Wilhite, D., and Wood, D.: Agricultural Drought Indices. Proceedings of an Expert Meeting, WMO, 2011.

Spearman, C.: The Proof and Measurement of Association Between Two Things., in: Studies in individual differences: The search for intelligence., edited by Jenkins, J. J. and Paterson, D. G., pp. 45–58, Appleton-Century-Crofts, East Norwalk, https://doi.org/10.1037/11491-005, 1961.

Stagge, J. H., Kohn, I., Tallaksen, L. M., and Stahl, K.: Modeling drought impact occurrence based on meteorological drought indices in Europe, J. Hydrol., 530, 37–50, https://doi.org/10.1016/j.jhydrol.2015.09.039, 2015.

Stahl, K., Kohn, I., Blauhut, V., Urquijo, J., De Stefano, L., Acácio, V., Dias, S., Stagge, J. H., Tallaksen, L. M., Kampragou, E., Van Loon, A. F., Barker, L. J., Melsen, L. A., Bifulco, C., Musolino, D., de Carli, A., Massarutto, A., Assimacopoulos, D., and Van Lanen, H. A. J.: Impacts of European drought events: insights from an international database of text-based reports, Nat. Hazards Earth Syst. Sci., 16, 801–819, https://doi.org/10.5194/nhess-16-801-2016, 2016.

Sutanto, S. J., van der Weert, M., Wanders, N., Blauhut, V., and Van Lanen, H. A. J.: Moving from drought hazard to impact forecasts, Nat. Commun., 10, 4945, https://doi.org/10.1038/s41467-019-12840-z, 2019.

Svoboda, M., Fuchs, B. A., Integrated Drought Management Programme, World Meteorological Organization, Global Water Partnership, University of Nebraska–Lincoln, and National Drought Mitigation Center: Handbook of drought indicators and indices, http://www.droughtmanagement.info/handbook-drought-indicators-and-indices/, oCLC: 954624216, 2016.

Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, Geographical Review, 38, 55, https://doi.org/10.2307/210739, 1948.

Vicente-Serrano, S. M. and López-Moreno, J. I.: Hydrological response to different time scales of climatological drought: an evaluation of the Standardized Precipitation Index in a mountainous Mediterranean basin, Hydrol. Earth Syst. Sci., 9, 523–533, https://doi.org/10.5194/hess-9-523-2005, 2005.

Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, J. Clim., 23, 1696–1718, https://doi.org/10.1175/2009JCLI2909.1, 2010.

Vicente-Serrano, S. M., Azorin-Molina, C., Sanchez-Lorenzo, A., Revuelto, J., Morán-Tejeda, E., López-Moreno, J. I., and Espejo, F.: Sensitivity of reference evapotranspiration to changes in meteorological parameters in Spain (1961-2011), Water Resour. Res., 50, 8458–8480, https://doi.org/10.1002/2014WR015427, 2014a.

Vicente-Serrano, S. M., Lopez-Moreno, J.-I., Beguería, S., Lorenzo-Lacruz, J., Sanchez-Lorenzo, A., García-Ruiz, J. M., Azorin-Molina, C., Morán-Tejeda, E., Revuelto, J., Trigo, R., Coelho, F., and Espejo, F.: Evidence of increasing drought severity caused by temperature rise in southern Europe, Environ. Res. Lett., 9, 044 001, https://doi.org/10.1088/1748-9326/9/4/044001, 2014b.

Vicente-Serrano, S. M., Azorin-Molina, C., Peña-Gallardo, M., Tomas-Burguera, M., Domínguez-Castro, F., Martín-Hernández, N., Beguería, S., El Kenawy, A., Noguera, I., and García, M.: A high-resolution spatial assessment of the impacts of drought variability on vegetation activity in Spain from 1981 to 2015, Nat. Hazards Earth Syst. Sci., 19, 1189–1213, https://doi.org/10.5194/nhess-19-1189-2019, 2019.

Vide, J. M.: Diez características de la pluviometría española decisivas en el control de la demanda y el uso del agua, Asoc. de Geogr. Espanoles, 18, 9–16, 1994.

Wilhite, D. A. and Glantz, M. H.: Understanding: the Drought Phenomenon: The Role of Definitions, Water Int., 10, 111–120, https://doi.org/10.1080/02508068508686328, 1985.