Comparison of meteorological and satellite-based drought indices as yield predictors of Spanish cereals

David García-León a,b,⁎, Sergio Contreras c, Johannes Hunink c

a Department of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University of Venice, Via Torino 155, 30172, Venice, Italy
b Euro-Mediterranean Center on Climate Change, Via della Libertà 12, 30175, Venice, Italy
c FutureWater, Calle Azuena, 23, 30205, Cartagena, Spain

ARTICLE INFO

Keywords:
Cereal yields
Agricultural drought
NDVI
LST
InfoSequia
ESYRCE

ABSTRACT

In the context of global warming, as drought episodes become increasingly frequent, it is crucial to accurately measure the impacts of droughts on the overall performance of agrosystems. This study aims to compare the effectiveness of meteorological drought indices against satellite-based agronomical drought indices as crop yield explanatory factors in statistical models calibrated at a local scale. The analysis is conducted in Spain using a spatially detailed, 12-year (2003–2015) dataset on crop yields, including different types of cereals. Yields and drought indices were spatially aggregated at the agricultural district level. The Standardised Precipitation Index (SPI), computed at different temporal aggregation levels, and two satellite-based drought indices, the Vegetation Condition Index (VCI) and the Temperature Condition Index (TCI), were used to characterise the dynamics of drought severity conditions in the study area. Models resting on satellite-based indices showed higher performance in explaining yield levels as well as yield anomalies for all the crops evaluated. In particular, VCI/TCI models of winter wheat and barley were able to explain 70% and 40% of annual crop yield level and crop yield anomaly variability, respectively. We also observed gains in explanatory power when models for climate zones (instead of models at the national scale) were considered. All the results were cross-validated on subsamples of the whole dataset and on models fitted to individual agricultural districts and their predictive accuracy was assessed with a real-time forecasting exercise. Results from this study highlight the potential for including satellite-based drought indices in agricultural decision support systems (e.g. agricultural drought early warning systems, crop yield forecasting models or water resource management tools) complementing meteorological drought indices derived from precipitation grids.

1. Introduction

Climate change is expected to increase the frequency and severity of droughts in continental Europe and the Mediterranean countries in the next few years (Lehner et al., 2017). These extreme events threaten global food security and local agri-food markets (Lesk et al., 2016; Hadebe et al., 2017). It is therefore necessary to detect droughts at an early phase and, more importantly, assess their impacts on the productivity of cropping systems. Only then, risk control and mitigation measures can be taken (IPCC, 2012; UNISDR, 2013). However, for an accurate estimation of drought impacts on crop productivity, there are still various data-related and methodological challenges that should be addressed (Siebert et al., 2017). In particular, drought impact assessments require: a) time series of local drought conditions, with sufficient spatial detail to account for the differences between agricultural areas in terms of climatic and biophysical conditions; and b) accurate, high-resolution, long-term yield databases.

Meteorological drought indices – e.g. the Standardised Precipitation Index (SPI) (McKee et al., 1993) or the Standardised Evapotranspiration Index (SPI) (Vicente-Serrano et al., 2010) – reported by regional Drought Monitoring Systems, such as the US Drought Monitor (http://droughtmonitor.unl.edu/) or the European Drought Observatory (http://edo.jrc.ec.europa.eu), are usually derived from climate gridded datasets. The use of gridded data may bias the estimation of drought impacts on crop yields (Aufhammer et al., 2013). This is because this data usually rely on a coarse network of weather stations that is not able to capture the local climate and biophysical heterogeneities of the landscape. To overcome this limitation, it is increasingly common to use remotely sensed indices as a mean to provide a high resolution measurement of the current status of the vegetation

⁎ Corresponding author at: Department of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University of Venice, Via Torino 155, 30172, Venice, Italy.
E-mail addresses: david.garcialeon@unive.it (D. García-León), s.contreras@futurewater.es (S. Contreras), j.hunink@futurewater.es (J. Hunink).
https://doi.org/10.1016/j.agwat.2018.10.030
Received 16 May 2018; Received in revised form 6 September 2018; Accepted 20 October 2018
0378-3774/ © 2018 Elsevier B.V. All rights reserved.
European countries, such as France or Germany (López-Querol et al., 2014), Cereal yields in Spain are generally lower than those observed in other regions. In 2014 these regions accounted for 96% of the total national production of cereals (MAPAMA, 2015). According to Rembold et al. (2013) showed that linear models outperform in many cases more complex statistical models, isolated or combined, helps to quantify the severity of droughts over a specific location and time period (Unganai and Kogan, 1998; Singh et al., 2003; Rembold et al., 2013; WMO and GWP, 2016).

At the same time, global (e.g. FAOSTAT) and regional (e.g. NASS-USDA in US, or FADN in Europe) crop yield data are usually aggregated at regional levels, making it difficult to retrieve reliable yield-drought relations at a finer scale (van Ittersum et al., 2013; FAO and DWFI, 2015). National yield statistics available in most European countries have also this limitation. Quantifying the local-scale impacts of drought severity on crop yields requires data, at least, at the plot or county level. In Spain, ESYRCE, a database on crop acreage and yields (Encuesta sobre Superficies y Rendimientos de Cultivos), comes at a plot-level resolution, which facilitates its incorporation into local drought-yield analyses.

To assess the relationship between drought conditions and crop yields, either process-based methods or statistical (data-driven) methods can be used. Process-based methods, also known as mechanistic methods, are typically used for studies at the plot level but they are data- and computationally demanding. At the European level, the Joint Research Centre implemented the process-based WOFOST model (Boogaard et al., 2014), whose outputs are channelled through an agricultural drought early warning bulletin (MARS) (https://ec.europa.eu/jrc/en/mars/bulletins). However, in recent years statistical (or empirical) methods have been applied with success by a wide range of researchers (e.g. Schlenker and Roberts, 2009; Lobell et al., 2011; Tack et al., 2015; Ceglar et al., 2016; Schauburger et al., 2017). More specifically, multiple linear regression models have proven to be successful (Quiring and Papakryiakou, 2003; Lobell et al., 2014; Bachmair et al., 2018), as they are flexible and easily applicable. Michel and Makowski (2013) showed that linear models outperform in many cases more complex statistical crop models in explaining yield variability.

This work focuses on the main cereal crops cultivated in Spain, i.e. wheat, barley, grain maize, oat, and rye, as they correspond to the largest cultivated area and are used for human and animal consumption. In 2014 these five classes covered 96% of the total harvested area and 93% of the total national production of cereals (MAPAMA, 2015). Cereal yields in Spain are generally lower than those observed in other European countries, such as France or Germany (López-Querol et al., 2016), constrained by more adverse heat and drought conditions and shorter growing cycles. With the exception of maize, typically grown in summer under irrigation regimes, cereals in Spain are predominantly rainfed (MAPAMA, 2015) and thus highly vulnerable to drought events, even though they appear to be more resistant to water deficit than other crop species (Daryanto et al., 2017).

The aim of this study is to locally quantify the ability of meteorological and satellite-based drought indices to explain factors of cereal yield variability in water limited environments using high-resolution agricultural and drought data. Results were validated by considering different climate zones, pre-processing methods and temporal aggregation levels, and by assessing the models’ forecast performance with a real-time exercise.

2. Materials and methods

2.1. Crop yield data

Farm-level yields and harvested areas spanning the years 2003 to 2015 were retrieved from the ESYRCE (Encuesta sobre Superficies y Rendimientos de Cultivos) dataset (see Table S1 for descriptive statistics of this dataset). ESYRCE is an annual, comprehensive, field-level survey providing data on agricultural surface and crop yields in Spain. ESYRCE serves as the Spanish input for the European Union’s Farm Accountancy Data Network (FADN) (http://ec.europa.eu/agriculture/rica/), which describes in detail the income and productivity of agricultural holdings throughout the European Union and represents a tool for evaluating EU’s Common Agricultural Policy.

ESYRCE is based on a conglomerate stratified sampling method. The Universal Transversal Mercator (UTM10) coordinate system is used to build 10km-sided blocks covering the entire country, as is shown in the left panel of Fig. 1. Each of these blocks is then subdivided in 100 equivalent cells of 1 km × 1 km (or 100 ha). Squared segments of 700 m side anchored in the southwest extreme of each cell are the basic units that are sampled every year (see centre and right panel of Fig. 1). Blocks are sampled according to their production intensity: three segments are visited in extensive blocks, and from nine to fifteen in intensive or very intensive production areas.

Around 6 million hectares of cereals are planted on average in Spain every year. Production is widespread around the whole territory but intensively located at the central plateaus and the banks of the main rivers (Ebro, Duero, and Guadalquivir), as evidenced in Fig. 2 (top) for wheat and barley (see Fig. S1 for oat, rye and grain maize). There are about a dozen species of cereals grown in Spain every year. In this study, we focus on the five most representatives, namely, winter wheat (including common wheat and durum wheat), barley, oat, rye and grain maize. The four first species are eminently rainfed whilst grain maize is mostly irrigated. These five classes cover 96% of the total area and 93%
of the total production of cereals in Spain.

As shown in Fig. 2 (bottom), winter wheat and barley yields have remained relatively stable since the beginning of the sample period (see Fig. S2 for oat, rye and grain maize), corroborating the generalised evidence of the stagnation of crop yields across Europe during the last two decades (Lobell and Moore, 2015; Ceglar et al., 2016).

2.2. A meteorological drought index: the SPI

The Standardized Precipitation Index (SPI-\(n\)) (McKee et al., 1993) compares the total rainfall observed at a particular location during a window period of \(n\) months with the long-term rainfall distribution for the same period at that location. SPI is computed on a monthly basis for a moving window of fixed length in which rainfall is accumulated. The length of the accumulated period typically takes values from 1 to 24 months depending on which type of drought needs to be evaluated: for example, short accumulation periods (e.g., SPI-1 to SPI-3) are better related with drought impacts on seasonal herbaceous systems, while larger SPIs (SPI-9 or SPI-12) are better connected with impacts on perennial crops (Contreras and Hunink, 2015).

We concentrated on shorter accumulation periods (1, 3, and 6), more linked to agricultural impacts, and calculated SPI based on a monthly precipitation gridded dataset from the Climatic Research Unit of the University of East Anglia (CRU TS, version 3.25) (Harris et al., 2014). This updated version covers the period 1901–2016 at the global scale and has a spatial resolution of 0.5°.

For each accumulation period, precipitation frequencies were fitted a gamma probability distribution. The resulting cumulative distribution function was then transformed into that of a standardised normal variable with zero mean and variance equal to one. SPI values were measured in units of standard deviation from the long-term mean of the standardised distribution and range from -3 (extreme dry) to 3 (extreme dry).
2.3. Satellite-based drought indices

Three satellite-based drought indices, the Vegetation Condition Index (VCI), the Temperature Condition Index (TCI), and the average of the two, the Vegetation Health Index (VHI) (Kogan, 1995, 1997), were used to quantify local drought severity. These indices were retrieved from the InfoSequia Drought Monitoring System (www.infosequia.eu), a satellite-based web-mapping climate service for the operational monitoring of droughts at the Spanish national level.

The VCI is retrieved from the Normalized Difference Vegetation Index (NDVI), and provides a normalised relative index in which the current observed value of NDVI is compared against the extreme limits (minimum and maximum values) observed during a reference period. TCI is similarly computed, but using instead Land Surface Temperature (LST) as input. While VCI provides a good surrogate of the vegetation stress due to wetness conditions, the TCI is more closely related to the thermal stress of vegetation. VHI, computed as the average of both indices, provides a combined effect of both stressors and thus represents a good overall indicator of the health status of vegetation. VCI, TCI and VHI values range between 0 (highest level of drought severity) and 100 (lowest level of severity). These indices are widely employed in drought impact assessments on agriculture (WMO and GWI, 2016; Bokusheva et al., 2016).

For any desired time step \( t \), VCI and TCI are computed as

\[
VCI_t = 100 \times \frac{NDVI_t - NDVI_{min}}{NDVI_{max} - NDVI_{min}}
\]

\[
TCI_t = 100 \times \frac{LST_{max} - LST_t}{LST_{max} - LST_{min}}
\]

\[
VHI_t = \frac{VCI_t + TCI_t}{2}
\]

where \( NDVI_{min} \) (\( LST_{min} \)) and \( NDVI_{max} \) (\( LST_{max} \)) refer to the historical minimum and maximum values at each desired time step over the study period. Composite values to calculate both indices were extracted from the Collection 6.0 of MODIS products. In particular we used MOD13A2 and MYD13A2 (Didan, 2015a, 2015b) for NDVI and MYD11A2 (Wan et al., 2015) for LST.

2.4. Spatial aggregation

The basic spatial unit used in this analysis was the agricultural district. This classification stems from an initiative of the Spanish Ministry of Agriculture in the 1970s to respond to the need of implementing and designing more efficient agricultural policies. The country was divided in non-legislative medium-sized spatial units (larger than a municipality and smaller than a province), with homogenous agricultural characteristics, such as similar potential output, agro-management practices and agro-climatic patterns. The total number of agricultural districts in Spain is 346 (excluding Canary Islands), with an average area of 14*10^4 Ha. and a standard deviation of 9.4*10^4 Ha. Fig. S3 shows a geographical description of the agricultural districts classification.

The motivation for this spatial aggregation is twofold. First, the existence of limitations inherent to survey data, usually answered vaguely and incompletely. Potential farm-level uncertainties as well as data inaccuracies and gaps can be largely smoothed out by aggregating individual farm-level data to a wider spatial unit. Second, the mismatch in spatial resolution between the SPI (0.5°) and the remotely sensed variables from InfoSequia (1 km) can a priori bias the regression results in favour of the latter. By upscaling the spatial level of analysis, a fairer comparison between all candidate predictors can be done. In this study, geo-referenced, area-weighted crop yields and drought indicators were thus spatially aggregated at the agricultural district level.

2.5. Statistical crop models

Statistical models of crop yield typically express either yield levels or yield anomalies as the dependent variable. Models based on yield levels usually account for a time trend to control for the combined effects of changes in agro-management practices, environmental and socio-economic factors and climatic changes and a weather-related component driving crop yield variability. We used the following pooled yield level model for each crop:

\[
y_{it} = \alpha_i + \beta_t \cdot t + \sum_{i} y_{it} X_{it} + \varepsilon_{it}
\]

where \( Y_{it} \) is the crop yield level in agricultural district \( i \) at growing season \( t \), \( \alpha_i \) represents an agricultural district fixed effect, \( \beta \) is a vector of loadings of \( t \), \( X_{it} \) represents the vector of candidate predictor at different frequencies over the growing season and \( \varepsilon_{it} \) are the residuals. Considering the Spanish annual calendar for crops (Oteros et al., 2015), a growing season between October and June was used for winter wheat, barley, oat and rye. For these crops, candidate predictors were quarterly aggregated to capture three different growing stages of the plant: early or tillering stage (October-December), mid or anthesis stage (January-March) and late or grain filling stage (April-June). For grain maize, a growing period between March and September was considered, with candidate predictors aggregated in equally-spaced two-month periods.

Additionally, yield anomalies relating inter-annual changes of crop yields to inter-annual changes of weather variables were calculated. To calculate yield anomalies, yield levels for every agricultural district, \( i \), were standardised using the Z-score formula as follows:

\[
Z_{it} = \frac{Y_{it} - \bar{Y}_{it}}{\sigma_{it}}
\]

being \( Y_{it} \) the sample average yield of agricultural district \( i \) and \( \sigma_{it} \) its standard deviation. \( Z_{it} \) has a mean of 0 and a standard deviation of 1. The resulting rescaled variables represent the dependent variable in the following model:

\[
Z_{it} = \sum_{i} \delta_{i} X_{it} + u_{it}
\]

where yield anomalies \( Z_{it} \) of agricultural district \( i \) at growing season \( t \) are linearly related to the candidate predictors \( X_{it} \) with coefficients \( \delta_{i} \) and residuals \( u_{it} \). As the dependent variable is standardised, no constant term is required. Model residuals should be normally distributed, independent and homoskedastic. We tested these assumptions using the Shapiro, Durbin–Watson and Breusch-Pagan tests. Also, an alternative method to calculate yield anomalies was considered. More specifically, yield anomalies were obtained by estimating a temporal trend with linear and quadratic terms. Alternative detrended yields were obtained as follows

\[
Z'_{it} = Y_{it} - \hat{\delta}_1 \cdot t - \hat{\delta}_2 \cdot t^2
\]

where parameters \( \hat{\delta}_1 \) and \( \hat{\delta}_2 \) were estimated via OLS by regressing \( Y_{it} \) on \( t \) and \( t^2 \).

The robustness of our results was evaluated for all crops and modelling methods. Several measures of quality fit and prediction accuracy were calculated, such as variable-adjusted R², Root Mean Squared Error (RMSE), Akaike Information Criterion (AIC) or Model Efficiency Index (NS). In addition, a cross-validation estimation experiment was designed. In particular, a repeated (\( N = 1000 \)) sequence of regressions from a random subset of the whole sample (70%) was estimated and tested in the remaining (30%) observations. Robust RMSE values were calculated for all crops and models.
2.6. Predictive ability

A real-time forecasting exercise was designed to assess the effectiveness of the prediction models to describe actual yield variability. The region of Castilla y León, which includes a total number of 59 agricultural districts, was selected for this statistical experiment. Castilla y León has the largest proportion of cereal production in Spain (see Fig. 2). Modelling was performed on a rolling window over a 5-year period (2011–2015). Models of crop yield anomalies (Eq. (6)) were re-calibrated after a complete vintage of data (yield and drought indices) was released. In-season yield anomaly forecasts were updated as new weather data became available. In quarters (months) where weather data was not yet released, average historic values were used. The three different SPI-based models were compared with the benchmark VCI-TCI model and an optimised version of the latter, fitted to Continental climate regions including only agricultural districts with less than 5% of urban land, according to the most updated version of the CORINE Land Cover project (https://land.copernicus.eu/pan-european/corine-land-cover). A false alarm yield anomaly detection rate (ratio of false positives and false negatives to total number of forecasts) was computed to compare the performance ability of each model.

3. Results

3.1. Explaining crop yield levels and anomalies

Pooled-OLS estimation for representative agricultural districts (more than 10 years of observed yields) was performed for every crop and candidate predictor. In Table 1, variable-adjusted coefficients of determination (R²-adj) for all the competing models are presented. Satellite-based (VCI, TCI, and VCI + TCI) models performed systematically better than precipitation-based (SPI-n) models in explaining both crop yield levels and crop yield anomalies. Seventy percent of winter wheat yield level variability was explained when quarterly values of satellite indices were combined. The overall fit was slightly better than that obtained with only the VCI (69.3%) or the TCI (66.1%). In all cases, models built with satellite-based indicators showed superior fit than those based upon precipitation-based. The use of satellite data improved R²-adj figures by 7% to 10%. As shown in the second part of Table 1, the performance of the competing models to describe wheat yield anomalies decreased overall, but relatively less in the case of satellite-based models. R²-adj percentage gains from using satellite-based indices (and their combination) were substantial (above 40%). Results for barley, oat and rye were quantitatively comparable to those shown above for winter wheat. Yield level explanatory power of satellite-based models ranged between 50%–70% and was always greater than SPI-based models. R²-adj gains were more marked in yield anomaly models, with satellite-based models almost doubling the explanatory power of precipitation-based models in some cases. Among the SPI models, SPI-3 models performed better than the others. The three-month SPI provides a comparison of the precipitation over a specific 3-month period with the precipitation totals from the same 3-month period for all the years included in the historical record, thus making this indicator particularly suitable for the chosen temporal (three-month) aggregation scheme.

Several other goodness-of-fit and model evaluation statistics were calculated, such as Root Mean Squared Error (RMSE), Nash-Sutcliffe efficiency index (NS) and Akaike Information Criterion (AIC). All of them support the messages delivered in this subsection with regards to the R²-adj analysis. Detailed results for all competing models can be seen in Tables S2-S4.

Of the winter crops considered in this study (winter wheat, barley, oat and rye), 83% are rainfed (MAPAMA, 2015). The few irrigated fields were excluded from the analysis. Fitting Eqs. (4) and (6) to grain maize data yielded notably different results. The explanatory power of both types of models decreased substantially compared to the crops considered before. The combination of VCI and TCI explained a maximum yield level variability of 25%.

3.2. Robustness of the results

To assess the consistency of the main results, several experiments were conducted. Firstly, a cross-validation procedure was designed, as described in the last part of Section 2.5. The prediction error of each model was tested by calculating the RMSE both within the training dataset and the test dataset. Table S5 summarises these average robust statistics, for each crop species and candidate model. These results confirm those presented in Section 3.1.

Secondly, individual models were fitted to each agricultural district. Due to the short studied period (2003–2015), local estimations could suffer from small sample size problems. To reduce the impact of this limitation, a LOOCV exercise was applied to individual districts with more than 10 observed yields. Also, to avoid potentially overfitted models, the VHI was used. A summary of the findings is presented in Fig. 3 and Fig. 4. The panels in Fig. 3 show for each district two interesting features: firstly, the already mentioned better performance of crop yield level models against crop yield anomaly models and, secondly, the confirmation of the better fit of satellite-based models versus SPI1 (left), SPI3 (centre) and SPI6 (right) at the local level. Fig. 4 describes geographically where individual R²-adj gains are located, which are concentrated to the west and south hemisphere of the country. Better fit of satellite-based models against SPI models is also observed for the cases of barley, oat and rye (Fig. S4-S6). Meanwhile, localised R²-adj gains show uneven and sparse, as seen in Fig. S7-S9.

Lastly, an alternative method to detrend yields was considered. Following Eq. (7), a quadratic time trend was estimated and new yield anomalies were obtained. Eq. (6) was re-calculated with this new dependent variable. The findings in Section 3.1 regarding the explanatory power of the candidate models and the better performance of satellite-based models are confirmed by this scenario. No measurable differences stem from the detrending method chosen (see Table S6 for detailed results).

3.3. Efficient temporal aggregation of predictors: Quarterly versus monthly aggregation

An operational statistical model for forecasting crop yields should be based on drought indicators at a high frequency. Info sequia variables are provided at a maximum of 8-day frequency whereas accumulated precipitation data is only available on a monthly basis.

Eqs. (4) and (6) were adapted at this stage to accommodate monthly

<table>
<thead>
<tr>
<th>dep. var: yield level</th>
<th>Wheat</th>
<th>Barley</th>
<th>Oat</th>
<th>Rye</th>
<th>Maize</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI</td>
<td>0.693</td>
<td>0.674</td>
<td>0.646</td>
<td>0.509</td>
<td>0.103</td>
</tr>
<tr>
<td>TCI</td>
<td>0.661</td>
<td>0.607</td>
<td>0.629</td>
<td>0.457</td>
<td>0.132</td>
</tr>
<tr>
<td>VCI + TCI</td>
<td>0.700</td>
<td>0.677</td>
<td>0.654</td>
<td>0.511</td>
<td>0.254</td>
</tr>
<tr>
<td>SPI-1</td>
<td>0.636</td>
<td>0.583</td>
<td>0.668</td>
<td>0.456</td>
<td>0.084</td>
</tr>
<tr>
<td>SPI-3</td>
<td>0.656</td>
<td>0.621</td>
<td>0.629</td>
<td>0.488</td>
<td>0.167</td>
</tr>
<tr>
<td>SPI-6</td>
<td>0.655</td>
<td>0.609</td>
<td>0.622</td>
<td>0.473</td>
<td>0.168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dep. var: yield anomaly</th>
<th>Wheat</th>
<th>Barley</th>
<th>Oat</th>
<th>Rye</th>
<th>Maize</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCI</td>
<td>0.323</td>
<td>0.387</td>
<td>0.271</td>
<td>0.254</td>
<td>0.090</td>
</tr>
<tr>
<td>TCI</td>
<td>0.254</td>
<td>0.300</td>
<td>0.226</td>
<td>0.150</td>
<td>0.085</td>
</tr>
<tr>
<td>VCI + TCI</td>
<td>0.350</td>
<td>0.412</td>
<td>0.299</td>
<td>0.260</td>
<td>0.176</td>
</tr>
<tr>
<td>SPI-1</td>
<td>0.203</td>
<td>0.269</td>
<td>0.261</td>
<td>0.175</td>
<td>0.054</td>
</tr>
<tr>
<td>SPI-3</td>
<td>0.249</td>
<td>0.342</td>
<td>0.247</td>
<td>0.221</td>
<td>0.122</td>
</tr>
<tr>
<td>SPI-6</td>
<td>0.245</td>
<td>0.318</td>
<td>0.229</td>
<td>0.204</td>
<td>0.126</td>
</tr>
</tbody>
</table>
The fit of the new models is described in Table S7. The addition of more regressors created a generalised and proportional increase in the overall fit of the available models. Satellite-based models continued to outperform SPI-based models, both for crop yield levels and yield anomalies. Interestingly, the SPI-1 family of models behaved better than models based on SPI-3 because SPI-1 figures describe monthly precipitation anomalies better (by construction).

### 3.4. Accounting for different climate zones

The use of climate zones to upscale process-based or statistical crop models is widespread at the global and national scale. Geographic areas are configured based on homogeneity in weather variables that have the greatest influence on crop yield. In Spain, due to its large size, there are marked differences between provinces close to the coast and those belonging to the vast central plateau, or Meseta, a continental influenced climate with hot, dry summers and cold winters, beneficial to cereals seeds requiring vernalisation processes. In the north, the Cantabrian Mountains, the Basque Country, Cantabria, Asturias and Galicia have a maritime climate, with cool summers and mild winters. On the Mediterranean coast, the climate is moderate with rain in spring and autumn. Following Fernández-Díaz et al. (1987), three different climate zones were distinguished, namely, Mediterranean, Continental and Atlantic. The geographical distribution of each zone can be seen in Fig. S10. Crop yield level and yield anomaly models were fitted separately for each climatic area. Results were partially conditioned by the

---

**Fig. 3.** Scatterplots of regression model performance (R²-adj) of models fitted to winter wheat yields at single agricultural districts. X-axis denotes fit of yield level models whereas Y-axis denotes fit of yield anomaly models. Black solid line denotes 1:1 equal fit line. Green dots indicate fit of VHI-based models. Blue dots show the fit of SPI-1 models (top left), SPI-3 models (top right) and SPI-6 models (bottom left). A 95% confidence ellipse for the mean of each bi-variate distribution is provided. Results are based on LOO independent estimations (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

**Fig. 4.** Agricultural district level model fit (R²-adj) of VHI-based models (top left), SPI-1-based models (bottom left) and R²-adj gains from using VHI-based against SPI-1-based (right). Grey-shaded districts were not included in the analysis.
than those observed for winter wheat and barley. Goodness of
creases in explanatory power of oat and rye modelling were weaker
and 20.5% for yield level and yield anomaly models, respectively. In-
able nearly all models become more accurate. The optimised VCI-TCI
winter wheat yield anomalies better than SPI-based models in real time.

3.5. Out-of-sample performance: real-time forecasting exercise

As seen in the left panel of Fig. 5, satellite-based models predict
winter wheat yield anomalies better than SPI-based models in real time.
In general, as more information about the current season comes available
nearly all models become more accurate. The optimised VCI-TCI model (Continental climate zone and < 5% of urban land) delivers a
better performance skill than the benchmark VCI-TCI model, benefiting
from cleaner pixels and climate-specific parameterisation. Similar conclusions are drawn from the forecast derived from monthly models
(right). Almost for all lead times the optimised VCI-TCI model provides
better forecasts on average, with increasing performance as the cropping
season unfolds and more data comes available. Results for barley
are qualitatively similar to those of winter wheat (see Fig. S11). So are
those for oat and rye but the relatively reduced number of observations
of these two last crops limits the value of the forecast exercise, as seen
in Fig. S11-S13.

4. Discussion and conclusions

Satellite-based models showed better performance than SPI-based
models, for both crop yield levels and crop yield anomalies and for all
considered crops. Yield level models were able to explain about 70% of
winter wheat and barley yield level variability, a remarkable figure
given the simplicity of the modelling approach and the high degree of
spatial resolution. Crop yield anomaly models of wheat and barley
explained 35%–40% of total yield variability, a similar figure to that
obtained by other authors in different study regions (Quiring and
Papakryiaokou, 2003; Lobell et al., 2014; Ceglar et al., 2017). For oat
and rye, similar results were obtained in qualitative terms but estima-
tions suffered from sparse data and reduced sample size. Our results for
grain maize, predominantly irrigated in Spain, were very poor for both
precipitation-based and satellite-based models, probably conditioned
by the smaller range of variability of maize yields and because of as-
suming a fixed agronomic calendar for all cropping systems. It is rea-
sonable to think that growing cycles are adjusted to latitude (heat) and
other climatic factors. Incorporating this feature into our models may
well result in an improvement of the overall fit. Considering alternative
drought indicators, such as total evapotranspiration, may increase the
fit of maize models (Zdeněk et al., 2017). For the case of wheat, dis-
tinguishing between common and durum wheat, the latter regarded as
being more resistant to drought and heat stress, could also enhance our
model performance.

We also applied crop models to different climate-geographical
zones, on the premise that different agroclimatic regions at the national
scale may have different model statistics, due to large climate hetero-
gegeneity between zones (van Wart et al., 2013; Ribeiro et al., 2018). We
found that there was an R²-adjust gain in concentrating our modelling on
Continental climate areas, where national cereal is dominant. Further
investigation is required, though, to assess the effect of agronomic cal-
dendar and phenology in our estimations. For the sake of simplicity, we
imposed in our models a fixed agronomic calendar for all climate zones.
Running each climate zone model adjusted to the timing (calendar) and
length (phenology) of the growing cycle would shed some more light
into our final results. This statement will be tested in future research.
Also, the overall performance of our approach to assess drought impacts
in more humid climates may improve if the water balance of these areas
is taken into account (Vicente-Serrano et al., 2010) or if non-local hy-
drology (e.g. river discharge) is included (Zampieri et al., 2018).

Table 2
Explained variability (R²-adjusted) of statistical crop models for the five cereal
species considered. Dependent variable: crop yield level (top) and crop yield
anomaly (bottom). Drought indicators: VCI-TCI. Models are fitted to each cli-
matic zone.

<table>
<thead>
<tr>
<th>dep. var: yield level</th>
<th>Wheat</th>
<th>Barley</th>
<th>Oat</th>
<th>Rye</th>
<th>Maize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mediterranean</td>
<td>0.622</td>
<td>0.569</td>
<td>0.476</td>
<td>–</td>
<td>0.372</td>
</tr>
<tr>
<td>Continental</td>
<td>0.746</td>
<td>0.728</td>
<td>0.685</td>
<td>0.531</td>
<td>0.538</td>
</tr>
<tr>
<td>Oceanic</td>
<td>0.315</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dep. var: yield anomaly</th>
<th>Mediterranean</th>
<th>Oceanic</th>
</tr>
</thead>
<tbody>
<tr>
<td>yield anomaly</td>
<td>0.270</td>
<td>0.141</td>
</tr>
<tr>
<td>yield anomaly</td>
<td>0.305</td>
<td>0.279</td>
</tr>
</tbody>
</table>

Fig. 5. Wheat yield anomaly false alarm average detection rate for 5-year (2011–2015) real-time forecasting exercise. SPI-1-based, SPI-3-based, SPI-6-based and VCI-
TCI-based models are compared at quarterly (left) and monthly (right) frequencies. The VCI-TCI* model is optimised to Continental climate districts and incorporates
the removal of districts with > 5% of urbanised area.
Finally, we demonstrated through a simple real-time forecasting exercise the higher performance of satellite-based models as operational drought assessment tools, both at the quarterly and monthly temporal scale, with satellite-based models better avoiding false positives and false negatives yield anomalies. More work and alternative measures, however, are required to validate the forecast accuracy of these models. The use of a probabilistic approach that incorporates short-term drought forecasts can be also considered for future exercises (Madadgar et al., 2017).

Results derived from this analysis demonstrate that: (1) finer spatial resolution provided by satellite products capture local drought conditions better than drought measures based on climate gridded datasets, which usually rest on data provided by a coarse network of weather stations (Aufhammer et al., 2013); (2) remote sensing data related to vegetation health give a better picture of local drought conditions than drought indices that only rely on rainfall, mainly because vegetation-related indices inherently consider the local biophysical (soil, slope) and climate conditions.

The current study used two common indices describing the vegetation status of the plant. Future local-scale assessments could expand on the number of climate/land surface variables included in our models and use it to more comprehensively understand crop-specific sensitivities to drought. Other available satellite-based indicators not currently used for operational drought monitoring, such as near-surface air relative humidity data, may improve the performance of our models and tell us more of the spatial patterns and temporal trends of crop yields (AghaKouchak et al., 2015). Also, remotely sensed soil moisture from the Soil Moisture and Ocean Salinity (SMOS) or the Soil Moisture Active and Passive (SMAP) missions may be incorporated as additional indicators of agricultural drought (Sánchez et al., 2016). A geographically weighted combination of the available indicators could also result in an increase of the explanatory power of our models (Páscoa et al., 2017).

Even though we advocate in this study for the use of satellite-based measures to detect agronomical drought stress because of the limitations of climate gridded data, the recent proliferation of local, good quality, high-resolution precipitation gridded datasets (Quintana-Seguí et al., 2017) has yet to be tested in local-scale drought assessments. In the case of Spain, the Version 4 of the Spain02 dataset (Herrera et al., 2016), shows a horizontal resolution of up to 0.11°, five times larger than the CRU dataset and very similar to that of the remote sensing products used in the current study. There also exists a novel trend in the literature claiming for coupling statistical models with seasonal forecasts stemming from climate models or an ensemble of climate models (Hao et al., 2018). This (hybrid) modelling strategy opens an entire avenue for research and has yet to be further explored in regional and local assessments (Ceglar et al., 2018) as, for example, in our case study area.

There is an increasing need for weather and climate services that provide user-focused information on anticipated drought impacts to improve agricultural water management (van den Hurk et al., 2016). However, there is still a gap between what stakeholders in the agricultural sector need in terms of spatial resolution (i.e. at least district level) and the meteorological drought information available from observations or climate models (i.e. coarser information potentially useful at the provincial level). The current study confirmed that high-resolution satellite-based drought information portrays drought impacts at the district scale better than meteorological drought information. Still, for drought impact forecasting, seasonal forecasts from dynamic climate models based on coarse meteorological information (Iizumi et al., 2013) may still have an added value.

The predictive quality of seasonal climate models is still increasing and has shown to be useful for certain areas in Europe (Turco et al., 2017). Potentially, satellite-based information on current drought conditions can be used to downscale coarser forecasts of crop yield impacts. New guidelines for updating Agricultural Water and Drought Management Plans in all the Spanish River Basins recognise the need to include SPI as an index to be monitored for agricultural drought emergency plans. This study suggests that satellite-based drought indices may complement SPI, as they have been demonstrated here to be more directly related to crop yield impacts and thus to the economic consequences of droughts in Spain.

Acknowledgements

David García-León acknowledges financial support from the European Commission (H2020-MSCA-IF-2015) under REA grant agreement no. 705408. InfoSequia has been partially funded by the Spanish Ministry of Economy and Competitiveness through a Torres-Quevedo grant, and by the EU-funded BRIGAID project (grant agreement no. 700699). Sergio Contreras and Johannes Hunink also acknowledge financial support from EU-funded IMPREX project (grant agreement no. 641811). We thank the Spanish Ministry of Agriculture for providing the ESYRCE dataset. We are grateful for helpful comments and suggestions received from two anonymous reviewers and seminar participants at the Fondazione Eni Enrico Mattei and the European Commission’s Joint Research Centre.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.agwat.2018.10.030.

References
