Agricultural and Forest Meteorology 216 (2016) 58–67

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Impact of meteorological drivers on regional inter-annual crop yield variability in France

Andrej Ceglar*, Andrea Toreti, Rémi Lecerf, Marijn Van der Velde, Frank Dentener

European Commission, Joint Research Centre, Institute for Environment and Sustainability, Via Enrico Fermi 2749, 21027 Ispra, Varese, Italy

ARTICLE INFO

Article history:
Received 27 March 2015
Received in revised form
30 September 2015
Accepted 2 October 2015
Available online 11 November 2015

Keywords:
Climate variability
Grain maize
Winter wheat
France
Partial least square regression

ABSTRACT

The impact of intra-seasonal climate variability on inter-annual variation in winter wheat and grain maize yields over 92 French administrative regions is assessed. Observed monthly time series of temperature, precipitation and solar radiation during the growing season are analysed together with reported annual crop yields with a statistical approach based on partial least square regression. Results highlight remarkable spatial differences in the contribution of the main meteorological drivers to crop yield variability and in the timing of the maximum impact. Overall, temperature and global solar radiation are identified as the most important variables influencing grain maize yields over the southern, eastern and northern parts of France, while rainfall variability dominates yields over the central and north-western parts of the country. Positive rainfall anomalies during the summer months lead to an increase in maize yields, while positive temperature and radiation anomalies have the opposite effect. Extensive irrigation suppresses the rainfall signal in dry years. Winter wheat yields are predominantly influenced by temperature variations in eastern France and by rainfall variations over the northern, north-western and south-eastern France. In general, variation in global radiation plays a more important role in the southern than in the northern part of the country. Our study contributes to a better understanding of the impact of intra-seasonal climate variability on crop yields. Potential applications of the inferred models are discussed, especially in terms of seasonal crop yield forecasting and validation of dynamic crop model simulations.

1. Introduction

Prevailing climatic conditions underpin the suitability of agriculture to produce food, feed, fuel and fibre. At the same time agricultural production is greatly affected by weather extremes and climate variability (e.g. Cantelaube and Terres, 2005), the latter referring to the variations beyond synoptic timescales of the mean state and other properties of the climate system (Cubasch et al., 2013). Disentangling the influence of climatic variability on recorded crop yield variability has been an age-old activity of farmers, agronomists, and agro-meteorologists (e.g. Porter and Semenov, 2005). A renewed impetus for this research has emerged due to concerns related to climate change and the associated expected changes of the principal climatic factors determining crop growth. Understanding the relationship of climate variability with past crop production is of high importance to assess the resilience of our agricultural production systems to future climate conditions as well as the identification of adequate measures to adapt to climate change.

Intra-seasonal climate variability can affect crop production during all phases of the crop growing cycle: directly through the effects of temperature, water availability, radiation interception, and carbon fixation; indirectly by modulating nutrient availability and the occurrence of diseases and pests (Olesen et al., 2000). The sensitivity of optimal crop growth and development to specific weather conditions depends on the crop and growth stage. For instance, during the early growth stages of grain maize, unfavourable weather conditions (e.g. wet and cold weather) can limit the size of the leaves and therefore the photosynthetic capacity. In the later stages, adverse conditions (e.g. heatwave and drought) can reduce the number of silks produced, resulting in poor pollination of the ovules and restricting the number and/or the size of the developing kernels (Ritchie et al., 1993). Importantly, the same extreme weather event can also lead to contrasting responses

* Corresponding author. Tel.: +39 0332786256; fax: +39 0332783033.
E-mail address: andrej.ceglar@jrc.ec.europa.eu (A. Ceglar).

http://dx.doi.org/10.1016/j.agrformet.2015.10.004

of crops, for example as a function of growth stage (e.g. Van der Velde et al., 2012). Finally, crop quality and field workability are highly weather dependent.

Two approaches to identify the driving factors behind crop yields can be distinguished: mechanistic (dynamic) crop modelling (e.g. Van der Velde et al., 2012) and statistical modelling (aiming at relating reported crop yields variability to a set of explanatory variables; e.g. Lobell and Burke, 2010), even though hybrid approaches have also been developed (e.g. Lobell, 2013). The main advantage of using a crop model is the comprehensive characterization of the crop production system. Once properly calibrated, and evaluated with observed data, crop models can also provide information on possible management interventions to better cope with expected changes in temperature and precipitation (e.g. Laux et al., 2010; Balkovič et al., 2014). The main disadvantage of applying dynamic crop models (originally developed to run at the field scale) at the larger scale is the requirement of an extensive set of input data with information on soils, meteorological variables, agro-management practices (sowing and harvesting date, fertilization rate and irrigation) and eco-physiological parameters describing the crop variety. Since these models have generally been calibrated to local field conditions, their use in other regions would require a recalibration (Folberth et al., 2012). Nevertheless, large scale implementations of crop models can reproduce average observed yields (Liu, 2009; Balkovic et al., 2013), often by using agricultural management inputs that are centred around one year. In any case, detailed and complete spatially resolved datasets with information on crop rotation, spatial distribution of varieties and cultivars, spatial and temporal heterogeneity of fertilization rates as well as irrigation practices (required for a complete model parameterization) are rarely available (Lobell et al., 2008). Recent advances in the ability of ensembles of field-scale crop models to reproduce the effects of climate variability on crop yield provide a promising way ahead (Asseng et al., 2013).

Over the last decade, the use of statistical approaches to characterize the relationship between yields and meteorological variables has increased with the increasing availability and improved quality of observed data (e.g. from remote sensing and reported statistics). Lobell (2010) suggested a superior performance of statistical approaches w.r.t. crop models to identify this relationship. However, most statistical analyses of inter-annual crop yield variability have been focused on the seasonal or growing-season time scale (e.g. Lobell et al., 2008; Tebaldi and Lobell, 2008; Schlenker and Roberts, 2009; Lobell and Burke, 2010; Ceglar and Kajfež-Bogataj, 2012; Michel and Makowski, 2013). Therefore, opportunities exist to use statistical approaches to better characterize intra-seasonal impacts on inter-annual crop yield variability. This is especially pertinent given the improved quality of weather forecasts during the last decades (e.g. ECMWF, 2013). Modern agriculture has been increasingly using information from operational weather forecasts, for instance to: plan the preparation of fields (ploughing), sow or plant, apply agricultural chemicals, schedule irrigation, weed, crop harvest and storage, prevent damages due to chilling, frost and freezes, forestry operations, etc.

The main objective of our study is to identify the key meteorological variables and their period of maximum influence on the inter-annual variability of grain maize and winter wheat yields during crop growth. We propose a statistical approach that is able to: tackle the problem of co-variation and provide information on the main intra-seasonal driving meteorological factors of crop yield inter-annual variability. Subsequently, we evaluate whether the results from the statistical approach are in agreement with the agronomic knowledge on the principal meteorological drivers and timing with respect to sensitive growth stages. Our analysis requires statistical crop yield and meteorological information for sufficiently long time scales at sub-national spatial scales. Therefore, we analyse winter wheat and grain maize yield variability by using reported data from 92 French administrative regions (hereafter called *départements*). Four major climate types (maritime, Mediterranean, continental and mountainous) meet and interweave in France (Joly et al., 2010; Peel et al., 2007); this allows a comparison of climate variability-crop yield relationships over different climate types.

2. Data

Time series (from 1989 to 2014) of grain maize and winter wheat yields from 92 French *départements* (Fig. S.1) were provided by AGRESTE Ministère de l'Agriculture (AGRESTE, 2015). Wheat is predominantly produced in the northern part of France, while grain maize is more predominantly cultivated in south-western France (Van der Velde et al., 2012). Weather data were retrieved from the MARS Crop Yield Forecasting System (MCYFS) database, established and maintained by the Joint Research Centre for the purpose of crop growth monitoring and seasonal forecasting (Biavetti et al., 2014). In short, daily meteorological data are obtained every day from around 4000 weather stations and interpolated into a regular 25×25 km grid over Europe and neighbouring countries (635 stations are located in France). In this study, we use monthly mean temperature, monthly cumulated precipitation and global solar radiation for the entire growing season. The analysed period, therefore, stretches from October to July for winter wheat and from April to September for grain maize. Gridded meteorological data are spatially aggregated at the *département* level, only considering the agricultural areas as provided by the Global Land Cover 2000 project (GLC2000; Bartholome and Belward, 2005) within each *département* (Fig. S.1). Even though the quality of the French agricultural statistics is very high, few issues (unlikely related to climate or agronomical practices) were identified in several grain maize time series. These suspicious time series exhibited either one or a combination of the following issues:

(a) equal crop yield values in three or more consecutive years,
(b) biophysically implausible yield values (e.g. yields higher than 14 t/ha),
(c) sudden drops or increases in yield values, i.e. a break point in the time series associated with an almost negligible inter-annual variability afterwards.

The affected *départements* for grain maize are: Haute-Garonne, Aude, Isère, Bouches-Du-Rhône, Var, Alpes-De-Haute-Provence, Lozère, Alpes Maritimes and Cantal. They were discarded from further analysis. All *départements* were kept for the winter wheat analysis.

3. Methods

3.1. De-trending

In order to analyse the impact of climate variability on crop yield inter-annual variability, time series must be de-trended. Crop yields are strongly influenced by intra-seasonal and inter-annual climate variability, but also by improvements and responses in agro-management practices and other socio-economic factors. To maximize and enhance ecosystem service benefits from agricultural fields (including crop yield), appropriate agro-management is often required, for instance, to control soil erosion and reduce nutrient applications. Generally, it takes several years before new crop varieties or new agro-management practices come into practice. Here, we assume that the influence of these factors is mainly reflected in the multi-annual trend component of the yield time series. However, as mentioned in the introduction, such methodology cannot identify and filter out ad-hoc adaptation measures taken by farmers to counterbalance adverse weather conditions (e.g. Van der Velde et al., 2010). At the same time, certain developments are hard to isolate; as an example, while the maize area stayed constant between 2000 and 2013, the irrigated maize area in France decreased by 12% (DISAR, 2014).

The locally weighted polynomial regression (LOESS; Cleveland, 1979) is here applied to de-trend the crop yield time series. The same procedure is also applied to all the other explanatory variables (temperature, precipitation and global radiation) as the PLSR regression, described in Section 3.3, assumes stationarity of time series. Moreover, this study focuses on inter-annual variability of time series rather than on trends.

3.2. Spatial clustering of crop yield time series

A hierarchical clustering method (Murtagh, 1985) is used to identify spatially homogenous areas in terms of inter-annual crop yield variability. A correlation-based dissimilarity measure is implemented in the clustering procedure.

This spatial classification can aid in the interpretation of the dominant climatic drivers and possibly prevalent agro-management techniques.

3.3. Inter-annual crop yield variability

A Partial Least Squares Regression (PLSR; Garthwaite, 1994; Wold et al., 2001; Rosipal and Kramer, 2005) approach is used to estimate the relationship between meteorological variables and crop yield time series. PLSR is a flexible method for multivariate data analysis that has been already applied in numerous previous studies (e.g. Mehmood et al., 2012). It is very useful especially when the number of explanatory variables is similar or higher than the sample size. In this study, the number of explanatory variables amounts to 18 (3 meteorological variables for 6 months of the growing season) and 30 (3 meteorological variables for 10 months of growing season) for grain maize and winter wheat, respectively. In addition, the PLSR can deal with cases when the explanatory variables are strongly correlated (i.e. there is strong collinearity). Here, significant correlations (which are stronger in the warm half of the year) appear between monthly meteorological variables.

PLSR generalizes and combines features from principal component analysis (Jolliffe, 2002) and multiple-regression and can be interpreted as a form of Canonical Correlation Analysis (Rosipal and Kramer, 2005). The PLSR model is mainly based on the extraction of a sub-set of latent variables (i.e. inferred, not directly observed variables, to have the best predictive power) from the full set of predictors **X**. These latent variables are derived by maximizing the covariance of the transformed **X** (here, meteorological data) and **Y** (here, the observed crop yield data). Briefly, independent normalized variables **X** and **Y** are decomposed as

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{T} + \boldsymbol{E}$$

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^{T} + \boldsymbol{F}$$

where $\boldsymbol{P}$ and $\boldsymbol{Q}$ represent weight matrices, $\boldsymbol{T}$ and $\boldsymbol{U}$ denote the latent variable matrices, $\boldsymbol{E}$ and $\boldsymbol{F}$ are the matrices of residual terms. Latent variables $\boldsymbol{T}$ are inferred from the explanatory meteorological variables $\boldsymbol{X}$ and can correspond to an aspect of physical reality, in our case the dominant monthly weather patterns having the highest impact on inter-annual crop yield variability. Thus, the model to be used for the estimation becomes $\boldsymbol{Y} = \boldsymbol{T}\boldsymbol{D}\boldsymbol{Q}^{T} + \Gamma$ assuming $\boldsymbol{U} = \boldsymbol{T}\boldsymbol{D} + \boldsymbol{H}$ where $\boldsymbol{D}$ is a diagonal matrix, Γ and $\boldsymbol{H}$ matrices of residuals**.** For a complete overview of the methodology, the reader is referred to Wold et al. (2001) and Rosipal and Kramer (2005). To fully take advantage of both Canonical Correlation Analysis (e.g. Izenman, 2008) and PLSR, the canonical powered partial least square regression (Indahl et al., 2009) is used in this study. PLSR is applied to de-trended crop yield time series ($\boldsymbol{Y}$) and the aforementioned monthly meteorological variables ($\boldsymbol{X}$).

Bootstrap is used to determine the number of relevant latent variables as well as the importance of the explanatory meteorological variables on the prediction of crop yield anomalies. 500-time resampling is found to provide very consistent results with stable bootstrap uncertainty intervals. The number of relevant latent variables in the PLSR regression is determined by the ordinary bootstrap validation approach (Efron and Tibshirani, 1993). The importance of each variable is assessed by using the Variable Importance of Projection (VIP; Mehmood et al., 2012). This method measures the contribution of each explanatory variable according to the variance explained by each PLSR latent variable. More details can be found in the Supplementary Online Material (SOM).

4. Results

4.1. Analysis of de-trended crop yield time series

Grain maize and winter wheat are widely grown crops in France. Yields of grain maize have increased in the last 20 years (Fig. 1) although a levelling off can be observed after 2000. An increased frequency of heat spells has been suggested as a potential explanation (Hawkins et al., 2013). Stagnating yields of winter wheat (Fig. 1) can be attributed to different causes, such as changes in agricultural practices (especially in the application rates of fertilizers), higher frequency of heat stress and drought events during the stem elongation period (Brisson et al., 2010; Michel and Makowski, 2013). High spatial variability characterizes the observed yields for both crops, suggesting that better information on the impact of climate variability on yields can be gained by working at the regional scale rather than at the national one.

The inter-annual variability of yields, expressed as the variance of the de-trended time series, is highly variable across France (Fig. S.2). The highest grain maize variability can be observed over Haute–Vienne (69, see Fig. S.1) and surrounding *départements* in the western part of France. In contrast, in central-northern France the inter-annual variability is lower. As for winter wheat, slightly higher variability can be observed in several *départements* in the northern, western and southern parts of the country. The spatial distribution of inter-annual variability of winter wheat yields is more homogeneous.

Figs. S.3 and S.4 display a cross-correlation analysis of de-trended yields across all *départements* for grain maize and winter wheat, respectively. Homogeneous regions are chosen according to hierarchical clustering of de-trended crop yield time series. Five homogeneous regions for winter wheat are identified (Fig. 2): south-western—SWF_{WW}, south-eastern—SEF_{WW}, central-east—CEF_{WW}, central-west—CWF_{WW} and north-west—NWF_{WW}. As expected, low correlation values between yield anomalies can be observed when comparing wheat yields of northern and southern France and of the maritime and Mediterranean zones. This suggests that different climatic processes influence crop yield variability over these regions. This is not the case for central and eastern France, where higher spatial coherence between yield time series can be observed. In the case of grain maize, slightly different regions were obtained (Fig. 2): south—SF_{GM}, eastern—EF_{GM}, central—CF_{GM}, central-north—CNF_{GM} and north-west—NWF_{GM}. A clear difference can be observed between yields from southern France, with the most intensive maize production, and the rest of the country. The

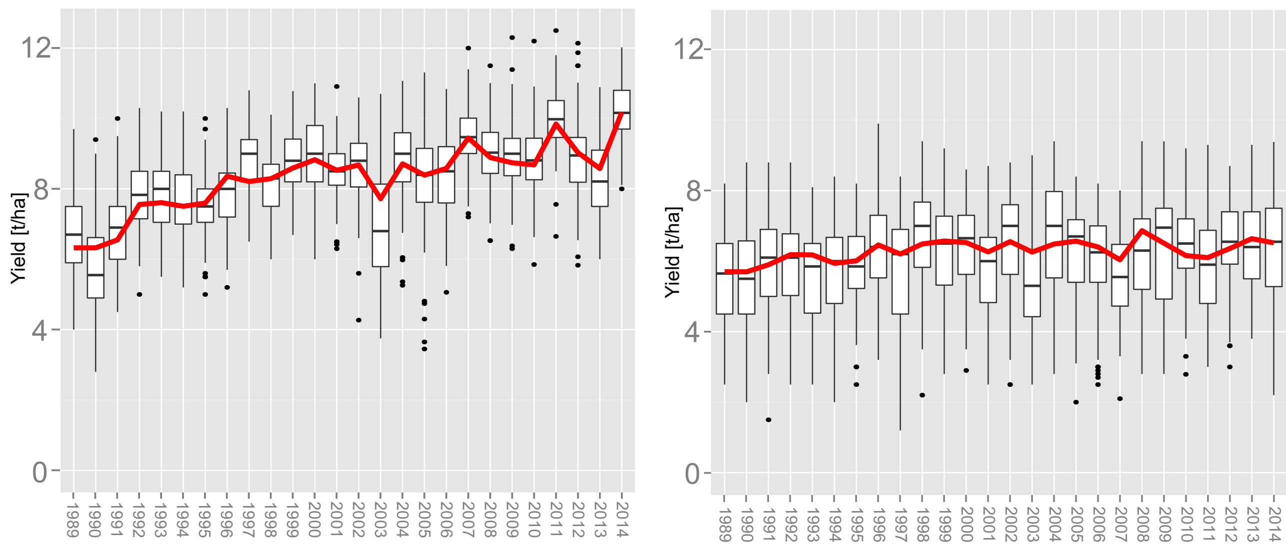

Fig. 1. Box-plots of grain maize (left) and winter wheat (right) yield time series over France for the *départements* where the inferred PLSR regression model has a prediction skill (see Fig. 6). The bold horizontal bars represent the median value of observed yield series, boxes represent the interquartile range, whiskers the range between 10th and 90th percentile and dots the extreme values. The red lines represent the median of the simulated values over the same *départements* using the inferred PLSR models.

decorrelation spatial scale is shorter for grain maize than for wheat. This could be explained by larger dependency of grain maize on water availability during the most sensitive growth stages, which is related to irrigation intensity as well as prevailing climate type. Regions that are irrigated report yields that are temporally more stable. For example, there is a clear distinction in percentage of irrigated grain maize between southern, central and northern parts of France (AGRESTE, 2015; Van der Velde et al., 2010). Intensively irrigated areas are located over the Mediterranean, the south-western, central and the eastern-most parts of France (Fig. S.5), where also more homogeneous grain maize yield time series can be observed. In addition, those regions are characterized by lower inter-annual maize yield variability (Fig. S.2).

4.2. Assessment of PLSR regression predictive skills

The ordinary bootstrap cross validation (see SOM) is used to determine the optimal number of latent variables as well as the significance of the derived PLSR models. Cross-validated Mean Square Error of Prediction ($MSEP_{boot}$) values indicate that two latent variables are sufficient in 70 *départements*, whereas in the remaining *départements* three latent variables lead to optimal PLSR model in the case of grain maize. As for winter wheat, two latent variables are sufficient for 68 *départements*, whereas three are chosen for the rest of the *départements*. However, the most significant reduction in $MSEP_{boot}$ is already achieved by using the first latent variable (Fig. S.6).

After the optimal number of latent variables is determined for each *département* separately, the significance of the optimal PLSR models is assessed (for more details the reader is referred to SOM). The analysis reveals that optimal PLSR models significantly explained a fraction of the yield inter-annual variability over the major part of France for both crops. However, few exceptions are identified, where the derived PLSR models do not appear to be significant in terms of capturing an intra-seasonal climate signal in the crop yield time series (Fig. 2). These *départements* are excluded from further analysis.

Although the derived PLSR models significantly explained a fraction of variability in the majority of maize yield time series, spatial differences can be observed in the amount of explained variability as well as in the prediction skill. The reason for this can be attributed to several factors. First, the quality of input data (meteorological as well as crop yield statistics) determines the quality of the inferred models. The density of the meteorological stations as well as the applied interpolation methods (from the station scale to the regional scale) can also influence the quality of the inferred model. Second, the magnitude of the de-trended variability plays an important role. As for grain maize, the proposed

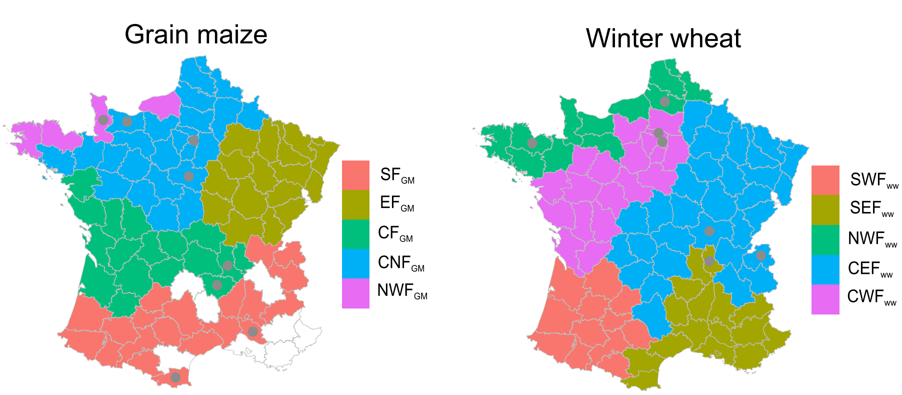

Fig. 2. Homogeneous regions obtained by using hierarchial clustering of de-trended grain maize (left) and winter wheat (right) yield time series. Dots identify *départements* where the inferred PLSR models are not significant in terms of capturing the intra-seasonal climate signal in crop yield time series. White regions denote areas where data was discarded.

statistical model can explain a large fraction of variability when the yield series are characterized by high inter-annual variability (Fig. S.2). On the contrary, the statistical model explains less variability in regions with lower crop yield inter-annual variability (mainly in the northern part of France).

As stated in the introduction, only de-trending is applied here to filter out the effect of changes in agro-management. However, intensity and changes in these practices could still affect the inter-annual crop yield variability (Jourdain et al., 2001; Branca et al., 2011). For instance, the effects of irrigation practices on grain maize yields can go beyond a simple trend. Fig. S.7 shows de-trended inter-annual variance of rainfed and irrigated maize yields for *départements* reporting both. Inter-annual yield variability of irrigated maize is generally lower than rainfed maize variability, with the exception of some *départements* in EF_{GM}, where variability is comparable. Irrigation mitigates the influence of weather (especially precipitation) by decreasing the leaf temperature during hot days, and therefore lowering the negative impact of hot spell on biomass production (e.g. Van der Velde et al., 2010). Here, the effect of intra-seasonal climate impact on crop yields is not investigated separately for irrigated and rainfed maize (due to the shorter time series available). Therefore, in *départements* with prevailing irrigated agriculture, the aggregated maize yield time series mainly reflect the sub-regional irrigated yield time series. This can be clearly observed in our analysis, where the rainfall signal during the summer months is not relevant (see Section 4.3.1). In addition, the amount of explained variability of derived PLSR models is lower for *départements* where the irrigated area is higher than the one of the rainfed area.

The derived PLSR models are able to explain more than 44% and 53% of crop yield inter-annual variability for grain maize and winter wheat, respectively (Figs. 3–5). For grain maize, the explained variability is highest over south-western, western and eastern-most France. A similar spatial pattern in model performance is found for winter wheat, with slightly higher explained variability over southern and western France. In general, explanatory power of the derived PLSR models is slightly higher for winter wheat. This could be related to the fact that irrigation, which can mask the climate signal, is generally not used for winter wheat. It is noteworthy that over regions with higher model skill, there is a clear distinction between explanatory variables importance. Over the regions with lower, but still significant, model skill, the distinction in the importance of explanatory variables becomes less pronounced. This can indicate that the PLSR models perform well over regions where a few key weather factors define the potential yield (e.g. precipitation and temperature during the sowing period and anthesis period). In order to obtain more stable PLSR regression models (with relatively low residual variability), longer yield time series are needed. Longer time series of crop yield would increase the sample of different seasonal weather patterns and would make the PLSR regression models more stable. In addition, different aggregation periods for the meteorological variables could be useful (e.g. development stages or sub-monthly time aggregates). However, accurate phenological observations would need to be available for the entire period of analysis.

4.3. The importance of intra-seasonal climate variability

Bootstrapped VIP scores and standardized regression coefficients of the PLSR regression models are used to assess the influence of intra-seasonal climate variability on crop yield anomalies. Standardized regression coefficients reflect both the magnitude and the direction of the climate influence. They represent how many standard deviations the crop yield changes for each standard deviation unit change in the explanatory variables. Values above 0 indicate positive contributions, whereas below-zero values indicate negative contributions. Figs. 3–5 show the resulting standardized regression maps for grain maize and winter wheat, respectively. The coefficients are only shown for those explanatory variables identified as important by the VIP measure. Standardized regression coefficients only provide information on how the anomaly of a particular explanatory variable is related to yield anomaly. Crop growth, however, is a cumulative process, integrating the influence of climate anomalies during the whole period of the growing season. In order to assess the relative importance of temperature, precipitation and global solar radiation for the whole growing period, the cumulative VIP maps are needed (Figs. 4–6). They show the relative contribution of these variables (integrated over the growing season) to the explained crop yield variability.

4.3.1. Grain maize

The sensitivity of crop growth to severe weather anomalies differs during the growing season. Grain maize is more sensitive to stress factors during the anthesis (flowering) period and this is well reflected in the regression map, showing the most influential meteorological variables mainly in July and August (Fig. 3). The grain-filling period in August is also highly sensitive to mainly temperature and precipitation. However, spatial differences can be observed in the timing of impact as well as in the meteorological variables having higher relevance. Generally, positive temperature and global radiation anomalies in July and August lead to lower maize yield, whereas positive rainfall anomalies in the same months have the opposite effect.

Maize yields in the NWF_{GM}, CNF_{GM} and CF_{GM} are sensitive to meteorological conditions in August, while weather conditions in July have a substantial impact on maize yields in SF_{GM} and EF_{GM}. Among other factors, discussed below, this may be associated with an earlier planting date over southern and eastern parts of the country (Fig. S.8). The cumulative importance over the growing season indicates that global radiation and precipitation play a more important role than air temperature over the western half of the country (Fig. 4). As a consequence of the mild maritime climate over those regions, temperature is not the most important explanatory variable. In several *départements* of the NWF_{GM} and CNF_{GM} regions, weather in the early season (May) seems to have important effects on maize yields (Fig. 3) and the associated standardized regression coefficients are positive. In the early season, during the vegetative stage of grain maize, the leaf area grows exponentially and the intercepted radiation increases. Changes in solar radiation in the early season therefore play a crucial role as they affect the capability of crop to intercept light later in the growing season, when the grain yield formation occurs. Temperature in September, with positive influence on yields, is identified as an important variable over several *départements* of NWF_{GM} and CNF_{GM}. Over EF_{GM}, yields are dominantly impacted by weather conditions in July; weather in August is less relevant in explaining inter-annual maize yield variability. This could be due to the fact, that the climatic water balance during August in these areas is closer to zero.

In the major grain maize producing areas of south-western France, all meteorological variables in July and August have an influence on grain maize yields, but the cumulative importance maps show that global radiation and air temperature are more important than the others (Fig. 4). Irrigation during dry years could suppress the precipitation signal, as can be observed in southern (SF_{GM}), eastern (EF_{GM}) and central France (CF_{GM}) of Fig. 3. Rainfall in July is not relevant in extensively irrigated regions, such as *départements* 59, 61, 66, 83, 84, 86 and 92 in SF_{GM}. In the *départements* of SF_{GM}, where rainfall in July is an influential variable, the regression coefficient magnitude is lower than in regions with prevailing rainfed agriculture. In CF_{GM}, July rainfall is an influential variable only over *départements* with almost no irrigated cropland. Rainfall in August

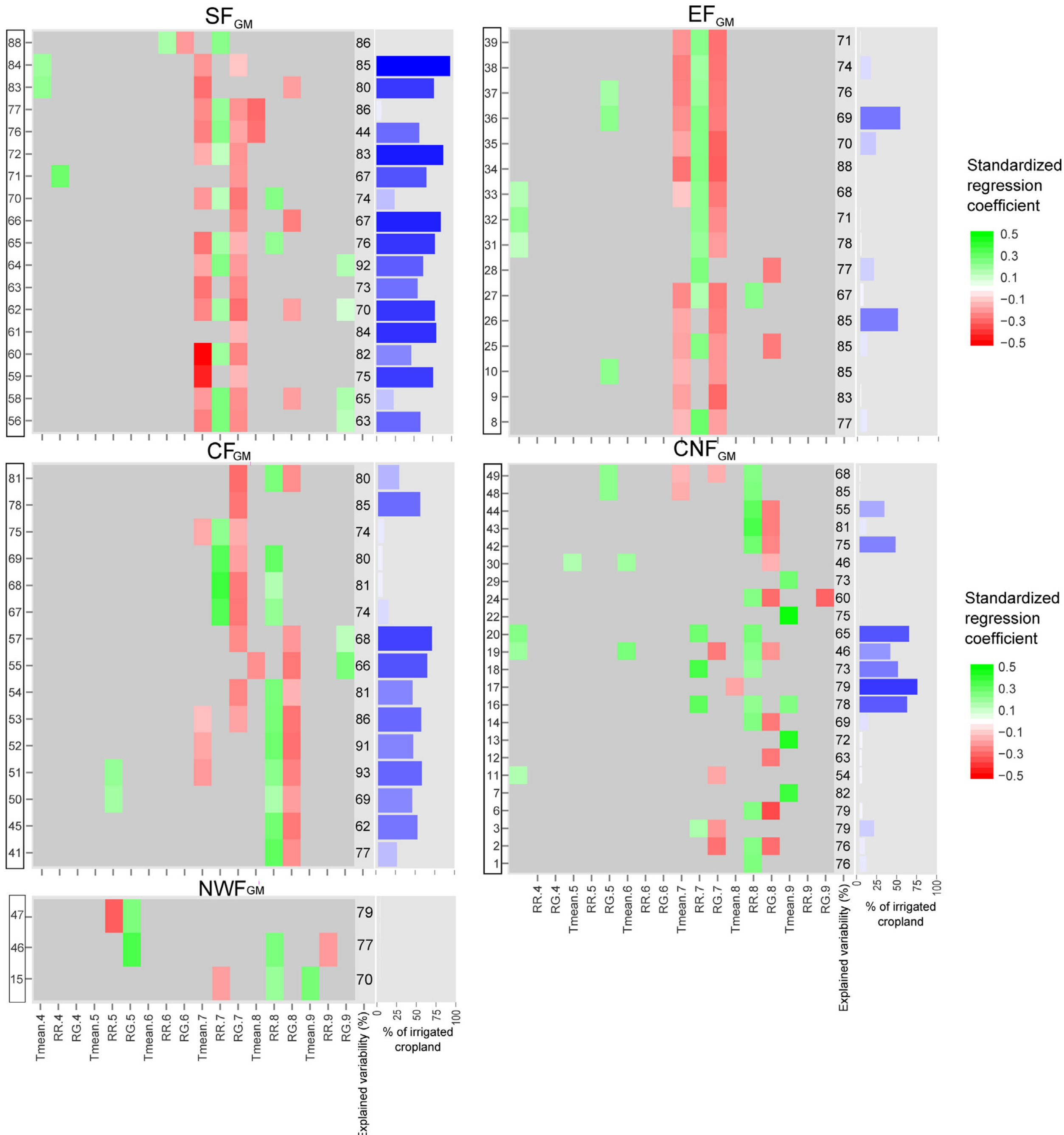

Fig. 3. Standardized regression coefficients of the explanatory meteorological variables for the identified homogeneous regions. The coefficients indicate how many standard deviations of grain maize yield will change for every standard deviation unit change of the explanatory variables. Shown are only the coefficients for the explanatory variables having a bootstrapped VIP value higher than 0.83 (i.e. in at least 90% of bootstrap samples). On the *x*-axis, labels are associated with: mean air temperature (T_{mean}), cumulative rainfall (RR) and cumulative global radiation (RG) for each month of the grain maize growing season (from April until September, months are denoted with numbers; e.g. T_{mean} 4 is the mean air temperature of April). The explained variability of the PLSR regression models is indicated on the *y*-axis on the right side of each plot. Horizontal blue bars on the right side of the plots represent the amount of irrigated maize cropland (Fig. S.5) for each *département*. Numbers on the left side of each plot represent the *département* identifier.

is of higher relevance in CF_{GM}, also over irrigated cropland. This might be the consequence of less intensive irrigation during the grain filling period of August and/or the fact that grain maize time series are aggregated by using the irrigated and the rainfed maize yield time series. In areas where grain maize is extensively irrigated, the impact of heat stress on leaf growth is probably lowered as irrigation decreases the ambient temperature and, thus, the heat stress experienced by the leaves. Hawkins et al. (2013) report that the relative importance of precipitation variability for maize yields has decreased since the 1960s, while simultaneously the heat stress variability has gained in importance. Although the crop yield time series in our analysis are shorter than the ones used by Hawkins et al. (2013), our results support their findings during July. This pattern is likely related to heat wave occurrence, since these events can significantly influence the monthly mean air temperature.

4.3.2. Winter wheat

The importance of meteorological variables on winter wheat is regionally much more variable and more disperse across the growing season than for grain maize (Fig. 5). In large areas of

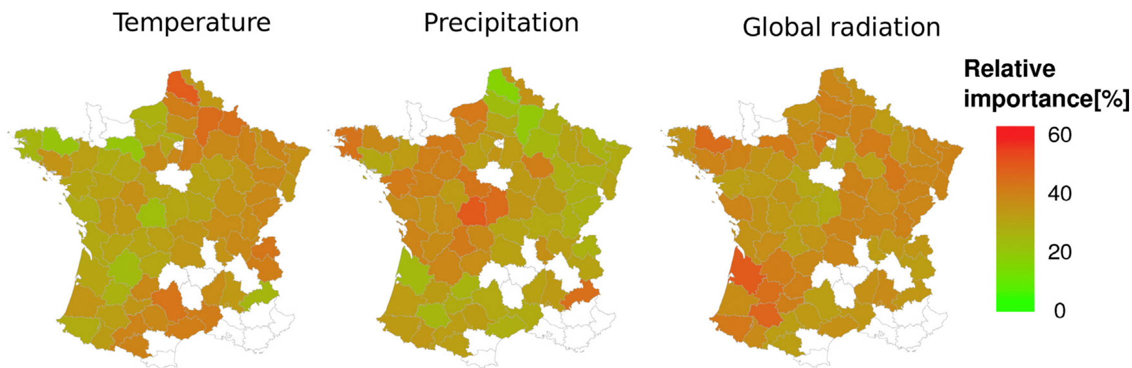

Fig. 4. Cumulative importance of temperature, precipitation and global solar radiation for the explained variability of grain maize yields, expressed in relative terms. The white colour denotes areas where either the PLSR regression model is not significant or regions where data was discarded.

the country, weather conditions at the beginning of the season (October/November/December) are important. Nevertheless, spatial differences can be observed in terms of timing of the largest impacts. This is related to timing of sowing that usually takes place between October and November in France (Fig. S.8, depending on the region and the weather conditions) as weather conditions affect the preparation of seedbed and subsequent sowing. Adverse conditions for planting are linked to over-wet conditions. Moreover, the timing of tillering, which is the most important process governing the canopy formation, depends on the sowing date. In general, positive global radiation anomalies in October increase yields in SWF_{WW}. In other regions of France, positive rainfall anomalies in November/December are related mainly to negative yield anomalies. December positive temperature anomalies in CEF_{WW} contribute to lower crop yields.

Weather conditions during the first two months of the year are important in CWF_{WW}, SWF_{WW} and NWF_{WW}; however, high spatial variability can be observed in terms of impacted regions. VIP values highlight the importance of global radiation and precipitation over the central and western-most parts of France in February (Fig. 5). Given that winter climate over these areas is rather mild and humid, excessive precipitation during the first two months of the year can lead to waterlogging during the dormancy period. On the contrary, positive global radiation anomalies are related with positive yield anomalies. It is worth to note that positive air temperature anomalies during winter months (identified as important ones) impact the yield negatively. This can be related to the reduced frost resistance (vernalization) caused by warmer weather.

Over CEF_{WW} and SEF_{WW}, positive temperature anomalies in March, when winter wheat usually enters the heading stage, lead to negative yield anomalies. Indeed, warmer temperatures lead to sooner breaking of winter dormancy and expose crops to spring freeze injuries. The probability of occurrence of these events is significantly higher over eastern France. Nevertheless, the direct effect of frost events cannot be fully captured here, since monthly averages are considered, while frost kill occurs over shorter time scales. Other factors need to be considered to better understand the significant impact of temperature in March. The sowing date has little influence on the heading or flowering date when the wheat plant is exposed to sufficiently low temperatures (to be completely vernalized by the time plant breaks dormancy in spring; Hu et al., 2005). However, warmer conditions during the vernalization period can result in fewer heads, and poorer yields, especially in varieties with longer vernalization requirements (Ortiz et al., 2012).

Development of winter wheat after heading is modified by temperature changes and photoperiod sensitivity. Depending on the region and variety sown, winter wheat is generally entering the flowering period in the second half of April or May, when it is highly sensitive to weather anomalies. During this period, inadequate soil moisture or heat stress may result in flower abortion and, thus, reduced yields (Barnabas et al., 2008). Weather conditions in April seem to affect yields over SWF_{WW}, SEF_{WW} and CWF_{WW}, while May seems to play a more important role over CEF_{WW}. High sensitivity to temperature in April is found for SWF_{WW}, with positive temperature anomalies contributing to reduced yield levels. Positive precipitation anomalies in April have positive influence on yield. Over the NWF_{WW}, CEF_{WW} and part of CEF_{WW}, weather conditions in June and especially July are important because precipitation shortage might reduce the yield formation during the grain filling period. Over CEF_{WW} and SWF_{WW}, July weather conditions are less relevant, since grain maturity is usually reached before.

The cumulative importance of meteorological variables indicates that temperature has a higher impact on inter-annual crop yield variability over south-western and eastern France (Fig. 6). Precipitation appears to be of greater relevance over western France and the Mediterranean region. In the latter one also global radiation provides a significant contribution.

5. Discussion

To understand and improve the low model quality especially over the central-northern and northern part of France (which are important for winter wheat production) further investigation is required. Efforts should focus on trying to detect and attribute changes in the variability of wheat yield caused by changes in agro-management practices. Indeed, long-term effects of changes in agro-management practices can influence not only the trend in crop yields, but also their inter-annual variability. In addition, longer time series are required to obtain more stable PLSR regression models over regions with low prediction skill. The inferred PLSR models can be potentially used for prediction purposes, especially over *départements* where low $MSEP_{boot}$ values have been estimated. This is clearly shown in Fig. 1. However for both crops, simulated yields tend to overestimate observed values during years with extremely negative observed yield anomalies (e.g. grain maize in 1990, 2003 and 2013; winter wheat in 2003 and 2007). The current statistical approach could be complemented with insights from other (statistical) modelling solutions as, for instance, the proposed statistical model does not explicitly take into account extreme events. Indeed, weather extremes (e.g. warm and cold spells, strong wind, hail, heavy precipitation leading to flooding or local water logging) act on a shorter time scale and may sometimes lead to complete crop failure (e.g. Hawkins et al., 2013). By using the monthly temporal resolution, those events are partially reflected in the monthly averages or sums. The identified periods of higher relevancy during the growing season can be used to plan adaptation measures in order to reduce the negative influence of extreme weather events.

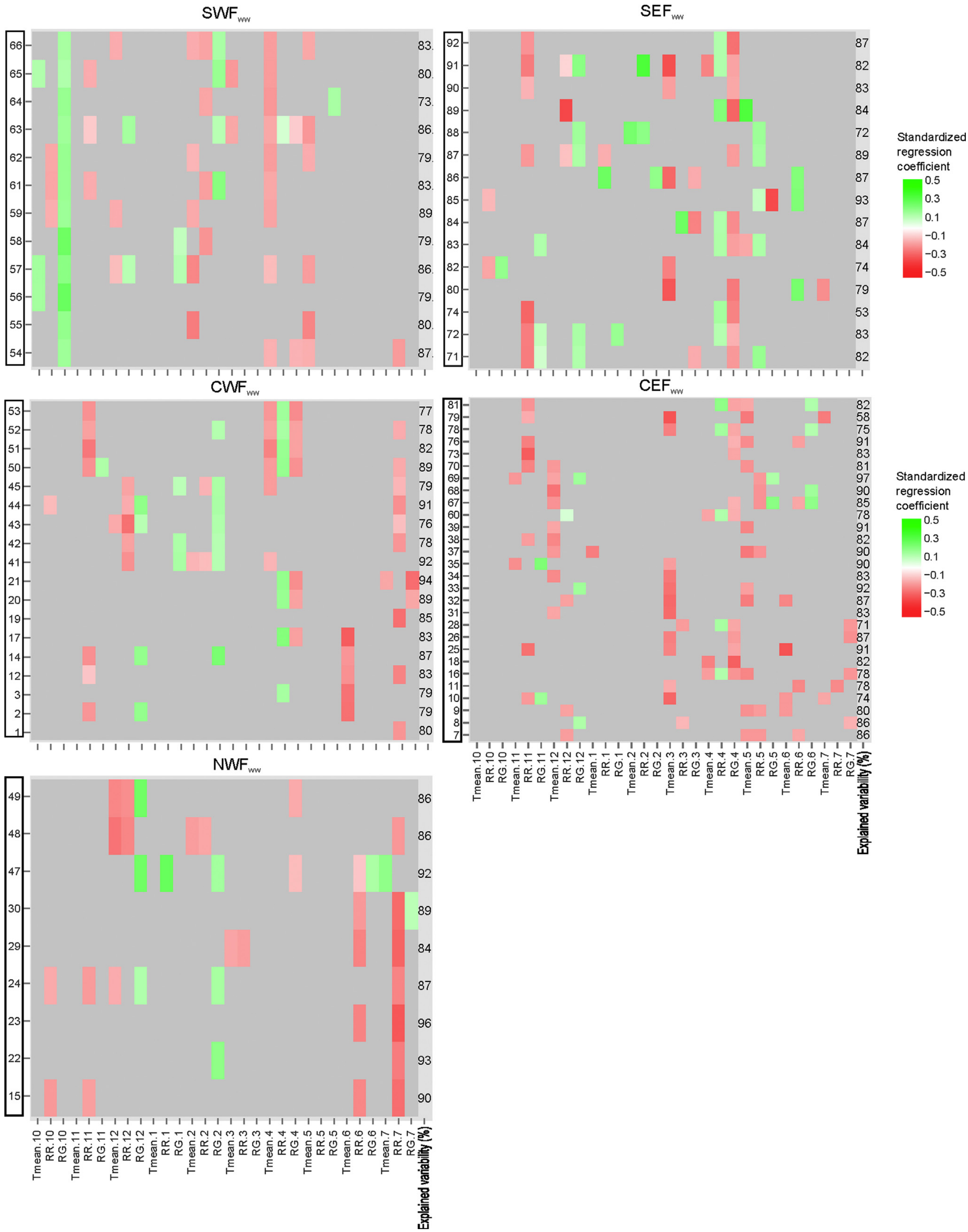

Fig. 5. As Fig. 3 but for winter wheat.

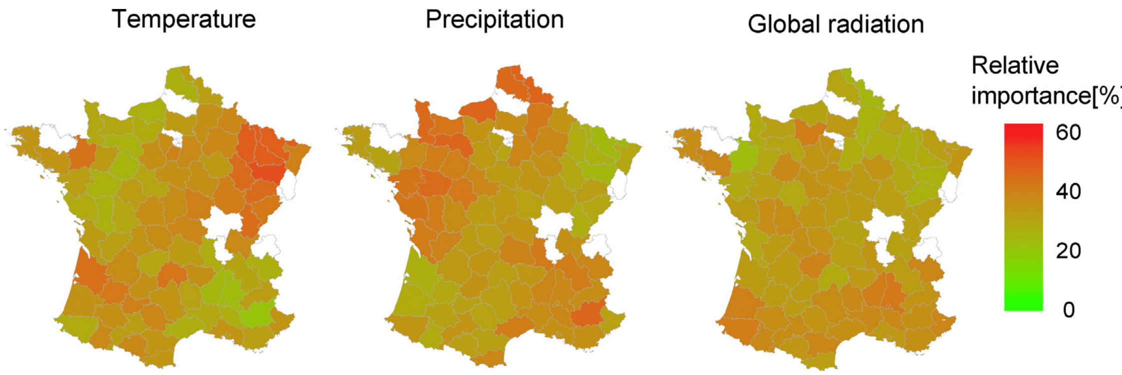

Fig. 6. As Fig. 4 but for winter wheat.

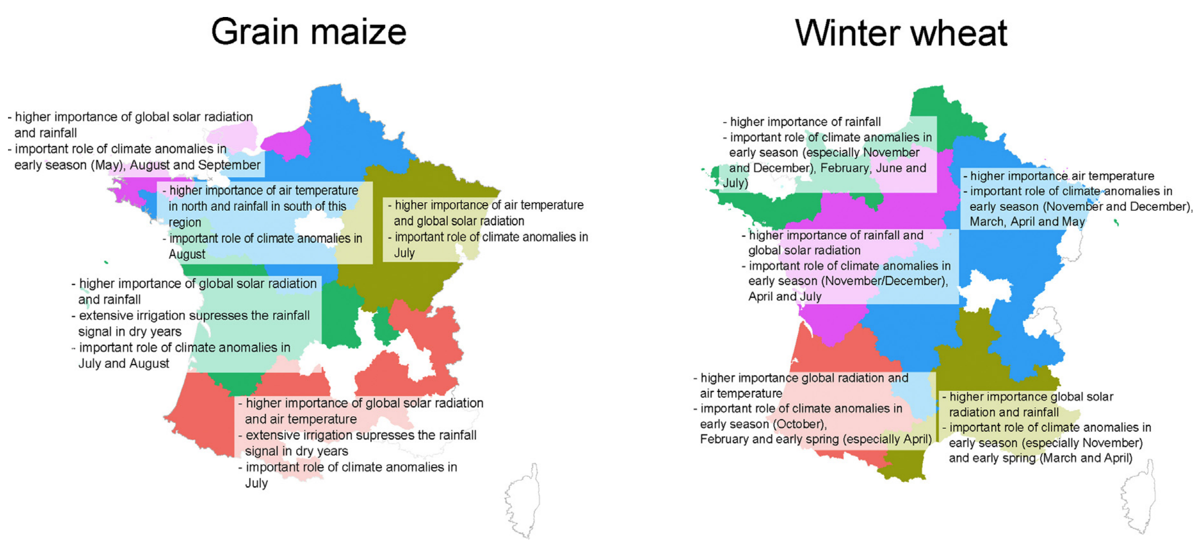

Fig. 7. Identified meteorological variables and their significant influence on inter-annual variability of winter wheat and grain maize yields.

The results of this study can contribute to the improvement of operational crop yield forecasting systems, such as the one used by the Joint Research Centre (MCYFS, 2015). This study can also contribute to the development/improvement of long-term (e.g. seasonal) crop yield forecast. The derived statistical models could be easily integrated in a seasonal crop yield forecasting system by using the probabilistic weather forecasts of the identified meteorological variables. The proposed approach can be also used for climate change studies, providing a robust framework to assess the impact of climate change, as projected by climate models, on crop yields. Finally, we envision that similar analysis of dynamic crop model results could reveal important information on the accuracy of the dominant processes included in mechanistic crop models to model yields at regional level. The proposed statistical framework could be applied for evaluating dynamic crop model simulations by building an emulator of dynamic crop model. A similar approach to analyse the impact of intra-seasonal climate variability on crop yields can also be extended to other regions of the world, provided sufficient length of crop yield and explanatory meteorological variables time series.

6. Conclusions

We have assessed the impact of intra-seasonal climate variability on regional crop yields by analyzing multi-annual time series of winter wheat and grain maize yields in France at *département* level. Partial least square regression has been used to identify the key intra-seasonal meteorological factors driving inter-annual crop yield variability. For both crops, apparent spatial differences have been observed in the timing of impact as well as in the meteorological variables having (air temperature, precipitation and solar radiation) the highest relevance. A graphical summary of the main findings is presented in Fig. 7.

In the case of grain maize, crop yields are mainly influenced by weather in July and August, even in irrigated regions. In large parts of southern, eastern and north-eastern France, summer temperature has been identified as the most important factor, with positive temperature anomalies leading to reduction in crop yields. Global radiation in the early growing season is the main factor over the westernmost part of France. Grain maize yields in eastern France are not strongly affected by climate conditions in August. The rainfall effect on crop yield is difficult to detect in irrigated regions. Indeed, global radiation and temperature are the dominant climatic variables affecting inter-annual maize yield variability over extensively irrigated areas of south-western and southern France. Less irrigated regions in south-western, western-most and central parts of France are more sensitive to rainfall and global radiation variations.

Winter wheat in most regions is more sensitive to weather conditions in late autumn/early winter (with the exception of some *départements* located in central and northern France), spring and in some cases early summer. The exact timing of the sensitivity, however, is highly variable across the country. Weather in autumn affects the preparation of seedbed and subsequent sowing, when adverse conditions for planting are linked to over-wet conditions. Global radiation and precipitation in February seem to have an important positive influence on wheat yields over the central and western-most parts of France. Weather conditions in April and May, coinciding with the flowering period, have a relevant impact over the whole country, except the northern and north-western *départements*, where flowering occurs later. Indeed, over those regions weather conditions during June and July have been recognized as the most influential. Overall, temperature has a substantial influence on winter wheat yields in the south-western and eastern parts of France, while rainfall is especially important over the northern and southern parts of the country. Finally, the significant role of radiation over the southern part of France can be deduced from the analysis.

Acknowledgements

We thank AGRESTE for maintaining and providing a long-term database of crop yields in France. We thank two anonymous referees for their comments and suggestions that helped to significantly improve this study.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agrformet.2015.10.004.

References

AGRESTE, 2015. La statistique l'évaluation et la prospective agricole. Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt, ⟨https://stats.agriculture.gouv.fr/disar/faces/report/tabDocBySource.jsp⟩ (accessed: 1 April, 2015).
Asseng, S., Ewert, F., Rosenzweig, C., et al., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Change, 10.1038.NCLIMATE1916.
Balkovic, J., Van der Velde, M., Schmid, E., Skalsky, R., Khabarov, N., Obersteiner, M., Sturmer, B., Xiong, W., 2013. Pan-European modelling with EPIC: implementation, up-scaling and regional crop yield validation. Agric. Syst. 120, 61–75.
Balkovič, J., van der Velde, M., Skalsky, R., Xiong, W., Folberth, C., Khabarov, N., Smirnov, A., 2014. Global wheat production potentials and management flexibility under the representative concentration pathways. Global Planet. Change 122, 107–121.
Barnabas, B., Jager, K., Feher, A., 2008. The effect of drought and heat stress on reproductive processes in cereals. Plant Cell Environ. 31, 11–38.
Bartholome, E., Belward, A.S., 2005. GLC2000: a new approach to global land cover mapping from Earth observation data. Int. J. Remote Sens. 26 (9), 1959–1977.
Biavetti, I., Karetsos, S., Ceglar, A., Toreti, A., Pananos, P., 2014. European meteorological data: contribution to research, development, and policy support. In: Proc. SPIE 9229, Second International Conference on Remote Sensing ad Geoinformation of the Environment, Cyprus, 10.1117.12.2066286.
Branca, G., NcCarthy, N., Lipper, L., Joleiole, M.C., 2011. Climate smart agriculture: a synthesis of empirical evidence of food security and mitigation benefits from improved cropland management. In: MiCCA Working Papers. FAO, Rome.
Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.X., Huard, F., 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crops Res. 119, 201–212.
Cantelaube, P., Terres, J.M., 2005. Seasonal weather forecasts for crop yield modelling in Europe. Tellus A 57, 476–487, http://dx.doi.org/10.1111/j.1600-0870.2005.00125.x.
Ceglar, A., Kajfež-Bogataj, L., 2012. Simulation of maize yield in current and changed climatic conditions: addressing modeling uncertainties and the importance of bias correction in climate model simulations. Eur. J. Agron. 37, 83–95.
Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74 (368), 829–836.
Cubasch, U., Wuebbles, D., Chen, D., Facchini, M.C., Frame, D., Mahowald, N., Winther, J.G., 2013. Introduction. In: Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom, New York, NY, USA, p. 119.
DISAR, 2014. La statistique. l'évaluation et la prospective agricole. Ministère de l'Agriculture, de l'Agroalimentaire et de la Forêt, ⟨https://stats.agriculture.gouv.fr/disar/#⟩ (accessed: 10 July, 2014).
ECMWF, 2013. ECMWF annual report 2013. In: ECMWF Technical Reports, ⟨http://www.ecmwf.int/sites/default/files/Annual_Report_2013.pdf⟩ (accessed: 14 October, 2014).
Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York, NY.
Folberth, C., Gaiser, T., Abbaspour, K.C., Schulin, K., Yang, H., 2012. Regionalization of a large-scale crop growth model for sub-Saharan Africa: model setup, evaluation, and estimation of maize yields. Agric. Ecosyst. Environ. 51, 21–33.
Garthwaite, P.H., 1994. An interpretation of partial least squares. J. Am. Stat. Assoc. 89, 122–127.
Hawkins, E., Fricker, T.E., Challinor, A., Ferro, C.A.T., Ho, C.K., Osborne, T.M., 2013. Increasing influence of heat stress on French Maize yields from 1996s to 2030s. Global Change Biol. 19, 937–947.
Hu, C.S., Delgado, J.A., Zhang, X.Y., Ma, L., 2005. Assessment of groundwater use by wheat (*Triticum aestivum* L.) in the Luancheng Xian Region and potential implications for water conservation in the North-western North China Plain. J. Soil Water Conserv. 60, 80–88.
Indahl, U., Liland, K., Naes, T., 2009. Canonical partial least squares—a unified PLS approach to classification and regression problems. J. Chemom. 23, 495–504.
Izenman, A.J., 2008. Modern Multivariate Statistical Techniques. Springer, New York, pp. 733.
Jolliffe, T., 2002. Principal Component Analysis. Springer-Verlag, New York, pp. 488.
Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., Wavresky, P., 2010. Les types de climats en France, une construction spatiale. Cybergeo: Eur. J. Geogr. [En ligne], ⟨http://cybergeo.revues.org/23155⟩ (accessed: 18 December 2014). (Cartographie, Imagerie, SIG, document 501).
Jourdain, D., Scopel, E., Affholder, F., 2001. The impact of conservation tillage on the productivity and stability of maize cropping systems: a case study in Western Mexico. In: CIMMYT Economics Working Paper 01–02. CIMMYT, Mexico, DF.
Laux, P., Jacket, G., Tingem, R., Kunstmann, H., 2010. Impact of climate change on agricultural productivity under rainfed conditions in Cameroon—a method to improve attainable crop yields by planting date adaptations. Agric. For. Meteorol. 150, 1258–1271.
Liu, J., 2009. A GIS-based tool for modeling large scale crop water relations. Environ. Modell. Softw. 24, 411–422.
Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.M., Falcon, W.P., Naylor, R.L., 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319, 607–610, 10.1126-science.1152339.
Lobell, D.B., Burke, M.B., 2010. On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol. 105 (11), 1443–1452.
Lobell, D.B., 2010. Crop responses to climate: time-series models. In: Lobell, D., Burke, M. (Eds.), Climate Change and Food Security: Adapting Agriculture to a Warmer World. Springer, New York, NY, pp. 177–192.
Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field Crop. Res. 143, 56–64.
Mehmood, T., Liland-Hovde, K., Snipen, L., Saebo, S., 2012. A review of variable selection methods in partial least square regression. Chemometr. Intell. Lab. Syst. 118, 62–69.
MCYFS, 2015. MARS Crop Yield Forecasting System, ⟨http://m.arswiki.jrc.ec.europa.eu/agri4castwiki/index.php/Main_Page⟩ (accessed: 10 June 2015).
Michel, L., Makowski, D., 2013. Comparison of statistical models for analyzing wheat yield time series. PLoS ONE 8 (10), e78615.
Murtagh, F., 1985. Multidimensional Clustering Algorithms. COMPSTAT Lectures 4, Physica-Verlag, Wuerzburg.
Olesen, J.E., Jensen, T., Petersen, J., 2000. Sensitivity of field-scale winter wheat production in Denmark to climate variability and climate change. Clim. Res. 15, 221–238.
Ortiz, B., Tapley, M., van Santen, E., 2012. Planting date and variety selection effects on wheat yield. Alabama Cooperative Extension System, ANR-1442.
Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Koppen–Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644.
Porter, J.R., Semenov, M.A., 2005. Crop response to climatic variation. Philos. Trans. R. Soc. Lond., B: Biol. Sci. 360 (1436), 2012–2035.
Ritchie, S.W., Hanway, J.J., Benson, G.O., 1993. How a Corn Plant Develops. Cooperative extension service, Ames, IA, pp. 25.
Rosipal, R., Kramer, N., 2005. Overview and recent advances in partial least squares. In: Saunders, C., Grobelnik, M., Gunn, S., Shawe-Taylor, J. (Eds.), Subspace, Latent Structure and Feature Selection. Springer, New York, pp. 34–51 (LNCS 3940).
Schlenker, W., Roberts, M.J., 2009. Nonlinear temperature effects indicate severe damages to U.S. Crop yields under climate change. Proc. Natl. Acad. Sci. U.S.A. 106 (37), 15594–15598.
Tebaldi, C., Lobell, D.B., 2008. Towards probabilistic projection of climate change impacts on global crop yields. Geophys. Res. Lett. 35, L08705 (10.1029.2008GL033423).
Van der Velde, M., Wriedt, G., Bouraoui, F., 2010. Estimating irrigation water use and effects on maize yield during the 2003 heatwave in France. Agric. Ecosyst. Environ. 135, 90–97.
Van der Velde, M., Tubiello, F.N., Vrieling, A., Bouraoui, F., 2012. Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. Clim. Change 113, 751–765.
Wold, S., Sjostrom, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. Syst. 58, 109–130.