Agricultural and Forest Meteorology 209–210 (2015) 1–10

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Agrometeorological analysis and prediction of wheat yield at the departmental level in France

David Gouache [a,*], Anne-Sophie Bouchon [a], Elodie Jouanneau [b], Xavier Le Bris [c]

[a] ARVALIS – Institut du Végétal, Station Expérimentale, 91720 Boigneville, France
[b] ARVALIS – Institut du Végétal, Chemin des Bissonnets, 14980 Rots, France
[c] ARVALIS – Institut du Végétal, Station Expérimentale de La Jaillière, 44370 La Chapelle Saint Sauveur, France

ARTICLE INFO

Article history:
Received 24 July 2014
Received in revised form 20 January 2015
Accepted 27 April 2015
Available online 16 May 2015

Keywords:
Wheat
Yield
Model
Cross-validation
Prediction
Agrometeorology

ABSTRACT

Predicting annual crop yields is of interest for many agricultural applications. We present a prediction scheme at the departmental level, circa 100 km by 100 km, of winter wheat yields in France, applied for 23 departments, using official yield statistics from 1986 to 2010. Each model is a linear combination of 5–7 variables, selected from an initial pool of over 250 candidates. Candidate variables were generated using a phenological model and a crop water balance model, applied to a representative cropping situation for the department. Variable selection was carried out with forward stepwise regression methods. The variable selection process was cross-validated, so as to select robust variables. Model prediction performance was also evaluated by cross-validation. Satisfactory models were created for 20 departments, with root mean square error of prediction ranging from 0.25 t/ha to 0.39 t/ha. During use, whole season weather data is not available: this is complemented by frequential calculation over the past 20 years of historical weather data. We assessed the impact of time of prediction on model error by hindcasting yields for all 25 years of the dataset. We estimate that predictions can start 20 days after heading on average. We analysed predictive performance in an independent dataset and propose recommendations for use of these models outside their training dataset. The models give new insight as to the climatic factors that are key in determining yield in France.

1. Introduction

One of the main objectives of agrometeorological studies is to estimate crop production as a function of weather, soil and crop management parameters (Hoogenboom, 2000). Such studies have applications spanning very large temporal and spatial scales, from within-season adjustment of crop management practices to national and international forecasting of crop yields for political and economical stakeholders in famine early warning systems or commodity trading, from evaluation of strategic evolutions of crop management at the grower's scale to assessments of climate change impacts (Hoogenboom, 2000; Gommes et al., 2010). Predicting crop yields at international, national or sub-national levels has received attention from researchers for many years, especially for important staple crops, such as wheat (Hill et al., 1979; Feyerherm and Paulsen, 1981; Walker, 1989; Travaso, 1990; Hébel et al., 1993; Supit, 1997; Landau et al., 2000; Qian et al., 2009; Gobin 2010; de Wit et al., 2010). Although different approaches have been mobilized to develop crop yield prediction schemes, these can be broadly categorized into two approaches: regression models and process-based models. The first family attempts to link synthetic weather variables calculated over all or part of the growing season to regional yield variations, whereas the second makes use of crop simulation models that describe crop–soil–climate interactions with interwoven equations representing the response of crop functions to weather. Despite regular controversy between the two methods (Landau et al., 1998, 1999; Jamieson et al., 1999; Semenov and Shewry, 2011), both types of approaches have shown useful and convincing results. Common across both types of approaches are the difficulties in choosing relevant scale, as yields at aggregated scales depend on yields obtained across a wide variety of fields differing in soil properties, local weather, and crop management (Bakker et al., 2005; Challinor et al., 2003). Background tendencies for yield increases, or decreases, through improved management and cultivars, or reduced investments, respectively, also need to be accounted for, and this is mostly accomplished by detrending observed yields, although the choice of the trend model in and of itself is not straightforward (Supit, 1997; Lobell and Field,

Abbreviations: RMSE, root mean square error; RMSEP, root mean square error of prediction.

* Corresponding author. Tel.: +33 1 64 99 24 49; fax: +33 1 64 99 30 39.
E-mail address: d.gouache@arvalisinstitutduvegetal.fr (D. Gouache).

http://dx.doi.org/10.1016/j.agrformet.2015.04.027
0168-1923/© 2015 Elsevier B.V. All rights reserved.

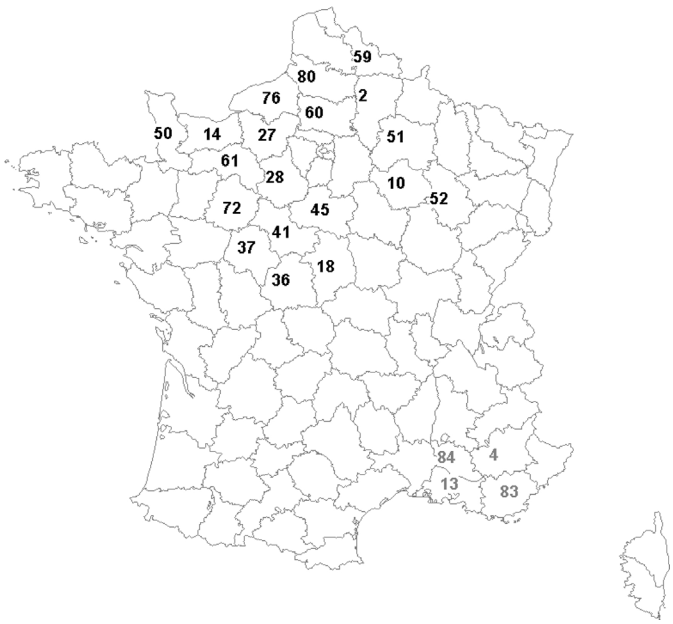

Fig. 1. Numbers of French departments for which winter wheat (in black: bread wheat; in grey: durum wheat) yield prediction models were established and tested.

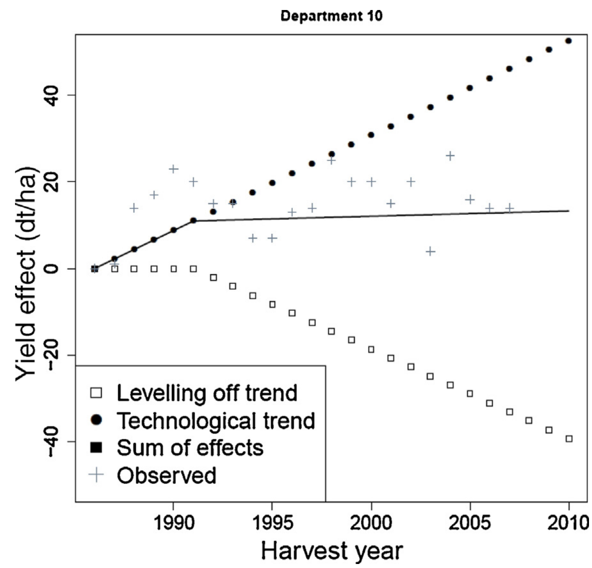

Fig. 2. Two types of trends potentially included in yield prediction models, example of department 10.

2007; Peltonen-Sainio et al., 2010). Recently, hybrid approaches have been developed where synthetic agrometeorological variables, including crop water balance indices, are calculated over relevant simulated phenological periods (Landau et al., 2000; Qian et al., 2009).

In the work presented here, we aim at developing a tool for regional prediction of winter wheat yields in different French regions, directly tailored to the needs and operational environment of the applied agricultural research and development institute Arvalis so as to maximise its effectiveness for the institute (Walker, 1989): best possible predictive ability, lead-time, simplicity of use and simplicity of communication. These constraints included readability of model results for dissemination purposes and ease in obtaining soil, weather and crop characteristics for calculations. These constraints lead to the use of a hybrid approach, constructing multiple regression models based on agrometeorological indicators calculated over phenological phases. Model performance based on root mean square error (RMSE) and root mean square error of prediction (RMSEP) assessed by cross-validation was measured so as to evaluate potential usefulness of the models. Lead-time for prediction was also assessed, as this is an important factor in determining the operational uses of the system (Hansen, 2002; Bannayan et al., 2003; Lawless and Semenov, 2005). The resulting models were also tested during four subsequent years (2011–2014), and this highlighted the need to provide tools to end-users that indicate if, and how far, outside of its range the model is being used. Results are discussed in terms of model performance, interpretation, and potential uses.

2. Materials and methods

2.1. Yield data

Winter wheat yield and cultivated areas used are the survey data collected by the French Ministry of Agriculture (AGRESTE: http://agreste.agriculture.gouv.fr) for harvest years 1986–2010. Data was obtained for 22 French departments (Fig. 1). Departments studied are mostly regrouped in the main bread wheat (*Triticum aestivum*) production area of the northern half of France, and cover 43% of its total bread wheat growing area. The last 4 departments correspond to durum wheat (*Triticum durum*) growing areas. Cultivated area was not used in the analyses, but was used to verify that trends in yield could not be explained by trends in increased cultivated area, that would imply that less fertile soils have recently been mobilized for wheat cultivation: no obvious link was ever found between the two. Cultivated area was also used to aggregate departmental yields to higher scales.

2.2. Agrometeorological variables and other predictors

Over 250 potential explanatory variables were calculated for each department and year. Two variables were included to account for technological trends in wheat yields. One is harvest year (HY). This variable creates a positive linear trend. The second trend variable is harvest year above breakpoint (HYB), calculated as max(0,HY-threshold). If integrated in the model, this variable allows for a bilinear trend (Fig. 2). Indeed, Brisson et al. (2010) have shown that yield trends in France have shown a significant levelling off since the mid 1990s. We used the method proposed in Brisson et al. (2010) to calculate the value of 'threshold' for each department, as the year for which the most significant levelling off of yields appears.

All other variables included in the analyses are agrometeorological variables reported to influence wheat growth in France (Gate, 1995; Lecomte, 2005), and worldwide (Landau et al., 2000; Qian et al., 2009; Gobin 2010; Peltonen-Sainio et al., 2010; Bouffier, 2014). These include: heat stress (Gate et al., 2010), drought stress, waterlogging, excess rain leading to poor disease control, freezing, and balance of climatic offer, such as photothermal quotient (Fischer, 1985). Given the number of studies and the breadth of agrometeorogical conditions in wheat growing environments, there is no unique set of variables to represent the totality of environmental factors affecting wheat production. We thus devised our set of variables based essentially on reported factors affecting French wheat yields (Gate, 1995; Lecomte, 2005), and so as to cover all phonological phases, from sowing until harvest. Variables were calculated as means, sums or number of days above or below thresholds for given phenological phases of wheat growth. The complete list of variables is available in Supplementary material S1. The phenological model used for calculations is a modified version of the AFRC-Wheat1 model (Weir et al., 1984; Gate, 1995; Gouache et al., 2012; see the third reference for a detailed

Table 1
Description of representative situations used to calculate agrometeorological variables for yield prediction models.

Department	Weather station	Longitude	Latitude	Soil water reserve (mm)	Sowing date	Variety
14	L'oudon	−0.01	48.99	120	20/10	Vivant
61	Alencon	0.10	48.43	100	1/11	Vivant
50	St. Hilaire Du Harcouet	−1.10	48.57	150	20/10	Vivant
72	Le Mans	0.20	47.95	150	20/10	Vivant
27	Breteuil Sur Iton	0.92	48.84	160	5/10	Bermude
76	Neuville Les Dieppe	1.10	49.94	190	20/10	Bermude
2	Saint Quentin	3.20	49.82	180	15/10	Bermude
60	Beauvais–Tille	2.12	49.46	180	15/10	Bermude
80	Oisemont	1.77	49.96	150	14/10	Bermude
59	Lille–Lesquin	3.10	50.58	180	15/10	Bermude
10	Troyes–Barberey-St.-Sulpice	4.02	48.33	180	10/10	Apache
51	Fagnieres	4.42	48.95	180	25/9	Caphorn
52	Chaumont Ville	5.15	48.10	80	25/9	Caphorn
18	Bourges	2.37	47.07	100	15/10	Apache
36	Chateauroux-Deols	1.72	46.86	100	15/10	Soissons
28	Chartres-Champhol	1.52	48.47	150	15/10	Premio
37	Tours–Parcay-Meslay	0.75	47.44	180	10/10	Premio
41	Orleans-Bricy	1.78	47.98	120	15/10	Altigo
45	Orleans-Bricy	1.78	47.98	120	15/10	Altigo
4	Greoux	5.86	43.75	150	18/10	Claudio
13	Marignane-Marseille	5.23	43.44	150	25/10	Pescadou
83	Le Luc En Provence	6.38	43.38	150	25/10	Pescadou
84	Orange	4.85	44.14	150	25/10	Pescadou

description). The crop water balance model used to calculate drought stress and waterlogging indices (see Supplementary material S1 for details) is that presented by Choisnel (1985), and based on exactly the same principles as the reference water balance model described in Allen et al. (1998), parametrized with specific crop coefficients that vary with phenological stages.

These calculations imply that representative situations for each department are defined. For a given department, one to three representative situations were identified by the Arvalis regional wheat crop expert. A representative situation is defined by a weather station, a maximum soil water reserve, a sowing date, and a cultivar earliness. For departments where more than one representative situations were identified, this was narrowed down to one based on performance of the first set of models established (phase 1, see below and Fig 2). Characteristics of representative situations are presented in Table 1: the majority belong to the MétéoFrance national weather service, with stations for departments 04 and 51 belonging to Arvalis and INRA-AgroClim networks, respectively. It should be noted that the choice of weather stations was constrained by the need for a continuous 25 year weather series and for each station to provide real time daily updates to be usable in an operational setting. No such weather station could be identified in the available Arvalis database in department 62. In department 27, the 2nd criterion was not met. The weather stations provide daily measurements of minimum and maximum temperature, rainfall, radiation, and reference evapotranspiration.

2.3. Variable selection

The overall workflow of the modelling work is presented in Fig. 3. In phase 1, starting with the initial set of variables, a forward step-wise selection procedure, based on ordinary least squares, implemented in Delphi, allowed for sets of candidate models to be established. The objective of this phase was solely to identify a set of circa 20 variables that seemed to be the most promising candidates for a systematic evaluation in phase 2. This required the implementation of multiple forward stepwise selection procedures. This was done as follows: after carrying out a first forward selection, we analysed the coefficients of the variables in the model. When the sign of the coefficient was inconsistent with known effects in wheat (for example, a positive effect of waterlogging stress on yield is deemed inconsistent), a new instance of forward selection was implemented, after eliminating the inconsistent variable. This was repeated multiple times, until the models obtained displayed RMSE values that increased. This phase led to a number of candidate models, generally comprising 4–8 explanatory variables. This allowed for the initial 250 variables, covering the breadth of wheat phenological stages and possible agroclimatic parameters, to be narrowed down to approximately 20 variables per department (Supplementary material S2). The second phase then consisted in selecting a final model, based on the robustness of the variable selection procedure, as evaluated by the approach proposed in the MMIX package of R (Morfin and Makowski, 2009; Prost et al., 2008; Casagrande et al., 2009). Briefly, this consisted in bootstrapping and/or cross-validating the stepwise procedure. The bootstrapping procedure consisted in generating 500 random samples with replacement from the original dataset. The cross-validation approach consists in generating 25 samples from the data, removing each year one by one, i.e. leave-one-out cross-validation. Respectively, 500 and 25 models are obtained. For each variable included in this step, their frequency of selection in the models, and the distribution of their coefficients, can be calculated. Variables with the highest frequency of selection, with coefficients presenting a stable sign and a small spread, were considered most robust. No systematic threshold was applied in this phase, but generally, variables chosen for the final models were selected through the stepwise procedures at over 90% frequency, and had consistent positive or negative coefficients in circa 90% of situations. For departments 14, 50, 61, 37, and 72, both bootstrapping and cross-validation were carried out. As results were very similar (data not shown), only cross-validation was used for other departments.

2.4. Model evaluation: cross-validation

Once the final model had been chosen, a second leave-one-out cross-validation was applied to evaluate predictive ability of the models by calculation of RMSEP (Wallach and Goffinet, 1987). Briefly, this consisted in iteratively excluding each year from the dataset, refitting model coefficients (without any more variable selection), and predicting the year that was excluded.

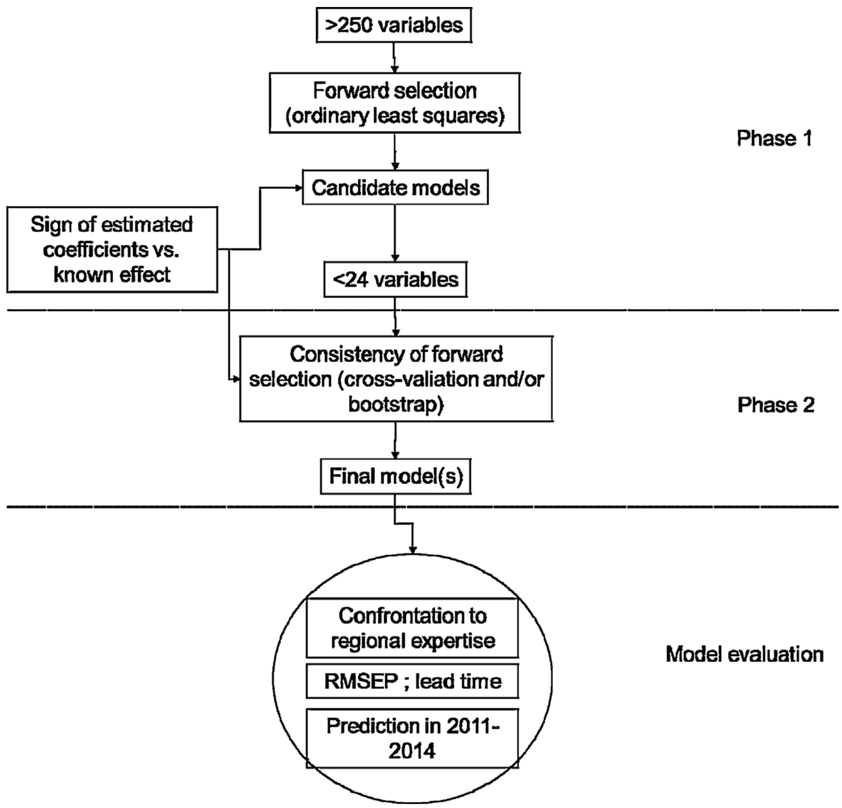

Fig. 3. Workflow of the modelling work.

2.5. Model evaluation: new data

Yield data from harvest years 2011–2014 was used to provide a second estimation of RMSEP. To analyse further, the prediction errors in these years that had not been used in the modelling process, we compared the distribution of the explanatory variables in the dataset used to fit the models to the values taken in the new dataset. To do so, for each department, we compared the range of explanatory variables in the new data set to that in the training data set. We also carried out, for each department, principal component analysis (PCA) on the explanatory variables with years as individuals, in the training dataset, and projected the years of the new dataset in this space. We then calculated the distance between the 4 new years and each year of the training dataset.

2.6. Hindcasts and model lead time

All models created here were established on observed weather data. However, in an operational context, observed weather data for the whole of the crop growing cycle is unavailable. Despite this, users of the predictions may want to obtain data from the model before all weather is observed. The method used to do so consists in completing the observed weather series with each of the observed weather series from the previous 20 years. Thus, the model proposes a distribution of future wheat yield, integrating both observed weather up to the date of prediction and historical weather data: the model prediction is the median of this distribution. We evaluated operational prediction capability by carrying out hindcasts (Challinor et al., 2005). We simulated situations in which observed weather data is absent, and prediction is carried out using the approach presented above. For each year and department, we simulated an absence of observed weather starting at heading date of the representative situation, then heading date + 10 days, and so forth until heading date + 60 days. We were thus able to calculate RMSE as a function of prediction date for each department. This RMSE was compared to 2 naïve predictors: the mean department yield or use of a prediction based on a positive yearly trend.

2.7. Other modelling strategies tested

For departments 14, 50, 61, 37, and 72, predictive ability of models obtained from stepwise selection were compared to model mixing approaches (Morfin and Makowski, 2009; Prost et al., 2008; Casagrande et al., 2009). Briefly, this consists of fitting every possible additive multiple regression model from the candidate variables. A prediction is then provided by each model. Finally, these predictions are averaged into one value, using different weighing criteria, for exemple Aikaike Information Criteria, of each model.

For these same departments, we also tested fitting a model to more than one department at a time. This consisted in grouping the 5 departments into 2 groups for which yields exhibited a very strong linear correlation with a slope not significantly different from 1: 14, 50 and 61 on one side, and 37 and 72 on the other. Explanatory variables remained calculated specifically for each department, but the variable selection and model fitting procedure was carried out on the grouped dataset. An additive departmental effect was included.

In department 02, a model was also created for a different representative situation (sowing date 10/11), given the fact that roughly 30% of cultivated wheat area is sown late, following late-harvest preceding crops such as sugar beet and potato.

Table 2
Predictive ability (root mean square error, RMSE, and root mean square error of prediction, obtained from cross-validation, RMSEP, both in dt/ha) of the selected departmental models.

Department	RMSE	RMSEP
14	2.37	3.26
61	2.33	3.37
50	2.83	3.86
72	2.56	3.42
2	1.88	2.81
60	2.52	3.47
80	2.13	2.87
59	2.12	3.15
10	3.07	4.09
51	2.52	3.94
52	1.89	2.65
18	2.39	3.34
36	1.89	2.48
28	2.83	3.7
37	2.36	3.12
41	2.70	3.34
45	1.70	2.7
4	1.85	2.43
13	2.24	3.29
84	2.65	3.55

3. Results

3.1. Model performance: RMSE and RMSEP

RMSEP of models ranged from 2.5 to 4.1 dt/ha (Table 2). This should be compared with mean bread wheat yields ranging from 60 to 80 dt/ha, mean durum wheat yields of 30–35 dt/ha, and observed range (maximum to minimum) of these yields spanning 20–40 dt/ha. Without a model, predicting yield based on average yield of the past 25 years would provide RMSEP values that are generally twice as high. RMSEP was generally 1 dt/ha above RMSE, indicating that models are not overfitted to the data.

No models are presented for 3 departments: 27, 76, 83. For department 27, this is due to the absence of a weather station with 25 years data and real time availability of data. In department 83, durum wheat growing areas are small, and closely located to those in department 84. Model results from the latter were judged to be sufficient to represent yield for both departments. Models created for department 76 were not judged satisfactory. The best model for this department obtained an RMSEP of 4 dt/ha, not above the highest values for the other departments, but, more importantly, displayed individual years with errors reaching 0.8 dt/ha, which was judged too high.

3.2. Model description: selected predictor variables

Almost all departments include a positive technological trend variable, with the exception of 13. Wheat cropping in this department is very particular, as it is carried out in rotations with flooded rice in the Camargue delta. Cultivated area and grower investment into their wheat crops in this sector are highly dependant on the relative economical opportunity presented by each crop, varying strongly year to year without any obvious trend. In 9 of 21 departments, a second, negative trend was also included in the model, more or less counterbalancing the technological trend (Fig. 2).

The significance and importance of the agrometeorological variables included in the model are presented in Table 3. In all but 4 departments, excess temperature during grain filling was selected as a variable influencing yields negatively. In these departments, heat stress during grain filling explained from 4 to over 30% of variance in yields, with an average of 13.6%. In all the northwestern, departments studied, excess rainfall during grain filling was also included in the models, explaining an average of 15.7% of variance in yields. Other key effects on winter wheat yields identified are winter waterlogging, and growing conditions during stem extension. Winter waterlogging appeared significantly in all but 6 departments, in which it explained an average of 17.5% of variance in yields. Departments in which this factor explained the highest proportions of variance in yields correspond to those in which soil types susceptible to waterlogging occupy a significant proportion of cultivated wheat areas. This limiting factor also appears strongly in the Mediterranean departments studied, where heavy autumn and winter rains are typical. Favourable growth conditions during stem extension appeared through three types of variables, representing excess temperatures, crop water balance, or ratio of radiation to temperature. Crop water balance during stem extension appeared in all but 6 departments, explaining a proportion of variance ranging from 1% to 35%. Crop water balance plays a less important role in departments on the northwest oceanic border, and becomes a more important predictor of yields in more continental and southern departments.

3.3. Model performance: hindcasts

Fig. 4 represents average prediction error (RMSE) as a function of the time of prediction, expressed in days from heading. In many departments, there is an appreciable decrease in error between 20 and 30 days after heading. RMSE values can be compared to the 2 naïve predictors of yield: average department yield or use of a linear trend: on average across departments, these naïve RMSE values are 6.7 dt/ha and 5.9 dt/ha, respectively. Most importantly, the RMSE value with the model is always inferior to the naïve predictor (see Supplementary material S3), even at heading. This implies that the model provides a substantial reduction in uncertainty of regional yield rapidly after heading. Departments from the Centre region are those for which prediction is most skillfull at heading. This is obviously linked to a much stronger weight of variables determined earlier on in the crop growing cycle in those models.

3.4. Model performance: comparison with a model-mixing approach

Table 4 presents predictions obtained through stepwise selection to those obtained from model mixing approaches, both according to Akaike Information Criterion (see Morfin and Makowski, 2009). Model mixing approaches provided less robust predictions, yielding RMSEPs 1.6 dt/ha to 2.3 dt/ha above those from stepwise selection. Results obtained using the other stepwise and model mixing approaches available in the MMIX package lead to equivalent conclusions.

3.5. Model performance: 2011–2014

We tested the model in harvest years 2011–2014. These 4 years were not involved in the model development process. Error levels in these 4 years proved to be much higher than those observed in the leave-one-out cross-validation. Indeed, RMSEP in this data set were, for years 2011–2014: 6.4, 11.7, 7.4, 9.2 dt/ha, respectively. The distribution in these errors is presented Fig. 5. The overall distribution of absolute errors shows a median value of 5.8 dt/ha, a third quartile value of 9.3 dt/ha, and an extreme value of 24.3 dt/ha. This median value remains acceptable, but is nonetheless close to double the value obtained through leave-one-out cross-validation. Most importantly, a significant proportion of situations displayed extreme errors, above 9 dt/ha. These extreme prediction errors displayed some important characteristics. First of all, year 2012 errors were skewed towards negative values, i.e. overprediction. Moreover, among the 27 cases with extreme errors (over 9 dt/ha), 10 of

Table 3
Signification, sign of effect on yield, and weight (percentage of variance explained) of different explanatory variables on departmental yields. Columns with more than one value indicate that more than one explanatory variable of the considered type is included in the model.

Department	Variables												
	Harvest year	Year above threshold	Cumulative temperature tillering	Winter waterlogging	Excess temperature during stem extension	Favorable crop water balance stem extension	Favorable radiation/ temperature ratio stem extension	Excess rain during stem extension	Excess temperature during grain filling	Excess rain during grain filling	Favorable crop water balance grain filling	Favorable crop water balance stem extension & grain filling	Favorable radiation/ temperature ratio grain filling
	Effect on yield												
	+	–	+	–	–	+	+	–	–	–	+	+	+
	Weight in selected models (% variance explained)												
14	26	17			6; 6				19	22; 5			
61	27	17		4					9; 4	38			
50	40	27		4					8	22			
72	48			12	9	7; 2			22				
2	24	9		19		6			21	19	2		
60	36					16	3		22	9	14		
80	34	15				1	10		14	27			
59	17			26	20	4			8	26			
10	33	28		33					6; 5	5			
51	8			22		32			18; 7	14			
52	17				6	27		19	21	10			
18	24	14			16	3		31					12
36	50	29		7		10				3	2		
28	24		3	6	26				19		23		
37	44	16		16	14	4			6				
41	41			13		19	19; 3	5					
45	13			16		35	18		15; 4				
4	34				8	3			29			26	
13				25	29	2	9		28; 6				
84	12			20	19	29			15; 6				

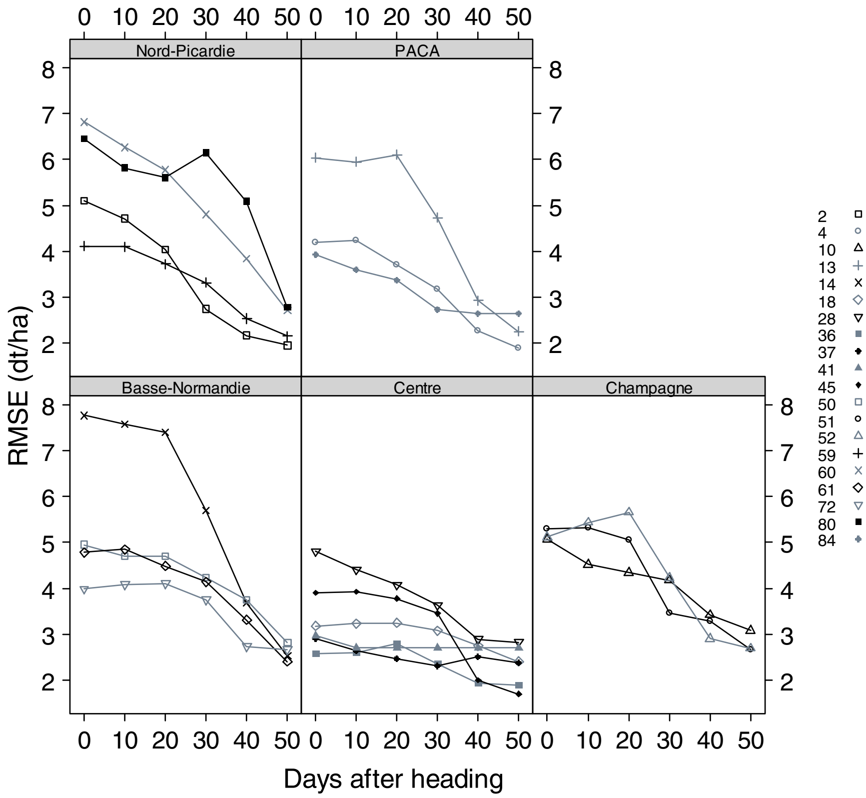

Fig. 4. Evolution of root mean square error (RMSE, in dt/ha) of predictions as a function of time of prediction (in number of days from heading) for all departments.

these correspond to 2012. Harvest year 2012 was characterized by very intense frost episodes that led to significant destructions in a great deal of fields across central, northern and eastern France (Deudon and Deswarte, 2012; Deswarte et al., 2012). This led to overturning of significant proportions of fields, that were re-sown to wheat, and/or abandonment of wheat fields. In these cases, the model is clearly out of its range of use. A second pattern is that certain departments were not affected by extreme errors. Departments 04, 37, and 59 had RMSE across these 4 years that was similar to that observed in cross-validation. Departments 50, 61 and 84 had RMSE levels in 2011–2014 not exceeding 6 dt/ha. Finally, we examined the distribution of all the individual climatic variables in 2011–2014, as compared to that observed in the training dataset. We noted that in 55% of the cases with extreme errors, at least one of the climatic variables in the model was outside of the range it had taken in the training data. We also compared the distribution of climate variables in 2011–2014 compared to the training set from a multivariate standpoint, by calculating the average of the distance of each year to all the years in the training set. Here, we noted an overall correlation between distance and absolute error and distance of 0.22. However, taken on a year by year basis, we observed correlation of 0.68, 0.32, 0.31, and 0.61 for 2011–2014, respectively (Fig. 5).

Table 4
Comparison of predictive performance (root mean square error of prediction, RMSEP, in dt/ha) of models obtained using stepwise selection of variables or model mixing approaches.

Department	Stepwise	Model mixing
14	3.09	4.73
61	3.72	6.03
50	2.67	4.78
72	2.66	4.54
37	3.89	5.63

3.6. *Model performance: aggregation at higher scales*

Models established at a higher scale of aggregation for departments 37 and 72, and 14, 50, and 61, respectively, presented RMSEPs higher than those provided by departmental models by approximately 1 dt/ha.

The predictions of each departmental model were aggregated together, using cultivated area of each department, to evaluate their capacity to predict the average yield over the whole bread wheat growing area they cover. Aggregating all the models leads to a very low prediction error (Fig. 6): 1.1 dt/ha RMSE and 1.3 dt/ha RMSEP.

4. Discussion and conclusions

4.1. *Advantages and disadvantages of the approach: model performance*

Predictive ability of the established models, as judged by RMSEP values, can be deemed high. RMSEP values are close to 5% of the mean yield values, and 10% of the observed range of variations. These values are equivalent to those reported by Qian et al. (2009) for Canada and Supit and van der Goot (1999) for France, who respectively used a similar hybrid approach and a crop model based approach, and appear better than those proposed by Balkovič et al. (2013) with a crop model based approach. RMSEP results are better than those provided by Hébel et al. (1993). This can be attributed to the fact that in their study, agrometeorological variables did not account for variation in phenological stages, but integrated variables according to calendar dates. Clearly, explicitly integrating the interaction between weather and phenological stages, whether through the use of a complete crop simulation model or through a hybrid approach, appears necessary to achieve the highest predictive ability. Our approach provides very

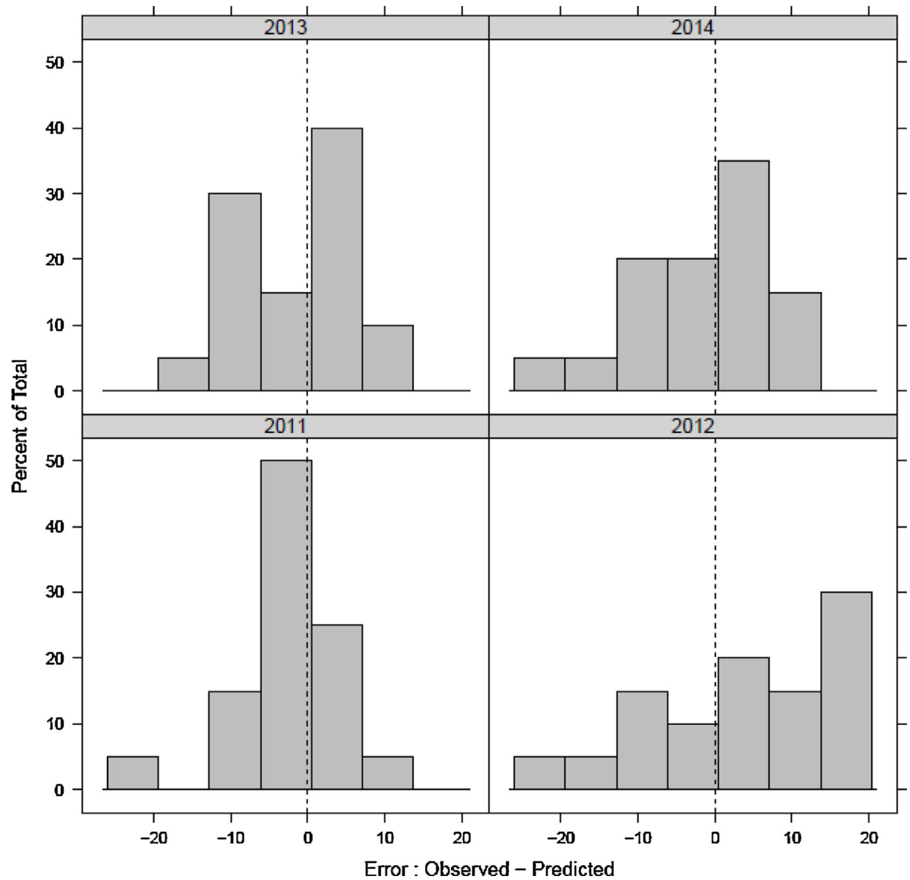

Fig. 5. Distribution of errors in harvest years 2011–2014 that were not used in the model fitting process.

similar results to those obtained by Qian et al. (2009) using a similar hybrid modelling, despite very big differences in agronomic context (Canada vs. France, spring vs. winter wheat), confirming the interest of such approaches.

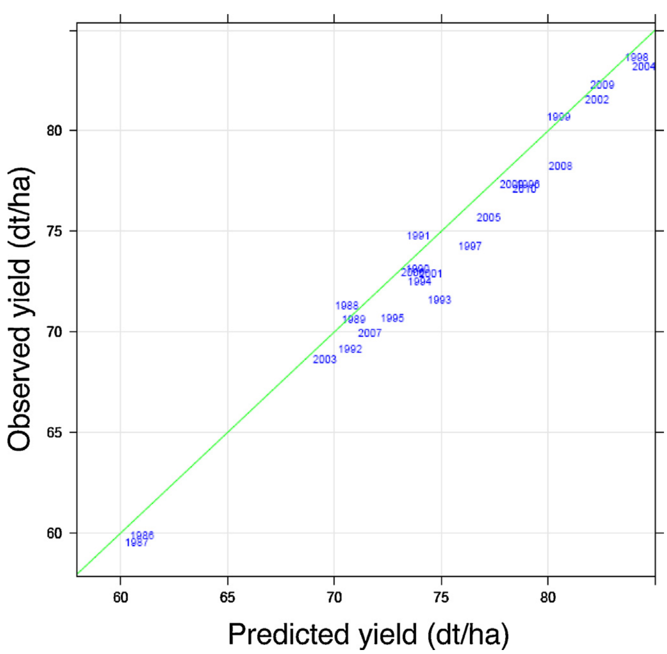

Fig. 6. Observed yield as a function of predicted yield (dt/ha) for the whole of the area covered by the bread wheat prediction models (43% of French cultivated area). 1.1 dt/ha RMSE, 1.3 dt/ha RMSEP, R^2 of 0.98. Data points are situated in the center of each tag.

Prediction errors were quite high for years 2011–2014. This raises a number of questions and operational recommendations for use of the models. The first question raised is the evaluation of the predictive ability of models. Indeed, leave-one-out cross-validation is a very much recommended and used validation method (Wallach et al., 2006). In our case, we observe a significant discrepancy between this method and introduction of an independent dataset. Both approaches have their advantages. Leave-one-out cross-validation allows to test the model on 25 years of climate data. However, this approach does not generate evaluation on a purely independent dataset since information from those years was used in the choice of variables used. An independent data set such as 2011–2014 solves this issue. However, it gives access to a much smaller sample size. At least, these results demonstrate that even rigorous cross-validation may provide a somewhat optimistic view on predictive ability of models. From a more operational standpoint, these results provide insight on how to account for the fact that models should not be used too far outside the range of conditions in which they were created. We propose 3 recommendations. First of all, the values of each variable should be systematically compared to the range in the training set: this is simple to carry out and consists in a strong warning that the model may provide poor predictions. Secondly, a more multivariate approach can be taken. Again, this is straightforward, and can help to identify situation in which individual variables may not be out of the training range, but their combination may be a particular case not well represented in the training data. Finally, any model result requires expertise before a prediction is given (Lazar and Genovese, 2004): the case of 2012 exemplifies this. Destruction of significant wheat surfaces during the winter is an obvious situation where the models will be out of range. In the same sense, recent years have all seen record extreme events (Deudon and Deswarte, 2011, 2012, 2013; Deswarte and Gouache, 2011; Gouache and Deswarte, 2012; Deswarte et al., 2012), and local, field based expertise and measurements can help in identifying situations in which the model will be outside its range. This should be all the more important under climate change (Iizumi et al., 2013; Fischer et al., 2014).

It can be argued that in situations outside the range of model development conditions, using a crop simulation model might be more robust, comparatively to the statistical approach presented here, in which variable selection limits the number of mechanisms accounted for, and in which the linear, additive approach clearly can not represent non-linearities and interactions known to exist (Jamieson et al., 1999; Semenov and Shewry, 2011). However, both types of approaches show relatively equivalent results, due to the difficulties in parametrizing crop models, the way processes are represented in them (Passioura, 1996; Wallach et al., 2006), and the complexity of obtaining and aggregating input data at the proper scale (Bakker et al., 2005). In our case, the objective of ease of communication and utilization (i.e. input data collection) is more easily served by a model with few explanatory variables than a fully integrated crop model: these same reasons have even stimulated work on building metamodels from crop models (Brooks et al., 2001). Improvements in predictive ability of the approach proposed here might be obtained by explicitely integrating non-linear reponses (Landau et al., 2000) or using more elaborate estimators for model parameters (Hébel et al., 1993).

At the highest level of aggregation, predictive ability is improved. This result is consistent with those presented by Supit and van der Goot (1999) and Bakker et al. (2005). Such a result can probably be attributed to compensation between errors at lower scales of aggregation.

4.2. Interpretation of climatic factors influencing wheat yields in France

The results presented allow for the some of the main climatic factors influencing yield at the departmental scale to be identified and weighed against each other. This allowed fruitful interactions with regional experts, which was key in obtaining the highest quality models possible.

Heat stress during grain filling appeared as a key factor influencing yields across a great majority of departments studied. Increasing temperatures during grain filling have indeed been found to negatively influence yields, both in France and worldwide (Brisson et al., 2010; Gate et al., 2010; Kristensen et al., 2011; Lobell et al., 2005). More importantly, it has been shown that part of the levelling off of wheat yields may be ascribed to an increase in adverse thermal conditions for grain fill (Brisson et al., 2010; Gate et al., 2010). Crop water balance during stem extension was also seen to be an important cause of yield variations in many departments in France. Increasing drought stress, like heat stress, has been shown to be a factor in the stagnation of French wheat yields (Brisson et al., 2010; Gate et al., 2010). Only half of the models obtained include a term specifying a trend for levelling off of wheat yields. When they do, the levelling off trend does not completely counterbalance the positive technological trend. This means that the agroclimatic variables included in these models explain all or part of the levelling off of wheat yields observed in France, reinforcing the hypothesis that a significant part of this worrisome trend is due to climate change. The variables identified in this work could also be specifically tested to see if they can be linked to stagnating wheat yields, as was done for drought stress and heat stress during grain filling in Brisson et al. (2010) and Gate et al. (2010).

Three other key phenomenon influencing wheat yields were put forward by this approach. The balance between incoming radiation and temperature during stem extension influences yields in a majority of departments studied. This is consistent with Fischer's (1985) findings and has already been reported as an important parameter for yield in France (Gate, 1995). Excess rain during grain filling was also found to be regularly detrimental to yield in the northernmost regions of France. This was for example the case in 2007 (Gate, 2007), when very frequent rains caused sever yield losses, due mostly to high disease pressure, but also to waterlogging that reduced photosynthetic capacity during grain filling. Negative effects of rainfall have also been reported to negatively influence yield of durum wheat in southeastern France around anthesis, due to increased disease pressure (Braun et al., 2010), although we did not identify such a variable here. Waterlogging was also identified as an important variable influencing yield potential early on. Interestingly, its relative weight in the models was more important in departments presenting significant proportions of waterlogging-prone soils. It was also a very important factor in the Mediterranean departments studied. This is probably in part due to the particularities of the Mediterranean climate, in which very heavy autumn rainfall is frequent. It may also be explained by the fact that drought stress during stem extension and grain filling in the region is very frequent and strong (Braun et al., 2010), in which case winter waterlogging that will have compromised efficient rooting can be particularly detrimental. Overall, our approach led to the confirmation of the importance of drought and heat stress in influencing French yields, but also to the addition of other factors, namely early waterlogging, radiation/temperature balance, and excess terminal rainfall.

4.3. Perspectives for use

The models developed in this study show important potential for use, thanks to high predictive ability, robustness, and relative ease in interpretation. They can be used to communicate to growers and advisers on the key limiting factors of yields in the region, and can provide an opportunity to discuss the importance of preserving soil structure in waterlogging-prone soils or to increase crop protection inputs in heavy disease pressure situations (Gate, 2007). They also show that final yield is most influenced by the grain filling phase in northwestern France, whereas in more central departments, earlier phases have a stronger influence. This has implications both for crop management and for breeding. From a crop management standpoint, this means that in northwestern France, factors influencing yield are determined once all crop husbandry has been carried out (nitrogen and fungicides especially), whereas in more central and southern regions, the yield potential of the year is significantly influenced by factors occurring at a time when crop management can still be adjusted. In other words, it appears that in central and southern regions, modifications in crop management could be advised based on the models developed here. Advising modifications in crop management obviously requires a very high degree of reliability. In certain extreme cases where the model is outside its initial range, this may be risky. Use of the model in 2011 showed that without proper expertise, use of the model may lead to wrong decisions. To counter this, we have developped some simple approaches to identify situations in which the model may be out of its range. From a breeding standpoint, it could be advised that raising yields in northernmost France will require a more important focus on improving the resistance of grain filling to heat stress, whereas in other regions a balanced approach between a number of abiotic stresses may be required.

Given their lead time, the models proposed may also be an important tool for grain collectors and marketers to better adjust their positions on commodity markets in response to predicted production potential. Finally, integrating seasonal weather forecasts with these models could increase lead time. This could add value both to the models developed here and to seasonal forecasts in France (Hansen, 2005).

Acknowledgements

The authors thank Arvalis regional staff for their expertise in defining representative situations and critical analysis of the agronomical significance of the identified agrometeorological variables: Anne-Sophie Hervillard-Courty, Anne Plovie, Thierry Denis, Alexis Decarrier, Gérard Briffaux, Jean-Charles Deswarte, Michel Bonnefoy, Stéphane Jézéquel. Authors are indebted to Antoine Bray and Philippe Gate for their initial impulse to this work in the Basse-Normandie region. Joan Schmidt is acknowledged for her guidance in English.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.agrformet.2015.04.027.

Bibliography

Allen, R.G., et al., 1998. Crop Evapotranspiration (Guidelines for Computing Crop Water Requirements). FAO, Rome http://www.fao.org/docrep/X0490E/X0490E00.htm
Bakker, M.M., et al., 2005. Variability in regional wheat yields as a function of climate, soil and economic variables: assessing the risk of confounding. Agric. Ecosyst. Environ. 110 (3–4), 195–209.
Balkovič, J., et al., 2013. Pan-European crop modelling with EPIC: implementation, up-scaling and regional crop yield validation. Agric. Syst. 120, 61–75, http://dx.doi.org/10.1016/j.agsy.2013.05.008.
Bannayan, M., et al., 2003. Application of the CERES-Wheat model for within-season prediction of winter wheat yield in the United Kingdom. Agron. J. 95 (1), 114.
Bouffier, B., 2014. Genetic and ecophysiological dissection of tolerance to drought and heat stress in bread wheat: from environmental characterization to QTL detection. PhD Thesis. Université Blaise Pascal, 318 pp.
Braun, P., et al., 2010. Rapport technique. Axe 1-potentialités et milieux. In: Rapport Technique Final du Projet FUI GARICC, 41 pp.
Brisson, N., et al., 2010. Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. Field Crops Res. 119, 201–212.
Brooks, R.J., et al., 2001. Simplifying Sirius: sensitivity analysis and development of a meta-model for wheat yield prediction. Eur. J. Agron. 14, 43–60.
Casagrande, M., et al., 2009. Factors limiting the grain protein content of organic winter wheat in south-eastern france: a mixed-model approach. Agron. Sustain. Dev. 29 (4), 565–574.
Challinor, A.J., et al., 2003. Toward a combined seasonal weather and crop productivity forecasting system: determination of the working spatial scale. J. Appl. Meteorol. 42, 175–192.
Challinor, A.J., et al., 2005. Probabilistic simulations of crop yield over western India using the DEMETER seasonal hindcast ensembles. Tellus 57A (3), 498–512.
Choisnel, E., 1985. Un modèle agrométéorologique opérationnel de bilan hydrique utilisant des données climatiques. In: Confnference of International CIID. Les Besoins en Eau des Cultures, Paris: éd. INRA, p. 115–132.
Deswarte, J.C., Gouache, D., 2011. Bilan 2010/2011. 2010/2011: les cultures ont su s'adapter au climat exceptionnel. Perspect. Agric. 383, 33–36.
Deswarte, J.C., 2012. Chaleur, gel et pluies : les céréales gagnent la course d'obstacles. Perspect. Agric. 394, 43–45.
Deswarte, J.C., Ducheyron, P., Gouache, D., 2012. Céréales d'hiver: un gel tardif modifie la capacité de résistance des plantes face au froid. Perspect. Agric. 395, 53–55.
Deudon, O., Deswarte, J.C., 2011. Bilan 2010/2011. Un climat atypique qui a contrarié la production de grains. Perspect. Agric. 381, 38–40.
Deudon, O., Deswarte, J.C., 2012. Bilan 2010/2011. Bilan météorologique 2011–2012: beaucoup d'inquiétudes suscitées par le grand froid et la pluie. Perspect. Agric. 392, 32–34.
Deudon, O., Deswarte, J.C., 2013. Bilan climatique: une campagne 2012/2013 avec de l'eau à profusion. Perspect. Agric. 403, 8–10.
Feyerherm, A.M., Paulsen, G.M., 1981. Development of a wheat yield prediction model. Agron. J. 73, 277–282.
Fischer, R.A., 1985. Number of kernels in wheat crops and the influence of solar radiation and temperature. J. Agric. Sci. 105 (2), 447–461.
Fischer, E.M., et al., 2014. Models agree on forced response pattern of precipitation and temperature extremes. Geophys. Res. Lett., 41, http://dx.doi.org/10.1002/2014GL062018.
Gate, P., 2007. Bilan écophysiologique du blé tendre. L'année de toutes les maladies. Perspect. Agric. 339, 52–55.
Gate, P., 1995. Ecophysiologie du Blé: de la Plante à la Culture. Lavoisier Editeur, Paris, France, 424 pp.
Gate, P., Brisson, N., Gouache, D., 2010. Les causes du plafonnement du rendement du blé en France: d'abord une origine climatique. Comptes-Rendus de l'Académie de l'Agriculture de France. 96 (3), 17–23.
Gobin, A., 2010. Modelling climate impacts on crop yields in Belgium. Clim. Res. 44 (1), 55–68.
Gommes, R., et al., 2010. Agrometeorological forecasting. In: WMO (Ed.), WMO/CAgM Guide to Agricultural Meteorological Practices. (Chapter 5).
Gouache, D., Le Bris, X., Bogard, M., Deudon, O., Pagé, C., Gate, P., 2012. Evaluating agronomic adaptation options to increasing heat stress under climate change during wheat grain filling in France. Eur. J. Agron. 39, 62–70.
Gouache, D., Deswarte, J.C., 2012. Bilan écophysiologique 2010/2011. La récolte sauvée par la capacité de compensation des blés tendres. Perspect. Agric. 385, 36–38.
Hansen, J.W., 2002. Realizing the potential benefits of climate prediction to agriculture: issues, approaches, challenges. Agric. Syst. 74 (3), 309–330.
Hansen, J.W., 2005. Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. Philos. Trans. R. Soc. B: Biol. Sci. 360 (1463), 2037–2047.
Hébel, P., et al., 1993. Shrinkage estimators applied to prediction of French winter wheat yields. Biometrics 49, 281–293.
Hill, J.D., et al., 1979. LACIE – an application of meteorology for United States and foreign wheat assessment. J. Appl. Meteorol. 19, 22–34.
Hoogenboom, G., 2000. Contribution of agrometeorology to the simulation of crop production and its applications. Agric. For. Meteorol. 103, 137–157.
Iizumi, et al., 2013. Historical changes in global yields: major cereal and legume crops from 1982 to 2006. Global Ecol. Biogeogr., http://dx.doi.org/10.1111/geb.12120.
Jamieson, P.D., et al., 1999. Comments on "Testing winter wheat simulation models' predictions against observed UK grain yields" by S. Landau et al. (1998). Agric. For. Meteorol. 96 (1), 157–161.
Kristensen, K., et al., 2011. Winter wheat yield response to climate variability in Denmark. J. Agric. Sci. 149, 33–47.
Landau, S., et al., 2000. A parsimonious, multiple-regression model of wheat yield response to environment. Agric. For. Meteorol. 101 (2–3), 151–166.
Landau, S., et al., 1998. Testing winter wheat simulation models' predictions against observed UK grain yields. Agric. For. Meteorol. 89, 85–99.
Landau, S., et al., 1999. Response to Comments on 'Testing winter wheat simulation models' predictions against observed UK grain yields by Landau et al.' by Jamieson et al. Agric. For. Meteorol. 96, 163–164.
Lawless, C., Semenov, M., 2005. Assessing lead-time for predicting wheat growth using a crop simulation model. Agric. For. Meteorol. 135 (1–4), 302–313.
Lazar, C., Genovese, G. (Eds.), 2004. The MARS Crop Yield Forecasting System. Methodology of the Mars Crop Yield Forecasting System. European Comission Joint Research Centre, p. 98.
Lecomte, C., 2005. Institut National Agronomique Paris-Grignon. In: L'évaluation Expérimentale des Innovations Variétales. Proposition D'outils D'analyse de L'interaction Génotype – Milieu Adaptés à la Diversité des Besoins et des Contraintes des Acteurs de la Filière Semences. PhD Thesis. https://tel.archives-ouvertes.fr/pastel-00003471/document
Lobell, D.B., Field, C.B., 2007. Global scale climate–crop yield relationships and the impacts of recent warming. Environ. Res. Lett. 2, 1–10.
Lobell, D.B., et al., 2005. Analysis of wheat yield and climatic trends in Mexico. Field Crops Res. 94, 250–256.
Morfin, M., Makowski, D., 2009. MMIX: Model Selection Uncertainty and Model Mixing R Package Version 1.1. http://CRAN.R-project.org/package=MMIX
Passioura, J.B., 1996. Simulation Models: Science, Snake Oil Education, or Engineering? Agron. J. 88 (5), 690–694.
Peltonen-Sainio, P., et al., 2010. Coincidence of variation in yield and climate in Europe. Agric. Ecosyst. Environ. 139 (4), 483–489.
Prost, L., et al., 2008. Comparison of stepwise selection and bayesian model averaging for yield gap analysis. Ecol. Modell. 219, 66–76.
Qian, B., et al., 2009. Statistical spring wheat yield forecasting for the Canadian prairie provinces. Agric. For. Meteorol. 149 (6–7), 1022–1031.
Semenov, M.A., Shewry, P.R., 2011. Comment on Lobell et al. Climate trends and global crop production since 1980. Nat. Precedings http://dx.doi.org/10.1038/npre.2011.6289.1
Supit, I., 1997. Predicting national wheat yields using a crop simulation and trend models. Agric. For. Meteorol. 88, 199–214.
Supit, I., van der Goot, E., 1999. National wheat yield prediction of France as affected by prediction level. Ecol. Modell. 116, 203–223.
Travaso, M.I., 1990. Evaluation des Facteurs Agroclimatiques Limitant le Rendement du blé Dans le Sud de Buenos Aires, Argentine. Essai de Modélisation à L'échelle Régionale. Thèse de Doctorat. Université de Paris-Sud, 96 pp.
Walker, G.K., 1989. Model for operational forecasting of Western Canada wheat yield. Agric. For. Meteorol. 44, 339–351.
Wallach, D., Goffinet, B., 1987. Mean squared error of prediction in models for studying ecological and agronomic systems. Biometrics 43, 561–573.
Wallach, D., et al., 2006. Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications. Elsevier Editors, 450 pp.
Weir, A.H., et al., 1984. A winter wheat crop simulation model without water or nutrient limitations. J. Agric. Sci. 102, 371–382.
de Wit, A., et al., 2010. Using ERA-INTERIM for regional crop yield forecasting in Europe. Clim. Res. 44, 41–53.