(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Estimating regional wheat yield from the shape of decreasing curves of green area index temporal profiles retrieved from MODIS data

Louis Kouadio [a,*], Grégory Duveiller [b,1], Bakary Djaby [a], Moussa El Jarroudi [a], Pierre Defourny [b], Bernard Tychon [a]

[a] *Department of Environmental Sciences and Management, Université de Liège, Arlon Campus Environnement, Avenue de Longwy 185, 6700 Arlon, Belgium*
[b] *Earth and Life Institute (ELI-e), Université catholique de Louvain, Croix du Sud 2, bte 16, 1348 Louvain-la-Neuve, Belgium*

ABSTRACT

Earth observation data, owing to their synoptic, timely and repetitive coverage, have been recognized as a valuable tool for crop monitoring at different levels. At the field level, the close correlation between green leaf area (GLA) during maturation and grain yield in wheat revealed that the onset and rate of senescence appeared to be important factors for determining wheat grain yield. Our study sought to explore a simple approach for wheat yield forecasting at the regional level, based on metrics derived from the senescence phase of the green area index (GAI) retrieved from remote sensing data. This study took advantage of recent methodological improvements in which imagery with high revisit frequency but coarse spatial resolution can be exploited to derive crop-specific GAI time series by selecting pixels whose ground-projected instantaneous field of view is dominated by the target crop: winter wheat. A logistic function was used to characterize the GAI senescence phase and derive the metrics of this phase. Four regression-based models involving these metrics (i.e., the maximum GAI value, the senescence rate and the thermal time taken to reach 50% of the green surface in the senescent phase) were related to official wheat yield data. The performances of such models at this regional scale showed that final yield could be estimated with an RMSE of 0.57 ton ha^{-1}, representing about 7% as relative RMSE. Such an approach may be considered as a first yield estimate that could be performed in order to provide better integrated yield assessments in operational systems.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

An early prediction of wheat yield prior to harvest at regional, national or international scales can play a crucial role in global markets and in policy and decision making. Many models for yield forecasting are available with varying levels of complexity and empiricism. One example is the Crop Growth Monitoring System (CGMS), currently used operationally at the Joint Research Centre of the European Union (EU) for predicting the total production of various crops in all EU Member States and neighbouring countries (Supit, 2000). However, uncertainty about the effect of drought due to limited weather station density and poorly known soil parameters, the lack of information about irrigation and the weighting of individual simulation results for administrative regions (Vossen and Rijks, 1995) have not been taken into account appropriately in this system. Remote sensing has often been mentioned as a way

of resolving some of these uncertainties, by retrieving biophysical variables from remote sensing observations and coupling them with mechanistic growth models (Delécolle et al., 1992; Moulin et al., 1998; Dorigo et al., 2007). Another option for yield forecasting at the regional level is to use simple algorithms that can deal with a strong spatial heterogeneity compared with more complex models that treat the surface as homogeneous (Franks et al., 1997).

Leaf area index (LAI) is a key variable that relates to both the structure and functioning of a vegetation canopy. Chen and Black (1992) defined LAI as half the total developed area of green leaves per unit of ground horizontal surface area. There are several methods for measuring LAI on the ground, using direct or indirect approaches (Gower et al., 1999; Jonckheere et al., 2004; Weiss et al., 2004). Ground measurements are time-consuming, rendering large geographic coverage almost impossible. Therefore, more attention has been paid to retrieving land surface variables, such as LAI, from satellite imagery. Retrieving LAI at the local level from high spatial resolution data has been done successfully in many studies (e.g., Launay and Guerif, 2005; Hadria et al., 2010; Duveiller et al., 2011b). But the types of images generally used in these studies cannot be acquired cloud-free with both a high revisit frequency (e.g., every 1–7 days) and over a large geographic area, which are

---

two requisites for crop growth monitoring at the regional scale. Current instruments meeting these criteria, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), generally have a spatial resolution of about 250 m. Some LAI products are available at the global scale based on these types of instruments (e.g., Myneni et al., 2002; Baret et al., 2007), but they have a spatial resolution of 1 km, which is too coarse to be crop specific over many landscapes in the world (Duveiller and Defourny, 2010). To overcome this problem, Duveiller (2011) proposed a method for selecting coarse spatial resolution pixels whose observation footprint matched the target land cover. By using high spatial resolution masks delineating the fields covered with a target crop combined with a model of the MODIS spatial response, crop specific pixel purity can be calculated and used to select pixels from which a reliable crop specific time series of biophysical variables can be derived. When the biophysical variable of interest is LAI, and all the main aerial plant organs (leaves, stems and ears) photosynthesize (as in winter wheat), Duveiller et al. (2011b) argue that it is more appropriate to use the concept of green area index (GAI) to refer to the biophysical variable retrieved through remote sensing because the electromagnetic radiation reflected from the crop canopy that is encoded in a satellite image is contributed by all the aerial plant organs.

The onset and rate of senescence are important factors for determining grain yield in winter wheat (*Triticum aestivum* L.). Studies over the past decade have shown that there is a close relationship between the maintenance of the photosynthetic life of flag leaves during grain filling and grain yield in wheat (Gooding et al., 2000; Reynolds et al., 2000; Richards, 2000; Dimmock and Gooding, 2002; Blandino and Reyneri, 2009; El Jarroudi et al., 2010). In addition, modelling approaches estimate that at least 45% of the grain carbohydrate content is derived from the flag leaf (El Jarroudi et al., 2009). Because only flag leaves are considered, it might be difficult to implement these methods in a large-area monitoring framework. Indeed, few studies directly link parameters quantifying GLA decrease to winter wheat yield. The aim of this study was to explore an approach for forecasting winter wheat yield at the regional scale, based on metrics derived from the senescence phase of crop-specific GAI time series retrieved from remote sensing data.

2. Materials and methods

2.1. Study area and target spatial administrative units

The EU has subdivided its territory into a harmonised set of administrative units with different hierarchical levels. These geographical regions are called *Nomenclatures des Unités Territoriales Statistiques* (NUTS), with NUTS0 being the entire national territory, NUTS1 the first division level of Member States and NUTS2 and NUTS3 being smaller territorial units, respectively. The level used for agricultural statistics produced by the Statistical Office of the European Communities (EUROSTAT) is the NUTS2 level, which was therefore the level most appropriate for this study. The NUTS regions, however, were designed to have comparable populations at the same NUTS level. The surface of the region, which would be appropriate for this study, is seen as a secondary criterion. The NUTS classification is also based on the administrative divisions applied in the Member States and therefore NUTS of the same level can have important differences in populations and surfaces. The method presented in this paper was applied to winter wheat for two crop seasons (2007 and 2009) over a study site comprising Belgium and northern France. In northern France, the NUTS2 regions are much larger than Belgian NUTS2 regions: the average area of French NUTS2 is 24,940 km^2 (standard deviation, σ, of 11,375 km^2), whereas the average Belgian NUTS2 is 2800 km^2 (σ = 1200 km^2). With regard to spatial extent, the French NUTS3
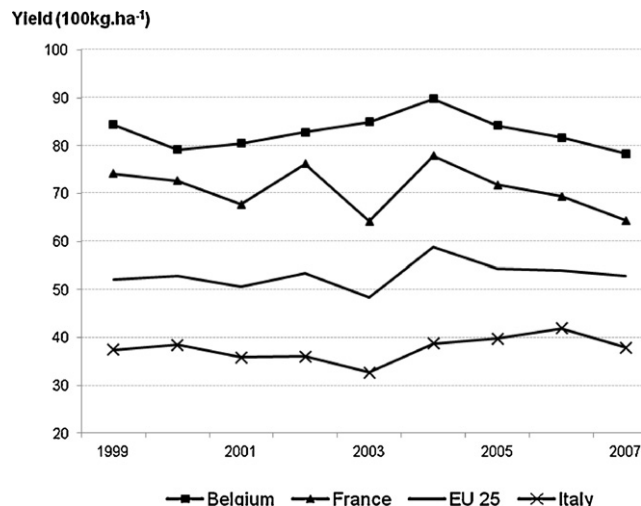

Fig. 1. Official statistics for wheat yield from 1999 to 2007. EU 25: European Union 25 countries.
(source: EUROSTAT)

regions (average area 5800 km^2, σ = 1800 km^2) are more comparable with Belgian NUTS2 and therefore it was these spatial entities that were considered in this paper. In these units, the area covered by winter wheat in 2000–2006 period ranged from 1% to 30% of the NUTS surface: between 1% and 20% for Belgian NUTS2 (EUROSTAT[2]) and between 11% and 30% for French NUTS3.[3] The average yield in these regions is high compared with the average yield at the EU level (Fig. 1).

2.2. GAI retrieval from MODIS time series

Crop-specific GAI time series can be retrieved from MODIS imagery over a fragmented agricultural landscape over several growing seasons using a combination of radiative transfer modelling, neural network inversion and thermal time interpolation (Duveiller et al., 2011a). To do this, a crop-specific mask is necessary to identify which cells in the MODIS grid should be used to provide adequate winter wheat time series. For this study, a winter wheat mask based on Landsat data at a 30 m spatial resolution was available from the GLOBAM[4] project. The overall accuracy of this crop-specific mask creation was close to 90%. As described by Duveiller et al. (2011a), a spatial response model for the MODIS imaging instrument was then convolved over the mask to provide a purity pixel map. This map indicated the proportion of the signal encoded in a MODIS pixel, which effectively comes from the target surface (in this case, this surface consisted of fields under winter wheat). The reflectance encoded in a pixel comes from a ground surface that is often larger than that desired (Cracknell, 1998). Since the size of the MODIS grid over the studied regions was close to the winter wheat fields (Duveiller and Defourny, 2010), "pure" pixels (i.e., whose signal comes 100% from winter wheat) are hard to find. Instead, the purity map was used to select the MODIS cells with at least 75% of pixel purity for winter wheat. For these selected cells, the necessary information for retrieving GAI (reflectance in the red and near-infra-red band, plus the viewing geometry) was compiled from level-2 MODIS Aqua and Terra collection 5 products
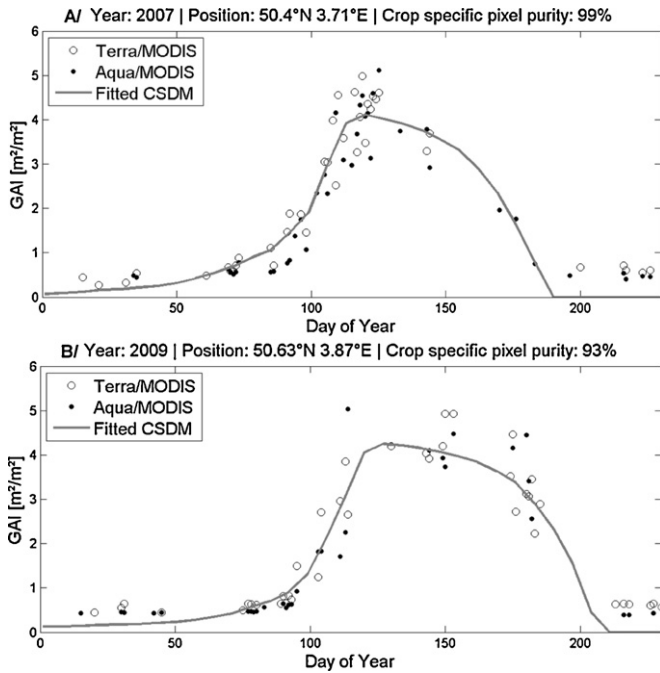
Fig. 2. Examples of interpolation using the canopy structural dynamic model (CSDM).

(available at https://lpdaac.usgs.gov/). The neural network procedure to extract GAI from these data is described by Duveiller et al. (2011a).

2.3. Characterizing the GAI dynamic

The GAI estimated from remote sensing can be affected by various sources of error. These include inaccurate atmospheric correction, contamination of the reflectance from non-wheat land cover when pixel purity is low and shortcomings in the retrieval method. The high temporal frequency of observations, however, is enough to characterize the overall GAI dynamic. Mathematical expressions have long been used to interpolate ground LAI estimations (Baret, 1986; Olesen et al., 2000). In our study, a canopy structural dynamic model (CSDM) was used to interpolate GAI in time and to describe the biophysical variable's time course continuously (Fig. 2). This semi-mechanistic model, developed by Baret (1986) and improved by Lauvernet (2005), is expressed as follows:

$$GAI(T) = A \left[\frac{1}{[1 + \exp(-a(T - T_0 - T_a))]^c} - \exp(b(T - T_0 - T_b)) \right] \quad (1)$$

where a and b define the rates of growth and senescence, respectively; c is a parameter allowing some plasticity for the shape of the curve; A is a scaling coefficient; and T_0, T_a and T_b are the thermal times of plant emergence, mid-growth and end of senescence. The CSDM is further parameterized so as to yield zero GAI once senescence is over. The senescent phase of the curve follows a simple $(1 - e^{bx})$ function. Fig. 2 shows an example of GAI curves fitted by the CSDM function. As described in the next section, this CSDM was used to filter out the GAI time series that did not have a temporally consistent dynamic.

Our main interest was in the senescent part of the curve which, according to Gooding et al. (2000), could be described using a

modified logistic model (Eq. (2)). The following model was therefore also fitted on the GAI estimations:

$$GAI(T) = \frac{A}{1 + [\exp(-k(T - m))]} \quad (2)$$

where A refers to the maximum value of GAI; m is the position of the inflection point in the decreasing part of the GAI curve; k is the relative senescence rate; and T is the thermal time expressed in growing degree-days.

The modified logistic function describes a response that is symmetrical around the inflection point. When A is fixed at 100%, the inflection point appears when GAI = 50% (i.e., m is the thermal time taken to reach 50% green area).

The quality of curve fitting was measured by the variance that accounted for (VAF): VAF(%) = 100 × (1 − (residual mean square/total mean square)). The VAF using the modified logistic function was compared with that using a modified Gompertz function in an equivalent way, expressed as follows:

$$GAI(T) = A \cdot \exp[-\exp(-k(T - m))] \quad (3)$$

The shape of the modified Gompertz model was similar to that of the modified logistic function. The main difference was that in Eq. (3) the inflection point m corresponded to the thermal time taken to reach 37% of the green area.

### 2.4. Filtering out sub-optimal GAI profiles

The temporal GAI profiles could be sub-optimal for regional yield characterization for several reasons. Some precautions were therefore taken to filter out those that did not have the expected shape of winter wheat or those with insufficient observations. The following criteria were used to retain GAI profiles: (i) the number of observations in the growing season had to be above 10; (ii) the temporal consistency: relative root mean square error (RRMSE) between CSDM fit and observations < 15% (for time-series type comparisons, RRMSE is sensitive to large deviations and relatively insensitive to small deviations); (iii) the day of the year when maximum GAI was reached by the CSDM had to be above 135; (iv) the pre-season remote sensing GAI had to be below 1.5 and the post-harvest remote sensing GAI below 1.0. During these periods, winter wheat GAI should be zero. If it is not, it might be an indication that the pixel is actually not (only) representing winter wheat. However, some tolerance is necessary since there might be some other vegetation influencing the signal, and the retrieval algorithm also posed difficulties for assessing low GAI values (often resulting in an erroneous residual GAI).

Fig. 3 presents the spatial distribution of pixels used in the process before and after filtering at the study site. For the two crop seasons 2007 and 2009, about 12% of pixels were efficiently used in this approach: before filtering, the number of pixels was 16,630 in 2007 and 34,936 in 2009; after applying filters the selected pixels dropped considerably (1820 in 2007 and 4076 in 2009). These pixels were globally distributed in 10 NUTS2 in Belgium and six NUTS3 in northern France for the 2009 crop growing season, and in 2007 crop season they were distributed in six NUTS2 and seven NUTS3 (Fig. 3).

2.5. Climatic variables and calculation of thermal time

Meteorological input for the thermal time calculation was based on 10 × 10 km grid cells. The methodology used for data interpolation was applied in CGMS (Supit, 2000). Weather data from weather stations were obtained from the Belgium CGMS system for the Belgian territory and from the European CGMS for areas outside Belgium. The spatial distribution of
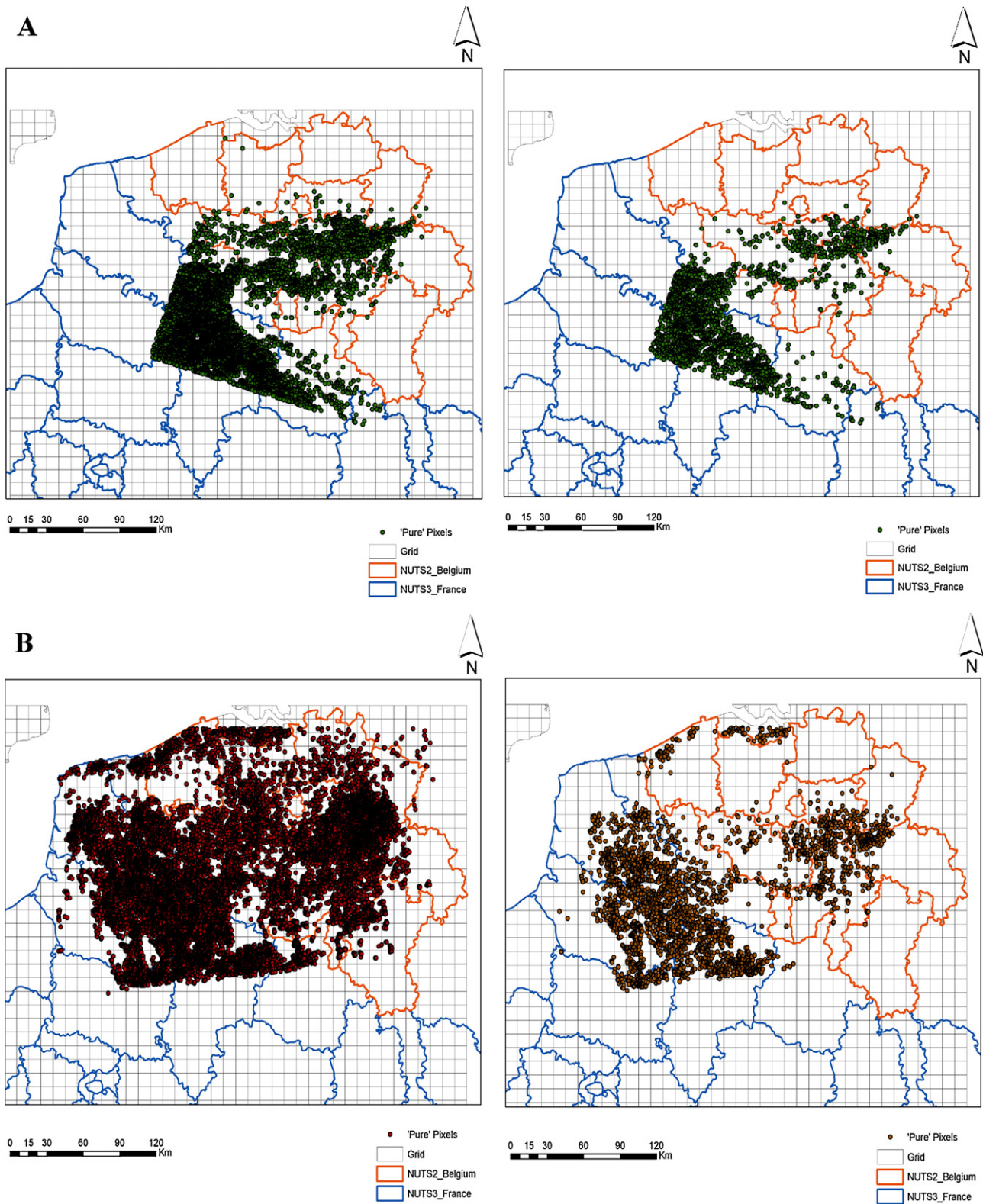
Fig. 3. Spatial selected pixels distribution at a northern European site. (A) Crop season 2007. (B) Crop season 2009. On the left: before applying filters (16,630 pixels in 2007 and 34,936 pixels in 2009). On the right: after applying filters (1820 pixels in 2007 and 4076 pixels in 2009).

selected pixels was intersected with the grid cells. This intersection was then used to derive the minimum and maximum temperatures for the calculation of the corresponding thermal time.

The thermal time, defined as the cumulative daily average temperature (above a crop specific temperature threshold) between two development stages (Masle et al., 1989), was used to express crop development. In our study, thermal time was counted from the

Table 1
Summary statistics and *t* tests for the VAF (%) using the modified Gompertz function and the modified logistic function.

VAF (%)	N	Method	Mean	95% CL	Mean	SD	95% CL	Std. dev.
VAF1	5896		72.25	71.61	72.90	25.27	–	–
VAF2	5896		72.65	72.00	73.31	25.56	–	–
		Pooled	−0.40	−1.32	0.52	25.42	25.09	25.75
		Satterthwaite	−0.40	−1.32	0.52			

Method	Variances	DF	*t* Value	Pr > |*t*|
Pooled	Equal	11,790	−0.86	0.3925
Satterthwaite	Unequal	11,788	−0.86	0.3925

VAF1 refers to the modified Gompertz function and VAF2 to the modified logistic function.

date of maximum GAI value and then cumulated for the following days.

2.6. Yield forecasting modelling

The models were calculated by multi-linear regression. The inputs were the derived metrics m, k, and GAI_{max} (the maximum GAI value). The aggregation of these values at the NUTS level was done by taking the mean value of parameters for each NUTS. For the mean calculation, only pixel profiles with a VAF above a given threshold (corresponding to the VAF of at least two-thirds of the total pixel profiles) were taken into account in the aggregation process at the NUTS level. Official statistical yield data were the reference for the calibration phase. French NUTS3 yield data are available at http://agreste.maapar.lbn.fr/ReportFolders/ReportFolders.aspx; Belgian NUTS2 yield data are available at http://statbel.fgov.be/fr/statistiques/chiffres/economie/agriculture/exploitations/.

Models selected for validation were those with at least two variables (variables were not correlated). Validation was performed using a bootstrap method because of the small number of NUTS used in the analysis. The program used to test the robustness of selected models is detailed in Appendix A. The use of the bootstrap method allows one to approximate the distribution of test statistics in settings where analytic calculations are intractable or in small samples where large-scale asymptotic results might not hold (Austin and Tu, 2004). The bootstrap method also provides less biased and more consistent results than the Jack-knife method (Fan and Wang, 1996).

Statistical indicators (RMSE and adjusted R^2 R_{adj}^2) were used to quantify the performances of models expressing the relationship between observed yields and metrics. The RMSE gave the weighted variations in errors (residual) between the predicted and observed yields. It is one of the most widely used error measures and can be converted in relative RMSE by dividing by the mean of observed yield. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{SSE}{n - p}}$$

where SSE indicates the sum of squared errors; n is the number of observations; and p is the number of parameters in the model.

The R^2 value for a regression can be made arbitrarily high simply by including more and more predictors in the model. The R_{adj}^2 value, however, is one of several statistics that attempts to compensate for this artificial increase in accuracy. It is calculated as follows:

$$R_{adj}^2 = 1 - \frac{(n - 1)(1 - R^2)}{n - p}$$

Statistical analyses relating to curve-fitting and modelling (generalized linear regression for calibration and the bootstrap method for validation) were carried out using the Proc NLIN and Proc GENMOD procedures, respectively, of SAS® software (version 9.1, SAS Institute Inc., Cary, NC, USA).

3. ResultsGAI decreasing curve fitting

By grouping the percentages of VAF after curve-fitting, the class with a high percentage was the one with a VAF greater than 67%. There were 68% of "pure" pixels profiles where the variances were better explained by the modified logistic function. The percentages of VAF for the other classes were 11% and 21%, respectively, for a VAF less than 33% and a VAF between 33% and 67%. For yield forecasting modelling, the threshold of 67% was taken into account in the aggregation of model inputs at the NUTS level.

A comparison with the VAF by a modified Gompertz function showed that there was no significant difference ($\alpha = 0.05$) with the VAF by the modified logistic function (Table 1).

A group test statistic for the equality of means was reported for both equal and unequal variances. Both tests indicated the lack of evidence of a significant difference between different VAF values ($t = -0.86$ and $p = 0.3925$ for the pooled test, $t = -0.86$ and $p = 0.3925$ for the Satterthwaite test).

3.1. Regression-based models

Correlation analyses between explanatory variables showed that there was no significant correlation between them. Therefore, four models were tested:

- Mod1: Yield = $f(m, GAI_{max})$
- Mod2: Yield = $f(k, GAI_{max})$
- Mod3: Yield = $f(m, k)$
- Mod4: Yield = $f(m, k, GAI_{max})$

The generalized linear regression generated multiple tables that allowed the quality of regression to be assessed. Table 2 presents a summary of the regression-based models. Globally, the models were highly significant ($P < 0.001$), except for model Mod3. The R_{adj}^2 varied between 0.47 and 0.57, with no difference in the RMSE (RMSE = 0.57 ton ha^{-1}). Converted in relative RMSE, this value represented about 7%. The scale parameter (Table 2) characterized the statistical dispersion of the probability distribution of the parameter of the model (Efron and Tibshirani, 1986). It was estimated according to maximum likelihood. The smaller was, the better the model prediction. Mod1 and Mod2 had similar performances in terms of R_{adj}^2, scale parameter and RMSE. These two models involved either metric m or metric k and GAI_{max} as explanatory variables. Only Mod3, which took metrics m and k into account, presented poor results relating to its small R_{adj}^2 and high mean of residuals and scale parameter (Table 3). When metrics m and k were used only as explanatory variables, wheat yield could not be adequately estimated. Comparing the four models, model Mod4 gave the best estimations for wheat yield: low value of the

Table 2
Statistical results of regression-based models.

	R_{adj}^2	RMSE (ton ha^{-1})	Model significance $(P > F)$[a]	Scale parameter
Mod1	0.47	0.57	***	0.78
Mod2	0.47	0.57	***	0.78
Mod3	0.03	1.08	ns	1.06
Mod4	0.51	0.57	***	0.76

Significance levels: *** = $P < 0.001$; ns = $P > 0.05$.

[a] P value associated with F value (this ratio compares variability explained by the regression line with variability not explained by the regression line).

Table 3
Performances of selected models after bootstrapping ($n = 11,000$).

	RMSE (ton ha^{-1})	Mean_residuals	SD of residuals	Lower 95% CL for Mean_residuals	Upper 95% CL for Mean_residuals
Mod1	0.56	0.59	0.45	0.58	0.60
Mod2	0.57	0.59	0.44	0.58	0.60
Mod4	0.55	0.56	0.42	0.55	0.57

mean of residuals and scale parameter, good RMSE, and good R_{adj}^2 (Tables 2 and 3). Fig. 4 is a graphical representation of official yields versus predicted yields by models. This figure shows the worst wheat yield estimation given by Mod3. All the other three models presented a satisfactory distribution in the same range of yield (6.5–11.0 ton ha^{-1}) around the line 1:1, whereas the range of predicted values with model Mod3 was about 8–9 ton ha^{-1} (Fig. 4) and inadequately distributed around this line.
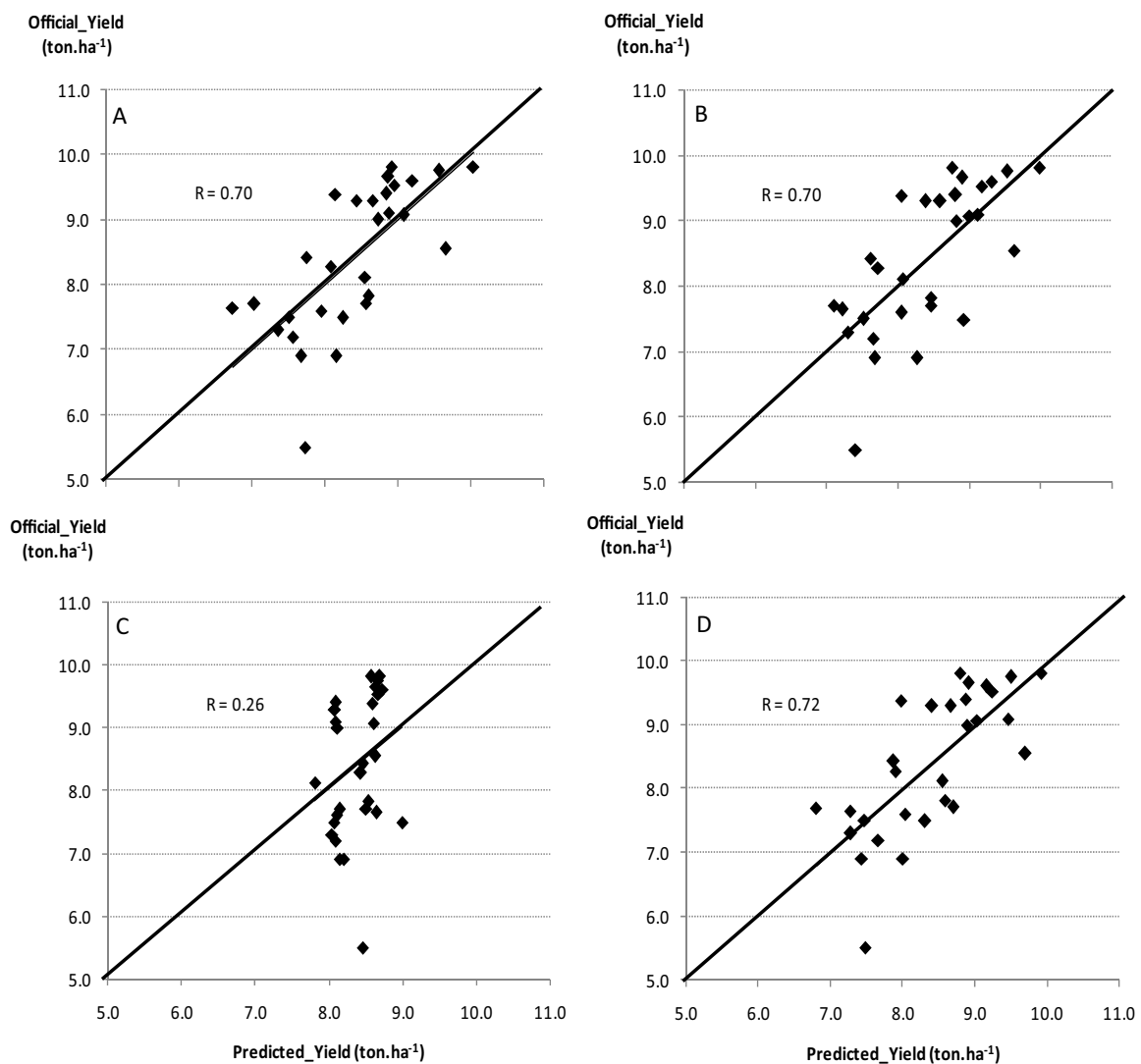
Fig. 4. Official yield versus predicted yield for NUTS3 (France) and for NUTS2 (Belgium) levels. (A) predicted yield using model Mod1; (B) predicted yield using model Mod2; (C) predicted yield using model Mod3; (D) predicted yield using model Mod4 (the line in bold is the line 1:1).

The maximum GAI value seemed to be one key variable in the estimation. A simple model between official yields and this metric gave no significant results (R_{adj}^2 0.34 and RMSE 0.90 ton ha^{-1}). The use of other parameters, such as metrics derived from the decreasing curves, however, significantly increased (in addition to the GAI$_{max}$) the accuracy of the final yield estimation.

4. Discussion

The aim of this study was to explore an approach for wheat yield estimation at the regional level, based on the senescence phase of crop-specific GAI retrieved from remote sensing data. The first hypothesis tested concerned the description of the senescence phase. Large scale studies of crop growth monitoring by remote sensing showed that a negative exponential model based on a simple combination of two parameters for leaf senescence modelling could be used to describe the senescence part of wheat growth (Baret and Guyot, 1986). At the field level, the decrease part of GAI was adequately described by a modified logistic function like the one used for flag leaves (Gooding et al., 2000; Kouadio et al., in press). The study therefore sought to verify the potential of such a modified logistic function in crop yield forecasting at the regional level by developing a regression-based model that relied on metrics of decreasing curves and the maximum GAI value extracted from remote sensing over a large geographic area. The results presented here show that final yield could be adequately estimated from these parameters. This conclusion partly reflects that reached by Baret and Guyot (1986), who showed that wheat production could be estimated from the senescence rate and other parameters. These preliminary results suggest that a simple but reliable yield prediction models could be developed. Wheat yield can be estimated from metrics derived from the GAI decreasing curve and its maximum value, based on an empirical approach. One of the main drawbacks of empirically based models for estimating crop yield has been that their application is valid only for the area for which they have been calibrated. Nonetheless, they are often the preferred approach because they are easy to implement and have limited data requirements (Doraiswamy et al., 2005; Moriondo et al., 2007; Becker-Reshef et al., 2010). By extending the senescence-based approach developed here, further studies will show the degree of precision that can be achieved compared with existing operational systems at the regional scale.

As has been well documented, for many empirical remote sensing based yield models it is assumed that the canopy vigour of winter wheat estimated by spectral measurements (e.g., NDVI) is directly related to final winter wheat yields (Ferencz et al., 2004). The fact that the maximum GAI value was listed as an explanatory variable for all three best models in this study also follows this assumption. As Becker-Reshef et al. (2010) noted in their study, one of the limitations of an empirical and remote sensing based regression model that relies on the maximum spectral NDVI measurements to estimate yields is that it cannot capture the impact of events that reduce yield but do not reduce the peak of green biomass. Metrics describing the senescence part of wheat GAI were calculated using the thermal time concept, which is more closely related to crop development stages. The factors that can lead to a loss of final yield after the peak GAI value could be taken into account through the shape of the decreasing curve. The good accuracy of models developed using this approach is therefore consistent with the concept that remote sensing data are a direct measurement of vegetation conditions that integrates the effect of all major environmental factors (Moriondo et al., 2007).

The lack of close correlation between the GAI$_{max}$ and observed yields on the one hand, and the results in terms of final yield estimation accuracy by adding metrics from the senescence part on the other hand, suggest that a regression-based model that relies on metrics derived from this part of the growth cycle could serve as a good indicator of yield prior to harvest. The relative simplicity of this approach gives it fundamental advantages for regional-scale operational applications.

Wheat (common and durum wheat) is by far the most important crop in Europe in terms of production. Its production represents almost half of all cereal production: in 2008 and 2009, wheat accounted for 48% and 47%, respectively, of EU cereal production (EUROSTAT, 2010). Our study sites were located in regions where average yields vary slightly (see Fig. 1). Based on the results obtained for two crop seasons, studies are being undertaken to assess the uses of this approach to help explain the inter-annual variability of wheat yields in these regions.

5. Conclusion

With the recent developments in earth observation data acquisition, the accurate and timely estimation of crop yields on regional, national and international scales has become increasingly possible in developing and developed countries. In this study, an approach for wheat yield estimates at regional level was explored. This approach was based on metrics derived from the decrease part of GAI temporal profiles. Wheat-specific GAI time series were retrieved from MODIS daily reflectance using a methodology that could be used in an operational monitoring context. A modified logistic function was then fitted to the decreasing GAI curves to calculate the metrics characterizing these shapes. Finally, regression-based models were developed using these metrics and the maximum GAI value. The results of this preliminary study show that the final yield could be estimated using this approach. Further research on applying the approach to other areas and years of GAI measurements is expected to confirm these results.

Acknowledgments

Appendix A. Example of program used to test the robustness of selected model

```
/* Macro used for the validation phase. The model assessed here involves
metrics m_gomp and GAI_max as explanatory variables */

%macro bootstrap (N_samples);

/* Bootstrap with case resampling. This step involves sampling with a
replacement from our original dataset to generate a new dataset
three-quarters the size of our original dataset. The bootstrapped datasets are
stored together as one big dataset called boot */

    proc surveyselect data=FL_NUTS_all
    method=urs
    samprate=0.75
    rep=&nsamples.
    out=boot;
    run;
/* Perform regression on the new dataset */
    proc genmod data=boot;
    class Country;
```

```
    model Yield=Country Mgomp GAImax/type3;
    freq numberhits;
    by replicate;
    output out=Estimation p=yhat resdev=residu; run;
    data estimation;
    set estimation;
    rest=abs(residu); run;
    proc means data=estimation n mean std clm;
    var rest;
    title 'bootstrap results'; run;
%mend;
%bootstrap(500);
```

References

Austin, P.C., Tu, J.V., 2004. Bootstrap methods for developing predictive models. J. Am. Stat. Assoc. 58, 131–137.

Baret, F., 1986. Contribution au suivi radiométrique de cultures de céréales. PhD Thesis. Université de Paris-Sud, France.

Baret, F., Guyot, G., 1986. Radiométrie de la maturation de couverts de Blé dans le visible et le proche infra-rouge. Agronomie 6, 509–516.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: Part 1: Principles of the algorithm. Remote Sens. Environ. 110, 275–286.

Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. Remote Sens. Environ. 114, 1312–1323.

Blandino, M., Reyneri, A., 2009. Effect of fungicide and foliar fertilizer application to winter wheat at anthesis on flag leaf senescence, grain yield, flour bread-making quality and DON contamination. Eur. J. Agron. 30, 275–282.

Chen, J.M., Black, T.A., 1992. Defining leaf area index for non-flat leaves. Plant Cell Environ. 15, 421–429.

Cracknell, A.P., 1998. Review article. Synergy in remote sensing – what's in a pixel. Int. J. Remote Sens. 19, 2025–2047.

Delécolle, R., Maas, S.J., Guérif, M., Baret, F., 1992. Remote sensing and crop production models: present trends. ISPRS J. Photogramm. Remote Sens. 47, 145–161.

Dimmock, J.P.R.E., Gooding, M.J., 2002. The effects of fungicides on rate and duration of grain filling in winter wheat in relation to maintenance of flag leaf green area. J. Agric. Sci. 138, 1–16.

Doraiswamy, P.C., Sinclair, T.R., Hollinger, S., Akhmedov, B., Stern, A., Prueger, J., 2005. Application of MODIS derived parameters for regional crop yield assessment. Remote Sens. Environ. 97, 192–202.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. Int. J. Appl. Earth Obs. Geoinform. 9, 165–193.

Duveiller, G., 2011. Crop specific green area index retrieval from multi-scale remote sensing for agricultural monitoring. PhD Thesis. Université catholique de Louvain, Belgium.

Duveiller, G., Baret, F., Defourny, P., 2011a. Crop specific green area index retrieval from MODIS data at regional scale by controlling pixel-target adequacy. Remote Sens. Environ. 115, 2686–2701.

Duveiller, G., Defourny, P., 2010. A conceptual framework to define the spatial resolution requirements for agricultural monitoring using remote sensing. Remote Sens. Environ. 114, 2637–2650.

Duveiller, G., Weiss, M., Baret, F., Defourny, P., 2011b. Retrieving wheat green area index during the growing season from optical time series measurements based on neural network radiative transfer inversion. Remote Sens. Environ. 115, 887–896.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1, 54–75.

El Jarroudi, M., Delfosse, P., Maraite, H., Hoffmann, L., Tychon, B., 2009. Assessing the accuracy of simulation model for Septoria leaf blotch disease progress on winter wheat. Plant Dis. 93, 983–992.

El Jarroudi, M., Kouadio, L., Martin, B., Giraud, F., Delfosse, P., Hoffmann, L., Tychon, B., 2010. Modelling plant diseases impact with the Belgian crop growth monitoring system. In: Wery, J., Shili-Touzi, I., Perrin, A. (Eds.), Proceedings of 'Agro2010 the XIth ESA Congress', Montpellier. August 29th–September 3rd, 2010, Agropolis International Editions. Montpellier, France, pp. 519–520.

EUROSTAT, 2010. Agricultural Statistics Main Results—2008–09, Eurostat Pocketbooks, 2010 Edition. Publications Office of the European Union, Luxembourg.

Fan, X., Wang, L., 1996. Comparability of jack-knife and bootstrap results: an investigation for a case of canonical correlation analysis. J. Exp. Educ. 64, 173–189.

Ferencz, C., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., Molnár, G., Pásztor, S., Steinbach, P., Székely, B., Ferencz, O.E., Ferencz-Árkos, I., 2004. Crop yield estimation by satellite remote sensing. Int. J. Remote Sens. 25, 4113–4149.

Franks, S.W., Beven, K.J., Quinn, P.F., Wright, I.R., 1997. On the sensitivity of soil-vegetation-atmosphere transfer (SVAT) schemes: equifinality and the problem of robust calibration. Agric. Forest Meteorol. 86, 63–75.

Gooding, M.J., Dimmock, J.P.R.E., France, J., Jones, S.A., 2000. Green leaf area decline of wheat flag leaves: the influence of fungicides and relationships with mean grain weight and grain yield. Ann. Appl. Biol. 136, 77–84.

Gower, S.T., Kucharik, C.J., Norman, J.M., 1999. Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems – a real or imaginary problem? Remote Sens. Environ. 70, 29–51.

Hadria, R., Duchemin, B., Jarlan, L., Dedieu, G., Baup, F., Khabba, S., Olioso, A., Le Toan, T., 2010. Potentiality of optical and radar satellite data at high spatio-temporal resolutions for the monitoring of irrigated wheat crops in Morocco. Int. J. Appl. Earth Obs. Geoinform. 12, S32–S37.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. Agric. Forest Meteorol. 121, 19–35.

Kouadio, L., Djaby, B., Duveiller, G., El Jarroudi, M., Tychon, B. Cinétique de décroissance de la surface verte et estimation du rendement du blé d'hiver. Biotechnol. Agron. Soc. Environ., in press.

Launay, M., Guerif, M., 2005. Assimilating remote sensing data into a crop model to improve predictive performance for spatial applications. Agric. Ecosyst. Environ. 111, 321–339.

Lauvernet, C., 2005. Assimilation variationnelle d'observations de télédétection dans les modèles de fonctionnement de la vegetation: utilisation du modèle adjoint et prise en compte de contraintes spatiales. PhD Thesis. Université Joseph Fourier, Grenoble I, France.

Masle, J., Doussinault, G., Farquhar, G.D., Sun, B., 1989. Foliar stage in wheat correlates better to photothermal time than to thermal time. Plant Cell Environ. 12, 235–247.

Moriondo, M., Maselli, F., Bindi, M., 2007. A simple model of regional wheat yield based on NDVI data. Eur. J. Agron. 26, 266–274.

Moulin, S., Bondeau, A., Delecolle, R., 1998. Combining agricultural crop models and satellite observations: from field to regional scales. Int. J. Remote Sens. 19, 1021–1036.

Myneni, R.B., Hoffman, S., Knyazikhin, Y., Privette, J.L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G.R., Lotsch, A., Friedl, M., et al., 2002. Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data. Remote Sens. Environ. 83, 214–231.

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.

Reynolds, M.P., Delgado, B.M.I., Gutiérrez-Rodríguez, M., Larqué-Saavedra, A., 2000. Photosynthesis of wheat in a warm, irrigated environment: I: genetic diversity and crop productivity. Field Crops Res. 66, 37–50.

Richards, R.A., 2000. Selectable traits to increase crop photosynthesis and yield of grain crops. J. Exp. Bot. 51, 447–458.

Supit, I., 2000. An exploratory study to improve the predictive capacity of the crop growth monitoring system as applied by the European Commission. Treemail Publishers, Heelsum, The Netherlands.

Vossen, P., Rijks, D., 1995. Early crop yield assessment of the EU countries: the system implemented by the Joint Research Centre, Report EUR 16318 EN of the Office for Official Publications of the E.U. Luxembourg, pp. 15–39.

Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. Agric. Forest Meteorol. 121, 37–53.