Predicting sunflower grain yield using remote sensing data and statistical models

Arnaud Micheneau, Luc Champolivier, Jean-François Dejoux, Al Bitar Ahmad, Célia Pontet, Ronan Trépos, Philippe Debaeke

Predicting sunflower grain yield using remote sensing data and statistical models

A. Micheneau[1], L. Champolivier[1], J.-F. Dejoux[2], A. Al Bitar[2], C. Pontet[1], R.Trepos[3], P.Debaeke*[4]

1. *Terres Inovia, Auzeville-Tolosane, France*
2. *CESBIO, Toulouse, France*
3. *INRA, UR MIAT, Auzeville-Tolosane, France*
4. *INRA, UMR AGIR, Auzeville-Tolosane, France*

Keywords: yield prediction, remote sensing, statistical models, sunflower, production basin

Abstract

Context and state of the art - Predicting grain yield a few weeks before harvest is of strategical interest for the cooperatives which collect and store grains. With such information, they can optimize their logistics (allocation of storage cells, transfers between silos) and thus reduce the financial and environmental costs of storage. In the literature, a range of methods have been developed to predict grain yield and quality (oil, protein) for various field crops at different spatial and temporal scales.

The first type of method is based on the observation (survey) of reference fields which sample the diversity of soils and crop practices at landscape level (Doré *et al.*, 2008 ; Champolivier *et al.*, 2011 ; Hall *et al.*, 2013). Generally, this method is used more to understand yield build-up and analyze yield gap at harvest than to predict yield within season. However, quantifying the yield-limiting factors and measuring physiological indicators (e.g. leaf chlorophyll content) may help the experts to revise their prediction throughout the season (Le Bail *et al*., 2005). This method requires a suitable sampling of the production area to minimize the prediction error at an aggregated level. A second type of method is based on the upscaling at territorial level of models (dynamical or statistical) predicting yield or quality at field level. Numerous studies were reported using simple agroclimatic models (Potgieter *et al*., 2005) or process-based models (Jagtap and Jones, 2002; Launay and Guérif, 2003). These models are more often used for assessing attainable yields and their spatio-temporal variability over long climatic series than for predicting yield at an annual basis over a production area. Correctly estimating the cropping practices is more critical to achieve than to get accurate data of soils and weather at field (or grid) level. A third approach combines observations from remote sensing (satellite) and models (statistical or dynamic crop models). Green area index (GAI) and soil moisture data derived from reflectance interpretation and microwave radiometry respectively are assimilated into the crop models to correct the errors due to model structure and input data, thus improving the predictive quality of the model (Launay and Guérif, 2005; de Wit and van Diepen, 2007 ; Ines *et al*., 2013 ; Chipanshi *et al.*, 2015). This approach is more adapted to the full description of a production basin in real-time conditions.

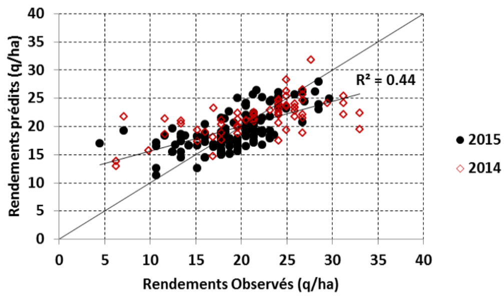
Sunflower crop has been seldom used in remote-sensing studies (Claverie *et al.,* 2012) in spite of the plasticity of GAI and its key role in yield achievement. GAI at anthesis (GAImax) is a good indicator of plant density, N status and early water stress during the vegetative phase; GAI duration after anthesis (GAD) is a good indicator of grain filling and oil accumulation in sunflower (Merrien *et al*., 2012). It has been shown in previous studies that GAImax and GAD were two candidates for predicting grain yield early in the growing season. In this study, we will combine remote sensing and statistical models to predict sunflower yield during grain filling.

Materials & Methods - In 2014 and 2015, 187 sunflower fields (6.7 ha in average) were surveyed in Haute-Garonne and Gers departments (SW France) at different periods of the growing season. Commercial yields (q/ha) were provided by farmers. Green area index (GAI) was estimated by the inversion of radiative transfer model with BVnet (Baret *et al*., 2007) from satellite images (Landsat 8 and Deimos in 2014; Landsat 8 and Spot 5 in 2015) at 6 (2014) and 11 (2015) dates in average throughout the growing season. From these GAI estimations, two variables were calculated: GAImax and GAD (Green Area Duration), which is the area under

GAI curve decreasing from anthesis to physiological maturity. According to the professionals, a yield prediction would be relevant about 3 weeks before harvest. As sunflower is harvested in early September, canopy information was no more used after 10 August. Different regression models were compared (Merrien and Grandin, 1990): (1) Yield = f(GAImax) ; (2) Yield = f(GAD) ; (3) Yield = f (GAImax, GAD). Linear, quadratic, linear-plateau, and quadratic with plateau models were tested for (1) and (2) relationships and ‘plateau’ models were omitted for (3) relationship. Cross validation was based on the leave-one-out approach but also by dividing the dataset into two subsets (2014, 2015) for calibration and evaluation. The Root Mean Square Error of Prediction (RMSEp) was calculated to compare the performance of the 10 models.

Results & Conclusion - In Table 1, models (5) and (9) based on GAD or GAD + GAImax were among the best performing and the simplest ones to apply in practice. RMSEp (2014 + 2015) was < 4 q/ha in both cases (leave-one-out approach). The performance of model 9 was displayed on Fig.1 based on cross validation.

	RMSEp (q/ha) (2014, 2015)	Best regression models
Yld = f(GAD)	3.98	(5) Y = 0.154 GAD + 8.425
Yld = f(GAI max, GAD)	3.90	(9) Y = 1.101 GAImax + 0.131 GAD + 7.713



The model performance was greater in 2015 (RMSEp = 3.36 for model 9) than in 2014 (RMSEp = 4.69) probably because of less fungal diseases and more observed GAI values in 2015. Average observed grain yields were 21.7 and 19.2 q/ha in 2014 and 2015 (drier year, lower plant density) respectively. Between 3 and 8 images were exploitable in 2014 while GAI was calculated with 7 to 15 images in 2015 depending on the fields. This should have played on the accuracy of GAI estimation which was greater in 2015. These simple statistical models using remote sensing data at high temporal and spatial resolution provide a baseline for the comparison with process-based crop models such as SUNFLO (Casadebaig *et al.*, 2011) whether or not assimilating satellite information (as GAI for instance). Predicting oil production will be the next issue to address with this approach.

References

Baret F. et al., 2007. *Remote Sensing of Environment* 110, 275–286
Casadebaig P. et al., 2011. *Agricultural Forest* Management 151, 163-178.
Claverie M. et al., 2012. *Remote Sensing of Environment* 124, 844–857.
Champolivier L. et al., 2011. *Innovations Agronomiques* 14, 39-57.
Champolivier L. et al., 2012. *Proc.18th Int. Sunflower Conf.*, Mar del Plata (Argentina).
Chipanshi A. et al., 2015. *Agricultural and Forest Meteorology* 206, 137–150
Doré T. et al., 2008. *Agronomy for Sustainable Development* 28, 151-161.
Ines A.V.M. et al., 2013. *Remote Sensing of Environment* 138, 149–164
Jagtap S.S, Jones JW, 2002. *Agriculture, Ecosystems and Environment* 93, 73–85.
Hall et al., 2013. *Field Crops Research* 143, 119-129.
Launay M., Guérif M., 2003. *Agronomie* 23, 135–146.
Launay M., Guérif M., 2005. *Agriculture, Ecosystems and Environment* 111, 321–339.
Le Bail M. et al., 2005. *European Journal of Agronomy* 23, 379–391.
Merrien A., Grandin G., 1990. In : *Le tournesol et l’eau*. Ed. Cetiom (Paris), 75–90.
Merrien A. et al., 2012. *Proc.18th Int. Sunflower Conf.*, Mar del Plata (Argentina).
Potgieter A.B. et al., *Agricultural and Forest Meteorology* 132, 143–153.
de Wit A.J.W., van Diepen C.A., 2007. *Agricultural and Forest Meteorology* 146, 38–56.