Europ. J. Agronomy 54 (2014) 84–96

Contents lists available at ScienceDirect

European Journal of Agronomy

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

Prediction of sunflower grain oil concentration as a function of variety, crop management and environment using statistical models

Fety Nambinina Andrianasolo[a,c,d,*], Pierre Casadebaig[a,d], Elie Maza[b,d], Luc Champolivier[c], Pierre Maury[a,d,1], Philippe Debaeke[a,d,1]

[a] INRA, UMR AGIR, BP 52627, 31326 Castanet-Tolosan Cedex, France
[b] INRA, UMR GBF, BP 52627, 31326 Castanet-Tolosan Cedex, France
[c] CETIOM, Centre INRA de Toulouse, BP 52627, 31326 Castanet-Tolosan Cedex, France
[d] Université de Toulouse, INP, ENSAT, BP 52627, 31326 Castanet-Tolosan Cedex, France

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form
27 November 2013
Accepted 5 December 2013

Keywords:
GAM
Genotype by environment interaction
Regression model
Sunflower oil concentration

ABSTRACT

Sunflower (*Helianthus annuus* L.) raises as a competitive oilseed crop in the current environmentally friendly context. To help targeting adequate management strategies, we explored statistical models as tools to understand and predict sunflower oil concentration. A trials database was built upon experiments carried out on a total of 61 varieties over the 2000–2011 period, grown in different locations in France under contrasting management conditions (nitrogen fertilization, water regime, plant density). 25 literature-based predictors of seed oil concentration were used to build 3 statistical models (multiple linear regression, generalized additive model (GAM), regression tree (RT)) and compared to the reference simple one of Pereyra-Irujo and Aguirrezábal (2007) based on 3 variables. Performance of models was assessed by means of statistical indicators, including root mean squared error of prediction (RMSEP) and model efficiency (EF). GAM-based model performed best (RMSEP = 1.95%; EF = 0.71) while the simple model led to poor results in our database (RMSEP = 3.33%; EF = 0.09). We computed hierarchical contribution of predictors in each model by means of R^2 and concluded to the leading determination of potential oil concentration (OC), followed by post-flowering canopy functioning indicators (LAD2 and MRUE2), plant nitrogen and water status and high temperatures effect. Diagnosis of error in the 4 statistical models and their domains of applicability are discussed. An improved statistical model (GAM-based) was proposed for sunflower oil prediction on a large panel of genotypes grown in contrasting environments.
© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Worldwide vegetable oil consumption is expected to grow by 2% per year as a result of increasing edible oil and renewable energy demands (FAO, 2012). In the 2011/2012 campaign however, oilseed grains production was greatly reduced because of adverse cropping conditions, then leading to a negative balance between supply and demand. The use of deemed tolerant oilseed crops, such as sunflower (*Helianthus annuus* L.), should be thus given consideration. The latter shows some agronomic and industrial potentialities (Ayerdi-Gotor et al., 2008; Aguirrezábal et al., 2009; Pilorgé, 2010) as a promising competitive oilseed crop.

Sunflower cultivation could be particularly improved in France, where it is often grown in limited, shallow soils, non-irrigated and poor-nutrient sites (Debaeke et al., 2006; Casadebaig, 2008). In those situations, genotype × environment × management interactions were evidenced (Grieu et al., 2008) since genotypes do not exhibit the same strategies to cope with stress in restrictive conditions (Gallais, 1992; Denis and Vear, 1994).

Obtaining higher-oil concentration varieties appeared to be an alternative track for enhancing sunflower production, and could become a plus-value for French producers (Vear et al., 2003; Roche, 2005). Sunflower oil concentration was reported to be a conservative genetic component (Fick and Miller, 1997; Ruiz and Maddonni, 2006); however, recent studies highlighted differential responses of sunflower genotypes in contrasting cropping conditions; greater variability of oil concentration was whether linked to management and environmental conditions (Champolivier et al., 2011), or to genotypic and environment interactions (Andrianasolo et al., 2012). In both cases, a good understanding of oil concentration elaboration and effects of genotype and environmental factors raised to be essential for proposing convenient management strategies targeting both grain yield and oil content.

Sunflower oil is composed of 98% fatty acids (Berger et al., 2010; Echarte et al., 2010), which are produced from two potential sources; main originates from post-flowering photosynthetic

* Corresponding author at: INRA, UMR AGIR, BP 52627, 31326 Castanet-Tolosan Cedex, France. Tel.: +33 681354270; fax: +33 561735537.
E-mail address: fandrian@toulouse.inra.fr (F.N. Andrianasolo).
[1] Co-advisors of the first author Ph.D. thesis.

1161-0301/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.eja.2013.12.002

carbon (Merrien, 1992), supplemented with carbon assimilates stored in vegetative parts before flowering that will be remobilized thereafter (Hall et al., 1990; Merrien, 1992). Plant parts that provide carbon after flowering are considered as "source" (source pool: leaves, stems) whereas those requiring carbon at this period are denoted "sink", namely grains. Reported determinants of sunflower oil concentration are genotype and environmental factors (Connor and Hall, 1997; Champolivier et al., 2011), among which intercepted radiation, nitrogen availability, high temperatures and water stress are often cited. These factors could play on both source and sink components, though only few studies explicitly separate effects on source and sink or make the link with oil concentration.

Genotype effect – *i.e.* genotypes with intrinsic high or low-oil concentration – was described to play through kernel to hull proportion (López Pereira et al., 2000; Izquierdo et al., 2008). At source level, genotype effect could play through contrasting strategies in mobilizing pre-flowering and post-flowering available carbon (Sadras et al., 1993).

Cumulative intercepted radiation between 250 and 450 degrees days after flowering was found to be the main determinant of oil concentration ($R^2 \sim 80\%$) among sunflower hybrids in Argentine (Aguirrezábal et al., 2003). Higher plant densities could have a positive effect on source before flowering (Ferreira and Abreu, 2001) and on sink after flowering; Diepenbrock et al. (2001) suggested that the variation of oil concentration could be partly linked to negative impact of higher plant densities on final grain weights. However, Rizzardi et al. (1992) observed genotype × plant density interactive effects on final oil concentrations when comparing two contrasting genotypes.

Nitrogen effect is often described through the negative relationship between oil and protein concentration (Connor and Sadras, 1992); highest oil concentrations were met in non-fertilized treatments. Nitrogen doses that are brought during vegetative period permit to optimize dry matter at flowering (Hocking and Steer, 1983; Debaeke et al., 2012) thus potential quantity of mobilized pre-flowering assimilates during grain-filling.

High temperatures after flowering were reported to shorten grain filling duration; depending on authors, we identified various temperature thresholds: 30 °C (Aguirrezábal et al., 2009), 34 °C maximum temperatures (Chimenti et al., 2001; Rondanini et al., 2003) or 17 °C mean temperature (Angeloni et al., 2012).

Little evidence exists about the effect of water availability on oil concentration; Santonoceto et al. (2003) observed significant differences in oil concentration in the final phase of oil accumulation under water stress, with an obvious lower rate of grain oil accumulation for non-irrigated modality. Before flowering, water stress could affect leaf expansion (Casadebaig et al., 2008), while it could limit green leaves photosynthesis and duration in post-flowering period (Aguirrezábal et al., 2009).

Literature-based knowledge about sunflower oil concentration determination is illustrated in a schematic conceptual framework (Fig. 1).

To help understanding crop physiology and yield determinism, crop models are tools that are increasingly developed. These can be used for multiple purposes, either to describing complex biological systems, or to interpreting experimental results, making a diagnosis of limiting factors and providing advices and predictions toward farmers for better crop and policy management (Boote et al., 1996). Statistical/empirical models, particularly, have been of great use in the history of science. Their easiness of computing and usability enhanced their attractiveness among decision-makers and practitioners (Razi and Athappilly, 2005), while they allow highlighting relative importance of variables when much is uncertain (Lobell et al., 2005; Tittonell et al., 2008; Tulbure et al., 2012). Statistical models could be divided into two main subgroups: parametric and non-parametric. Parametric models (*e.g.* simple or multiple linear

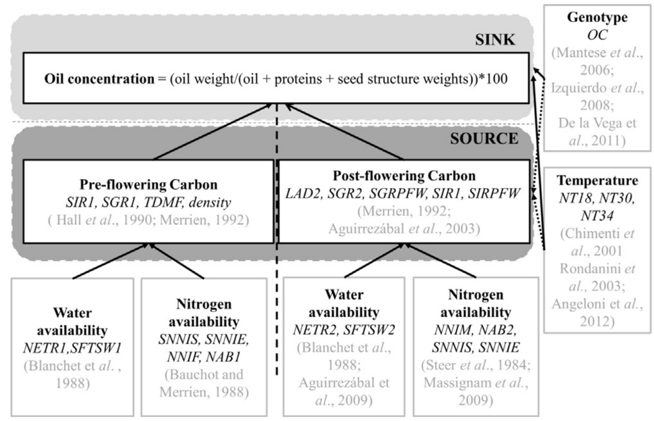

Fig. 1. Schematic framework of sunflower oil concentration elaboration as described in section 1 and relative selected predictors used for statistical modeling. Meanings of abbreviations are given in Table 2. Continuous arrows indicate literature reported relationships which were used to compute the selected predictors. Dotted arrows indicate known relationships that were not used in this study.

regression) have the advantage to be quantifiable, and assessable, but the form of the relationship between dependent and independent variable(s) should be known *a priori* to avoid misleading results; non-parametric ones (*e.g.* GAM, regression trees and neural networks) do not assume neither any *a priori* model structure nor any formal distribution of the data. They permit to bring out non-linear relationships but often lead to heavy parameterized models. Wullschleger et al. (2010) used non-parametric models to establish equations of parametric ones for switchgrass yield prediction. Other non-parametric models (regression trees, Breiman et al., 1984) were utilized to analyze yield variability in maize (Tittonell et al., 2008), wheat (Lobell et al., 2005), soybean (Zheng et al., 2009), sugarcane (Ferraro et al., 2009) or switchgrass (Wullschleger et al., 2010; Tulbure et al., 2012).

Few statistical models exist for seed oil prediction; those existing are mostly parametric. For instance, multiple linear regressions were used to model palm oil (Khamis et al., 2006; Keong and Keng, 2012), though their predictive performances were not assessed. For sunflower, a non-linear empirical model was established by Pereyra-Irujo and Aguirrezábal (2007) relating actual oil concentration to genotypic oil concentration, radiation cumulated during the post-flowering specific period (Aguirrezábal et al., 2003) and plant density. However, the model was parameterized in sites where nitrogen was non-limiting, and where water stress could be likely moderate or non-existing.

For specifically predicting oil concentration, the crop model SUNFLO (Casadebaig et al., 2011) uses a multiple linear regression model linking oil concentration with some simulated genotype, environmental stress and post-flowering canopy functioning indicators. Following oil model evaluation on an independent dataset, it was hypothesized that the acceptable though improvable RMSEP (predictive root mean squared error: ~4 oil points) was due to the narrowness of ranges of situations represented in the database, and the choice of predictors that failed to take into account physiologically-based responses of sunflower.

Therefore, the objectives of this paper are the following: (1) build statistical models based on physiologically-sound predictors and compare their predictive performance for sunflower grain oil concentration on a large dataset; (2) highlight essential features of grain oil elaboration by assessing variable importance and unraveling interactions; (3) compare the performance of these statistical models with the reference one from Pereyra-Irujo and Aguirrezábal (2007). The latter was chosen as reference model given its simplicity (low number of variables, simple equation), easiness of use

Download English Version:

https://daneshyari.com/en/article/4508980

Download Persian Version:

https://daneshyari.com/article/4508980

Daneshyari.com