Predicting wheat yield and protein content at the plot scale with machine-learning and mechanistic models

Fety Andrianasolo, Jean-Charles Deswarte, Cécile Garcia

Predicting wheat yield and protein content at the plot scale with machine-learning and mechanistic models

Andrianasolo Fety*[1] (f.andrianasolo@arvalis.fr), Deswarte Jean-Charles[2], Garcia Cécile[2]

[1] *Arvalis Institut du Végétal, Boigneville, France*; [2] *Arvalis Institut du Végétal, Villiers-le-Bâcle, France*

Introduction

Providing the earliest reliable model predictions of crop production is of a major interest since it allows to optimize crop management and strategic positioning in the market. For wheat in particular, the recurrence of drought episodes due to climate change and variability challenges crop models on their ability to take into account genotype x environment effects at the plot scale.

Materials and Methods

Three yield component-based models were proposed to predict the grain number, grain weight and protein content respectively, for soft and durum wheat. These models were calibrated on historical database of field experiments conducted at plot scale (from 1996 to 2016), on more than 50 varieties and 40 sites in France. Models simulating grain number and protein content took into account varietal effects, agro-climatic indicators (e.g water stress and photothermal quotient around heading) and variables describing canopy state (nitrogen nutrition index, aboveground biomass) by the use of random forest methods. Grain weight model corresponded to a logistic curve describing the needs in temperatures (degree-days) of each variety, modulated by daily temperature, radiation and water stress effects during grain filling. Models were calibrated using indicators such as model efficiency, bias, and relative root mean squared error (RRMSE).

Results and Discussion

The models calibration results are illustrated in Figure 1. They were further evaluated on an independant dataset with very promising performances (relative error of 15% and 8%) for yield and proteins, respectively. Models highlighted differential varietal response to water stress between soft and durum wheat. These models will be soon deployed as part of a decision support tool, which combines the use of a dynamic crop model (CHN model) and LAI sensor data used to readjust model predictions during the campaign.

Conclusions

Based on this study, the combination of machine learning and mechanistic models appears to be a promising way for simulating wheat yield and protein at plot scale.

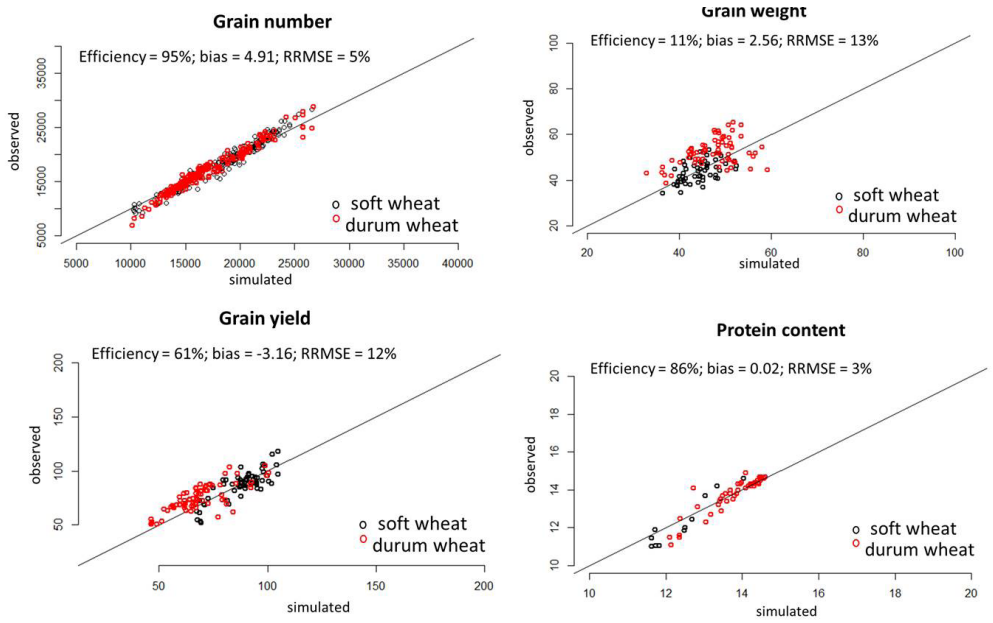

Figure 1. Results of wheat yield components model calibration (Grain number, grain weight, yield and protein content at harvest).

Keywords: wheat yield, random forest, logistic, water deficit, CHN crop model.