Predicting herbage yield in perennial ryegrass breeding trials using UAV derived data and machine learning

Pranga J.[1,2], Borra-Serrano I.[1], Aper J.[1], Ghesquiere A.[1],
Roldán-Ruiz I.[1,3], Janssens I.[2], Ruysschaert G.[1], Lootens P.[1]

[1]Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Melle, Belgium
[2]Department of Biology, University of Antwerp, Wilrijk, Belgium
[3]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium

Introduction

(Close) remote sensing technologies can be used to monitor crops. In recent years, there has been a growing interest in the development of UAV-based methods for non-destructive estimations of grassland biomass. Here, we examine whether the prediction of herbage yield in perennial ryegrass can be improved using canopy height (CH) information combined with spectral data derived from two sensors.

Materials and Methods

A breeding trial established in May 2019 (Merelbeke, Belgium) was investigated. It comprises diploid and tetraploid varieties and populations (468 plots in total). Four UAV flights (April 1, April 15, April 23, May 4 2020) were carried out with two sensors, a 10-band multispectral and an RGB camera system. The acquired data were used to predict the dry matter yield (DMY) of the first spring cut in May 2020. To model DMY, five different feature combinations and three machine learning (ML) algorithms were used. A schematic workflow (Fig. 1) shows the procedures applied for data processing and modeling.

Fig. 1. Schematic workflow

Results and Conclusions

- On average, tetraploids produced more DMY than diploids during the spring cut (Fig. 2)
- Based on the CH data, the lowest mean rRMSE values of 10.7% and 11.0% were reached for diploids and tetraploids, respectively (Fig. 3).
- For diploids, the CH information already resulted in a low mean rRMSE, while for tetraploids using or adding spectral bands to the CH data slightly improved the performance.
- In terms of rRMSE, no clear advantage of one machine learning algorithm over another was identified when applied to the same dataset (Fig. 3).

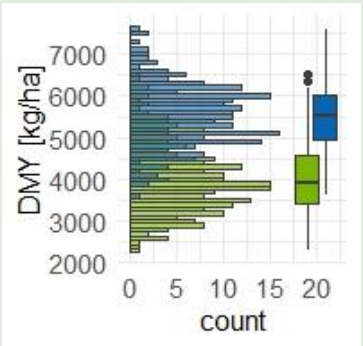

Figure 2. Distribution of DMY in the spring cut, separated for diploids (in green) and tetraploids (in blue).

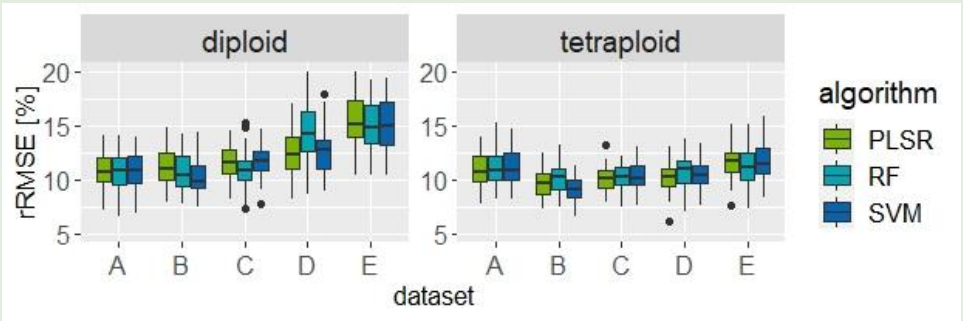

Figure 3. Box plot representing a distribution of model performance estimates (rRMSE), fitted by three ML algorithms: PLSR, RF, and SVM. DMY was set as a target variable, while (A) CH data; (B) CH combined with multispectral bands; (C) CH combined with RGB and HIS (hue, intensity, saturation) data; (D) multispectral bands; and (E) RGB and HIS data were set as predictors.

Acknowledgments

The research was a part of the *FutureArctic* project funded by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 813114. Special thanks to Thomas Vanderstocken and Aaron Van Gehuchten for performing the drone flights.