Validation of a workflow based on Sentinel-2, Sentinel-1 and meteorological data predicting biomass in pastures

Charles Nickmilder1, Anthony Tedde1, Isabelle Dufrasne2, Françoise Lessire2, Bernard Tychon3, Yannick Curnel4, Jérome Bindelle1, and Hélène Soyeurt1

1 TERRA Research Centre, Passage des Déportés 2, BE-5030 Gembloux   
2 Centre des Technologies Agronomiques, rue de la Charmille, 16, BE-4577 Strée-Modave   
3 Département des sciences et gestion de l’environnement,avenue de Longwy 185, BE-6700 Arlon  
4 Centre wallon de Recherches agronomiques (CRA-W), Rue de Liroux, 9, BE-5030 Gembloux

Abstract

This study develops the validation of the four most promising models resulting from a workflow processing Sentinel-1, Sentinel-2 and meteorological data through 13 different machine learning algorithms. They led to 124 models predicting biomass measured as compressed sward height on square sub-samples of the parcels (i.e., pixel-based estimation with a resolution of 10 m). The training and validation data were acquired in 2018 and 2019 in the Walloon Region of Belgium with a rising platemeter equipped with a GPS. The cubist, perceptron, random forest, and general linear models had a validation root mean square error (RMSE) around 20 mm of CSH. However, the information relevant for the farmer and for an integration in a decision support system is the amount of biomass available on the whole pasture. Therefore, those models were also validated at a parcel-scale using data from another farm (122 CSH records acquired with a different rising platemeter) based on predictions made on input variables expressed at the parcel-scale or at a subparcel-level and then aggregated at the parcel-scale. The resulting RMSE were higher than before. To improve the quality of prediction, a combination of the outputs of the models might be needed.

Keywords : remote sensing, compressed sward height, pastures, machine learning

Introduction

The advantages of grazing in agroecosystems are well recognized: lower feeding cost (Hennessy et al., 2020), higher animal welfare and increased milk quality (Elgersma, 2015). However, intensive dairy farmers tend to turn away from grazing (Lessire *et al*., 2019). One of the main reasons is the need for a frequent assessment of the available standing biomass. Some decision support systems (DSS) were developed previously on the basis of mechanistic models (e.g. Romera *et al.*, 2010 and Ruelle *et al.*, 2018) to address this issue. Another approach to develop such DSS is the use of remote sensing. This study used space borne synthetic-aperture radar (SAR) and optical imagery from Sentinel-1(S-1) and Sentinel-2(S-2) constellations in combination with meteorological data. A total of 124 models were developed previously on data acquired over three farming area located in Wallonia (Southern part of Belgium) and four were considered as promising: a cubist, a generalised linear model with elastic regularisation of the Gaussian family(glmnet) and a random forest (rf) using 160 variables and a neural network (nnet) using 47 variables. The reference biomass measurements were performed with a Jenquip EC20G rising platemer (RPM) equipped with a GPS yielding compressed sward height (CSH). The objective of this study was to test our models on a totally independent dataset, acquired with a different RPM of the same producer (Jenquip EC-01) based on the same recording technology (i.e. ratchet) that was not equipped with a GPS. If the models were able to achieve good performances, it would enlarge the scope of input data and possible collaborations, else it would mean that the workflow has to be developed further.

Material and methods

One-hundred-and-twenty-two field CSH data were acquired in Eastern Wallonia from May 2019 to October 2019 during farm walks (Delhalle and Knoden, 2019). A total of 33 records out were discarded due to the presence of clouds on the satellite images. Two types of validation were applied: the satellite data (sigma nought σ0 for S-1 and level 2A reflectance for S-2) were extracted either at the parcel level either at a subparcel level. On the one hand, the extraction at the parcel level consisted in the collection of the satellite data over the parcel for each date and then the median of the nearest non-cloudy image within a 5 days-time window was computed. The integration of the time window avoided the dismissal of a part of the CSH data. The predictions were then made on these aggregated data. On the other hand, the parcels were divided in pixel-like subplots and the extraction and prediction steps were performed on each subplot. The resulting predicted CSH were then aggregated at the parcel level for each date with a mean. The predictions were analysed through the prism of a graphical comparison of the predictions against the reference CSH, and the values of the RMSE of validation.

Results and discussion

The predictions at the parcel-level and the subparcel-level behaved similarly compared to the original dataset (figure 1) although some points were not represented due to their prediction value being out of the range of values the RPM can measure (from 0 to 250 mm of CSH). More precisely: the glmnet model predicted 5 values out of the range for the prediction at the parcel level and 7 in the other case and the other models did not predict out of range for the prediction at the parcel level. For the subparcel-level, the neural network failed to produce a correct value once and the random forest thrice. The similarity in behaviour and the effect of the extreme prediction values on the RMSE can also be seen in table 1.

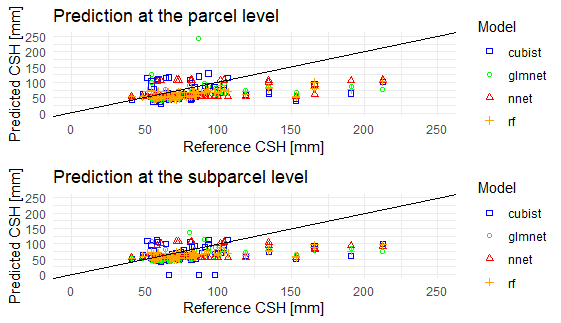
The models achieved much better performances during the previous independent validation (Nickmilder). It might be due to the high occurrence of clouds over this area, the difference in grassland management practices and pasture composition. Although the RMSE is worse than what was found in the literature (Cimbelli and Vitale, 2017), the RPD for the non-glmnet models are acceptable (near 1) given that the standard deviation is 30.9 mm of CSH. Another possible source of difference is the use of models developed at a subparcel level for prediction at the parcel level. Two types of combination could help increase the precision: a combination of the prediction led to higher R squared with similar RMSE or the combination of models developed at both the subparcel and the parcel level.

Figure 1: Representation of the predictions VS the original CSH dataset

Table 1: RMSE of validation of the four models applied at the parcel or at the subparcel level.

|  |  |  |
| --- | --- | --- |
| types | model | RMSE |
| parcel | cubist | 39 |
| parcel | glmnet | 491441 |
| parcel | nnet | 35 |
| parcel | rf | 33 |
| subparcel | cubist | 38 |
| subparcel | glmnet | 103698 |
| subparcel | nnet | 35 |
| subparcel | rf | 33 |

Conclusion

The 4 models that were developed and selected previously to predict biomass did not show good performances once applied to data acquired at the parcel level with a different RPM of the same manufacturer on a different location. It stressed the need to develop the combination prospective to increase the robustness of the models given that one of them diverged completely during this validation.

Acknowledgements

The authors would like to thank the “Wallonie agrigulture SPW” for funding the ROAD-STEP project and this study. This work contains modified Copernicus Sentinel data (2020). The authors would like to thank Fourrages Mieux ASBL (Delhalle and Knoden, 2019) for providing the validation data.

References

Cimbelli, A.; Vitale, V. (2017) Grassland height assessment by satellite images. *Advances in Remote Sensing*, 06, 40–53. <doi:10.4236/ars.2017.61003>.

Delhalle R., Knoden D. (2019) “Mesures d’hauteur d’herbe de prairies pâturées en Ardenne.” Fourrages Mieux ASBL

Elgersma, A. (2015) Grazing increases the unsaturated fatty acid concentration of milk from grass-fed cows: A review of the contributing factors, challenges and future perspectives*. European Journal of Lipid Science and Technology*, 117, 1345–1369. <doi:10.1002/ejlt.201400469>

Hennessy, D.; Delaby, L.; Dasselaar, A.V.D.P.v.; Shalloo, L. (2020) Increasing grazing in dairy cow milk production systems in Europe. pp. 1–15.

Lessire, F.; Jacquet, S.; Veselko, D.; Piraux, E.; Dufrasne, I. (2019) Evolution of grazing practices in Belgian dairy farms: Results of two surveys. *Sustainability*, 11, 1–15. <doi:10.3390/su11153997>

Nickmilder C., Tedde A., Dufrasne I., Lessire F., Tychon B., Curnel Y., Bindelle J., Soyeurt H. Development of machine learning models to predict compressed sward height in Walloon pastures on the basis of Sentinel-1, Sentinel-2 and meteorological data using multiple data transformations. [Manuscript submitted for publication]. TERRA Research Centre, University of Liège.

Romera, A.J.; Beukes, P.; Clark, C.; Clark, D.; Levy, H.; Tait, A (2010). Use of a pasture growth model to estimate herbage mass at a paddock scale and assist management on dairy farms. *Computers and Electronics in Agriculture*, 74, 66–72. <doi:10.1016/j.compag.2010.06.006>.

Ruelle, E.; Hennessy, D.; Delaby, L. (2018) Development of the Moorepark St Gilles grass growth model (MoSt GG model): A predictive model for grass growth for pasture based systems. *European Journal of Agronomy*, 99, 80–91. <doi:10.1016/j.eja.2018.06.010>.