Deployment of models to predict compressed sward height at a large scale: results and feedback.

Nickmilder, C.[1], Tedde, A.[1], Dufrasne, I.[2,3], Lessire, F.[2,3], Tychon, B.[4], Curnel, Y.[5], Bindelle, J.[1], Soyeurt, H.[1]
[1] TERRA Research Centre, Passage des Déportés 2, BE-5030 Gembloux
[2] Faculté de Médecine vétérinaire / Département de gestion vétérinaire des Ressources Animales (DRA) / Nutrition des animaux domestiques, Uliège, Place du 20 Août 7, 4000 Liège
[3] Centre des Technologies Agronomiques, Rue de la Charmille, 16, BE-4577 Strée-Modave
[4] Département des sciences et gestion de l'environnement, Avenue de Longwy 185, BE-6700 Arlon
[5] Centre wallon de Recherches agronomiques (CRA-W), Rue de Liroux, 9, BE-5030 Gembloux

Abstract

Currently, there is a high interest to integrate data linked to remote sensing and methods from the machine learning domain to develop tools helping pastures management. In this context, over the past two years, we published models predicting the available compressed sward height (CSH) in pasture using Sentinel-1, Sentinel-2, and meteorological data. Those scalable models could be the basis of a decision support system (DSS) available for Walloon farmers. A platform performing the CSH prediction was developed and this paper aims to provide some insights in its prediction capabilities and tackle the challenge of using data acquired at different moments in time times. Predictions were made from the beginning of January until the end of October 2021 using our most promising published models. After data cleaning, the coefficient of variation of CSH predictions, calculated for each studied date (N=35) and parcel (N=192,862), ranged from 0 to 986. This extreme variation suggests some prediction imperfections. Before the integration of the platform in a DSS, the main task to solve is the issue of missing or non-operational S1 or S2 data. Indeed, even if a gap filling method was applied, only 62% of potentially exploitable dates were usable.

Keywords: machine learning; decision support system; dairy cows; grazing management; pasture

Introduction

Recently published papers (Shalloo *et al.,* 2018; Shalloo *et al.*, 2021) underline the "work in progress" nature of the integration of data linked to remote sensing and methods from the machine learning (ML) domain in the ecosystem of tools available for managing pastures. Our team developed ML models predicting the compressed sward height (CSH) available on pastures from satellite and meteorological data (Nickmilder *et al.*, 2021). To bridge the gap between these research models and their potential use by farmers, a platform performing the prediction was implemented to be the data provider for a decision support system (DSS). This paper aims to provide some insights in the prediction capabilities of this platform and tackle the challenge of using data acquired at different moments in the season.

Material and methods

The models implemented in the prediction platform were the three best performing ones (*i.e.,* a cubist, a neural network perceptron (nnet) and a random forest (rf)) based on a previous work done by Nickmilder et al. (2021) with a RMSE of CSH estimated from an independent validation around 20 mm. Those models use meteorological, Sentinel-1 (S1), and Sentinel-2 (S2) data to predict CSH. The workflow of the prediction platform is the following. First, the platform acquires daily the newly available data and launches the pre-processing when needed. Then, it performs a spatial standardisation, realizes the prediction process, and finally makes some post-processing if needed. The data treatment was similar to Nickmilder et al. (2021). The S1, S2 and meteorological data were acquired in a form that covered the whole area of Wallonia (the Southern part of Belgium) and thus all its 194,657 parcels of agricultural area with pastures and was collected from mid-January 2021 to the end of October 2021. The meteorological data came from the Agromet platform (CRA-W, 2021) and

were aggregated on daily basis. The S1 data were acquired from the European space agency. The S2 data were acquired from the Theia platform under the form of level-2A products. All these datasets were resampled on a raster grid with 10m resolution. Each parcel was thus constituted of pixels and each time there was S1 and/or S2 data in the pixel, it was considered as a record. Some filtering on S2 tiles were made: removal of tile with too much missing values/ saturated pixels or cloud coverage (75% threshold each time).
To deal with missing acquisition, a gap filling method was applied. For every day of the year, a check was made to confirm the availability of S2 data. If it was confirmed, the date was considered as usable (UD). All the pixels were filled with available data at this UD and the incomplete pixels were filled with data acquired the day before and so on until 4 days before the UD. Thus a dataset for one UD is in fact a composite dataset gathering S1 and S2 data up to 4 days before the actual UD. To assess the relevance of the CSH predictions and the reliability of the prediction platform scaling local models to a greater scale (i.e., entire Walloon Region), we have studied: the occurrence of concurring data acquisition, the raw values, the presence of outliers, and the descriptive statistics for each date and parcel.

Results and discussion

Theoretically, the S2 satellites have a revisit frequency over Wallonia of 3 to 5 days. Considering the worst case scenario of 5 days, we should have at least 58 dates (i.e., 290/5 days) usable for the period studied. However, even with the application of a gap filling methodology, only 35 dates (62%) had enough S2 data of sufficient quality to be further processed and 25 of these dates covered the grazing season (April – October). Without the gap filling application, only 17 dates would have been considered (29%). On these 35 dates, 99% (192,866) of the parcels were represented at least once and the total number of records was 201,875,534. Unfortunately, there was a huge number of non-usable S2 data. This might be due to a combination of edgy position of pixels relatively to the satellite orbits, poor weather and out of range/ missing input values. These values cannot deliver reliable information. Therefore, a post processing filter was applied after the prediction step to remove all the non-finite values. This decreased the predicted set to 92,782,075 records. The data distribution of those predictions per model is summarized in Table 1. Some values (less than 2%) were out of the range of expected CSH values (*i.e.*, [0 mm; 250 mm]). After deletion, the dataset was composed of 92,757,937 records. Given that a parcel is composed of several pixels, therefore, the estimation of the coefficient of variation of CSH is a measure of its CSH heterogeneity. The cubist CSH predictions gave the highest variability, the nnet and rf models were more stable although the trends of higher CSH were asynchronous as shown in Figure 1. This meant that the models used extracted different part of the information in the dataset and a combination of these information must be accounted in the future. The observed null values were due to the presence of only one pixel (Table 1). To visualize the evolution of CSH throughout the year, the average CSH was calculated for each date (Figure 1). As expected due to the response of plants to increasing temperatures, we observed a slight increase during the spring (April – June) and then a decrease. However, the sparse data acquisition due to the poor weather conditions during the summer blurred the trends. Another point underlined in Figure 1 is the sensitivity of the models to cloudiness: for example, the 11th February was cloudy with very thin clouds that were not detected as such, and thus decreasing the quality of the predictions for this specific date.

Conclusion

From a technological point of view, the platform is operational and now usable to predict on a daily-routine basis the CSH in Wallonia. However, given the low proportion of exactly concurring data, we had to implement a time lag tolerance in the platform for its future use in a DSS. This means that for each S2 acquisition date, predicted datasets were completed with data going back up to four days. This methodology managed to decrease the impact of non-concurring data in a context of predicting with all the datasets but there is still work to do. Indeed, only 62% of dates were exploitable. Hence, the models still need to be improved given the occurrence of those quite extreme values.

Acknowledgements

The authors would like to thank the 'Wallonie agriculture SPW' for funding the ROAD-STEP project and the National Research Fund for the grant "T.0221.19". This work contains modified Copernicus Sentinel data (2020).

References

CRA-W 2021 Agromet.be; https://agromet.be/

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 Based on Sentinel-1, Sentinel-2 and Meteorological Data Using Multiple Data Transformations. Remote Sens. 2021, 13, 408. https://doi.org/10.3390/rs13030408

Shalloo, L., Donovan M. O., Leso L., Werner J., Ruelle E., Geoghegan A., Delaby L., and Leary N. O.. 2018. "Review: Grass-based dairy systems, data and precision technologies." *Animal* 12 (s2): S262–S271. https://doi.org/10.1017/S175173111800246X

Shalloo L., Byrne T., Leso L. et al. A review of precision technologies in pasture-based dairying systems. Irish Journal of Agricultural and Food Research. DOI: 10.15212/ijafr-2020-0119

Theia, 2021: https://www.theia-land.fr/

Tables

	Raw CSH predictions (N=92,782,075)			Coefficient of variation per parcel and date (N = 92,757,937)		
	Cubist	Nnet	Rf	Cubist	Nnet	Rf
Minimum	5.39	45.26	25.89	0.00	0.00	0.00
1%	31.58	45.26	37.25	1.42	0.00	1.22
1st quartile	47.83	56.23	53.00	6.50	0.00	5.65
Median	56.68	56.23	60.09	10.21	9.19	9.09
Mean	62.41	64.74	64.39	12.03	11.40	10.46
3rd quartile	70.832	62.08	70.92	15.26	17.78	13.85
99%	145.86	130.64	130.36	41.89	42.18	30.91
Maximum	331.88	130.64	203.15	906.95	68.65	74.62

Table 1: Data and coefficient of variation distribution of CSH predictions (in mm) obtained by the three tested models on the 192,866 parcels, delivering 92,782,075 records with some extreme associated predictions on the 35 dates of acquisition.

Figures

Figure 1: representation of the mean of the parcel's mean height over each acquisition date.