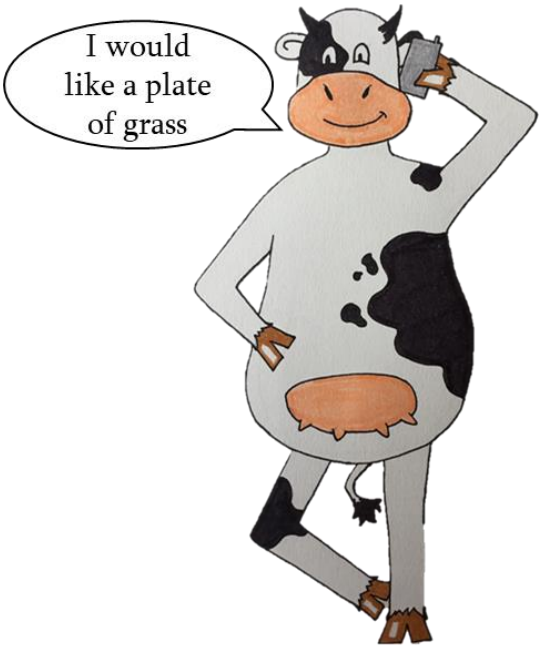
LIÈGE université
Gembloux
Agro-Bio Tech

Development of machine learning algorithms fed by meteorological and remote sensing data to assess the available grass on pastures.

Charles NICKMILDER



COMMUNAUTÉ FRANÇAISE DE BELGIQUE
UNIVERSITÉ DE LIÈGE – GEMBLOUX AGRO-BIO TECH

Development of machine learning algorithms fed by meteorological and remote sensing data to assess the available grass on pastures.

Charles NICKMILDER

Dissertation originale présentée en vue de l'obtention du grade de docteur en sciences agronomiques et ingénierie biologique

Promoteur : Prof. Hélène Soyeurt

Année civile : 2023

Abstract

Charles Nickmilder, 2023, Development of machine learning algorithms fed by meteorological and remote sensing data to assess the available grass on pastures. PhD Dissertation. Gembloux, Belgique, Gembloux Agro-Bio Tech, University of Liège, 223 p.

Grasslands cover a large part of emerged lands. This large soil occupancy also occurs in Wallonia. This profusion of grasslands is the result of a combination of advantages that could be called Ecosystem Services (ES). These ES involve loads of stakeholders, from politics to farmers, also including industrial actors. Focusing the analysis on the agronomic prospect of grasslands, their inherent interest is underlined: grasslands constitute a relevant feed source for cattle breeding. Indeed, it is well suited for bovine breeding, bovine being the most represented animal in the Walloon ruminant livestock. Over the past years, the number of farms decreased with a steady state. Meanwhile, the total agricultural area did not change that much. Therefore, and due to a high level of splitting up of the Walloon parcels, some farms use parcels far apart for cattle breeding. Therefore, there is an interest to offer a set of tools to ease the remote management of grazing – tools that can be called decision support systems (DSS). This PhD thesis aims to contribute to the development of such a tool and more specifically its input information.

So, this thesis covers the development of a standardised process to develop machine learning (ML) models using meteorological data and remotely sensed data, acquired through the Sentinel-1 (S1) and Sentinel-2 (S2) constellations of satellites, to predict a proxy of the available feed on pastures: the compressed sward height (CSH). These 124 ML models are intended to be at the heart of the DSS. Beside the ML algorithms, data pre-treatments included feature transformations. In the end, four models were found to be the most promising for the creation of generalisable models and the root mean square error (RMSE) of these models was around 20 mm of CSH when measuring at a pixel/sub-parcel level. After the model development, this thesis work focusses on the process of creating a platform performing the CSH prediction at the scale of Wallonia, with a 10-meter resolution. The handling of the massive amount of data required to get back to the basic concepts of big-data science and engineering. In order to assess the reliability of the platform, a retrospective analysis was performed over the grazing seasons from 2018 to 2021. It passed the tests searching for dates completely utterly predicted (less than 1% of the data were poorly predicted) and for instabilities in

the prediction (coefficient of variation of the CSH per parcel ranged from 0 to 312%).

In conclusion, it appears that the ML models and the associated platform, in their current state, passed the test for fine spatial representation, accuracy of the prediction and rapidity to deliver the information. However, they do not completely pass the criterion of temporal regularity due to the frequent unavailability of the satellite information. To solve this issue, complementary modelling might be needed to fill the missing data.

Résumé

Charles Nickmilder, 2023, Développement d'algorithmes d'apprentissage automatique alimentés par des données météorologiques et issues de la télédétection afin d'estimer le disponible herbager en pâtures. Thèse de doctorat. Gembloux, Belgique, Gembloux Agro-Bio Tech, Université de Liège, 223 p.

Les herbages couvrent une large partie des terres émergées. Cette large occupation des sols se voie aussi en Wallonie. Cette abondance d'herbage est le résultat d'une combinaison d'avantages qui peuvent être appelés des services écosystémiques (SE). Ces SE impliquent une floppée de parties prenantes, du politique à l'agriculteur, en passant par des acteurs industriels. En focalisant l'analyse sur l'intérêt agronomique des herbages, leur intérêt est souligné : ils constituent une source de nourriture pertinente pour les animaux d'élevage, particulièrement adaptée aux bovins qui constituent la majorité du cheptel ruminant wallon. Ces dernières années, le nombre d'exploitations agricoles a diminué à un rythme soutenu, tandis que la surface agricole n'a que peu changé. De ce fait et dû à un morcellement élevé du parcellaire agricole wallon, certaines exploitations agricoles exploitent des parcelles relativement éloignées pour l'élevage. Par conséquent, il y a un intérêt à offrir un ensemble d'outils pour faciliter la gestion, à pareille distance, du pâturage – entrant sous l'appellation d'outils d'aide à la décision (OAD). Cette thèse de doctorat vise à contribuer au développement de pareil outil et, plus spécifiquement, des informations l'alimentant.

Cette thèse couvre le développement d'un processus standardisé pour le développement de modèles d'apprentissage autormatique (machine learning - ML) utilisant des données météorologiques et issues de la télédétection, acquises par les constellations de satellites Sentinel-1 (S1) et Sentinel-2 (S2), pour prédire un proxy du disponible fourrager en pâtures : la hauteur d'herbe compressée (CSH). Ces 124 modèles de ML seront au cœur de l'OAD. Au-delà des algorithmes de ML, le processus de pré-traitement développé inclue aussi des transformations de variables. Au bout du compte, quatre modèles ont été défini comme étant les plus prometteurs pour une généralisation/un passage à plus grande échelle. L'erreur quadratique moyenne (RMSE) de ces modèles s'élève à plus ou moins 20 mm lorsque l'échelle pixellaire/sous-parcellaire est considérée. Après le développement des modèles, cette thèse se focalise sur le processus de création d'une plateforme de prédiction de la hauteur d'herbe compressée à l'échelle de la

Wallonie, avec une résolution de 10 mètres. La manipulation des quantités importantes de données engendrées par ce processus a nécessité de revenir aux bases de la science et l'ingénierie des données. Afin d'estimer la fiabilité de la plateforme, une analyse rétrospective a été réalisée sur les saisons de pâturage de 2018 à 2021. La plateforme a réussi les tests concernant la recherche de prédictions anormales (moins d'un pourcent des données ont été mal prédites) et concernant la stabilité des prédictions (les coefficients de variation de la CSH par parcelle et par date variaient de 0 à 312%).

En conclusion, les modèles de ML et la plateforme associée, en l'état actuel, répondent aux critères de finesse spatiale, d'ajustement de la prédiction et de rapidité de délivrance de l'information. Par contre, ils ne répondent pas complètement au critère de régularité temporelle suite au fréquent manque d'information satellitaire. Pour résoudre ce problème, des modélisations complémentaires doivent être envisagées afin de combler les données manquantes.

Acknowledgements

Ce document est le résultat d'un travail de longue haleine et je n'aurais pu parcourir ce chemin seul. Je tiens à mettre en lumière et remercier ces compagnons de route. Mais auparavant, je souhaite remercier le service public de Wallonie d'avoir financé le projet ROAD-STEP (D31-1393/S1). Merci aux institutions ayant aidé à fournir des données dont le Centre des technologies agronomiques (CTA) et l'asbl FourragesMieux.

En premier lieu, je tiens à te remercier Hélène. Ta supervision s'est plus apparentée à un mentorat qu'à une relation de hiérarchie classique. Avant de commencer cette thèse, je pensais ne plus jamais m'impliquer dans le domaine de la recherche scientifique. Lorsque je t'ai fait part de ce dégoût envers ce panier de crabes, c'est la première fois que j'ai été confronté à tes qualités qui m'ont poussées à continuer à relever le défi, challenge après challenge. En effet, ton approche basée sur la place centrale de l'Humain, la collaboration, l'échange, l'apprentissage et la considération de tous les points de vue m'a poussé à me dépasser, à me former et à tenir le coup lors des coups de down.

Cette place centrale de l'Humain se retrouve aussi chez tous les collaborateurs du SIMa. Merci à vous, Lionel et Dominique, de veiller à ce que le service tourne, et dans la bonne humeur. Merci à Yves pour ses cours de stats et d'avoir éveillé en moi, avec l'aide de l'assistante de l'époque c-à-d Kékette, la curiosité pour les stats, cette matière supportant la plupart des avancées scientifiques récentes. Merci à mes co-doctorants, Anne-Catherine, Pauline, Sébastien, Mireille, Lei, Killian, Nikolaï pour les discussions super enrichissantes. Et non, il n'y a pas un impair, grosse dédicace à Anthony pour m'avoir supporté dans son bureau et m'avoir partagé sa passion pour les stats. Merci à tous pour les pauses café qui permettent de s'élargir les idées.

Merci aux membres du comité de thèse d'avoir guider et encadrer les recherches, que ce soit Jérome Bindelle pour la compréhension de la relation complexe entre l'animal et la pâture ainsi que la présidence du jury, Isabelle Dufrasne pour son approche très concrète des variables modélisées et la fourniture de base de données, ou encore Bernard Tychon et Cozmin Lucau pour la gestion des aspects de télédétection. Merci également à Adrien Michez d'avoir rejoint le jury de thèse afin d'apporter son expertise dans le domaine de la télédétection appliquée aux prairies.

Merci à vous tous que j'ai côtoyé à Gembloux, vous qui m'avez aidé à grandir, mûrir et m'épanouir. En particulier merci à vous les Aigris, si vous voulez savoir vos attributs, on en discutera autour d'une bière débat. Du coup, gros betch à Tintin, Jacquy, Michel, Scro, Dylan, Chouuu et Beatch. Grosse bise aussi à mes anciens co-kotteurs dont Balou, Tanguy, Cléms exposant 3.

D'un point de vue technique, merci à Sci-hub pour avoir mis à disposition tous ces articles scientifiques kidnappés et gardés en otage par ces gangsters d'éditeurs scientifiques. Merci aussi à toute la communauté scientifique et informatique pour la promotion de l'open access et de l'open science.

Me supportant depuis bien avant que je rentre à Gembloux et bien parti pour me supporter encore longtemps, la palme de la patience revient à ma famille. Merci papa, merci maman de m'avoir supporté toutes ces années. Quand je voyais toute la galère et le boulot que vous abattiez, je me disais que j'allais éviter de faire un métier qui bouffe autant de temps. Au final, me voilà à finaliser un doctorat en y ayant consacré autant d'énergie que vous dans vos activités. Comme quoi des chats ne font pas des chiens (heureusement car j'aime pas les chiens). Courage à Pierre et Eléonor qui doivent encore en finir, l'un avec un doc', l'autre avec des études que je n'arrive toujours pas à comprendre.

Et finalement, à celle qui m'a supporté au cours de ces 4 dernières années et demi, au cours de ces interminables confinements, au cours de toutes ces phases durant lesquelles j'étais invivable (le TFE, la préparation des conférences, les articles, la rédaction de la thèse, …), à ma compagne avec qui j'ai la joie de partager pleins de moments inoubliables, de fou-rires et d'amour, à celle qui est une source d'inspiration pour moi par sa volonté de toujours faire mieux, à toi Marine, je te dis merci, je te félicite pour ta patience et j'ai hâte de partager pléthore de nouvelles aventures avec toi. PS : non, Acrylique, tu es bien gentille de vouloir m'aider avec tes patounes sur le clavier mais on n'ajoute pas d'emoji dans un document aussi sérieux.

Table of Contents

1 Chapter 1 Introduction 3
1.1 Outline 3
1.2 Story 3
1.3 Grassland overview and stakeholders prospective 4
1.3.1 Grasslands VS pastures 4
1.3.2 Climate: IPCC & FAO 5
1.3.3 Ecosystem services 6
1.3.4 Politics: From Europe to Wallonia 9
1.3.5 Society prospective 10
1.3.6 Industry prospective 11
1.4 Grassland: back to an agronomic prospective 13
1.4.1 Grassland direct and indirect use for feed 13
1.4.2 Grassland modelling 19
1.5 Grassland and remote sensing 25
1.5.1 Story 25
1.5.2 Categories description 25
1.5.3 Practical characteristics of remote sensing platforms 28
1.5.4 Applications of satellite remote sensing on pastures 31
1.6 Thesis objective 34
1.7 Plan 36
1.7.1 Story version 36
1.7.2 Scientific version 36
1.8 Chapter bibliography 38
2 Chapter 2 Model creation 51
2.1 Outline 51
2.2 Abstract 51
2.3 Keywords 52
2.4 Introduction 52
2.5 Materials and Methods 54
2.5.1 Datasets 56
2.5.2 Grid 59
2.5.3 Fusion and Data Transformation 59

2.5.4 Split the Dataset 60
2.5.5 Variable Selection 60
2.5.6 Model Training 61
2.5.7 Model Validation 62
2.6 Results 65
2.6.1 Description of the Datasets 65
2.6.2 Variable Selection 67
2.6.3 Prediction of Compressed Sward Height (CSH) 70
2.6.4 Effect of the Pixellation/Aggregation of Data 74
2.6.5 Best Performing Models 75
2.7 Discussion 80
2.8 Conclusions 84
2.9 References 85
3 Chapter 3 Platform development 93
3.1 Outline 93
3.2 Abstract 93
3.3 Keywords 94
3.4 Introduction 94
3.5 Material and methods 97
3.5.1 Study area 97
3.5.2 Global design 97
3.5.3 Data acquisition 98
3.5.4 Data pre-processing 100
3.5.5 Spatial standardization 100
3.5.6 Merging datasets 102
3.5.7 Model 103
3.5.8 Prediction 104
3.5.9 Analysis of the predictions 104
3.6 Results and Discussion 106
3.6.1 Practical constraints 106
3.6.2 Model accuracy 107
3.6.3 Data fusion 109
3.6.4 Prediction relevancy 114
3.6.5 Spatial heterogeneity analysis 117
3.6.6 Model selection 120
3.6.7 On the choice of working with Compressed Sward Height 121

3.7 Conclusion 122
3.8 References 124
3.9 Supplementary material 128
3.9.1 Overview of recent scientific papers related to the understanding of pastures 128
3.9.2 Details on the model creation 141
3.9.3 Evolution of the Walloon grasslands 144
3.9.4 Variable importance of the newly trained models. 145
3.9.5 Distribution of the predicted CSH depending on the time lag tolerance considered 148
3.9.6 Proposition of representation of the probability to get x predictions a year 149
3.9.7 Summary of the predicted CSH values at the pixel level 150
4 Chapter 4 Would I pack the same way? 155
4.1 Story 155
4.2 Outline 155
4.3 Objective and modelled feature 157
4.4 Data origin 161
4.4.1 CSH data 161
4.4.2 Meteorological data 162
4.4.3 Sentinel-1 163
4.4.4 Sentinel-2 163
4.4.5 Complementary data 164
4.5 Machine learning modelling 165
4.5.1 Technical foreword 165
4.5.2 Algorithm description 166
4.5.1 Validation set holding-out 168
4.5.2 Feature transformation 168
4.5.3 Tuning 168
4.5.4 Validation 169
4.6 Data treatment 174
4.6.1 CSH data 174
4.6.2 Parcel definition 174
4.6.3 S1 data 176
4.6.4 S2 data 179
4.6.5 Degree-day (DD) computation 180

4.6.6 Joining the data 180
4.7 Gap filling 183
4.7.1 Prior the modelling 183
4.7.2 During the modelling 185
4.7.3 After the modelling 186
4.8 Use beyond the scope of the thesis 189
4.8.1 Models transfer 189
4.8.2 Platform: SIMBA 190
4.8.3 Decision support system (DSS) 191
4.9 Chapter bibliography 198
4.10 Supplementary material 202
4.10.1 Detailed year per year descriptive statistics of the CSH dataset 202
4.10.2 Correspondence table between the hyper-parameters of the models in R and python 203
4.10.3 Descriptive statistics of the completed dataset 208
4.10.4 Iterations of the DSS development 211
5 Chapter 5 Conclusion and perspectives 219
5.1 Story 219
5.2 Answer to the question 219
5.3 Perspectives summary 221
5.4 My takeaway 223

Table of Figures

Figure 1-1 : Protein and energy content of white clover and English ryegrass along the phenological status. Figure translated from FourragesMieux (2017) [i-148].....16

Figure 1-2: Illustration of the difference between the reflectance sensors in terms of parts of the electromagnetic spectrum covered and the number of samplings performed in that spectrum, reproduced from Zabalza (2015) [i-126] 26

Figure 1-3: Graphical representation of the organisation of the thesis.....37

Figure 2-1: Framework developed in the study to process Sentinel-1, Sentinel-2 and meteorological data to predict standing biomass in grazed pastures. The first steps are the pre-treatment and the (re-)sampling of the different dataset according to the same reference grid. Then the datasets are merged to get a tabular dataset. Transformations of the variables are then computed. Afterwards, machine learning models are trained and validated on distinct parts of the dataset. Based on the results of the validation, the most promising models are determined. 55

Figure 2-2: Location the studied farms in the agricultural areas of Belgium and their grazing parcels. 56

Figure 2-3: Distribution of the CSH acquired during the farm walks on each recorded farming area following the day and the year of acquisition, the day of the year equals one being the first of January. The error bars (whiskers) extend from the upper/lower hinge of the box to the largest/smallest value within 1.5 times the interquartile range. The theoretical frequency of acquisition was one acquisition a week. The data were aggregated through all the parcels of each farm. 65

Figure 2-4: Variable position on the two first principal components for the calibration dataset. Transformed variables were pooled with their corresponding raw ones. 67

Figure 2-5: Correlation plot of the variable rankings between the different machine learning methods used. 69

Figure 2-6: Ranking score of the 100 most relevant variables distinguished according to the type of data (shape) and the transformation applied (colour)....69

Figure 2-7 : Adjusted R-Squared curve trained on a cumulative number of variables sorted according to their ranking score. The breakpoints are represented with vertical lines. The dot-dashed lines correspond to the breakpoints taken into account, while the plain lines correspond to the other detected breakpoints. 70

Figure 2-8: Cross-validation root mean square error (RMSEcv) of the best-predicting models. 71

Figure 2-9: Validation root mean square error (RMSEv) of the best models predicting compressed sward height (mm). 72
Figure 2-10: Distribution of the original and predicted CSH values. All models that managed to provide a usable expression were used to predict CSH (mm) on the independent validation dataset. The y-axis was divided into parts of 5 mm and for each part the number of records were counted. To ease the representation, the number of observations was transformed into its logarithm of base 10. 73
Figure 2-11: Correlations of the predictions of the 20 best models on the basis of the RMSE of independent validation. 77
Figure 2-12: Distribution of the original and predicted CSH values of the independent pixellated validation dataset for the 11 best models, one per method. Each set of predictions of the models occupied a spot on the x-axis and the original pixellated validation dataset was on the left. The order of the models was the following: cubist, enet, glmnet, lasso, nnet, pcr, pls, rf, ridge, svmPoly, svmLinear2. The vertical axis corresponds to the CSH distribution: the axis was divided into parts of 5 mm of CSH and for each part the number of records was counted. 78
Figure 3-1: Design of the prediction platform 98
Figure 3-2: Workflow of the joining process 103
Figure 3-3: Repartition of the predicted dates throughout the years 110
Figure 3-4: Data acquisition dates (points) and their link to the mean daily solar radiation (line and shown y values, in J/cm²) received on all the Agromet stations. 111
Figure 3-5: Number of parcels per number of dates with data available for prediction. This may be interpreted as the probability of getting x prediction a year. The vertical lines denote the median number of occurrences..... 113
Figure 3-6: Mean per date of acquisition of the mean per parcel of the CSH predicted at the pixel level. 117
Figure 3-7: Geographical representation of the predictions for parcels known by our team. 119
Figure 3-8: Concomitance of research areas and methods, with the raw values. 132
Figure 3-9: Concomitance of research areas and methods, with values relative to the amount of research paper concerning that field. 133
Figure 3-10: Impact of the merging tolerance (time lag), applied on the S1 dataset, on the prediction compressed sward height value. 148
Figure 3-11: Impact of the merging tolerance (time lag), applied on the S2 dataset, on the prediction compressed sward height value. 149

Figure 3-12: Probability of getting x number of prediction per year for each year studied. 150
Figure 4-1: Illustration of the organisation of the discussion chapter in regard to the formula definition of the models in R. 156
Figure 4-2: Density of the distribution of the CSH per year on the left. On the right each plot represents a different parcel-time of acquisition combination... 162
Figure 4-3: Modelling process workflow. 165
Figure 4-4: Example of wide & deep & convolutional NNet thought as a perspective model to explore. MOD means a transition layer with a transfer function, conv designates a convolution layers and CSH is the final layer that translates the values into actual CSH values. 167
Figure 4-5: Parameters for the projection and resampling of the ASTER-GDEM to a 3 m resolution into the CRS EPSG 31370. 178
Figure 4-6: Summary of the gap-filling methodologies depending on their intervention location during the modelling process. 183
Figure 4-7: illustration of the meta-model concept 186
Figure 4-8: Example of wide & deep & convolutional neural networks thought as a perspective model to explore. MOD means a transition layer with a transfer function, conv designates a convolution layers, the pre-out layer could be a classical MOD and the output of the model is the CSH layer. The orange Evolution part represents the beheading process at the heart of transfer learning in NNet. 190
Figure 4-9: Mock-up of the first version of the decision support system 211
Figure 4-10: Overview of the mock-up of the two windows composing the first version of the GUI of the DSS. The left part relates to the spatial representation of the predictions and the right one to the definition of the constraints (cattle load) on the parcel to finally get a feed wedge. 212
Figure 4-11: Pop-up window for the definition of the parameters on which the server data retrieval requests are based 213
Figure 4-12: Pop-up window for parcel drawing/definition. 213
Figure 4-13: Pop-up window for the cattle need definition 213
Figure 4-14 : Blueprints of the second version of the GUI of the DSS 216

Index of Tables

Table 1-1: Ecosystem services (ES) grouped by theme. Some studies not cited in [i-138] original table were added to illustrate the actual use of these ecosystem services 7
Table 1-2: Labels and protected designation of origin promoting grazing, inspired from Soyeurt et al. (2022) [i-158] 12
Table 1-3: Cattle needs terminology. 14
Table 1-4: Energy and proteins needs, and quantity ingested by cattle category. 15
Table 1-5: Energy and protein content of feeds (mean ± standard deviation) inspired from [i-187] 15
Table 1-6 : Amounts of models reviewed per modelling algorithms reviewed, between 2000 and 2021 23
Table 1-7 : Remote sensing satellite platforms main characteristics. 29
Table 1-8 : Scientific use of satellite remote sensing platform in an agricultural context, with a focus on papers related or including pastures. 32
Table 2-1: Presentation of the machine learning algorithms explored during the training processes. The column 'Abbreviation and references' gathers the abbreviations used in the text, corresponding to the name of the method in caret R package, and some articles using the methods in related fields. 63
Table 2-2: Presentation of the machine learning algorithms explored in variable selection and training processes. The column 'Abbreviation and references' gathers the abbreviations used in the text, corresponding to the name of the method in caret R package, and some articles using the methods in related fields. 64
Table 2-3: Summary of the compressed sward height [mm] of both training and validation datasets. The aggregated lines correspond to the mean compressed sward height (CSH) per parcel per acquisition date. 65
Table 2-4 : Summary of the standardised Mahalanobis distance (GH) of the training and validation datasets compared to the training dataset. Values above five are considered as potential outliers. 66
Table 2-5: Cross-validation performances of machine learning models run on the selected subsets. 71
Table 2-6 : Validation performances of models which converged during the cross-validation for all selected subsets. 73

Table 2-7: Summary of the RMSE [mm] of the models on the CSH data aggregated at the parcel level. The difference between "Training aggregated" and "Aggregated training" was that the first case corresponded to a prediction on the aggregated training dataset and the second to an aggregation of the prediction made on the training dataset at a pixel scale. The same goes for the validation items. 74
Table 2-8 : Description of the main features of the best-performing models from the RMSE of validation on pixellated independent validation point of view. 76
Table 2-9 : Percentage of positive (%>0) and negative (%<0) residuals values for each model tested. 79
Table 3-1 : List of software and package used 97
Table 3-2: Hyper-parameters tested and final values for the models used..... 104
Table 3-3: Performances of the newly trained models with the new data and feature definition. (RMSEcv = cross-validation root mean squared error (RMSE); RMSEcal = calibration RMSE; RMSEval = validation RMSE; RPD = validation standard deviation (SD)/validation RMSE) 108
Table 3-4: Descriptive statistics of the available dataset. (*) the number of available dates = number of dates for S2 data acquisition 110
Table 3-5: Percentage of data acquired within a given time lag (dt) from the computed date..... 112
Table 3-6: Descriptive statistics of the cleaned dataset using the five studied models predicting the compressed sward height (mm). CV=coefficient of variation, the ratio of the standard deviation (SD) by the mean multiplied by 100..... 115
Table 3-7: Descriptive statistics of the coefficient of variation of the CSH computed for each date for each parcel..... 118
Table 3-8: Ranking of the most promising models according to multiple criteria 121
Table 3-9: Classification criteria and details for the models reviewed. 129
Table 3-10: References on which the short review is based 134
Table 3-11: Hyper-parameter explored during the hyper-parameter tuning . 143
Table 3-12: Performances of the models used for the platform creation 144
Table 3-13: Number of pastures and area covered between 2015 and 2020. The SAU corresponds to the global area dedicated to agricultural activities. 144
Table 3-14: Added and removed pasture areas between 2015 and 2020. Values were rounded to ha units 145
Table 3-15: Variable ranking according to their importance, from the highest to the lowest. 146

Table 3-16: Summary of the predicted CSH values at the pixel level. The left part relates to raw finite predictions (N= 1,426,156,171; 1,047,054,399; 1,482,392,774; 764,039,165 in 2018, 2019, 2020 and 2021) and the right one to data filtered to the]0:250]mm of CSH range (N= 1,425,852,569; 1,046,797,529; 1,481,945,618; 763,510,335 in 2018, 2019, 2020 and 2021). All values are in mm of CSH. 151
Table 4-1: Characterization of the stratification feature depending on the features considered. 169
Table 4-2: Summary of the criterion regarding the model selection, sd designates the standard deviation of the actual values, N the number of values, y the actual value, y the predicted value, and y the mean of the actual values. 172
Table 4-3: Input features of the considered mechanistic model for gap-filling, courtesy of J. Bindelle and E. U. Kokah , [ch-063] 187
Table 4-4: Highlights of the analysed DSS regarding the input data, the methodology, the output features, the level of user involvement, the fine spatial and temporal resolution, the intended public and the cost found. A"1" means that the DSS presents at least a partial response to the column. Other values in the cells are explained altogether with the abbreviations meaning in Table 4-5.....194
Table 4-5: Meaning of the abbreviations used in Table 4-4.....195
Table 4-6: Year per year descriptive statistics of the CSH dataset used 202
Table 4-7: Correspondence table between the hyper-parameters of the models in R and python (at the time of writing, i.e. beginning of 2023). 203
Table 4-8: Descriptive statistics of the CSH dataset with the meteorological, S1 and S2 data added. 208
Table 4-9: Meaning of the feature abbreviation of Table 4-8. 210
Table 4-10: Equations and parameters used to define the cattle need in the first mock-up of the GUI of the DSS.....214

List of abbreviations

Abbreviation	Signification
Agromet	Research project aiming at providing a network of meteorological station and thus recording at the scale of Wallonia
ANFIS	Adaptative neuro fuzzy inference system
C	Carbon
CAP	Common agricultural policy
CH_4	Methane
CNN	Convolutional neural network
CNSW	Carte numérique des sols de Wallonie
CO_2	Carbon dioxide
COP	Conference of the Parties
CRA-W	Centre de recherche agronomique wallon, research partner
CSH	Compressed sward height
CTA	Centre for agronomical technologies, research partner
CV	Cross-validation
DA	Discriminant analysis
DD	Degree-day
DEM	Digital elevation model
DSS	Decision support system
dt	Delay/time lag between the date of interest and the acquisition date of the remotely sensed data
ES	Ecosystem services
ESA	European Space Agency
FAO	Food and agriculture organization
FJA	"Fédération des jeunes agriculteurs"
FourragesMieux	Research partner, farmer adviser
GAM	Generalised additive mode
GH	Global H distance, based on Mahalanobis distance
GiS	Geographic information system
glm	Generalised linear model
GPS	Global Positioning System
gpt	Geoprocessing toolbox from the SNAP software
GRD	Ground range detected products, related to S1
GUI	Graphical user interface
hoevd	held-out external validation dataset
IDB	Index database
IoT	Internet of Things
IPCC	Intergovernmental Panel on Climate Change
L2A – BOA/TOC	Level of treatment of S2 data, corresponding to reflectance corrected to reflect bottom of atmosphere/Top of canopy values
LAI	Leaf area index
LiDAR	Light detection and ranging

lm	Linear model regression
LSTM	Long-short term memory
MAE	Mean absolute error
ML	Machine learning
MODIS	Moderate Resolution Imaging Spectroradiometer
N	Number
N_2O	Nitrous hemioxyde
NA	Not a number/invalid value
NDVI	Normalised difference vegetation index
Nnet	Neural network
NPK	Nitrogen, Phosporous and Potassium fertilisers
PAA	Anonymised agricultural parcels
PCR	Principal component regression
PDO	Protected designation of origin
PLS-R	Partial least square regression
PSwPAC	Walloon strategic plan for the CAP
RF	Random forest
RGB	Red-green-Blue colours
RMSE	Root Mean Square Error
RNN	Recurrent neural network
ROAD-STEP	Research project “Réseau d’outil d’aide à la décision pour la surveillance des troupeaux en pâtures”
RPD	Residual predicted deviation
RPM	Rising platemeter
RTK	Real-time kinematic positioning
S1	Sentinel-1
S2	Sentinel-2
SAR	Synthetic-aperture radar
SPOT-4	“Système probatoire d’observation de la Terre”
SPW	Walloon public services
SRTM	Shuttle Radar Topography mission
SVM	Support vector machine
TOF	Time of flight
UAA	Useful agricultural area
UAV	Unmanned aerial vehicle/drone
UNet	U-Shaped neural network
WalleSmart	Research project and hosting platform output of the research project
Xgboost	Extreme gradient boosted model

Preamble

Dear reader,

This thesis was a demanding journey that revealed rich in emotions, encounters, and learnings. I wanted to have you travel alongside while I relate you the beforehand steps that led to the answer to the research question entitling this thesis. Therefore, you will find "Story" paragraphs, and parts where I speak directly with you, alongside the text to have you feel this adventure-like/fairy tale atmosphere.

Chapter 1

Introduction

1.1 Outline

The goal of this introduction is to present the general prospects related to this thesis. It will start with a global context related to the research topic. Then, there will be a presentation of the vocabulary related to the scientific area studied. Afterwards, the interests of stakeholders will be presented. From these interests, the relevancy of studying grassland areas from an agronomical point of view will appear clearly. Factors influencing these grasslands will be presented. Then, modelling of grasslands will be detailed. Given the inherent constraints to pasture modelling, the scope of disciplines constituting the context of this thesis will be widened to remote sensing and its relationship to grassland related researches.

1.2 Story

An usual tale would begin with a cliché expression like “Once upon a time” or “a long time ago”, or even “in ages now long forgotten”. This story shall not begin with such platitude. First and foremost, its subject is part of the most trending topics among scientists although paradoxically its origins are rooted thousands of years ago. Indeed, grazing has been done for years, decades, centuries and even millennia. Please, keep calm and let me explain before you begin arguing about the needlessness of working again with practises that are so deeply sown in our habits that it could be called as a sacred and basic custom by now. I know, the thrill of magical feeling that some other scientific areas might provoke does not come that easily when we speak of grassland, pasture, grazing, herds, cows… Yet this subject is back on the table and I suggest that we go through it together. I might have more arousing prospects to tell you afterwards.

1.3 Grassland overview and stakeholders prospective

According to the Food and Agriculture Organisation (FAO), the percentage of grassland in the worldwide land area use was assessed to be around 25% [i-007] & [i-008]. Whilst O'Mara (2012) [i-009] also gives similar percentage of coverage, a more recent paper estimated the pasture part of the worldwide landcover, estimated as the land area without permanent ice cover, to be around 40% [i-010]. In Wallonia, according to the data provided through the anonymous agricultural repartitions of parcels (PAA) available on *Walonmap* [i-011], pastures covered in 2020 around 23.9% of the total land area and 49.9% of the total Walloon useful agricultural area (UAA), although an estimation provided by the Public Services of Wallonia |i-163] indicates that 42.6% of the UAA are permanent grasslands and 4.9% are temporary grasslands. To dig a bit further, the trends over the past 6 years were to have a decrease of the total UAA, an increase of the number of pastures, a decrease of the individual pasture size and an increase of the total pasture-dedicated area. Meanwhile this evolution of land occupancy, there is a decrease in the number of farms using these lands [i-184]. Therefore, and due to the high level of splitting up of the Walloon parcels, some farms use parcels far apart. To offer a set of tools to ease the remote management of grazing – tools that can be called decision support systems (DSS) – the Walloon Region government funded the research program ROAD-STEP ("Réseau d'outil d'aide à la décision pour la surveillance des troupeaux en pâtures", meaning network of DSS to monitor cattle herds on pasture). One of the prospects of that project is the core idea of this thesis: create a DSS aiming at assessing the available feed on pastures.

1.3.1 Grasslands VS pastures

Before going any further, I would like to make sure we understand well each term of this lexical field. Therefore, I suggest we rely on the definitions provided by Allen *et al.* (2011) [i-006] whose paper was used as a definition reference by the FAO and was updated in 2017, 2018, and 2020. As underlined in its preface, the definitions and terms used are going to change with the evolution of concepts, methods, and techniques. Therefore, it should be seen as a picture of the definitions as of now, end of 2022 - beginning of 2023. The main takeaways from this definition paper are highlighted below and some precisions are also provided.

- The term **grassland** is synonymous with **pastureland** when referring to an imposed grazing-land ecosystem. The vegetation of grassland in this context is broadly interpreted to include grasses, legumes, and other forbs, and at times woody species may be present. In this work, grasslands are thus synonymous to pastures.

- The **temporary grasslands** differ from **permanent grasslands** by their age/duration of exploitation. In the context of this thesis, the threshold is set to five years, below grasslands are considered temporary. In the context of the area covered, around 5% of the total Walloon useful agricultural area were temporary in 2019, which means around 10% of the Walloon pastures are temporary [i-015] & [i-163].
- The **food** is human edible stuff while **feed** represents animal edible stuffs that do not suit for human.
- The **forage** corresponds to the edible parts of plants, other than separated grain, that can provide feed for grazing animals or that can be harvested for feeding.
- The **sward** represents a population or community of herbaceous plants characterised by a relatively short growth and relatively continuous ground cover, including both above- and below-ground parts.
- The **leaf area index** (LAI) is the ratio of the area of green leaf (one side only) per unit area of ground. It refers to leaf only or to lamina plus half the surface area of exposed sheaths and petioles depending on the study-case.
- The **biomass** is the total dry weight of vegetation per unit area of land above a defined reference level, usually ground level, at a specific time. In the case of this thesis, the reference level is a key component for the whole system as there is a gradient of biomass and forage quality along the height of the plants (cf. the further section 1.4.1.4).
- The verb "**to graze**" means to consume predominantly herbaceous forage in situ by animals.

1.3.2 Climate: IPCC & FAO

The climate change phenomenon is nowadays widely acknowledged and the first working group of experts of the Intergovernmental Panel on Climate Change (IPCC) recently highlighted the high importance of understanding better the greenhouse gases (GHG) emissions in pastures [i-001]. Indeed, Roe *et al.* (2019) [i-002] asserted that the main GHG – namely carbon dioxyde (CO_2), methane (CH_4), and nitrous hemioxyde (N_2O) - emissions could be mitigated by grasslands. This mitigation catalyst role is further underlined by the involvement of grasslands in the global biomass stocks and subsequent carbon (C) storage [i-003]. Indeed, Gourlez de la Motte *et al.* (2016) [i-004] showed that grassland is a carbon sink and provided a calculated assessment of the average net annual productivity of this biome of $161 gC\ year^{-1} m^{-2}$ on a Walloon pasture. This estimation reflects a

current trend and Henderson *et al.* (2015) [i-005] highlighted that better grazing management could lead to higher C sequestration rates.

Unlike the recent dates of publication of the papers cited before might indicate, the awareness of the key role of grasslands in the C sequestration strategy has been arising for decades as reports published around 2009 – 2010 by the FAO underline. Indeed, they regroup the key investigations of dozens of papers back to the nineties and before like [i-012] & [i-013] & [i-014].

1.3.3 Ecosystem services

Carbon storage/sequestration is in fact only one of the many interactions of grasslands with the human society, also called ecosystem services (ES). The structure of the presentation of the ecosystem services of grasslands in this thesis is adapted from Zhao *et al.* (2020) [i-138], Sollenberger *et al.* (2019) [i-140] and the FAO [i-161]. Four main classes represent the type of contribution with the human well-being:

- Provisioning ES are related to products obtained from the ecosystem;
- Supporting ES are the ecosystem perspectives related to habitat for other species;
- Regulating ES mean the benefits due to the regulation of ecosystem processes. It is often related to the environment "raw material" such as the water, the soil and the air;
- Cultural ES are the non-material and non-consumptive benefits, often related to intellectual description and subjective perception.

Each class is then broken down in specific themes (Table 3-1), each associated with one or more ES. Some of these ES, depending on their formulation, may in fact be classified in multiple classes but in the context of this thesis, a singular relationship is considered. For instance, gene pools are described as part of provisioning ES in Zhao *et al.* (2020) [i-138] while they belong to supporting ES according to the FAO [i-160]. The relative importance of the number of ES cited in Table 3-1 is consistent with Guerra (2016) [i-135] assertion on the prospective: between multiple farming system, pastures provide the most soil preservation-related ES [i-135].

Table 1-1: Ecosystem services (ES) grouped by theme. Some studies not cited in [i-138] original table were added to illustrate the actual use of these ecosystem services

"Class"	Theme	Detailed ES	Complementary sources to [i-138]
Provisioning	Food supply	Forage production	
		Livestock production	
		Meat/milk production	
		Honey production	
	Fresh water supply	Water yield/supply	[i-134]
	Fuel supply	Biofuel supply	
	Habitat supply	Habitat for wildlife species	
	Genetic library/seed bank	Gene diversity	
		Biological diversity	
	Other biotic material (e.g. medicinal, fibers)	Raw materials	
		Sources of natural medicines	
		Wool production	
		Fibre production	
Regulating	Pest control	Prevention and control of parasites	
		Pest control	
	Waste treatment		
	Wildfire control	Control of the amount of fuel available for the fire by grazing	[i-033]
	Air quality regulation		
	Carbon storage/sequestration		[i-141]
	Water flow regulation		
	Water purification		
	Erosion regulation	Wind erosion mitigation	
		Water erosion mitigation	
		Barrier against desertification through the root structure and soil cover	[i-032]
	Soil accumulation		
	Pollination service		[i-141]

	Climate regulation		
	Nutrients delivery/retention		
Cultural	Recreation, recreational space and ecotourism	Satisfaction of visitors	
		Ecotouristic suitability	
	Aesthetic appreciation	Flower cover; canopy and shrub layer cover	
	Ecological knowledge/ education		
	Spiritual and religious		
	Horticulture/cultural identity		
Supporting		Forage production	[i-140]
		Nutrient cycling	[i-140]
		Crop pollination	[i-140]

As a side note, the relative confusion between grasslands and pasturelands that comes from the grazing status should be monitored carefully in the context of ES assessment. Although the water infiltration is higher in grasslands compared to other soil occupation, this advantage is mitigated by grazing [i-134].

The increased knowledge about the ES led to the will to compute the inherent financial implications as in Liu *et al.* (2022) [i-139]. They assessed an **annual economic value of grassland per hectare** ranging from $3,955 for semidesert grasslands to $5,466 for tropical grasslands (US dollars values). This value was a combination of avoided cost, benefit transfer, direct market value, replacement cost, and travel cost.

1.3.4 Politics: From Europe to Wallonia

The economic prospective of ES, together with the rise of awareness about the importance of pastures to mitigate levers of climate change led to politic trends. At the global level, we can speak about the Conference of the Parties (COP) and others international protocols. However, if we want to have the grassland included and stated in the recent policy, we must get down from a worldwide scale to a continental one.

At the European level, Lessire *et al.* (2019) [i-016] underlined the high political interest of the European Union to deploy greening policies with pastures being a significant lever, as stated in European texts [i-017], [i-018], [i-019] and by Tamm *et al.* (2016) [i-020]. Amidst the repercussions of this rise of awareness, the new Common Agricultural Policy (CAP) was presented in December 2019. It was written to include general guidelines and it was up to the states or relevant regional governments to write specific policy including the answers to the guidelines. In this context, studies like Henits *et al.* (2022) [i-137] in Hungary have been conducted for crop classification.

To answer the requirements of the new CAP, the Public Services of Wallonia (SPW) wrote the Walloon strategic plan for the CAP (PSwPAC) and submitted it in March 2022 to the European Commission [i-021]. The PSwPAC included focuses on the definition of pastures/grasslands, the safeguarding of permanent pastures in terms of area and dedicated a focus on the Natura 2000[1] related areas. To provide fulfilments to many sub-levels of the requirements of the CAP, the main action regarding pastures is: “preserve the permanent pastures and keep them in a good agronomical and environmental state”. In this context, studies like [i-

[1] Natura 2000 is network of nature protection areas in Europe that aims at providing protection to threatened animals and habitats [i-185]

136] have been conducted to assess the management and therefore the preservation status of the pastures.

1.3.5 Society prospective

As suggested in the ES section, grasslands and pastures have impact on the human society. Therefore, a strong implication of the public is to be expected and thus grasslands and pastures users such as farmers should account for this. For instance, the preservation of landscape and the perception of consumers are points of attention for farmers as highlighted in Lessire *et al.* (2019) [i-016] and Michaud *et al.* (2020) [i-022]. Despite the importance of grasslands, grazed or not, in the landscape, questions arise from a general resilience perspective: when we let aside the biodiversity richness brought by pastures, how can we justify using arable lands to "produce"/let grow grass to feed cattle instead of using them to produce food? Furthermore, do we need to modify these habitats and thus interfere with their potential as habitat for wildlife?

The answer to the second question is highlighted by the FAO [i-142]: without agriculture, the unmanaged natural systems can feed only 600 million people. The agriculture is thus needed to support our human society. Regarding the protein production competition prospect, complex feed/food debates exist, as highlighted by Mottet *et al.* (2017) [i-143]. On the one hand, pasture fed cattle like beef, lamb, and mutton and derived cheeses are related to the highest GHG emissions to produce 100 G of proteins once compared to chicken, pigs and plant-based alternatives, as shown in Poore and Nemecek (2018) [i-144]. The main drivers for this difference being the enteric fermentation and the feed management, i.e. the inclusion of concentrate in the feed and the improvement of pasture yields [i-144], pastures being a feed source. Adding itself to an already complex subject, the "biofuel" subjects also joined the party to compete for land use and especially plant biomass use, as highlighted by Muscat *et al.* (2020) [i-145]. This last addition led to the proposition of a priority list for the use of biomass from human food to feed to fuel [i-145]. However, this priority list is no "holy grail", parameters must be taken into account, and this leads to the other hand of the problem: the biomass and land usability.

Knowing the efficiency of meat protein production compared to the amount of protein consumed by animals, the competition for land resources that implies a clear disadvantage to the grazing strategy. However, human being can only eat and take advantage of a part of these proteins (the transfer ratio of human edibility can reach a 10-time multiplication between without and with the animal acting as a "transformer" according to Laisse *et al.* (2019) [i-177]) thus making the animals as human edible protein producers. Another prospect tipping the scale in favour

of grazing is related to the complexity of the land-use. In some areas like the Ardennes, the problem appears more clearly. Indeed, it is very hard to crop with the current techniques, technologies, and materials. For instance, using a tractor on steep areas, as found in the Ardennes, is complicated as tractors might lack power to tract the rest of the material or even tip over/fall on their side. Therefore, cattle grazing is the most sensitive choice in such areas. Regarding the provisioning needs for the health maintenance of the cattle, during the winter month in a full-grazing scenario, the non-human edible parts of the plants grown for food can actually be valued as feed to enhance the biomass valorisation and improve the ratio between the energy invested for harvest and cropping and the energy and protein content retrieved. Therefore, cattle breeding could be a way to improve the efficiency of farming.

1.3.6 Industry prospective

1.3.6.1 Farmers prospective

From an economic point of view, depending on the geographical study area, the competitive advantage of cattle grazing is more or less pronounced. Dillon *et al.* (2008) [i-026] performed a comparative analysis of the milk production cost through nine countries having different levels of grazed grass in the cow diet and a clear diminishing cost appeared along the increase of the amount of grass in the diet. Peyraud *et al.* (2010) [i-024] added the environmental prospect to this analysis and also highlighted the advantage of this production system. Lessire *et al.* (2019) [i-016] focused on an area (i.e., Wallonia) not taken in the previous analyses and ended up with the same conclusions, as did the more recent paper by Murphy *et al.* (2021) [i-025]. Reijs *et al.* [i-027] also assessed the economic viability of grazing but this time for one of the countries identified as with a low amount of grazed grass in the diet in Dillon *et al.* (2008) [i-026], The Netherlands, and they tended to find the advantage of this practice not that much pronounced, as Papadopoulou *et al.* (2020) [i-028] also found in Greece. This variability of competitiveness across areas and periods (milk prices crisis occurred between 2008 and 2021) probably influenced the perception of the advantage of grazing as it could be related to the intensive factor developed by Dalcq *et al.* (2020) [i-117].

The reasons for the grass advantage in areas like Wallonia are: a production cost of the feed relatively low, depending on the fertilisation and work levels; the high level of feed autonomy this resource provides as the nutrient profile is quite complete and only a minimal supplementation is required for demanding animals like dairy cows, although this assertion should be mitigated with the changes observed during the grazing season, as illustrated in section 1.4.1.3; the transformation of agricultural constraints (e.g. topography and thus slopes) into

opportunities, e.g. the Walloon case of the Ardennes where cropping is harder than cattle breeding given that the topological conditions don't allow for tractors safe usage.

1.3.6.2 Transformation industry

The grazing impacts also the milk characteristics which have a beneficial effect on the human health. Indeed, Chilliard *et al.* (2000) [i-041] reviewed the impact of the diet on the fatty acid profile of milk and underlined especially the link between the grass and C18:3 which translates in a change of the organoleptic properties of the products when grazing is performed. An illustration of the differences of the fatty acid profiles between grazing cows and cows fed with a total mixed ration is provided by Elgersma *et al.* (2006] [i-042], Kelly *et al.* (1998) [i-043] and Frelich *et al.* (2012) [i-044]. Beside the fatty acid profile, there is also an impact of grass-based diet on the protein content [i-045].

Therefore, for these nutritional advantages and the potential increase in well-being associated to grazing, the industry promotes grazing. Soyeurt *et al.* (2022) [i-158] identified quality labels and protected designation of origin (PDO) including the need for grazing (Table 1-2). These specifications are mostly based on the access to the pastures.

Table 1-2: Labels and protected designation of origin promoting grazing, inspired from Soyeurt et al. *(2022) [i-158]*

Name	Type	Detail
CANDIA	Specifications of a milk producer	150 days of grazing year for at least 6 hours per day
"Lait de pâturage" (aka "Milk from pastures")	Label	150 days of grazing year for at least 6 hours per day with at least 10 m² available per cow per days on the parcel [i-035]
"Grand Pâturage" (aka "Big grazing")	Label	At least 180 days of access to pastures per year [i-182]
"MARGUERITE HAPPY COW"	Specifications of a milk producer	180 days of grazing with at least 2500 m² available per cow & at least 70% of the ration must be grass [i-036]
"Abondance"	PDO specifications for cheese	Herd's ration includes >50 % grass (summer) or mostly hay (winter) [i-037] & [i-158]
"Comté"	PDO specifications for cheese	Put the cows on pastures as soon and as long as possible [i-038] & [i-158]
"Laguiole"	PDO specifications for cheese	>120 days of grazing [i-039] & [i-158] & [i-183]
"Beurre et crème d'Isigny"	PDO for dairy products	210 days of grazing [i-040]
"Camembert"	PDO specifications for cheese	180 days of grazing [i-040]

1.4 Grassland: back to an agronomic prospective

Now that the interests of the different stakeholders have been explained, diving into a traditional agronomic view of the pastures might be relevant. This "traditional" calling encompasses the following prospect: the use of grassland as feed providers. Falling under this generic scope, one can mention the following research areas: the assessment of the adequation of the grassland-sourced feed (through a direct or indirect consumption) to the animals' needs; the impact of the composition of the grassland on this adequation. Understanding the underlying relationships and, more importantly in the context of this thesis, transferring the resulting knowledge require the creation of abstracted representations of the reality, i.e. models. Multiple modelling paradigms are detailed and the temporal evolution between those is illustrated.

1.4.1 Grassland direct and indirect use for feed

This thesis is focused on the use of grassland resources to feed cattle, and more specifically through pasture grazing. However, grazing is, luckily for the winter months in Wallonia, not the only way to feed cattle. Grass-based alternatives are the dried hay (on the field or in an on-farm drier) or silage of the freshly cut grass (in windrow or plastic balls).

For the sake of completeness, it should be mentioned that the Walloon breeding systems are not all completely based on grass. Although this is a trend and a specification often required in organic farming, as far as the Walloon status, published on Biowallonie (2020) [i-110], compares to the total numbers of the "Service public de Wallonie" (2023) [i-111], organic only represents 10% of the total bovine livestock. This means that 90% of the Walloon farms could remain prone to more "classical"/intensive approaches although the use of pastures is often considered as a sound – from an economic point of view, as illustrated by Dillon *et al.* (2008) [i-026] that observed a decrease of the ration cost with the increase of the grass proportion- feed source. This so-called "classical"/intensive approach tends to consider pastures as playgrounds, or as nurseries for heifers or as feed supplement to a balanced diet. This diet should include cereals, oleaginous, protein crops, roots and the direct or industry produced derivative of these products. Theoretically, the balance between these products is the result of the research for an economic optimum while satisfying the cattle needs. However, Dalcq *et al.* (2018) [i-159] performed a study, based on the accounting sheets of 390 Walloon herds. It indicated that Walloon dairy farms used grass in their feed balance and also demonstrated a different degree of grass use depending on the level of intensification. This corroborates the highlights of Lessire *et al.* (2019) [i-016]'

survey, covering 926 exploitations. Indeed, in that survey, 96.5% of the lactating dairy cows were grazing.

In Wallonia, it is more common to find bovine than ovine and goats, as their respective population were ± 1,114,000 (in 2018) and ± 51,000 (in 2016) [i-163]. Therefore, the rest of this state-of-the-art chapter will be illustrated with bovine related features.

Cattle needs are associated with metrics that are defined differently across the world. The terms, units and definition used in Belgium for ruminants, inspired from Sillon Belge (2018) [i-146] and Beckers (2019) [i-147], are developed in Table 1-3. Beside fulfilling these needs, the feed balance should also aim to minimise noxious compounds content and offset the potential quality variability.

Table 1-3: Cattle needs terminology.

Need	Unit	Definition
Energy	VEM [VEM/kgDM]	Net energy that an animal can use for its own health and its milk production. There is no need for conversion of the VEM values between the feed. Expressed in 'VEM' per kg of dry matter
Energy	VEVI [VEVI/kgDM]	Similar to VEM, instead of relating to the energy available for milk production, it relates to the net energy available to the animal for intensive growth/fattening
Proteins	DVE [gDVE/kgDM]	Total amount of proteins that could be digested in the intestine. The amino acids might come from the direct digestion of the feed or be synthesized by the bacteria of the rumen Expressed in g of DVE per kg of dry matter
	MAT/MPT [%DM]	total protein matter/ crude protein content Expressed in % of the dry matter
	OEB	Balance between proteins and energy usable for rumen synthesis of proteins and volatile fatty acids
	VS	Measurement of the fibre content of the feed
Macro-nutrients		Ca, P, K, Na, Mg, S
Micro-nutrients		Mn, Cu, Co, Se, Zn …
Vitamins		A, D, E
Water		

From a practical point of view, VEM and DVE are the most observed indicators. However, the quality indications of these values, related to the chemical composition and the digestibility of the ration, might be hindered by sub-optimal OEB, i.e. poor balance between energy and proteins could hinder the synthesis of amino-acids and volatile fatty acid by the bacteria living in the rumen. Another hindering prospect is a sub-optimal VS, i.e. the fibre content. High enough fibre content in the feed induces increased chewing and thus increased saliva production that would regulate the pH of the rumen and ensure an appropriate rumen mobility.

1.4.1.1 Numerical approach

Before venturing into the assessment of the fulfilment of these needs, attributing numbers to these needs is required and done in Table 1-4, inspired from Beckers (2019) [i-147]. Globally, with the exception of the rapidly growing young cattle, the characteristics of the intake stay the same.

Table 1-4: Energy and proteins needs, and quantity ingested by cattle category.

Cattle category	Energy need	Proteins need (g DVE/kgDM)	Ingestion
Dairy cow	940 VEM/kg DM	80-90	9-20 kg DM/day
Young beef /bull, fast growth and fattening	1050 VEVI/kg DM	80-90	1.5% to 2% of the bodyweight

To fulfil these needs, feeds show a wide variety of characteristics. The Requasud association [i-186] provides an overview of feed analysis performed between 2006 and 2018 [i-187]. Table 1-5 illustrates the mean and standard deviation computed across all the monitored municipality. This overview does not provide definitive numbers, as Beckers (2019) [i-147] underlined the variability of the feed characteristics, altogether with the impact of the storage quality. Depending on the needs of the cattle, there might be a need to combine feed sources to fulfil all the criterion without excesses.

Table 1-5: Energy and protein content of feeds (mean ± standard deviation) inspired from [i-187]

Feed	Energy content (VEM, kgDM)	Protein content - MPT (%DM)
Maize	930.65 ± 44.32	7.45 ± 0.86
Pure ryegrass	850.5 ± 77.68	13.91 ± 3.9
Mixed grass	838.38 ± 83.53	14.21 ± 3.83

Beside the feed composition and the storage adequation, another prospect to take into account to create the feed mix is the impact of the phenological status, as illustrated in Figure 1-1 translated from FourragesMieux (2017) [i-148]. This change along the phenological status of the plant and the heterogeneity of the characteristics for different plant-based feeds advocate for a management based on mixing the feed for a better resiliency of the feed. However, understanding factors related to these differences is of high scientific interest. These are: the floristic composition; the time; the vegetation status; the plant inner heterogeneity; and the pasture inherent heterogeneity.

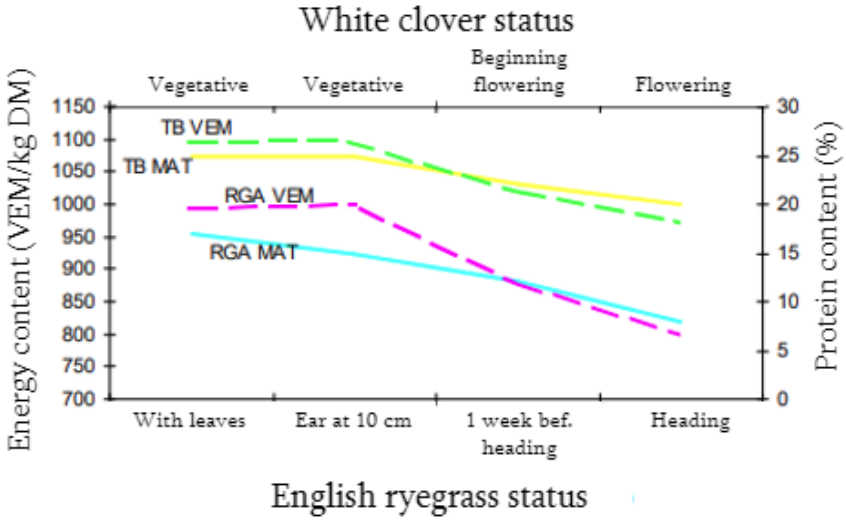


Figure 1-1 : Protein and energy content of white clover and English ryegrass along the phenological status. Figure translated from FourragesMieux (2017) [i-148].

1.4.1.2 Impact of the floristic composition

To sum up this numerical parenthesis, assessing the nutritional quality of the grass is of interest to achieve the objective of fulfilling the cattle needs. However, it is challenging as this quality will change following different factors. The main one, illustrated in Figure 1-1 with the different protein and energy content between two species, is related to the floristic composition. Indeed, as mentioned earlier, grasslands are not strictly defined by the vegetation. As long as the vegetation can be used as forage by grazing and/or cutting, there is no constraint on the exact floristic composition and grassland could include grasses, legumes, forbs, or even woody species [i-006].

This absence of constraints on the pasture composition makes more complex their modelling as models fitted for one floristic composition could not suit other conditions of applications. This characterisation of pasture composition is so important that Lavorel *et al.* (2007) [i-104] called the **classification of plants, the "Holy Grail",** for pasture modelling as it conditions the parametrisation of the models reviewed. This subject of classification is still widely discussed. As Perez-Harguindeguy *et al.* (2016) [i-107] explained, the classification could be based on taxonomic, morphological, physiological, and phenological attributes to help make more homogeneous the functions and properties of the plants, their response to environmental factors, the way they affect other trophic levels, and influence ecosystem properties. Similarly, Funk *et al.* (2017) [i-106] highlighted the high

diversity of classification strategies and their implication regarding the understanding and modelling of grasslands and the related ecological processes.

Illustrating these different prospective with different approaches might better highlight the inherent complexity of pasture plant composition.

Taking the prospective of a Walloon farmer's advisers association, FourragesMieux [i-108], the taxonomic approach is the most relevant. Indeed, in one the documents they distribute (FourragesMieux (2017) [i-148]), they recommend a mixture approach of the floristic composition: a pasture should be made of: more than 75% of Poaceae; between 10 and 20 % of legumes; and contain less than 15% of other dicotyledons, although they do not mention whether this should be percentage of dry matter or surface cover or number of plants per square metre.

Switching to a more scientific prospective, Cruz *et al.* (2002) [i-105] adopted a two-level approach: first-off they discriminated 4 main **plant forms** in grasslands i.e., Poaceae, legumes, "rosettes", and the other dicotyledons; then they focused on Poaceae and discarded the rest of the plants, despite scientific illustrations of interaction resulting in nutrient transfer between plants [i-178]. As Graux (2017) [i-055] mentioned, Cruz *et al.* (2002) [i-105] discriminated 4 groups within that Poaceae family were defined based on the combination of the trend to capture loads or only small amounts of resource and on rapid or slow change in the foliation level, classification that was used in models such the one developed by Jouven *et al.* (2006) [i-109]. Cruz *et al.* (2010) [i-115] enlarged this discrimination of Poaceae to 6 groups and included the notion of potential use. Cruz *et al.* (2002) [i-105] and Cruz *et al.* (2010) [i-115] based their classification on the species level and they associate a species to a group.

This association of a species to a "group" and thus this association of taxonomic classification to functions is still being discussed and another type of classification, i.e., plant classification based on "traits", is gaining more and more interest from the scientific community. Violle *et al.* (2007) [i-102] attempted to fix the confusion in "traits" terminology for plant studies and defined multiple levels to discriminate traits and parameters. The smallest unit was the individual level and traits were defined at that individual level without reference to the environment, the functional traits being morpho-physio-phenological traits which have an impact on fitness [i-102]. In the meantime, they proposed a standardisation of the equation related to the plant traits computation stressed the importance of integration functions to scale from individuals to ecosystems [i-102]. Based on these functional traits Diaz *et al.* (2013) [i-120] developed a framework to assess species' effects on ecosystems and their tolerance of environmental changes.

Similarly, based on these functional traits, Diaz *et al.* (2015) [i-188] explored the spectrum of functional traits and provided a standardised tool to compare plants.

1.4.1.3 Time and vegetation status relationships

As illustrated in Figure 1-1 translated from FourragesMieux (2017) [i-148], the nutritional quality of the grass "mixture" does not stay the same along the time. Therefore, computing the nutritional quality of the grass based on the sole floristic composition is not possible. The specific behaviour illustrated in Figure 1-1 leads to advice to farmers such as intensive grazing and mowing promotion to avoid advance in the phenological status. In more concrete terms, it is often recommended to place the cattle on pastures when the sward height is between 10 and 15 cm FourragesMieux (2017) [i-148]. Higher sward cover translates into higher number of ears and thus a decrease of the nutritional quality and the palatability of the grass mixture.

At a smaller scale, the **nutritional quality of each plant** depends on meteorological features such as: the temperature, the water availability (and thus the precipitation and the soil geopedological and topological characteristics), and the solar radiation for the photosynthesis. This translates, at the larger scale of Wallonia, into a season with intensive grass growth from the end of March until the beginning of November, although, as underlined in FourragesMieux (2017) [i-148], there could be a drop in the grassland production during the summer (around August) as some grasses (alike some ryegrass) don't like hot weather. This drop is dependent on the species and functional groups present in the pasture [i-152], and might also express itself as a structural change inside the plant organism [i-164].

Resulting from this temporal production-feed need difference, feeding cattle requires to plan its grazing and include mowing events and grass drying or silage, or even include other feed sources for the winter months. This planning requires models to assess the available feed.

1.4.1.4 Heterogeneity inside the plant

Moreover, before starting to talk about the pasture modelling, a key element has to be defined now: a plant is not a uniform organism. It has roots, stem and leaves. Within these compartments there are variations in terms of chemical constituent and properties. In the case of the stem, there is a lignification gradient, with a higher level of lignification at the basis and lower at the higher end. Similar gradients exist for nitrates [i-097] and other chemical constituents. It matters in the context of this thesis as cows mechanically don't have access to the full height of the stem and mowing events need to define a lower cut height threshold. Therefore, considering the full height might not always be relevant, especially in

the context of assessing the available feed for cows, in terms of quantity and quality.

1.4.1.5 Pasture spatial heterogeneity

As suggested, pastures are heterogeneous in terms of species and the dependency towards local conditions contributes to a spatial heterogeneity of the sward cover. Notwithstanding topological specificities, this spatial heterogeneity is in fact inherent to grazed pastures as (1) the animals show spatial and specific selectivity when they graze [i-151], (2) there is an asynchronism between the relatively steady feed consumption by animals and the seasonal growth dynamics, (3) plant defoliation lasts for a few seconds while the growth afterwards takes weeks. As a result, extensive grazing practices tend to increase heterogeneity compared to intensive practices [i-153]. Beside the will to seize the heterogeneity of the feed available, better understanding and reflecting the spatial repartition of the feed is critical for the farm management, as Şahin Demirbağ *et al.* (2009) [i-154] showed that areas with a sward height superior to 12 cm represent more than 50% of the total pasture growth at the beginning of the growing season.

1.4.2 Grassland modelling

1.4.2.1 Ways to understand grasslands

As stated earlier, grasslands grazing has been done for centuries. Human's experience added up and trends were progressively understood and represented under the form of abstracted representation of the reality i.e. models. These models can be distinguished based on the level of knowledge of the studied phenomenon and on the acceptation of randomness. Regarding the level of knowledge, models are discriminated between mechanistic or empirical models. The first ones use a theory to predict a phenomenon whilst the second ones study real-world events. These two categories are loose and there exists a gradient of inclusion of theories and experiments in the modelling. Regarding the acceptation of randomness, models that do not accept randomness in the modelling process, and thus always deliver the same prediction for the same data entering the process, are called deterministic. These deterministic models are opposed to stochastic ones that are characterised by the consideration of uncertainty and randomness. The two categorisations (mechanistic VS empirical and deterministic VS stochastic) can be combined. A further layer of characterisation, that also can be added to the two other characterisation dimensions, is the automation of the model creation through informatic support, also called machine learning. Lastly, another layer of characterisation is the consideration of time that define static (no consideration)

and dynamic (consideration of time with e.g., differential equations) models. [i-055]

The **importance of the models**, no matter their categorisation, is twofold. On the one-hand these are crucial for a **transmission of the information**. On the other hand, if they are developed to perform the **assessment of a feature** (i.e. a category of information, also called a variable in some contexts) of interest from another one called a proxy, they allow for the recording and analysis of the proxy to determine the actual feature of interest. In the case of grasslands grazing, as explained earlier, the amount of biomass, and the content of energy, protein, micro- and macro-nutrients and water are to be accounted. Although some variations have been shown and demonstrated earlier regarding the content of the different component in the dried feed, the scientific standard of the feed availability assessment is the dry biomass. The method to assess this standard is 1) cut 3 or 4 sub-parcels of the parcel of interest (the area is often a quarter of a square meter), 2) dry and weight the mown grass to assess the dry biomass in kg, and 3) extrapolate that biomass to the total area of the parcel of interest, with the assumption that the entirety of the pasture is covered the same way [i-123], [i-149] and [i-150]. This method presents the inconvenient to be destructive (dried mown grass is not valued anymore), time-consuming, and there is a potential bias coming from the assumption of the parcel homogeneity, although grazed areas are known to be heterogeneous [i-151] (section 1.4.1.5). Furthermore, these measurements need to be repeated during the grazing season [i-165]. A model using rapidly and non-destructively acquired data, such as compressed sward height (CSH) – recorded with a rising platemeter (RPM) or a falling one, as proxies allows to bypass some of the disadvantages [i-162], notably with the combination with a GPS [i-166]. However, this better representation of the spatial heterogeneity requires a specific attention to the number and the repartition of the sampling points to minimize the sampling error [i-171]. Another advantage of this CSH proxy is its wide acceptation by farmers and the scientific community [i-131], [i-066] and [i-149].

The biomass is not the only **feature** modelled to assess the **feed availability** on the pastures. During a recent review of the literature covering articles between 2000 and 2021 searching for ways to understand and manage pastures, we discriminated mechanistic and deterministic models and found 11 published models corresponding to these combined designations. Most models aimed at modelling one or more of the following grass-focused indicators: the biomass [i-124], [i-119] and [i-125]; a form of grassland height [i-114]; the LAI [i-063], [i-116] and [i-068]; the quality of the grass indicators like the structural composition [i-

118], [i-109], [i-119] and [i-074]. One of the reviewed models, called "SEPATOU", focused on evaluating multiple farm management strategies through their simulation [i-112] and [i-113]. The last main prospect modelled were the resulting soil and environmental conditions [i-121], [i-122] and [i-123] of the pasture management. Studies based on other model categorisation focused on the other feed indicators: the response of the sward cover to an ultrasonic exposure [i-167], [i-168] and [i-169]; the response of the sward cover to an electrical exposure [i-170]; the height deduction from ground-taken pictures [i-094].

Focusing on the **modelling** process revealed that most models used linear equations, some included feature transformation, based on the **biophysical processes**. The input features of those models were mainly meteorological data (temperature, precipitations, solar radiation, evapotranspiration, wind speed and direction, and the relative humidity), soil related data (drainage, geopedological conditions and soil water content), pasture management indicators (NPK input, management event like mowing and grazing) and grass properties (including the initial biomass/LAI status). As Shalloo *et al.* (2018) [i-029] and Shalloo *et al.* (2021) [i-030] stated, grass-based livestock ruminant production did not completely leverage the whole range of advances in precision technologies to better understand and manage pasture, most probably due to the constraints inherent to outdoor applications. In this context, Shishodia *et al.* (2020) [i-031] proposed a wide range of technologies, methods and applications related to remote sensing and the internet of things currently used in precision agriculture in general. However, only three studies out of the pool of mechanistic and deterministic studies reviewed related to biophysical modelling [i-063], [i-068] and [i-074] tried to get out of that pure biophysical modelling and linked the models to visible/hyperspectral remotely sensed data (acquired through SPOT-4, Formosat-2 and Planetscope satellites) although there has been a rise of the use for this type of data in recent studies focused on grassland understanding. A next section will deepen this aspect of remote sensing data implication in grassland studying.

Concerning the **modelling prospect** of the papers reviewed, the main trends were a change in the representation of the categories: in the recent years, there was an increase of the proportion of empirical models altogether with an increase of the stochastic modelling and with a rise of machine learning. These paradigm shifts altogether with the increasing use of remote sensing represent a challenge regarding the data science and engineering approach as the related databases grew in size and largely outgrew the methods and algorithms developed for less demanding modelling. This change in size induced a need for revising algorithms and tests as the hardware needed to satisfy the old tests had to evolve (e.g. to

compute a covariance matrix requires a matrix inversion and therefore loads of RAM once you get in big datasets, like >50.000 rows) and the power of the previous tests applied on huge datasets was too high and therefore, even non-meaningful information could be flagged as significant, therefore impacting the results.

For the sake of better understanding pastures and the different prospective/use-cases, **models** and algorithms had to **evolve** and had the opportunity to evolve thanks to recent hardware advancements (increased number of CPU cores, CPU hyper-threading, increased CPU-IPC, increased clocks on the CPU and the RAM, addition of computation accelerators, …). This can be seen in our recent review as all the studies we found before 2014 relied on linear regressions with variable transformation (sometimes resulting in "exponential regressions") no matter the "categorisation" of the model. In 2014, we saw the first forms of neural networks (NNet) appear in Ali *et al.* (2014) [i-069], under the form of perceptron and adaptative neuro fuzzy inference system (ANFIS). In 2016, Monte-Carlo simulations joined the portfolio of algorithms used in pasture data science in Nakagami (2016) [i-127]. The same year the Cubist hybrid approach combining decision-tree principles to linear regression, resulting in iteratively corrected linear regressions, also joined the toolbox [i-073], alongside a version of the partial least square regression (PLS-R) modified to deal with sparse data [i-093]. The unmodified PLS-R [i-066], random forests (RF) [i-095] and the principal component regression (PCR) [i-070] only appeared in 2017 in our review. Support vector machines (SVM) only appeared in 2018 in Pullanagari *et al.* (2018)[i-128]. 2019 witnessed the rise of recurrent (RNN) [i-056], long short-term memory (LSTM) [i-129] and "U-shaped" (U-Net) [i-086] NNet and the use of the generalised linear model theory (glm) [i-131]. In 2020, extreme gradient boosted models (xgboost) [i-079] also joined the ball. Finally in 2021, the generalised additive model (GAM) [i-132] and the convolutional NNet (CNN) [i-133] were added to the inventory of modelling tools. The resulting number of models using each algorithm is shown in Table 1-6. These presented occurrence values should not cast shadow on the numerous studies combining the algorithms and on the fact that we reviewed models. Therefore, a team that made more than one paper on one model appears as only one entry in our 93-entry model review to avoid over-representing some technical algorithms mastered only by a small number of team.

Table 1-6 : Amounts of models reviewed per modelling algorithms reviewed, between 2000 and 2021

Modelling algorithm	Amount of models based on these algorithm
Exponential linear regression	7
Multiple linear regression (MLR)	31
Logistic regression	1
Generalised linear model (glm)	2
Mechanistic and deterministic model	11
Monte carlo simulation	1
Discriminant analysis (DA)	3
Principal component analysis/regression (PCR)	5
Generalised additive model (GAM)	1
Partial least square regression (PLS-R)	10
Sparse Partial least square (S-PLSR)	3
random forest (RF)	13
cubist	5
Support vector machine/regression (SVM/SVR)	5
Adaptative neuro fuzzy inference system (ANFIS)	3
Perceptron	10
Recurrent neural network (RNN)	3
U neural network (U-NET)	1
Convolutionnal neural network (CNN)	1
Long-short term memory neural network (LSTM)	2
Extreme gradient boosted algorithm (xgboost)	1

1.4.2.2 And now that we model pastures, what are we going to do?

Modelling grassland is great from the scientific, political, and economical prospective and for the technological opportunities it represents. However, science for science does not help the majority of the population, to do so requires transferring the knowledge to field actors. Therefore, beside the *in silico* approaches, more practical approaches have been developed. In Wallonia, the FourragesMieux association [i-108] gained importance in the recent years by offering information and advice based on actual values, e.g.: yields for different varieties of grasses or technical deed and management practices implications. To automate and widen the range of people helped and advised, the main alternative to this type of institution is to elaborate decision support system (DSS). This alternative should in fact be seen as a “and/or”/ non-exclusive solution, as it might be a complementary tool for such institution.

Unfortunately, only 11% of the 93 models studied in the literature review mentioned earlier are implemented in a DSS. The reasons for this poor transfer rate to the end-users might either be due to the poor performances of the underlying models as stated in Cockburn (2020) [i-048] or to a lack of adoption by the said end-users (here farmers). The latter option is justified in Eastwood *et al.* (2016) [i-049] with the DSS design and information delivery choices not matching

the farmers' expectations. Another prospect also hindering the actual decision-making is the time lag induced by the information overload and the integration of new data sources and treatments [i-050]. To sum up, to promote the adoption of a DSS, the **key points** are being **cheap** and **rapid** and provide **relevant** information. To enhance this prospect of relevancy of information and increase the adoption rate of this type of helping tool, modularity should also be noted as recommended in Power (2003) [i-051].

1.5 Grassland and remote sensing

1.5.1 Story

As you may have noticed, this research area is not yet much explored. I know, it is a bit disappointing and I don't want to slay your expectations. Hence, I suggest we move on to another mind arousing subject that is at the hearth of this phD thesis tale: remotely sensed data. It is commonly defined as data acquisition about an object or phenomenon without physical contact. Currently, it mainly refers to data acquired from satellites and drones/unmanned aerial vehicles. You are probably wondering on how these highly technological devices relate to pastures. I suggest we quickly review the current state of its use and implementation in the grassland and pasture scientific areas.

1.5.2 Categories description

As stated earlier, there is a rise in the use of remote sensing technologies, as highlighted in Reinermann *et al.* (2020) [i-046] and confirmed by a short up to date review we performed (the same that led to the previous grassland modelling section). Most papers using it underline its convenience to seize the spatial heterogeneity of the pastures and therefore have a better grasp of the grazing dynamics [i-047] and decrease the risk of sampling bias, spatially speaking.

In the recent years, there has been an increase in the use of remotely sensed data to study grasslands and other natural areas. Garioud *et al.* (2021) [i-065] offers a review of remote sensing application in a grassland context. Given the time of publishing of that review, we had already performed a review and now offer you another approach. We can mainly categorize the considered data according to the acquisition platform and the type of sensor. Moreover, the overview can be completed by the treatments applied to the data recorded and the sensitivity/weaknesses of the sensors.

Concerning the platforms, we identified the following, although we did not find scientific literature on grasslands application for every one of them:

- Satellites
- Unmanned aerial vehicles/drones
- Airplanes
- Balloons
- Kites

Concerning the types of sensors, we can distinguish them based on their passive or active operation, the spatial and sensor specific resolution and range. Therefore,

offering a classification of sensors is complex. Here we first distinguish passive from active sensors.

Passive sensors, in this remote sensing context, are associated to a measure of reflectance. As illustrated in Figure 1-2, further discrimination is based on the parts of the electromagnetic spectrum sampled and the number of samplings performed. With an increasing sampling frequency, one can cite visible, multispectral, and hyperspectral sensors. Visible sensors often operate in a tri-channel mode, i.e. they sample in Red-Green-Blue, hence they are sometimes called RGB sensors. Multi- and hyperspectral sensors sample spectrum outside of the visible, for instance ultraviolet and infra-red, Hence the usage of the Vis-NIR calling for some sensors that operate in the visible and near-infrared wavelengths. Beside these sensors types, thermal cameras are also passive sensors, they sample in the far-infrared area of the electro-magnetic spectrum.

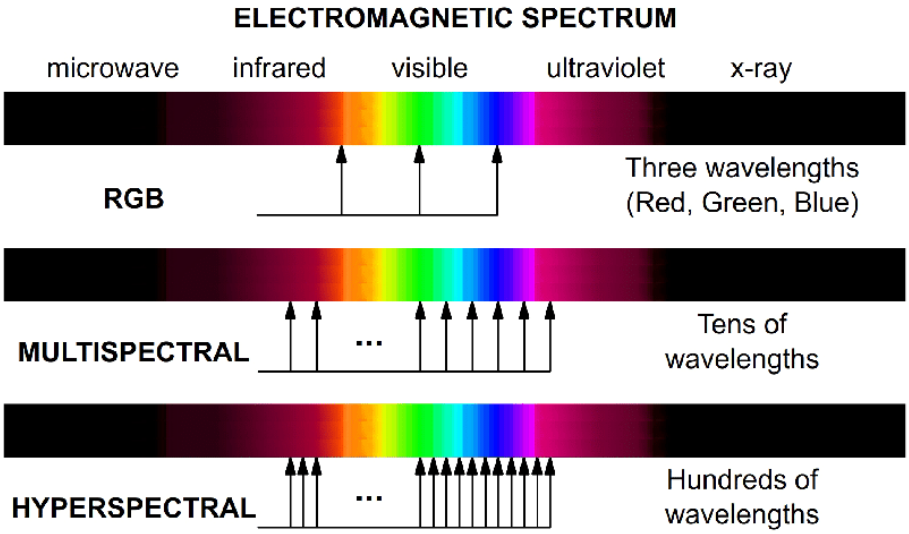


Figure 1-2: Illustration of the difference between the reflectance sensors in terms of parts of the electromagnetic spectrum covered and the number of samplings performed in that spectrum, reproduced from Zabalza (2015) [i-126]

Active sensors are based on the emission of electromagnetic waves and reception and analysis of the reflected signal. The type of signal emitted and received discriminates:

- **Radar** sensors emit electromagnetic signal. The most determining factor is the wavelength. In the case of this literature search, only the C and X bands were found to be used. The respective length wave and frequency range covered of both bands are: 3.75-7.5 cm and 2.5-3.75 cm, 4-8 GHz and 8-12 GHz, respectively.
- **LiDar** sensors emit a laser signal. The frequency range is beyond 10 THz while the wavelengths range from 250 nm to 10 µm depending on the laser.

Concerning the sensitivity of the sensors to exterior conditions, reflectance sensors are very sensitive to illumination conditions. Satellite-based reflectance sensors acquire irrelevant information for earth surface monitoring when clouds are on the way of the reflected light. Drone-based sensors could still provide relevant information for structure from motion photogrammetry under homogeneous cloudy conditions if they are equipped with a correction sensor measuring the incident light energy. The radar sensors are also sensitive to the air and soil humidity as shown in Tamm *et al.* (2016) [i-020] regarding the impact of the rain on the S1 signal.

Concerning the treatments, besides the utilisation "as it" of the spatialized imagery ("simple" feature extraction), other treatments exist in the remote sensing field:

- **Photogrammetry**: construction of three dimensional models to perform measurements from images (that could result from an image stitching process);
- **Image stitching**: construction of a global image from multiple small images. The goals can be to improve the resolution by introducing different view angles and increase the area covered. The most common algorithm, in the context of remote sensing and at the time of writing, is known as structure from motion (**sfm**);
- **Geocoding**: It is the step that transforms generated images into a map projection;
- **Convolution treatments** allows to loosen the focus and the analysis of a wider scope of data.

Another type of surface monitoring is proximal sensing. It consists in using sensors while being ground-based, these sensors being hand-held [i-181], pole-mounted [i-180] or mounted on rolling platforms [i-179]. Most sensors of the remote sensing area could be used here. We monitored only one "type" of sensor, not present in the remote sensing area: Red-Green-Blue-Depth (RGB-D) maps generating sensors. It includes time of flight (TOF) cameras and structured light

sensors. For the types of sensors already present in the remote sensing area, many of these sensors were in fact used to get "reference ground measurements" in combination with other data sources like in Lugassi *et al.* (2019) [i-099]. This data source is not considered in the rest of this work as it requires measurements from the ground.

1.5.3 Practical characteristics of remote sensing platforms

1.5.3.1 Comparative approach

A brief introduction about the characteristics of the main **satellite** constellations found in the literature related to pastures monitoring is proposed in Table 1-7. Although the acquisition cost column might indicate that some data are free to acquire, hidden costs are inherent to:

- the download: there is a need for a sufficient bandwidth to download rapidly enough the data to avoid signal and connection losses. Furthermore, the amount of data consumed is consequent.
- the storage: all the data downloaded must be stored at a certain point in time.
- the processing: some products have to be translated into actually usable products for the study.

To illustrate better the hidden cost, the specific case of Sentinel-2 (S2) products might be of interest. 8 S2 Tiles cover the area of Wallonia. Each weight around 2 GB. This means that this many data have to transit through the internet connection, consuming bandwidth and data quotas for each date of acquisition. Afterwards, depending on the data management strategy, these tiles should be stored under one or more format. This consumes disk space and thus requires investments in hardware. Afterwards, the S2 data don't require much pre-processing. Other products, such as S1 do. Indeed, S1 products need to be geocoded to become easily usable in GiS. This represents another hardware requirements based on the processing power. For all these reasons, freely acquired data are sometimes referenced to as low-cost data and not freely available.

Table 1-7 : Remote sensing satellite platforms main characteristics.

Name	Nature of the acquisition	Acquisition frequency	Acquisition width	Spatial resolution	Acquisition cost ?	Still acquiring ?
Sentinel-1	Radar C-band	3-5 days	+- 80 km	+- 5 m	NO	YES (1/2)
TerraSAR-X	Radar X-band	2.5 days		+-1 m	YES	NO
Spot-4	multispectral	/		10 m/20 m	YES	NO
Modis	multispectral	1 day	2330 km	250 m/1000 m	NO	YES
Landsat-8	multispectral	16 days	185 km	15 m/30 m/100 m	NO	YES
Formosat-2	multispectral	/	24 km*24 km	8 m		NO
WorldView-2	multispectral	1.1 day	16.4 km	0.46 m/1.84 m	YES	YES
Sentinel-2	multispectral	3-5 days	290 km	10 m/20 m/60 m	NO	YES
Planet : Dove	multispectral	1 day	24.6 km	3 m/5 m	YES	YES
Planet : Skysat	multispectral	5 to 7 times a day	11 km	0.5 m	YES	YES
Planet : Rapid Eye	multispectral	/	77 km	5 m	YES	NO

Concerning the use of unmanned aerial vehicles like **drones**, gathering these characteristics is not as easy as for satellites as each acquisition campaign could be different given the amount of combination possible. Indeed, in such study, a choice has to be made regarding: 1) the platform between a "fixed-wing" or a (multi-)rotor drone; 2) the type of sensor (Red-Green-Blue (RGB), multispectral, hyperspectral, thermal, …); 3) the spatial resolution needs for the study; 4) the battery availability; 5) the time and storage limitations for the acquisition. Given the huge amount of combination possible, comparative analysis such as Table 1-7 is not conduced for drones. However, for the record, using this technology has costs in terms of platform, human time and computation power after the acquisition that are partly dissolved in the case of satellite data as they are founded by public organisations, or the companies intend to share the cost between multiple users.

Airplanes are an air-borne platform mainly used in cases where the payload is too heavy for a drone or the range to cover is too big to get a proper resolution within a decent time, while the cost and recurrence of acquisition, or the payload, or technical constraints do not justify satellites. The last 2 airborne acquisition platforms (**balloons** and **kites**) were not found to be used in studies related to grassland.

Given the will to get a decision support system with the lowest costs possible, it was chosen to discard the platforms requiring human time for the acquisition. As a result, only the satellites platform seemed sensitive for their higher automation level.

1.5.3.2 Complementary information for each platform

Sentinel-1 (S1) is a satellite constellation conducted by the European space agency in the context of the Copernicus program. Each satellite operates a C-band synthetic aperture radar instrument. At the time of writing, out of the two satellites operating during the first phase, only one remains operational and two more are planned. When the two first satellites were operational, one might have expected a recording every 2 to 3 days over Wallonia, accounting for overlaps in data acquisition. Nowadays, you should rather expect a 3 to 5 days revisit frequency. Given the increased interest since the launch, there has been an interest for standardising the geoprocessing, e.g. Filipponi *et al.* (2019) [i-052] and Filipponi *et al.* (2020) [i-053] presented a standardised framework based on the s1tbx which is part of the SNAP software edited by the ESA [i-054]. The underlying datasets used for corrections present the advantage of being widely available. There are multiple operational/acquisition modes and operating patterns for these satellites. Over the area of interest in this thesis, we focused on ground range detected (GRD) products, which means without the phase information with VV and VH polarization and a spatial resolution around 5 m.

TerraSAR-X is an older satellite operating in the X-band conducted by a public-private-partnership.

Spot-4 was part of the 7 satellites of the SPOT constellation (SPOT stands in French for "Système probatoire d'observation de la Terre", lit. "Satellite for observation of Earth"). It was operated by a French research institute, the CNES, and it stopped working in 2013. It acquired multispectral imagery: RG & Near- and mid-infrared; with a resolution of 10 to 20 m. According to Di Bella *et al.* (2004) [i-063] it acquired in the blue range while ESA does not mention that range of wavelength [i-062].

The Moderate Resolution Imaging Spectroradiometer (MODIS) sensors are onboard two satellites and data can be retrieved from the NASA. They record 36 spectral bands from 620 nm to 14.0385 µm with spatial resolution ranging from 250 m to 1.000 m.

Landsat-8 is a multi-spectral satellite acquiring 11 bands with medium spatial resolutions (15-30-100 m depending on the band) with a width of acquisition of 185 km and it is operated by a national organisation of the USA.

Formosat-2 was a satellite acquiring in the RGB-Near infrared with a spatial resolution of 8 m, sensor footprint of 24*24 km, temporal resolution of one day operated by a national organisation of Taiwan. It was decommissioned in 2016.

WorldView-2 is a commercial satellite providing panchromatic imagery with 0.46 m of resolution and a multispectral imagery, including the red-edge and the near-infrared with 1.84 m of resolution.

S2 is a satellite constellation conducted by the European space agency in the context of the Copernicus program. These are multi-spectral satellites acquiring 12 bands with medium spatial resolutions (10-20-60 m depending on the band) with a width of acquisition of 290 km.

Planet is an American private corporation operating two constellations of satellites now, a third was retired in 2020:

- Dove: also called cubesats. These satellites have a 3 to 5 m resolution and an acquisition width of 24.6 km
- Skysat: these 21 satellites have nowadays a resolution of approximately 0.5 m and an acquisition width of 11 km; they acquire RGB, and near infrared bands.
- RapidEye: now retired. The satellites were recording reflectance in RGB, red-edge and near-infrared. It had a spatial resolution of 5 m and an acquisition width of 77 km

Although the 3 constellations of Planet satellite are represented separately in Table 1-7, they will be designed as the Planet satellites in the rest of this section dedicated to the description of the remotely sensed data as they present similar characteristics, especially regarding the cross-sensor radiometric consistency.

1.5.4 Applications of satellite remote sensing on pastures

Regarding the application of remotely sensed data, we identified 9 main scientific areas of interests more or less closely related to pastures: the detection of mowing events (MowEve); the validation and exploration of other data sources/modelling tools, also sometimes called “data fusion” (Val_df); the pasture phenology determination (Phen); the LAI estimation; the estimation of the pasture biomass (Biom); the estimation of any form of grass height (HH); the estimation of the grass quality (QH); the longitudinal -temporally speaking- monitoring of pastures (MonLon); the land cover classification (Clc). Per platform per study area classification of scientific papers is represented in Table 1-8.

The impressive number of references in the “Val_df” column, representing the data fusion related studies, is explained by the scientific curiosity regarding the possibility to use data acquired at different moments, the use of data of different natures, the combination of multiple modelling paradigms, and the integration of spatialised data to model more precisely parameters than with simple extrapolation. This last justification is also shown in the “Biom” column.

Table 1-8 : Scientific use of satellite remote sensing platform in an agricultural context, with a focus on papers related or including pastures.

Name	MowEve	Val_df	Phen	LAI	Biom	HH	QH	MonLon	Clc
S1	[i-020] [i-056] [i-010] [i-097] [i-136]	[i-059] [i-097] [i-060] [i-136]	[i-058] [i-060]	[i-060]	[i-060]	[i-130]		[i-059]	
TerraSAR-X			[i-061]						
Spot-4		[i-063]	[i-064]		[i-064]				
Modis		[i-073] [i-101]		[i-088]	[i-069] [i-070] [i-072] [i-088]	[i-101]			[i-071]
Landsat-8		[i-073] [i-067] [i-085] [i-087] [i-088] [i-095] [i-101] [i-060]	[i-060]	[i-095] [i-060]	[i-072] [i-078] [i-088] [i-060]	[i-094] [i-101]			[i-090] [i-064]
Formosat-2		[i-068]							
WorldView-2		[i-067] [i-066]			[i-067] [i-066]				
S2	[i-097] [i-136]	[i-067] [i-085] [i-093] [i-096] [i-097] [i-099] [i-100] [i-060] [i-103] [i-136]	[i-077] [i-058] [i-060]	[i-100] [i-060]	[i-076] [i-077] [i-130] [i-080] [i-082] [i-084] [i-092] [i-093] [i-098] [i-060] [i-103]	[i-130] [i-089] [i-094]	[i-080] [i-084] [i-089] [i-098] [i-099]		[i-081] [i-090] [i-096] [i-137]
Planet	[i-136]	[i-073] [i-074] [i-085] [i-087] [i-095] [i-096] [i-100] [i-103] [i-136]		[i-095] [i-100]	[i-079] [i-083] [i-091] [i-103]	[i-079] [i-091]			[i-086] [i-096]

For the sake of completeness and although drones have been discarded for the lack of automation of the acquisition, it should be mentioned that this is an active area of research related to pastures and grasslands. The main trends found in the literature were the use of sfm algorithms to combine the images acquired into

bigger point-clouds and maps. The resulting modelling is a height map [i-172]. This output is often associated with an analysis of spectral indicators (direct reflectance values or composite indices combining multiple bands such as the normalised difference vegetation index (**NDVI**)) [i-173]. As for the other platforms, models have been developed and tested to confront these derived features to feed quality indicators such as the dry biomass, the protein content, … [i-174] and [i-175]. The question of the spatial representativity of the measurements through drones is also heavily discussed [i-174].

Given the low cost of S1 and S2 data, given the high number of studies using these data and given the quality of the products (e.g. high calibration of the spectral resolution), it was chosen to work with S1 and S2 data in the context of this thesis.

1.6 Thesis objective

This journey explored the grasslands stakeholders prospective (section 1.3), the agronomic point of view (section 1.4) and the relationship of grasslands with remote sensing (section 1.5). One of the main takeaways is the non-adequation of the feed need of the ruminants with the grass growth (section 1.4.1.3). Therefore, managing grazing is a complex task. This job is eased with an accurate knowledge of the available feed. The most used reference is the assessment of the available biomass (section 1.4.2.1). The scientific standard to perform the assessment is the cutting-drying-weighting-spatial interpolation process (section 1.4.1.5). Unfortunately, this methodology is destructive, requires to be repeated, is time consuming and thus costly, and the related spatialisation is limited (section 1.4.2.1). Therefore, there is an interest in using other indicators directly or indirectly related to the grass growth and the feed availability, such as CSH (section 1.4.2.1).

In the same context of acquisition constraints, remotely sensed data present the attractive advantage of a reduced recording time requirement compared to the reference method and the spatial heterogeneity of the parcels is directly considered, within the limits set by the resolution of the sensor. In the context of this thesis, the goal was to get fine spatial and temporal resolution at the lowest cost possible. The scope of remotely sensed data was scanned regarding optical/multi-spectral and SAR data. In that context, for their high-quality standard, no-cost, fine spatial and temporal resolution, and relative ease of access, S1 and S2 constellations were selected as data providers (section 1.5.3). Both have already been used in pasture related areas (section 1.5.4) such as mowing event detection [i-020], the assessment of the LAI or the above-ground biomass [i-060], [i-134], [i-067] and [i-076], the phenology of pastures determination [i-058]. Supporting this combination choice, some studies already included the combination of both data sources [i-097] and [i-058].

An additional source of information are the meteorological data as vegetation is sensitive to meteorological features such as the solar radiation, the temperature and the precipitations to grow, as illustrated in the choice for these data entering most mechanistic models (section 1.4.2.1). Therefore, meteorological data were used altogether with the remotely sensed data. The innovative prospect of this thesis is to mix innovation steps regarding remotely sensed data together with meteorological data to apply the resulting model on an under-studied research area, the pastures, using state of the art machine learning modelling algorithms, while keeping transferability to a larger scale in mind.

Indeed, as illustrated in section 1.4.2.1, modelling algorithms have evolved in the recent years. One could sum up the evolution as a transition from linear regression to automated learning models, also called the machine-learning area. A review of the literature highlighted the dominance of multi-feature linear regression, RF, and NNet with varying architectures. Beside these modelling algorithms, the nature of the relationship between the proxies and the modelled feature was also questioned as Oliveira *et al.* (2020) [i-155] tested multiple feature transformation within machine learning models (the range of models explored came from Fernández-Delgado *et al.* (2019) [i-156] and Garcia *et al.* (2020) [i-157]).

All these innovations were integrated to overcome the complexity to transfer the models to an actual decision support system, as illustrated with the 11% of achievement mentioned in section 1.4.2.2. The barriers identified for the specific pasture study-case were the complexity (time and meticulousness) of the data acquisition process; the complexity of the conversion of the acquired data to an actually meaningful feature, as some traits modelled need a second layer of conversion or the usage of coexisting features to seize the full meaning of the modelled trait; the need for seizing the spatial heterogeneity of pastures. In the context of this thesis, these barriers were clearly identified, and multiple steps were made towards the creation of the final DSS.

All these steps contributed to the answer this thesis aimed to bring to the following research question: “Is it possible to develop machine learning algorithms fed by meteorological and remotely sensed data to assess the available feed on pastures at the scale of Wallonia?”. The presentation of the answer will be sub-divided in smaller objectives that will each be developed in their own chapter (Figure 1-3).

1.7 Plan

1.7.1 Story version

Narrated tales begin either with an overview of the landscape on which the action will take place, either in the middle of the action. This story began to unfold with the approach of the scenery description. The next step will be the preparation of the bags we will use during the journey. Remaining at a local scale, we will pick up the best tools, a.k.a. the machine learning models assessing the available feed on pastures under the form of CSH from S1, S2 and meteorological data. Now that we are ready, a second part of the journey will consist in exploiting our tools on unencountered conditions across the region we wander. At night, near the campfire dreams of the future DSS will come. Drafts will be drawn and constraints directly appearing together with challenges underlined through discussions with other travellers will be noted down and these mock-ups will pave the way towards the DSS. The multiplicity of end-user possible will open the eyes for other exploitation of the learnings made during the exploitation of the tools at a larger scale. A particular study-case, the parcel management detection, will also be explored. After these exhausting parts of the journey, we spent the night at a tavern and took the time to think about the path travelled. Our shoes are worn out and not all our tools were useful, the last part of this tale will focus on what changes we should have made during the preparation and the unfolding of the trip.

1.7.2 Scientific version

As illustrated in Figure 1-3, now that the state of the art/introduction chapter set up the scenery in which the journey will take place, the second chapter will relate the process used to develop models predicting the CSH, a proxy of the biomass available on pastures, using remotely sensed S1 and S2 and meteorological data through multiple machine learning models and feature transformation.

The next chapter will narrate the development and implementation of a platform performing the assessment of this CSH at the scale of the whole Walloon Region through models developed using the pipeline described in the previous chapter. It will also discuss the constraints that were encountered and the limitations regarding the availability of these assessments. Indeed, given the transition from datasets with a limited number of records acquired over no more than fifty parcels at the time of writing to databases covering more than 190.000 parcels and almost 4.000 km² in 2018, the application of the models will require to manage properly the acquisition and fusion of data and to take into account software limitations.

Afterwards, a chapter will be dedicated to the presentation of specifications on which the final DSS should be based, and the first mock-ups attempt will also be presented.

The fifth chapter will be a discussion and perspective chapter that will include a feedback-type of discussion on the improvements that could be made. This will be followed by a conclusion chapter that will recall the answer to the research question, regroup the perspectives and develop my personal takeaways.

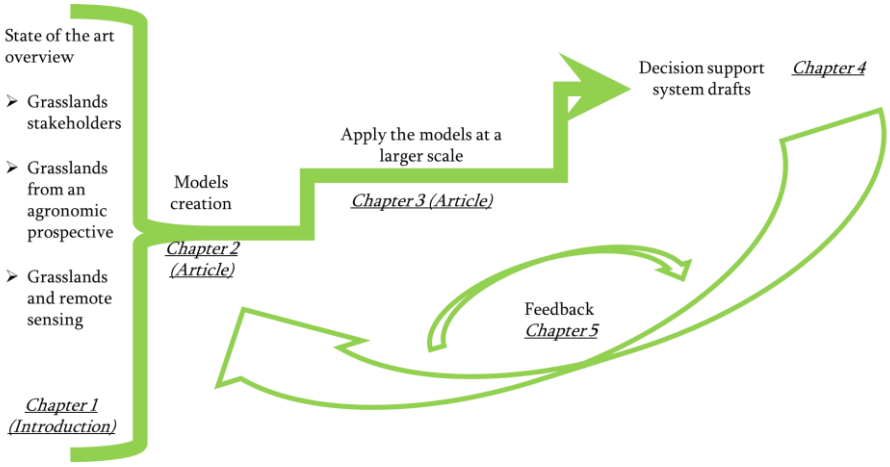


Figure 1-3: Graphical representation of the organisation of the thesis

1.8 Chapter bibliography

[i-001]: IPCC. 2021. Climate Change 2021. The Physical Science Basis. Contribution of Working Group 1 to Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 1195, 1266. URL https://www.ipcc.ch/report/ar6/wg1/.

[i-002]: Roe, S., Streck, C., Obersteiner, M., Frank, S., Griscom, B., Drouet, L., Fricko, O., Gusti, M., Harris, N., Hasegawa, T. and Hausfather, Z., 2019. Contribution of the land sector to a 1.5 C world. *Nature Climate Change*, *9*(11), pp.817-828. https://doi.org/10.1038/s41558-019-0591-9

[i-003]: Erb, K.H., Kastner, T., Plutzar, C., Bais, A.L.S., Carvalhais, N., Fetzel, T., Gingrich, S., Haberl, H., Lauk, C., Niedertscheider, M. and Pongratz, J., 2018. Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature*, *553*(7686), pp.73-76. https://doi.org/10.1038/nature25138

[i-004]: Gourlez de la Motte, Jérôme, E., Mamadou, O., Beckers, Y., Bodson, B., Heinesch, B. and Aubinet, M., 2016. Carbon balance of an intensively grazed permanent grassland in southern Belgium. *Agricultural and forest meteorology*, *228*, pp.370-383. https://doi.org/10.1016/j.agrformet.2016.06.009

[i-005]: Henderson, B.B., Gerber, P.J., Hilinski, T.E., Falcucci, A., Ojima, D.S., Salvatore, M. and Conant, R.T., 2015. Greenhouse gas mitigation potential of the world's grazing lands: Modeling soil carbon and nitrogen fluxes of mitigation practices. *Agriculture, Ecosystems & Environment*, *207*, pp.91-100. https://doi.org/10.1016/j.agee.2015.03.029

[i-006]: Allen, V.G., Batello, C., Berretta, E.J., Hodgson, J., Kothmann, M., Li, X., McIvor, J., Milne, J., Morris, C., Peeters, A. and Sanderson, M., 2011. An international terminology for grazing lands and grazing animals. *Grass and forage science*, *66*(1), pp.2-28. https://doi.org/10.1111/j.1365-2494.2010.00780.x

[i-007]: FAO, 2023, NSP - Grasslands, Rangelands and Forage Crops, URL https://www.fao.org/agriculture/crops/core-themes/theme/spi/grasslands-rangelands-and-forage-crops/en/ (accessed 17/02/2023)

[i-008]: FAO, 2023, Management of Grasslands and Rangelands, URL https://www.fao.org/agriculture/crops/thematic-sitemap/theme/spi/scpi-home/managing-ecosystems/management-of-grasslands-and-rangelands/en/ (accessed 17/02/2023)

[i-009]: O'Mara, F.P., 2012. The role of grasslands in food security and climate change. *Annals of botany*, *110*(6), pp.1263-1270. https://doi.org/10.1093/aob/mcs209

[i-010]: Taravat, A., Wagner, M.P. and Oppelt, N., 2019. Automatic grassland cutting status detection in the context of spatiotemporal Sentinel-1 imagery analysis and artificial neural networks. *Remote Sensing*, *11*(6), p.711. https://doi.org/10.3390/rs11060711

[i-011]: Région wallonne. 2022. "WalOnMap - Géoportail de la Wallonie." URL http://geoportail.wallonie.be/walonmap (accessed 17/02/2023)

[i-012]: FAO, 2023, Grassland carbon sequestration: management, policy and economics, URL https://www.fao.org/3/i1880e/i1880e.pdf (accessed 17/02/2023)

[i-013]: FAO, 2023, Coping with a changing climate: considerations for adaptation and mitigation in agriculture , URL https://www.fao.org/fileadmin/templates/agphome/documents/climate/B5_HLCPaper_web.pdf (accessed 17/02/2023)

[i-014]: FAO, 2023, Challenges and opportunities for carbon sequestration in grassland systems A technical report on grassland management and climate change mitigation , https://www.fao.org/fileadmin/templates/agphome/documents/climate/AGPC_grassland_webversion_19.pdf (accessed 17/02/2023)

[i-015]: de Thysebaert, D., Bottier, S., Bovy, C., Cuvelier, C., Dejemeppe, J., Généreux, C., Maes, E., Renard, V., Thiry, V., Wénin, M., 2021. L'ENVIRONNEMENT WALLON EN 10 INFOGRAPHIES., URL http://etat.environnement.wallonie.be/files/Infographie_2021/L'environnement%20wallon%20en%2010%20infographies-2021.pdf (accessed 17/02/2023)

[i-016]: Lessire, F., Jacquet, S., Veselko, D., Piraux, E. and Dufrasne, I., 2019. Evolution of grazing practices in Belgian dairy farms: Results of two surveys. *Sustainability*, *11*(15), p.3997. https://doi.org/10.3390/su11153997

[i-017]: European Commission, 2017, On the implementation of the ecological focus area obligation under the green direct payment scheme., Brussels, URL https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017DC0152&from=FR (accessed 17/02/2023)

[i-018]: European Commission, 2021, List of potential agricultural practices that eco-schemes could support, no. January: 5. https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/key_policies/documents/factsheet-agri-practices-under-ecoscheme_en.pdf (accessed 17/02/2023)

[i-019]: European Commission, 2019, The European Green Deal EN., Brussels, http://eur-lex.europa.eu/resource.html?uri=cellar:208111e4-414e-4da5-94c1-852f1c74f351.0004.02/DOC_1&format=PDF (accessed 17/02/2023)

[i-020]: Tamm, T., Zalite, K., Voormansik, K. and Talgre, L., 2016. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. *Remote Sensing*, *8*(10), p.802. https://doi.org/10.3390/rs8100802

[i-021]: Service public de Wallonie, 2021, Rapport du plan stratégique relevant de la PAC. Namur. https://agriculture.wallonie.be/documents/20182/21837/SFC2021_version+envoyee_17 mars2022.pdf/022ffe73-ba3f-46e2-8907-7d8c6ce1a434 (accessed 17/02/2023)

[i-022]: Michaud, A., Plantureux, S., Baumont, R. and Delaby, L., 2020. Les prairies, une richesse et un support d'innovation pour des élevages de ruminants plus durables et acceptables. *INRAE Productions Animales*, *33*(3), pp.153-172. https://doi.org/10.20870/productions-animales.2020.33.3.4543

[i-023]: Biowallonie, 2021, Les chiffres du BIO 2020. Namur, https://www.biowallonie.com/wp-content/uploads/2021/09/Biowallonie_ChiffresBio-2020-V2.pdf (accessed 17/02/2023)

[i-024]: Peyraud, J.L., Van den Pol, A., Dillon, P. and Delaby, L., 2010. Producing milk from grazing to reconcile economic and environmental performances. In *23th General Meeting of the European Grassland Federation, Kiel, Germany, 29 august-02 September, 2010* (pp. 163-164). URL https://library.wur.nl/WebQuery/wurpubs/402666 (accessed 17/02/2023)

[i-025]: Murphy, D.J., Murphy, M.D., O'Brien, B. and O'Donovan, M., 2021. A review of precision technologies for optimising pasture measurement on Irish grassland. *Agriculture*, *11*(7), p.600. https://doi.org/10.3390/agriculture11070600

[i-026]: Dillon, P.A.T., Hennessy, T., Shalloo, L., Thorne, F. and Horan, B., 2008. Future outlook for the Irish dairy industry: a study of international competitiveness, influence of international trade reform and requirement for change. *International Journal of Dairy Technology*, *61*(1), pp.16-29. https://doi.org/10.1111/j.1471-0307.2008.00374.x

[i-027]: Reijs, J.W., Daatselaar, C.H.G., Helming, J.F.M., Jager, J.H. and Beldman, A.C.G., 2013. *Grazing dairy cows in North-West Europe: economic farm performance and future developments with emphasis on the Dutch situation*. LEI Wageningen UR. URL https://library.wur.nl/WebQuery/wurpubs/fulltext/265398 (accessed 17/02/2023)

[i-028]: Papadopoulou, A., Ragkos, A., Theodoridis, A., Skordos, D., Parissi, Z. and Abraham, E., 2020. Evaluation of the contribution of pastures on the economic sustainability of small ruminant farms in a typical Greek area. *Agronomy*, *11*(1), p.63. https://doi.org/10.3390/agronomy11010063

[i-029]: Shalloo, L., O'Donovan, M., Leso, L., Werner, J., Ruelle, E., Geoghegan, A., Delaby, L. and O'Leary, N., 2018. Review: Grass-based dairy systems, data and precision technologies. Animal 12 (s2): S262-S271. https://doi.org/10.1017/S175173111800246X

[i-030]: Shalloo, L., Byrne, T., Leso, L., Ruelle, E., Starsmore, K., Geoghegan, A., Werner, J. and O'Leary, N., 2021. A review of precision technologies in pasture-based dairying systems. *Irish Journal of Agricultural and Food Research*, *59*(2), pp.279-291. https://doi.org/10.15212/ijafr-2020-0119

[i-031]: Sishodia, R.P., Ray, R.L. and Singh, S.K., 2020. Applications of remote sensing in precision agriculture: A review. *Remote Sensing*, *12*(19), p.3136. https://doi.org/10.3390/rs12193136

[i-032]: Rusanov, A.M., 2014. Recovery of natural vegetation and ecological functions of pasture ecosystems in arid steppes of the Cisural region. *Russian journal of ecology*, *45*, pp.243-248. https://doi.org/10.1134/S1067413614040109

[i-033]: Franca, A., Sanna, F., Nieddu, S., Re, G.A., Pintus, G.V., Ventura, A., Duce, P., Salis, M. and Arca, B., 2012. Effects of grazing on the traits of a potential fire in a Sardinian wooded pasture. *New approaches for grassland research in a context of climate and socio-economic changes. Options Méditerranéennes: Série A. Séminaires Méditerranéens*, (102), pp.307-311. URL https://om.ciheam.org/om/pdf/a102/00006933.pdf (accessed 17/02/2023)

[i-034]: Kolver, E.S., 2003. Nutritional limitations to increased production on pasture-based systems. *Proceedings of the Nutrition Society*, *62*(2), pp.291-300. https://doi.org/10.1079/PNS2002200

[i-035]: Lait de pâturage, 2018, URL https://www.lait-de-paturage.fr/ (accessed 17/02/2023)

[i-036]: Marguerite Happy Cow, 2022, URL https://www.margueritehappycow.be/ (accessed 17/02/2023)
[i-037]: Fromage-Abondance, 2016, URL https://www.fromageabondance.fr/ (accessed 17/02/2023)
[i-038]: Comité Interprofessionnel de Gestion du Comté, 2022, https://www.comte.com/ (accessed 17/02/2023)
[i-039]: Coopérative fromagère Jeune Montagne, 2022, https://www.jeune-montagne-aubrac.fr/fr/laguiole-aop-au-lait-cru/le-laguiole-jm.php (accessed 17/02/2023)
[i-040]: Caraes, C., Hebert, J., 2020. Agriculture et élevage de précision :Comptabiliser le temps de pâturage grâce à un collier connecté : Chronopâture. URL https://normandie.chambres-agriculture.fr/fileadmin/user_upload/Normandie/506_Fichiers-communs/PDF/INNOVATION-RECHERCHE/agri-connectee-chronopature.pdf (accessed 17/02/2023)
[i-041]: Chilliard, Y., Ferlay, A., Mansbridge, R.M. and Doreau, M., 2000, May. Ruminant milk fat plasticity: nutritional control of saturated, polyunsaturated, trans and conjugated fatty acids. In *Annales de zootechnie* (Vol. 49, No. 3, pp. 181-205). EDP Sciences. https://doi.org/10.1051/animres:2000117
[i-042]: Elgersma, A., Tamminga, S. and Ellen, G., 2006. Modifying milk composition through forage. *Animal feed science and technology*, *131*(3-4), pp.207-225. https://doi.org/10.1016/j.anifeedsci.2006.06.012
[i-043]: Kelly, M.L., Kolver, E.S., Bauman, D., Van Amburgh, M.E. and Muller, L.D., 1998. Effect of intake of pasture on concentrations of conjugated linoleic acid in milk of lactating cows. *Journal of dairy science*, *81*(6), pp.1630-1636. https://doi.org/10.3168/jds.S0022-0302(98)75730-3
[i-044]: Frelich, J., Šlachta, M., Hanuš, O., Špička, J., Samková, E., Zapletal, P., 2012 Seasonal variation in fatty acid composition of cow milk in relation to the feeding system. Animal Science Papers and Reports vol. 30 (2012) no. 3, 219-229 URL https://www.researchgate.net/profile/Martin-Slachta-2/publication/271439969_Seasonal_variation_in_fatty_acid_composition_of_cow_milk_in_relation_to_the_feeding_system/links/54d0b4ea0cf29ca81102d727/Seasonal-variation-in-fatty-acid-composition-of-cow-milk-in-relation-to-the-feeding-system.pdf (accessed 17/02/2023)
[i-045]: Gulati, A., Galvin, N., Lewis, E., Hennessy, D., O'Donovan, M., McManus, J.J., Fenelon, M.A. and Guinee, T.P., 2018. Outdoor grazing of dairy cows on pasture versus indoor feeding on total mixed ration: Effects on gross composition and mineral content of milk during lactation. *Journal of Dairy Science*, *101*(3), pp.2710-2723. https://doi.org/10.3168/jds.2017-13338
[i-046]: Reinermann, S., Asam, S. and Kuenzer, C., 2020. Remote sensing of grassland production and management—A review. *Remote Sensing*, *12*(12), p.1949. https://doi.org/10.3390/rs12121949
[i-047]: Pontes-Prates, A., de Faccio Carvalho, P.C. and Laca, E.A., 2020. Mechanisms of grazing management in heterogeneous swards. *Sustainability*, *12*(20), p.8676. https://doi.org/10.3390/su12208676
[i-048]: Cockburn, M., 2020. Application and prospective discussion of machine learning for the management of dairy farms. *Animals*, *10*(9), p.1690. https://doi.org/10.3390/ani10091690
[i-049]: Eastwood, C.R., Rue, B.D. and Gray, D.I., 2016. Using a 'network of practice'approach to match grazing decision-support system design with farmer practice. *Animal Production Science*, *57*(7), pp.1536-1542. https://doi.org/10.1071/AN16465
[i-050]: Armstrong, L., 2020. *Improving data management and decision support systems in agriculture*. Burleigh Dodds Science Publishing Limited. https://doi.org/10.19103/as.2020.0069.15
[i-051]: Power, D. 2003. What are the characteristics of a Decision Support System" http://dssresources.com/faq/pdf/13.pdf (accessed 17/02/2023)
[i-052]: Filipponi, F., 2019. Sentinel-1 GRD preprocessing workflow. In *International Electronic Conference on Remote Sensing* (p. 11). MDPI. https://doi.org/10.3390/ECRS-3-06201
[i-053]: Filipponi, F., 2020. Sentinel-1_GRD_Preprocessing_ Standard Workflow for the Preprocessing of Sentinel-1 GRD Satellite Data. *MDPI: Basel, Switzerland*.
[i-054]: ESA, 2022, SNAP—ESA Sentinel Application Platform v8.0.0, ESA: Paris, France, https://step.esa.int/main/toolboxes/snap/ (accessed 17/02/2023)
[i-055]: Graux, A.-I., 2017, From the functional analysis of soil-plant-animals interactions to the modelling of grasslands – example of the ModVege model. Master. Spécialité "Agroecology", URL https://hal.inrae.fr/hal-02790929 (accessed 10/04/2023)
[i-056]: Garioud, A., Giordano, S., Valero, S. and Mallet, C., 2019, August. Challenges in grassland mowing event detection with multimodal sentinel images. In *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)* (pp. 1-4). IEEE. doi: 10.1109/Multi-Temp.2019.8866914

[i-058]: Stendardi, L., Karlsen, S.R., Niedrist, G., Gerdol, R., Zebisch, M., Rossi, M. and Notarnicola, C., 2019. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. *Remote Sensing*, *11*(5), p.542. https://doi.org/10.3390/rs11050542

[i-059]: Garioud, A., Valero, S., Giordano, S. and Mallet, C., 2021. Recurrent-based regression of Sentinel time series for continuous vegetation monitoring. *Remote Sensing of Environment*, *263*, p.112419. https://doi.org/10.1016/j.rse.2021.112419

[i-060]: Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B. and Chang, Q., 2019. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *154*, pp.189-201. https://doi.org/10.1016/j.isprsjprs.2019.06.007

[i-061]: Ali, I., Barrett, B., Cawkwell, F., Green, S., Dwyer, E. and Neumann, M., 2017. Application of repeat-pass TerraSAR-X staring spotlight interferometric coherence to monitor pasture biophysical parameters: limitations and sensitivity analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), pp.3225-3231. https://doi.org/10.1109/JSTARS.2017.2679761

[i-062]: ESA, SPOT 4 (Satellite pour l'Observation de la Terre), URL https://earth.esa.int/eogateway/missions/spot-4 (accessed 17/02/2023)

[i-063]: Di Bella, C., Faivre, R., Ruget, F., Seguin, B., Guérif, M., Combal, B., Weiss, M. and Rebella, C., 2004. Use of SPOT4-VEGETATION satellite data to improve pasture production simulated by STICS included in the ISOP French system. *Agronomie*, *24*(6-7), pp.437-444. https://doi.org/10.1051/agro:2004034

[i-064]: Clementini, C., Pomente, A., Latini, D., Kanamaru, H., Vuolo, M.R., Heureux, A., Fujisawa, M., Schiavon, G. and Del Frate, F., 2020. Long-term grass biomass estimation of pastures from satellite data. *Remote Sensing*, *12*(13), p.2160. https://doi.org/10.3390/rs12132160

[i-065]: Wang, Z., Ma, Y., Zhang, Y. and Shang, J., 2022. Review of remote sensing applications in grassland monitoring. *Remote Sensing*, *14*(12), p.2903. https://doi.org/10.3390/rs14122903

[i-066]: Moeckel, T., Safari, H., Reddersen, B., Fricke, T. and Wachendorf, M., 2017. Fusion of ultrasonic and spectral sensor data for improving the estimation of biomass in grasslands with heterogeneous sward structure. *Remote Sensing*, *9*(1), p.98. https://doi.org/10.3390/rs9010098

[i-067]: Shoko, C., Mutanga, O. and Dube, T., 2018. Determining optimal new generation satellite derived metrics for accurate C3 and C4 Grass species aboveground biomass estimation in South Africa. *Remote Sensing*, *10*(4), p.564. https://doi.org/10.3390/rs10040564

[i-068]: Courault, D., Hadria, R., Ruget, F., Olioso, A., Duchemin, B., Hagolle, O. and Dedieu, G., 2010. Combined use of FORMOSAT-2 images with a crop model for biomass and water monitoring of permanent grassland in Mediterranean region. *Hydrology and Earth System Sciences*, *14*(9), pp.1731-1744. https://doi.org/10.5194/hess-14-1731-2010

[i-069]: Ali, I., Cawkwell, F., Green, S. and Dwyer, N., 2014, July. Application of statistical and machine learning models for grassland yield estimation based on a hypertemporal satellite remote sensing time series. In *2014 IEEE geoscience and remote sensing symposium* (pp. 5060-5063). IEEE. doi: 10.1109/IGARSS.2014.6947634

[i-070]: Ali, I., Cawkwell, F., Dwyer, E. and Green, S., 2016. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—A machine learning approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), pp.3254-3264. doi: 10.1109/JSTARS.2016.2561618

[i-071]: Hubert-Moy, L., Thibault, J., Fabre, E., Rozo, C., Arvor, D., Corpetti, T. and Rapinel, S., 2019. Mapping grassland frequency using decadal MODIS 250 m time-series: Towards a national inventory of semi-natural grasslands. *Remote Sensing*, *11*(24), p.3041. https://doi.org/10.3390/rs11243041

[i-072]: Li, F., Zeng, Y., Luo, J., Ma, R. and Wu, B., 2016. Modeling grassland aboveground biomass using a pure vegetation index. *Ecological Indicators*, *62*, pp.279-288. https://doi.org/10.1016/j.ecolind.2015.11.005

[i-073]: Houborg, R. and McCabe, M.F., 2016. High-Resolution NDVI from planet's constellation of earth observing nano-satellites: A new data source for precision agriculture. *Remote Sensing*, *8*(9), p.768. https://doi.org/10.3390/rs8090768

[i-074]: Liu, H., Dahlgren, R.A., Larsen, R.E., Devine, S.M., Roche, L.M., O'Geen, A.T., Wong, A.J., Covello, S. and Jin, Y., 2019. Estimating rangeland forage production using remote sensing data from a small unmanned aerial system (sUAS) and planetscope satellite. *Remote Sensing*, *11*(5), p.595. https://doi.org/10.3390/rs11050595

[i-076]: Punalekar, S.M., Verhoef, A., Quaife, T.L., Humphries, D., Bermingham, L. and Reynolds, C.K., 2018. Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sensing of Environment*, *218*, pp.207-220. https://doi.org/10.1016/j.rse.2018.09.028

[i-077]: Ara, I., Harrison, M.T., Whitehead, J., Waldner, F., Bridle, K., Gilfedder, L., Marques da Silva, J., Marques, F. and Rawnsley, R., 2021. Modelling seasonal pasture growth and botanical composition at the paddock scale with satellite imagery. *in silico Plants*, *3*(1), p.diaa013. doi: 10.1093/insilicoplants/diaa013

[i-078]: Jansen, V.S., Kolden, C.A., Schmalz, H.J., Karl, J.W. and Taylor, R.V., 2021. Using satellite-based vegetation data for short-term grazing monitoring to inform adaptive management. *Rangeland Ecology & Management*, *76*, pp.30-42. https://doi.org/10.1016/j.rama.2021.01.006

[i-079]: Dos Reis, A.A., Werner, J.P., Silva, B.C., Figueiredo, G.K., Antunes, J.F., Esquerdo, J.C., Coutinho, A.C., Lamparelli, R.A., Rocha, J.V. and Magalhães, P.S., 2020. Monitoring pasture aboveground biomass and canopy height in an integrated crop–livestock system using textural information from PlanetScope imagery. *Remote Sensing*, *12*(16), p.2534. https://doi.org/10.3390/rs12162534

[i-080]: Murphy, D.J., O'Brien, B., Askari, M.S., McCarthy, T., Magee, A., Burke, R. and Murphy, M.D., 2019. GrassQ-A holistic precision grass measurement and analysis system to optimize pasture based livestock production. In *2019 ASABE Annual International Meeting* (p. 1). American Society of Agricultural and Biological Engineers. doi:10.13031/aim.201900769

[i-081]: Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G. and Gilabert, M.A., 2020. Understanding deep learning in land use classification based on Sentinel-2 time series. *Scientific reports*, *10*(1), p.17188. https://doi.org/10.1038/s41598-020-74215-5

[i-082]: Zumo, I.M., Hashim, M. and Hassan, N.D., 2021. Mapping and estimation of above-ground grass biomass using Sentinel 2A Satellite Data. *International Journal of Built Environment and Sustainability*, *8*(3), pp.9-15. https://doi.org/10.11113/ijbes.v8.n3.684

[i-083]: Dos Reis, A.A., Silva, B.C., Werner, J.P.S., Silva, Y.F., Rocha, J.V., Figueiredo, G.K.D.A., Antunes, J.F.G., Esquerdo, J.C.D.M., Coutinho, A.C., Lamparelli, R.A.C. and Magalhães, P.S.G., 2020, March. Exploring the Potential of High-Resolution Planetscope Imagery for Pasture Biomass Estimation in an Integrated Crop–Livestock System. In *2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS)* (pp. 675-680). IEEE. doi: 10.1109/LAGIRS48042.2020.9165596

[i-084]: Mutanga, O. and Shoko, C., 2018, July. Monitoring the spatio-temporal variations of C3/C4 Grass species using multispectral satellite data. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 8988-8991). IEEE. doi: 10.1109/IGARSS.2018.8517685

[i-085]: Houborg, R. and McCabe, M.F., 2018. Daily Retrieval of NDVI and LAI at 3 m Resolution via the Fusion of CubeSat, Landsat, and MODIS Data. *Remote Sensing*, *10*(6), p.890. https://doi.org/10.3390/rs10060890

[i-086]: Parente, L., Taquary, E., Silva, A.P., Souza Jr, C. and Ferreira, L., 2019. Next generation mapping: Combining deep learning, cloud computing, and big remote sensing data. *Remote Sensing*, *11*(23), p.2881. https://doi.org/10.3390/rs11232881

[i-087]: Leach, N., Coops, N.C. and Obrknezev, N., 2019. Normalization method for multi-sensor high spatial and temporal resolution satellite imagery with radiometric inconsistencies. *Computers and Electronics in Agriculture*, *164*, p.104893. https://doi.org/10.1016/j.compag.2019.104893

[i-088]: He, L., Li, A., Yin, G., Nan, X. and Bian, J., 2019. Retrieval of grassland aboveground biomass through inversion of the PROSAIL model with MODIS imagery. *Remote Sensing*, *11*(13), p.1597. https://doi.org/10.3390/rs11131597

[i-089]: Serrano, J., Shahidian, S., Paixão, L., Marques da Silva, J., Morais, T., Teixeira, R. and Domingos, T., 2021. Spatiotemporal patterns of pasture quality based on ndvi time-series in mediterranean montado ecosystem. *Remote Sensing*, *13*(19), p.3820. https://doi.org/10.3390/rs13193820

[i-090]: Ghayour, L., Neshat, A., Paryani, S., Shahabi, H., Shirzadi, A., Chen, W., Al-Ansari, N., Geertsema, M., Pourmehdi Amiri, M., Gholamnia, M. and Dou, J., 2021. Performance evaluation of sentinel-2 and landsat 8 OLI data for land cover/use classification using a comparison between machine learning algorithms. *Remote Sensing*, *13*(7), p.1349. https://doi.org/10.3390/rs13071349

[i-091]: Barnetson, J., Phinn, S. and Scarth, P., 2021. Climate-Resilient Grazing in the Pastures of Queensland: An Integrated Remotely Piloted Aircraft System and Satellite-Based Deep-Learning Method for Estimating Pasture Yield. *AgriEngineering*, *3*(3), pp.681-702. https://doi.org/10.3390/agriengineering3030044

[i-092]: Chen, Y., Guerschman, J., Shendryk, Y., Henry, D. and Harrison, M.T., 2021. Estimating pasture biomass using Sentinel-2 imagery and machine learning. *Remote Sensing*, *13*(4), p.603. https://doi.org/10.3390/rs13040603

[i-093]: Sibanda, M., Mutanga, O. and Rouget, M., 2016. Comparing the spectral settings of the new generation broad and narrow band sensors in estimating biomass of native grasses grown under different management practices. *GIScience & Remote Sensing*, *53*(5), pp.614-633. https://doi.org/10.1080/15481603.2016.1221576

[i-094]: Cimbelli, A. and Vitale, V., 2017. Grassland height assessment by satellite images. *Advances in Remote Sensing*, *6*(01), p.40. https://doi.org/10.4236/ars.2017.61003

[i-095]: Houborg, R. and McCabe, M.F., 2018. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, *135*, pp.173-188. https://doi.org/10.1016/j.isprsjprs.2017.10.004

[i-096]: Gašparović, M., Medak, D., Pilaš, I., Jurjević, L. and Balenović, I., 2018. FUSION OF SENTINEL-2 AND PLANETSCOPE IMAGERY FOR VEGETATION DETECTION AND MONITORING. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*(1). https://doi.org/10.5194/isprs-archives-XLII-1-155-2018

[i-097]: Delforge, L., Sevrin, L., 2022, Formation de gaz nitreux en ensilage : peu courant mais ça existe !, Pleinchamp 22 septembre 2022, n°38, p9, URL https://centredemichamps.be/wp-content/uploads/2022/04/La-formation-de-gaz-nitreux-en-ensilage-2.pdf (accessed 17/02/2023)

[i-098]: Askari, M.S., McCarthy, T., Magee, A. and Murphy, D.J., 2019. Evaluation of grass quality under different soil management scenarios using remote sensing techniques. *Remote Sensing*, *11*(15), p.1835. https://doi.org/10.3390/rs11151835

[i-099]: Lugassi, R., Zaady, E., Goldshleger, N., Shoshany, M. and Chudnovsky, A., 2019. Spatial and temporal monitoring of pasture ecological quality: Sentinel-2-based estimation of crude protein and neutral detergent fiber contents. *Remote Sensing*, *11*(7), p.799. https://doi.org/10.3390/rs11070799

[i-100]: Darvishzadeh, R., Wang, T., Skidmore, A., Vrieling, A., O'Connor, B., Gara, T.W., Ens, B.J. and Paganini, M., 2019. Analysis of Sentinel-2 and RapidEye for retrieval of leaf area index in a saltmarsh using a radiative transfer model. *Remote sensing*, *11*(6), p.671. https://doi.org/10.3390/rs11060671

[i-101]: Tiscornia, G., Baethgen, W., Ruggia, A., Do Carmo, M. and Ceccato, P., 2019. Can we Monitor Height of Native Grasslands in Uruguay with Earth Observation?. *Remote Sensing*, *11*(15), p.1801. https://doi.org/10.3390/rs11151801

[i-102]: Violle, C., Navas, M.L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I. and Garnier, E., 2007. Let the concept of trait be functional!. *Oikos*, *116*(5), pp.882-892. https://doi.org/10.1111/j.0030-1299.2007.15559.x

[i-103]: Gargiulo, J., Clark, C., Lyons, N., de Veyrac, G., Beale, P. and Garcia, S., 2020. Spatial and temporal pasture biomass estimation integrating electronic plate meter, planet cubesats and sentinel-2 satellite data. *Remote Sensing*, *12*(19), p.3222. https://doi.org/10.3390/rs12193222

[i-104]: Lavorel, S., Díaz, S., Cornelissen, J.H.C., Garnier, E., Harrison, S.P., McIntyre, S., Pausas, J.G., Pérez-Harguindeguy, N., Roumet, C. and Urcelay, C., 2007. Plant functional types: are we getting any closer to the Holy Grail?. *Terrestrial ecosystems in a changing world*, pp.149-164. https://doi.org/10.1007/978-3-540-32730-1_13

[i-105]: Cruz, P., Duru, M.M., Therond, O., Theau, J.P.J., Ducourtieux, C., Jouany, C., Khaled, R.A.H. and Ansquer, P., 2002. Une nouvelle approche pour caractériser les prairies naturelles et leur valeur d'usage. *Fourrages*, (172), pp.335-354. URL https://hal.inrae.fr/hal-02683047 (accessed 17/02/2023)

[i-106]: Funk, J.L., Larson, J.E., Ames, G.M., Butterfield, B.J., Cavender-Bares, J., Firn, J., Laughlin, D.C., Sutton-Grier, A.E., Williams, L. and Wright, J., 2017. Revisiting the Holy Grail: using plant functional traits to understand ecological processes. *Biological Reviews*, *92*(2), pp.1156-1173. https://doi.org/10.1111/brv.12275

[i-107]: Perez-Harguindeguy, N., Diaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., Bret-Harte, M.S., Cornwell, W.K., Craine, J.M., Gurvich, D.E. and Urcelay, C., 2016. Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of botany*, *64*(8), pp.715-716. https://doi.org/10.1071/BT12225_CO

[i-108]: FourrgesMieux, 2022, URL https://www.fourragesmieux.be (accessed 17/02/2023)

[i-109]: Jouven, M., Carrère, P. and Baumont, R., 2006. Model predicting dynamics of biomass, structure and digestibility of herbage in managed permanent pastures. 1. Model description. *Grass and forage science*, *61*(2), pp.112-124. https://doi.org/10.1111/j.1365-2494.2006.00515.x

[i-110]: Biowallonie, 2022, Les chiffres du BIO 2021 en Wallonie., Namur URL https://www.biowallonie.com/wp-content/uploads/2022/07/Chiffres-du-bio-2021.pdf (accessed 17/02/2023)

[i-111]: Service public de Wallonie, 2023, Cheptel bovin, URL https://etat-agriculture.wallonie.be/contents/indicatorsheets/EAW-

A_II_c_1.html#:~:text=En%2030%20ans%2C%20le%20cheptel,051%20600%20t%C3%AAtes%20de%20b%C3%A9tail. (accessed 17/02/2023)

[i-112]: Cros, M.J., Garcia, F. and Martin-Clouaire, R., 1999, September. SEPATOU: a decision support system for the management of rotational grazing in a dairy production. In *Proceedings of 2nd European Conference on Information Technology in Agriculture, in: Schiefer G., Helbig* (pp. 549-557). URL https://www.academia.edu/download/49038202/SEPATOU_a_Decision_Support_System_for_th20160922-22472-154k7t8.pdf (accessed 17/02/2023)

[i-113]: Cros, M.J., Duru, M., Garcia, F. and Martin-Clouaire, R., 2003. A biophysical dairy farm model to evaluate rotational grazing management strategies. *Agronomie*, *23*(2), pp.105-122. URL https://hal.science/hal-00886164/ (accessed 17/02/2023)

[i-114]: Bittar, A., Meisser, M., Mosimann, E. and Calanca, P., 2018. Simulation of grass growth and pasture yields with ModVege. *Recherche Agronomique Suisse*, (6), pp.186-191. URL https://www.cabdirect.org/cabdirect/abstract/20183223352 (accessed 17/02/2023)

[i-115]: Cruz, P., Theau, J.P.J., Lecloux, E., Jouany, C. and Duru, M.M., 2010. Typologie fonctionnelle de graminées fourragères pérennes: une classification multitraits. *Fourrages*, (201), pp.11-17. URL https://usab-tm.ro/utilizatori/superadmin/file/2017/svf/Bibliographie/Typo%20multitraits%20Cruz%202010.pdf (accessed 10/04/2023)

[i-116]: Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Dürr, C. and Richard, G., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie*, *18*(5-6), pp.311-346. URL https://www.academia.edu/download/40107204/STICS_a_generic_model_for_the_simulation20151117-27862-1enl8ps.pdf (accessed 17/02/2023)

[i-117]: Dalcq, A.C., Dogot, T., Beckers, Y., Brostaux, Y., Froidmont, E., Vanwindekens, F. and Soyeurt, H., 2020. The Walloon farmers position differently their ideal dairy production system between a global-based intensive and a local-based extensive model of farm. *Plos one*, *15*(12), p.e0223346. https://doi.org/10.1371/journal.pone.0223346

[i-118]: McCall, D.G. and Bishop-Hurley, G.J., 2003. A pasture growth model for use in a whole-farm dairy production model. *Agricultural Systems*, *76*(3), pp.1183-1205. https://doi.org/10.1016/S0308-521X(02)00104-X

[i-119]: Hurtado-Uria, C., Hennessy, D., Shalloo, L., Schulte, R.P.O., Delaby, L. and O'CONNOR, D., 2013. Evaluation of three grass growth models to predict grass growth in Ireland. *The Journal of Agricultural Science*, *151*(1), pp.91-104. doi:10.1017/S0021859612000317

[i-120]: Díaz, S., Purvis, A., Cornelissen, J.H., Mace, G.M., Donoghue, M.J., Ewers, R.M., Jordano, P. and Pearse, W.D., 2013. Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecology and evolution*, *3*(9), pp.2958-2975. https://doi.org/10.1002/ece3.601

[i-121]: Maselli, F., Argenti, G., Chiesi, M., Angeli, L. and Papale, D., 2013. Simulation of grassland productivity by the combination of ground and satellite data. *Agriculture, ecosystems & environment*, *165*, pp.163-172. https://doi.org/10.1016/j.agee.2012.11.006

[i-122]: Paillette, C.A., Delaby, L., Shalloo, L., O'Connor, D. and Hennessy, D., 2015, June. Developing a predictive model for grass growth in grass-based milk production systems. In *18. Symposium of the european grassland federation* (Vol. 20, p. np). Wageningen Academic Publishers. URL https://hal.science/hal-01211031/

[i-123]: Ruelle, E., Hennessy, D. and 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*, pp.80-91. https://doi.org/10.1016/j.eja.2018.06.010

[i-124]: Barrett, P.D., Laidlaw, A.S. and Mayne, C.S., 2005. GrazeGro: A European herbage growth model to predict pasture production in perennial ryegrass swards for decision support. *European Journal of Agronomy*, *23*(1), pp.37-56. https://doi.org/10.1016/j.eja.2004.09.006

[i-125]: Johnson, I.R. and Thornley, J.H.M., 1983. Vegetative crop growth model incorporating leaf area expansion and senescence, and applied to grass. *Plant, Cell & Environment*, *6*(9), pp.721-729. https://doi.org/10.1111/1365-3040.ep11588103_6_9

[i-126]: Zabalza, J., 2015. Feature extraction and data reduction for hyperspectral remote sensing earth observation.

[i-127]: Nakagami, K., 2016. A method for approximate on-farm estimation of herbage mass by using two assessments per pasture. *Grass and Forage Science*, *71*(3), pp.490-496. https://doi.org/10.1111/gfs.12195

[i-128]: Pullanagari, R.R., Kereszturi, G. and Yule, I., 2018. Integrating airborne hyperspectral, topographic, and soil data for estimating pasture quality using recursive feature elimination with random forest regression. *Remote Sensing*, *10*(7), p.1117. https://doi.org/10.3390/rs10071117

[i-129]: Schulte, L.G., Perez, N.B., de Pinho, L.B. and Trentin, G., 2019, May. Decision support system for precision livestock: Machine learning-based prediction module for stocking rate adjustment. In *Proceedings of the XV Brazilian Symposium on Information Systems* (pp. 1-8). https://doi.org/10.1145/3330204.3330222

[i-130]: 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.

[i-131]: McSweeney, D., Coughlan, N.E., Cuthbert, R.N., Halton, P. and Ivanov, S., 2019. Micro-sonic sensor technology enables enhanced grass height measurement by a Rising Plate Meter. *Information Processing in Agriculture*, *6*(2), pp.279-284. https://doi.org/10.1016/j.inpa.2018.08.009

[i-132]: De Rosa, D., Basso, B., Fasiolo, M., Friedl, J., Fulkerson, B., Grace, P.R. and Rowlings, D.W., 2021. Predicting pasture biomass using a statistical model and machine learning algorithm implemented with remotely sensed imagery. *Computers and Electronics in Agriculture*, *180*, p.105880. https://doi.org/10.1016/j.compag.2020.105880

[i-133]: Rangwala, M., Liu, J., Ahluwalia, K.S., Ghajar, S., Dhami, H.S., Tracy, B.F., Tokekar, P. and Williams, R.K., 2021. Deeppastl: Spatio-temporal deep learning methods for predicting long-term pasture terrains using synthetic datasets. *Agronomy*, *11*(11), p.2245. https://doi.org/10.3390/agronomy11112245

[i-134]: Kumar, S., Anderson, S.H., Udawatta, R.P. and Kallenbach, R.L., 2012. Water infiltration influenced by agroforestry and grass buffers for a grazed pasture system. *Agroforestry Systems*, *84*, pp.325-335. https://doi.org/10.1007/s10457-011-9474-4

[i-135]: Guerra, C.A. and Pinto-Correia, T., 2016. Linking farm management and ecosystem service provision: Challenges and opportunities for soil erosion prevention in Mediterranean silvo-pastoral systems. *Land Use Policy*, *51*, pp.54-65. http://dx.doi.org/10.1016/j.landusepol.2015.10.028

[i-136]: De Vroey, M., de Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B. and Defourny, P., 2022. Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring. *Remote Sensing of Environment*, *280*, p.113145. https://doi.org/10.1016/j.rse.2022.113145

[i-137]: Henits, L., Szerletics, Á., Szokol, D., Szlovák, G., Gojdár, E. and Zlinszky, A., 2022. Sentinel-2 Enables Nationwide Monitoring of Single Area Payment Scheme and Greening Agricultural Subsidies in Hungary. *Remote Sensing*, *14*(16), p.3917. https://doi.org/10.3390/rs14163917

[i-138]: Zhao, Y., Liu, Z. and Wu, J., 2020. Grassland ecosystem services: a systematic review of research advances and future directions. *Landscape Ecology*, *35*, pp.793-814. https://doi.org/10.1007/s10980-020-00980-3

[i-139]: Liu, H., Hou, L., Kang, N., Nan, Z. and Huang, J., 2022. The economic value of grassland ecosystem services: A global meta-analysis. *Grassland Research*, *1*(1), pp.63-74. https://doi.org/10.1002/glr2.12012

[i-140]: Sollenberger, L.E., Kohmann, M.M., Dubeux Jr, J.C. and Silveira, M.L., 2019. Grassland management affects delivery of regulating and supporting ecosystem services. *Crop Science*, *59*(2), pp.441-459. https://doi.org/10.2135/cropsci2018.09.0594

[i-141]: Duru, M., Pontes, L.D.A.S., Schellberg, J., Theau, J.P. and Therond, O., 2019. Grassland functional diversity and management for enhancing ecosystem services and reducing environmental impacts: a cross-scale analysis. In *Agroecosystem diversity* (pp. 211-230). Academic Press. https://doi.org/10.1016/B978-0-12-811050-8.00013-3

[i-142]: FAO, 2023, How the world is fed, URL https://www.fao.org/3/Y4683E/y4683e06.htm (accessed 17/02/2023)

[i-143]: Mottet, A., de Haan, C., Falcucci, A., Tempio, G., Opio, C. and Gerber, P., 2017. Livestock: On our plates or eating at our table? A new analysis of the feed/food debate. *Global Food Security*, *14*, pp.1-8. https://doi.org/10.1016/j.gfs.2017.01.001

[i-144]: Poore, J. and Nemecek, T., 2018. Reducing food's environmental impacts through producers and consumers. *Science*, *360*(6392), pp.987-992. https://doi.org/10.1126/science.aaq0216

[i-145]: Muscat, A., De Olde, E.M., de Boer, I.J. and Ripoll-Bosch, R., 2020. The battle for biomass: A systematic review of food-feed-fuel competition. *Global Food Security*, *25*, p.100330. https://doi.org/10.1016/j.gfs.2019.100330

[i-146]: Sillon Belge, 2018, VEM, DVE, OEB et VS: vous souvenez-vous des systèmes alimentaires? URL https://www.sillonbelge.be/2153/article/2018-03-08/vem-dve-oeb-et-vs-vous-souvenez-vous-des-systemes-alimentaires (accessed 17/02/2023)

[i-147]: Beckers, Y. 2019, Equilibre des rations des ruminants et autonomie alimentaire URL https://fugea.be/wp-content/uploads/2019/05/Equilibre-des-rations-Ulg-Etalle-f%C3%A9vrier-2018.pdf (accessed 17/02/2023)

[i-148]: FourragesMieux, 2017, Notes de cours, URL https://www.fourragesmieux.be/Documents_telechargeables/Cours_A_2017_2018.pdf (accessed 17/02/2023)

[i-149]: Romera, A., Beukes, P., Clark, D., Clark, C. and Tait, A., 2013. Pasture growth model to assist management on dairy farms: Testing the concept with farmers. *Grassland Science*, *59*(1), pp.20-29. https://doi.org/10.1111/grs.12009

[i-150]: Romera, A.J., Beukes, P., Clark, C., Clark, D., Levy, H. and 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*(1), pp.66-72. https://doi.org/10.1016/j.compag.2010.06.006

[i-151]: Bindelle, J., Da Silva Neto, G.F., Kokah, E.U., de Faccio Carvalho, P. and Michez, A., 2021. Drone-based remote sensing of sward structure and biomass for precision grazing: state of the art and future challenges. In *21st Symposium of the European Grassland Federation*. URL https://orbi.uliege.be/bitstream/2268/261219/2/EGF%20presentation.pdf (accessed 17/02/2023)

[i-152]: Cruz, P., Theau, J.P.J., Lecloux, E., Jouany, C. and Duru, M.M., 2010. Typologie fonctionnelle de graminées fourragères pérennes: une classification multitraits. *Fourrages*, (201), pp.11-17. URL https://usab-tm.ro/utilizatori/superadmin/file/2017/svf/Bibliographie/Typo%20 multitraits%20Cruz%202010.pdf (accessed 17/02/2023)

[i-153]: Correll, O., Isselstein, J. and Pavlu, V., 2003. Studying spatial and temporal dynamics of sward structure at low stocking densities: the use of an extended rising-plate-meter method. *Grass and Forage Science*, *58*(4), pp.450-454. https://doi.org/10.1111/j.1365-2494.2003.00387.x

[i-154]: Şahin Demirbağ, N., Röver, K.U., Wrage, N., Hofmann, M. and Isselstein, J., 2009. Herbage growth rates on heterogeneous swards as influenced by sward-height classes. *Grass and Forage Science*, *64*(1), pp.12-18. https://doi.org/10.1111/j.1365-2494.2008.00665.x

[i-155]: Oliveira, R.A., Näsi, R., Niemeläinen, O., Nyholm, L., Alhonoja, K., Kaivosoja, J., Jauhiainen, L., Viljanen, N., Nezami, S., Markelin, L. and Hakala, T., 2020. Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sensing of Environment*, *246*, p.111830. https://doi.org/10.1016/j.rse.2020.111830

[i-156]: Fernández-Delgado, M., Sirsat, M.S., Cernadas, E., Alawadi, S., Barro, S. and Febrero-Bande, M., 2019. An extensive experimental survey of regression methods. *Neural Networks*, *111*, pp.11-34. https://doi.org/10.1016/j.neunet.2018.12.010

[i-157]: Garcia, R., Aguilar, J., Toro, M., Pinto, A. and Rodriguez, P., 2020. A systematic literature review on the use of machine learning in precision livestock farming. *Computers and Electronics in Agriculture*, *179*, p.105826. https://doi.org/10.1016/j.compag.2020.105826

[i-158]: Soyeurt, H., Gerards, C., Nickmilder, C., Bindelle, J., Franceschini, S., Dehareng, F., Veselko, D., Bertozzi, C., Gengler, N., Marvuglia, A. and Bayram, A., 2022. Prediction of Indirect Indicators of a Grass-Based Diet by Milk Fourier Transform Mid-Infrared Spectroscopy to Assess the Feeding Typologies of Dairy Farms. *Animals*, *12*(19), p.2663. https://doi.org/10.3390/ani12192663

[i-159]: Dalcq, A.C., Beckers, Y., Mayeres, P., Reding, E., Wyzen, B., Colinet, F., Delhez, P. and Soyeurt, H., 2018. The feeding system impacts relationships between calving interval and economic results of dairy farms. *animal*, *12*(8), pp.1662-1671. https://doi.org/10.1017/S1751731117003020

[i-160]: FAO, 2023, Supporting ecosystem services, URL https://www.fao.org/ecosystem-services-biodiversity/background/supporting-services/en/ (accessed 17/02/2023)

[i-161]: FAO, 2023, Ecosystem services, URL https://www.fao.org/ecosystem-services-biodiversity/background/en/ (accessed 17/02/2023)

[i-162]: Ondřej, C., Josef, H., Michal, H. and Pavel, C., 2018. The use of compressed height to estimate the yield of a differently fertilized meadow. *Plant, Soil and Environment*, *64*(2), pp.76-81. https://doi.org/10.17221/732/2017-PSE

[i-163]: Service Public de Wallonie, 2021, Evolution de l'économie agricole et horticole wallonne. URL https://agriculture.wallonie.be/evolution-de-l-economie-agricole-et-horticole-wallonne (accessed 17/02/2023)

[i-164]: Hennessy, D., Delaby, L., van den Pol-van Dasselaar, A. and Shalloo, L., 2020. Increasing grazing in dairy cow milk production systems in Europe. *Sustainability*, *12*(6), p.2443. https://doi.org/10.3390/su12062443

[i-165]: Scohy, D., 2020, L'herbomètre : un outil indispensable dans la gestion du pâturage ? URL https://www.web-agri.fr/herbe/article/168848/l-herbometre-est-il-un-outil-indispensable-dans-la-gestion-du-paturage (accessed 17/02/2023)

[i-166]: French, P., O'Brien, B. and Shalloo, L., 2015. Development and adoption of new technologies to increase the efficiency and sustainability of pasture-based systems. *Animal Production Science*, *55*(7), pp.931-935. https://doi.org/10.1071/AN14896

[i-167]: Bareth, G. and Schellberg, J., 2018. Replacing manual rising plate meter measurements with low-cost UAV-derived sward height data in grasslands for spatial monitoring. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, *86*, pp.157-168. https://doi.org/10.1007/s41064-018-0055-2

[i-168]: Fricke, T. and Wachendorf, M., 2013. Combining ultrasonic sward height and spectral signatures to assess the biomass of legume–grass swards. *Computers and Electronics in Agriculture*, *99*, pp.236-247. https://doi.org/10.1016/j.compag.2013.10.004

[i-169]: Legg, M. and Bradley, S., 2019. Ultrasonic arrays for remote sensing of pasture biomass. *Remote Sensing*, *12*(1), p.111. https://doi.org/10.3390/rs12010111

[i-170]: Rayburn, E.B., Lozier, J.D., Sanderson, M.A., Smith, B.D., Shockey, W.L., Seymore, D.A. and Fultz, S.W., 2007. Alternative methods of estimating forage height and sward capacitance in pastures can be cross calibrated. *Forage & Grazinglands*, *5*(1), pp.1-6. https://doi.org/10.1094/FG-2007-0614-01-RS

[i-171]: Ancin-Murguzur, F.J., Taff, G., Davids, C., Tømmervik, H., Mølmann, J. and Jørgensen, M., 2019. Yield estimates by a two-step approach using hyperspectral methods in grasslands at high latitudes. *Remote Sensing*, *11*(4), p.400. https://doi.org/10.3390/rs11040400

[i-172]: Borra-Serrano, I., De Swaef, T., Muylle, H., Nuyttens, D., Vangeyte, J., Mertens, K., Saeys, W., Somers, B., Roldán-Ruiz, I. and Lootens, P., 2019. Canopy height measurements and non-destructive biomass estimation of Lolium perenne swards using UAV imagery. *Grass and Forage Science*, *74*(3), pp.356-369. https://doi.org/10.1111/gfs.12439

[i-173]: Michez, A., Lejeune, P., Bauwens, S., Herinaina, A.A.L., Blaise, Y., Castro Muñoz, E., Lebeau, F. and Bindelle, J., 2019. Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system. *Remote Sensing*, *11*(5), p.473. https://doi.org/10.3390/rs11050473

[i-174]: Michez, A., Philippe, L., David, K., Sébastien, C., Christian, D. and Bindelle, J., 2020. Can low-cost unmanned aerial systems describe the forage quality heterogeneity? Insight from a timothy pasture case study in southern Belgium. *Remote Sensing*, *12*(10), p.1650. https://doi.org/10.3390/rs12101650

[i-175]: Lussem, U., Bolten, A., Menne, J., Gnyp, M.L., Schellberg, J. and Bareth, G., 2019. Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices. *Journal of Applied Remote Sensing*, *13*(3), pp.034525-034525. https://doi.org/10.1117/1.JRS.13.034525

[i-177]: Laisse, S., Baumont, R., Dusart, L., Gaudré, D., Rouillé, B., Benoit, M., Veysset, P., Rémond, D. And Peyraud, J.L., 2019. L'efficience nette de conversion des aliments par les animaux d'élevage: une nouvelle approche pour évaluer la contribution de l'élevage à l'alimentation humaine. *INRA Productions Animales*, *31*(3), pp.269-288. https://doi.org/10.20870/productions-animales.2018.31.3.2355

[i-178]: Louarn, G., Corre-Hellou, G., Fustec, J., Pelzer, E., Julier, B., Litrico, I., Hinsinger, P. and Lecomte, C., 2010. Déterminants écologiques et physiologiques de la productivité et de la stabilité des associations graminées-légumineuses. *Innovations agronomiques*, *11*, pp.79-99. https://hal.science/hal-01173245/document

[i-179]: Legg, M. and Bradley, S., 2019. Ultrasonic proximal sensing of pasture biomass. *Remote Sensing*, *11*(20), p.2459. https://doi.org/10.3390/rs11202459

[i-180]: Obanawa, H., Yoshitoshi, R., Watanabe, N. and Sakanoue, S., 2020. Portable LiDAR-based method for improvement of grass height measurement accuracy: comparison with SfM methods. *Sensors*, *20*(17), p.4809. https://doi.org/10.3390/s20174809

[i-181]: Lussem, U., Bolten, A., Gnyp, M.L., Jasper, J. and Bareth, G., 2018. Evaluation of RGB-based vegetation indices from UAV imagery to estimate forage yield in grassland. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*(3), p.1215. https://doi.org/10.5194/isprs-archives-XLII-3-1215-2018

[i-182]: Poirier D., 2019, L'Ortolan Grand Pâturage fromage "local et responsable", URL https://www.macommune.info/lortolan-grand-paturage-un-nouveau-fromage-local-et-responsable/ (accessed 28/02/2023)

[i-183]: Syndicat de défense et de promotion du fromage de laguiole A.O.C. - A.O.P., 2022, Cahier des charges, URL https://www.fromage-laguiole.fr/fr/le-laguiole-aoc-aop/cahier-de-ressources.php#bloc-88 (accessed 28/02/2023)

[i-184]: Service public fédéral Économie (Statbel), 2022, Etat de l'agriculture wallonne – Exploitations agricoles, URL https://etat-agriculture.wallonie.be/contents/indicatorsheets/EAW-A_II_b_1-1.html#:~:text=La%20Wallonie%20a%20perdu%20plus,an)%20constitue%20une%20tendance%20lourde. (accessed 17/02/2023)

[i-185]: European union, 2023, URL https://ec.europa.eu/environment/nature/natura2000/index_en.htm (accessed 17/02/2023)

[i-186]: Requasud, 2023, URL https://www.requasud.be/base-de-donnees/ (accessed 10/04/2023)

[i-187]: Requasud, 2023, URL https://www.requaconsult.requasud.be/requaconsult_nirf (accessed 10/04/2023)

[i-188]: Díaz, S., Kattge, J., Cornelissen, J.H., Wright, I.J., Lavorel, S., Dray, S., Reu, B., Kleyer, M., Wirth, C., Colin Prentice, I. and Garnier, E., 2016. The global spectrum of plant form and function. *Nature*, *529*(7585), pp.167-171. https://doi.org/10.1038/nature16489

Chapter 2

Model Creation

Adapted from: *Charles Nickmilder, Anthony Tedde, Isabelle Dufrasne, Françoise Lessire, Bernard Tychon, Yannick Curnel, Jérome Bindelle and Hélène Soyeurt. 2021. 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 Sensing, 13(3), p.408.* https://doi.org/10.3390/rs13030408

2.1 Outline

The main topic of this chapter is the generation of models predicting a proxy of biomass from Sentinel-1 radar satellite data, Sentinel-2 optical and multispectral satellite data and meteorological data. These models come from the statistical/machine learning area. The process workflow that led to the development of these model in this thesis is further developed in the article below.

2.2 Abstract

Accurate information about the available standing biomass on pastures is critical for the adequate management of grazing and its promotion to farmers. In this paper, machine learning models are developed to predict available biomass expressed as compressed sward height (CSH) from readily accessible meteorological, optical (Sentinel-2) and radar satellite data (Sentinel-1). This study assumed that combining heterogeneous data sources, data transformations and machine learning methods would improve the robustness and the accuracy of the developed models. A total of 72,795 records of CSH with a spatial positioning, collected in 2018 and 2019, were used and aggregated according to a pixel-like pattern. The resulting dataset was split into a training one with 11,625 pixellated records and an independent validation one with 4952 pixellated records. The models were trained with a 19-fold cross-validation. A wide range of performances was observed (with mean root mean square error (RMSE) of cross-validation ranging from 22.84 mm of CSH to infinite-like values), and the four best-performing models were a cubist, a glmnet, a neural network and a random forest. These models had an RMSE of independent validation lower than 20 mm of CSH at the pixel-level. To simulate the behavior of the model in a decision support system, performances at the paddock level were also studied. These were computed according to two scenarios: either the predictions were made at a sub-parcel level and then aggregated, or the data were aggregated at the parcel level and the predictions were made for these aggregated data. The results obtained in this study were more accurate than those found in the literature concerning pasture budgeting and grassland biomass evaluation. The training of the 124 models resulting from the described framework was part of the realization of a decision support system to help farmers in their daily decision making.

2.3 Keywords

Sentinel-1; Sentinel-2; machine learning; pastures; compressed sward height

2.4 Introduction

Grazed pasturelands play multiple roles in agroecosystems that can benefit the sustainability of ruminant-based agriculture. This includes lower feeding costs [1], higher animal welfare and lower occurrence of lameness and mastitis, and increased milk quality compared to indoor feeding [2]. However, despite these advantages, Walloon intensive dairy farmers increasingly turn away from grazing. This is because of the higher difficulty of managing grazed pastures as their main feeding method, especially with a decreasing workforce available on the farm and increases in herd sizes [3]. Indeed, due to the high variability of plant growth with weather conditions, grazing management requires a frequent assessment of the standing biomass available for animals to feed on. Tools like decision support systems (DSS) could ease this management from the farmer's perspective by providing an assessment of the standing biomass without having to travel to the pasture. Such a DSS, based on the simulation of the behavior of a dairy farm practicing rotational grazing, was for example developed by Cros et al. [4] and Amalero et al. [5] to help farmers plan their activities. Other examples of DSS, like Pasture Growth Simulation Using Smalltalk (PGSUS) [6,7] and PastureBase Ireland integrating the Moorepark St Gilles grass growth model (MostGG) [8], focus on the assessment of available forage in pastures. Both tools rely on reference field measurements used as inputs in mechanistic models. The reference measurement method consists of the cutting and drying of forage samples to get the actual dry biomass per area unit. This procedure was developed for researchers and is time-consuming, destructive, expensive and never applied by farmers. Moreover, the limited number of samples that can be taken strongly reduces the possibility of assessing the spatial variability of the pastoral resource. Objectives other than biomass might also be included in the DSS, such as leaf area index (LAI), real height or visual correspondence to standards.

Several alternatives to this reference measurement have been proposed, such as the indirect estimation of standing biomass [9,10,11] via the measurement of compressed sward height (CSH) using a rising platemeter (RPM). The CSH readings can be converted into biomass with varying levels of accuracy, depending on the structure of the assessed vegetation [12]. While this method can provide a high number of estimates and also spatialize the data if combined with a GPS

sensor [13], it requires time to perform the measurement on the pastures. It also requires consideration of the sampling pattern to capture the spatial variability of height on a given pasture, MacAdam and Hunt [14] recommends a “lazy W” sampling pattern while Gargiulo et al. [15] shows that for a homogeneous pasture, the impact of the sampling pattern is negligible. In addition, no consensus exists on which part of the biomass to consider. The CSH measurement implicitly considers all standing biomass, while some scientists, like Wang et al. [16], advise considering biomass above 1 cm. Others, like Hakl et al. [17] and Walloon farm advisers, use a threshold of 4 cm, while Crémer [18] and Nakagami [19] bound the limit to 5 cm. Other methods to assess standing biomass include the response of the sward canopy to ultra-sonic [20,21,22] or electric [23,24] signals, ground-level photography analysis [25]. All the methods cited previously require attention to be paid to the size, number and repartition of the sampling spots [26] in order to minimize sampling error. Aside from these ground-based methods, the estimation of standing biomass has also been explored through remote sensing from either satellite or airborne platforms (e.g., [27,28,29]). This offers already-spatialized data and reduces the risk of sampling bias. However, other constraints might appear, such as computation power requirements and the sensitivity of unmanned aerial vehicles to flight autonomy, weather and regulatory constraints. In the context of the current study, these constraints, together with the time-consuming aspect of acquiring data with unmanned aerial vehicles (UAV) led to the choice of not using this technology. However, it is worth mentioning that UAV based systems are at the core of current researches establishing relationships between the biomass and remotely sensed data like digital surface/elevation models derived from aerial pictures (using structure from motion), LiDAR pointclouds acquisition and treatment, and spectral vegetation indices data [11,30,31].

All these methods are part of a set of recent technological advancements that may help grazing management to embrace the smart farming approaches relevant to this sector [32], provided their adequate integration in DSS and a good level of acceptance by farmers. The latter requires DSS to be based on information routinely available at a large scale and at minimal cost. Moreover, it must be possible to improve the DSS in an iterative and interactive process [33].

In order to address the challenge of providing a tool offering a rapid estimation of pasture biomass under the above-mentioned constraints, the current study developed an analysis method to predict CSH measured on pasture using readily available meteorological data, space-borne synthetic-aperture radar (SAR) and optical imagery. Indeed, despite CSH being an indirect measurement of biomass, RPM has been widely used and benefits from widespread acceptance among

farmers and pasture scientists (e.g., [7,9,34]). Meteorological data have the advantage of being routinely available at the Walloon scale (i.e., for all the Southern part of Belgium which is the area of interest in this study) and provide insight into the key drivers of growth dynamics for various types of pasture plants growth dynamics [35,36,37]. Such models are at the heart of most DSS (e.g., STICS [38] and MostGG [39]).

Concerning the space-borne remote sensing data, a choice was made between all the existing optical and SAR sensors according to their spatial, temporal and spectral resolution, as well as their cost. The Sentinel-1 (S1) and Sentinel-2 (S2) constellations were chosen. S1 was used for mowing event detection [40], LAI and above ground biomass estimation [16], or to detect meadow phenology [41]. S2 mission was used for biomass estimation [42,43] or monitoring [44,45] or for LAI retrieval [46]. Some studies predicting standing biomass in grasslands have already included analysis of both S1 and S2 data [41,47]. However, they did not encompass any transformation of those signals, which might enhance some properties, such as how NDVI highlights the presence of chlorophyll in the vegetation. Nor did they test a wide range of machine learning (ML) methods that could catch different parts of the information. The methods appearing most frequently in these studies are multiple linear regression (lm), neural networks with or without recurrent layers, random forests (rf) and cubist, alone or together (e.g., rf and lm were used to predict the quantity and quality of grass swards in Alves et al. [48]). This study encompasses multiple variable transformations and ML methods (amidst the wide range of methods highlighted in Fernández-Delgado et al. [49] and in García et al. [50]) with the objective of extracting more information on those signals. The framework described in this study aims to predict biomass in Walloon pastures through the CSH proxy using more than 100 different ML models, in order to provide a tool offering a rapid estimation of pasture biomass on the basis of readily available meteorological data, SAR and optical imagery.

2.5 Materials and Methods

The structure of this section follows the workflow presented in Figure 2-1. The main steps are the following: data acquisition and pre-processing, data fusion to a grid, separation of training and validation datasets, selection of the most informative variables, model training and validation. The processing framework was made in R v3.6.2 using Rstudio IDE v1.2.5033 [51,52]. The framework can be summarised according to the following equation: $CSH = f(time, weather, S1, S2)$.

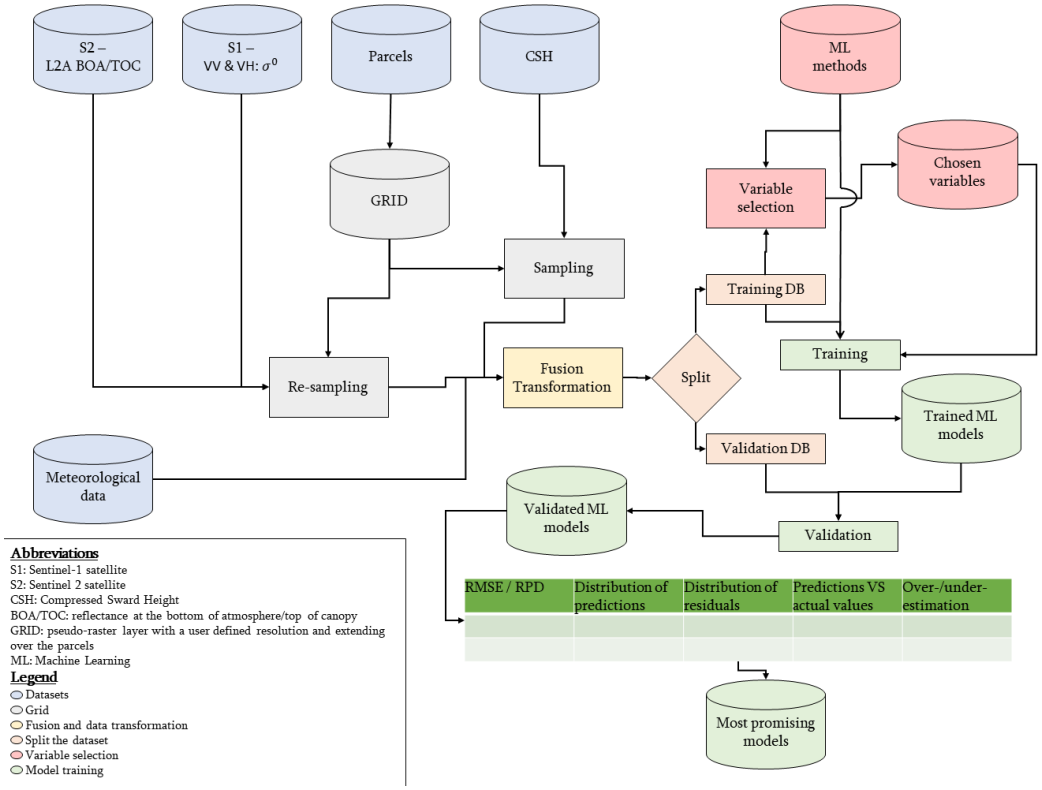


Figure 2-1: Framework developed in the study to process Sentinel-1, Sentinel-2 and meteorological data to predict standing biomass in grazed pastures. The first steps are the pre-treatment and the (re-)sampling of the different dataset according to the same reference grid. Then the datasets are merged to get a tabular dataset. Transformations of the variables are then computed. Afterwards, machine learning models are trained and validated on distinct parts of the dataset. Based on the results of the validation, the most promising models are determined.

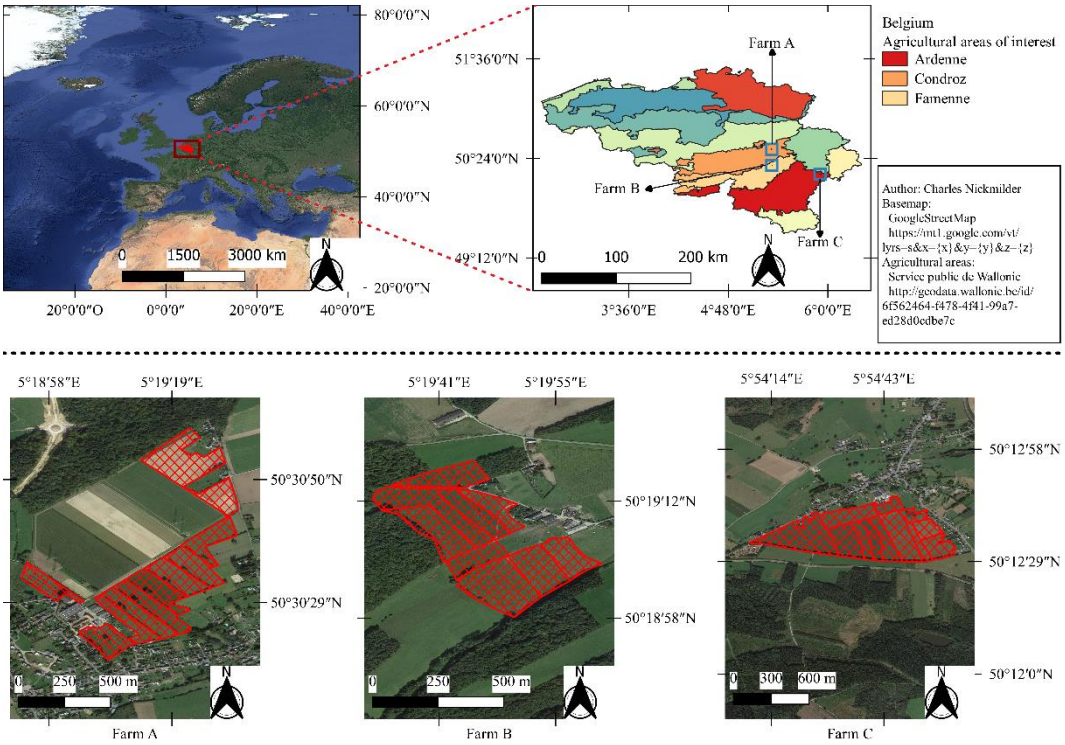


Figure 2-2: Location the studied farms in the agricultural areas of Belgium and their grazing parcels.

2.5.1 Datasets

2.5.1.1 Compressed sward height

A total of 72,975 CSH records were acquired approximately weekly during the grazing seasons of 2018 on Farm A (N = 13,753, 16 sampling dates) and Farm C (N = 9,309, 6 sampling dates), and 2019 on Farm A (N = 28,497, 27 sampling dates) and Farm B (N = 21,416, 22 sampling dates), using a Jenquip EC20 G rising platemeter (NZ Agriworks Ltd t/a Jenquip, New Zealand). The relationship between CSH and time was integrated by considering two variables: the number of the month of the year (e.g., 1 for January) and the day of the year (e.g., 1 for 01/01).

2.5.1.2 Meteorological data

The meteorological data used came from the meteorological station located on the experimental farm. It consisted of the daily rainfall and the degree-day-18 (DD_18): $DD_{18} = \frac{T_{max}+T_{min}}{2} - 18$, where T_{min} and T_{max} are the minimum and maximum temperature of the day. This formulation of the degree-day variables is

introduced in studies like the ones of Moot et al. [53] and Balocchi et al. [54]. In some cases, the minimum and maximum temperatures may have upper and lower threshold whose value depends on various factors including the ability of the plant to take water [55], family and photosynthetic pathways (see the sensitivity of the Rubisco to temperature in Salvucci et al. [56] and Greco et al. [57]). Given that the pastures are multispecific, it was chosen to ignore the threshold on minimum and maximum temperatures and the base temperature of 18°C was considered. In order to take the historical succession of meteorological events into account, the meteorological data were used in a cumulative way: the precipitation and the degree-day-18 were summed over periods of 3, 7 and 15 days and a mean of both variables since the last CSH measurement was also computed. Eight meteorological variables were thus considered in this study.

2.5.1.3 Sentinel-1 data

The S1 mission offers SAR data in the C band (C-SAR) with a mode-dependent spatial resolution of roughly 5 m and a revisit frequency of approximately three days over the studied area. It also has the advantage of providing data even in cloudy or night-time conditions. The S1 C-SAR data were acquired as level-1 GRD products with VV+VH dual polarisation from the Copernicus Hub [58,59] through the use of a dedicated R package: getSpatialData v0.0.4 [60]. The pre-processing was done in accordance with the standardized framework described by Filipponi [61,62], based on the s1tbx toolbox [63] from SNAP software v7.0.0. This gave products of roughly 5 m of spatial resolution, and consisted of:

- applying a precise orbit file: the real and precise orbit file was computed after the passage and is thus retrieved online to correct satellite position and velocity in the metadata;
- removing thermal noise;
- removing border noise: the artefacts at the image border were removed with the following two parameters: "borderLimit = 500" and "trimThreshold = 50";
- calibrating SAR backscatter: place to choose the output between σ^0, γ^0 and β^0. The second index is related to the first through the cosine of the reflection angle: $\gamma^0 = \frac{\sigma^0}{\cos(r)}$, with r being the reflection angle. The third one is described by equation: $\beta^0 = \frac{SER}{Srad}$, where SER represents the surface perpendicular to the beam, reflecting the totality of the power of the signal in an isotropic way and Srad the surface of the pixel in radar geometry [64]. As recommended by Rudant and Frison [64], only σ^0 was

considered in this study because β^0 is more suited for punctual targets and γ^0 for dense forests;

- speckle filtering: the noise coming from the interference of reflected waves was removed with a 'Refined Lee filter' with a filter size of 3x3.
- terrain correction: the correction was made using the Shuttle Radar Topography mission data (SRTM 3sec);
- converting to decibel.

The S1 data were represented in the dataset through two variables: σ^0 of VV and VH channels.

2.5.1.4 Sentinel-2 data

The S2 mission was chosen for its 13 optical bands, including the three in the red-edge spectral region. The resolution is band-dependent and is either 10 m, 20 m or 60 m with a revisit frequency of approximately three days over the studied area. The S2 data were acquired as L2A-BOA/TOC (Level 2A corresponds to the reflectance at the bottom of the atmosphere or top of canopy) products from the Copernicus Hub [58,59] through the use of a dedicated R package: sen2r v1.3.3 [65]. Some tiles were only available as L1C-TOA products. The pre-processing to transform them into L2A-BOA/TOC tiles was done with the Sen2Cor toolbox [59] and managed by the sen2r package. The main steps behind this transformation are precisely described in Mueller-Wilm [66] and are summarised as:

- scene classification based on band and band ratio values comparison to threshold into 12 classes: 0—No Data; 1—Saturated or defective; 2—Dark area pixels; 3—Shadows of cloud; 4—Vegetation; 5—Not vegetated; 6—Water; 7—Unclassified; 8—Cloud medium probability; 9—Cloud high probability; 10—Thin cirrus; 11—Snow;
- atmospheric correction, consisting of: retrieving the aerosol optical thickness and water vapor, removing the cirrus and retrieving the impact of terrain to get BOA reflectance. It was chosen to ignore the DEM impact;
- product formatting into JPEG2000.

Another processing step involved the resampling of the bands at the same resolution and the fusion of the tiles per acquisition date. The sen2r package managed gdal v 3.0.2 [67] for both operations. The resampling resolution was set to 10 m which was the smallest resolution of S2 images. The scene classification was also recorded. In order to avoid major bias in reflectance measured, only tiles with less than 75% of clouds signalled in the metadata were downloaded and processed. The S2 data were represented into 12 variables in the dataset (bands 1 to 12 without bands 9 and 10 and the scene classification layer). The first band was

included although it is supposed to be related only to atmospheric correction. Its integration in the scope of studied variables aimed at detecting potential artifacts related to the S2 pre-treatment.

2.5.2 Grid

A sub-division of the studied area was performed, as done by Higgins et al. [68] and Lugassi et al. [69]. In Ruelle and Delaby [70] and Ruelle et al. [8], a grid with a resolution of 2 m was the best compromise between precision and speed of computation. Here a resolution of 10 m was chosen, representing a compromise between the spatial resolution of satellite images (it corresponded to the highest resolution of S2 images) and the conservation of the variability of the CSH dataset.

The CSH records were first attributed to a pixel (i.e., a square unit of the grid) and when there was more than one record for a pixel, the CSH median value was considered. The S1 and S2 data were resampled using the same method (called "bilinear"): the pixel of the satellite image containing the centre of the grid pixel was identified, then the four nearest satellite pixels were also identified, and a median value of these four neighbouring values were allocated to the grid pixel.

The resampling was made independently for the different datasets in order to offer some modularity in the analysis. The R packages used for those operations were: data.table v1.12.8 [71] and dplyr v0.8.3 [72] which allowed different management of the data frames; sf v0.8-0 [73], sp v1.3-2 [74,75] and raster v3.0-7 [76] which allowed computation over spatial data and future v1.16.0 [77] and future.apply v1.4.0 [78] which were used to make the computation on parallel mode.

2.5.3 Fusion and Data Transformation

The fusion part consisted of gathering all the information in a single dataset with all the variables as columns and records as rows. First the gathering of all the spatial datasets was done according to their date of acquisition and pixel on the grid. In the case of non-simultaneous acquisition, the nearest acquisition to the CSH measurement was chosen within a 10-day time window. The second part consisted of attributing the meteorological data. To each record was attributed the value of the corresponding date. The non-linear relationship between the explanatory variables and the CSH may not be handled by all ML methods. To bypass those possible restrictions, some data transformations were computed, with some scaling when needed, and also integrated into the workflow:

- meteorological data: square (x^2) and exponential ($10^{\frac{x}{100}}$) transformations were applied on the cumulative data to compute 16 transformed variables;

- S1: the transformations applied on the 2 S1 variables were: square (x^2), exponential ($10^{\frac{x}{10}}$), inverse ($\frac{1}{x}$) and hyperbolic tangent ($\tanh\frac{x}{100}$). Eight transformed variables of the S1 dataset were added to the dataset;
- S2: the transformation applied on the reflectance variables were: square ($\left(\frac{x}{100}\right)^2$), cube ($\left(\frac{x}{100}\right)^3$), exponential ($10^{\frac{x}{1000}}$), inverse ($\frac{1}{x}$) and hyperbolic tangent ($\tanh\frac{x}{10000}$), square root ($\sqrt{x}$), logarithm of base 10 ($\log_{10} x$). A total of 77 transformed variables of the S2 dataset were included in the dataset

Besides these transformations applied to each variable independently, some calculations were made on pairs of variables. Indeed, spectral indices may emphasize some particular spectral signatures [79]. This was taken into account in this study by integrating the 138 non-redundant vegetation indices, computed on the basis of the S2 bands according to the formulas on the website "the Index DataBase" [80], IDB [81]. A total of 249 indices were in fact developed in this list, but a comparison of each index highlighted redundancies (same expression and different names). To avoid the introduction of meaningless collinearity in the dataset, only the first index in order of appearance in the list was used.

2.5.4 Split the Dataset

The total dataset had 16,577 records and 277 variables. Independence between the training and validation datasets was ensured by splitting the dataset at the farm level: Farms A and C were used to train the model, whilst Farm B was used to validate. The training dataset thus had 11,625 records and the validation dataset 4,952 records.

Potential outliers were highlighted by calculating the global H distance (GH) of those records from the principal components explaining 99% of the data variability. The principal component analysis (PCA) was performed with the FactomineR R package v2.1 [82]. Records with a GH value above 3 were considered with caution and records with a GH value above 5 were considered potential outliers. However, no value was discarded.

2.5.5 Variable Selection

As some of the ML methods explored in this study do not handle collinearity between features well, like the generalized linear model and the other methods presented in Table 2-1, a variable selection process was firstly performed, composed of three steps: (1) score determination, (2) definition of breakpoints, and (3) variable selection.

The working hypothesis was that the probability of selecting a relevant feature is higher if this one was considered as important by several ML algorithms. Thus, 12 models resilient to collinearity were developed based on the parametrisation presented in Table 2-2 on the standardized training set with a 19-fold cross-validation (CV). The final models had the lowest CV root mean square error (RMSEcv). Then, a variable importance ranking was established for each model through the use of a function of caret R package v6.0-85: VarImp() [83]. For each variable, the mean and median of their ranking in all models were computed. In order to penalize variables with a high variability of ranking between developed models, the statistical descriptors were standardized: the mean of the ranking of each variable was divided by its standard deviation ($mean_{std}$) and its median by its interquartile range ($median_{std}$).

The variables were ordered according to their decreasing ranking score. Then, multiple linear regressions were trained iteratively: initially, the most informative variable was used, then the first and the second most informative, and so on until all variables were included. For each regression, adjusted R-squared (R^2_{adj}) was computed, and its first derivative was calculated starting from the second value by subtracting the previous value from the actual. A rolling median filter with a window width of three records was applied to these first derivative values to smooth the signal. The negative values corresponded to the breakpoints. To prevent the occurrence of noise, only the first breakpoints were considered in this study. For each breakpoint considered, the variables having a higher ranking than that breakpoint were used to train models through a 19-fold cross-validation based on the acquisition dates of CSH. The variables with a lower ranking were not taken into account.

2.5.6 Model Training

The training phase consisted of a 19-fold cross-validation based on the acquisition dates of CSH on the standardized data. The selection of the best model was made according to the lowest RMSEcv value. For each breakpoint considered, 31 mL methods, including the 12 used in the variable selection process were explored, each resulting in one model. Although some models failed to converge towards a reasonably performant and finite solution within a decent timestep (one week), all models that managed to get a usable expression were kept for further validation and analysis.

The use of the caret R package [83] facilitated the exploration of the "hyper-parameters" of the ML algorithms explored in this study. Methods used in both variable selection and training processes were presented in Table 2-2 and those only used during the training process in Table 2-1.

2.5.7 Model Validation

The models that did not fail during the training process were tested on the independent validation dataset. The indicators observed were the RMSE of validation (RMSEv), the distribution of the validation predictions of CSH compared to the original distribution of CSH, the distribution of the residuals (computed as the predicted value minus the actual value) and the mean residual value per class of CSH with a class width of 5 mm of CSH. The limits of the sampling tool were 0 and 250 mm of CSH. Predicted values beyond these thresholds were brought back to the nearest class. This might have resulted in some flooring and ceiling effects in the representation of the prediction and the residuals. A perfect model would have a low RMSEv, show no difference between the distribution of the actual and the predicted CSH, have a centered distribution of residuals and show no relationship between residuals and actual CSH values.

The models developed and validated on the grid pixel basis described previously were also tested on datasets aggregated at the paddock-level for each acquisition date. Two combinations of aggregation and prediction were tested: either the prediction was made on the pixel and then averaged at the paddock level, or the data were averaged at the paddock level and then the prediction was computed. For the sake of completeness, the combination of prediction and aggregation was tested on both the training and the independent validation dataset. Only a summary of the residual prediction deviation ($RPD = \frac{sd}{RMSE}$) is presented in this paper.

Table 2-1: Presentation of the machine learning algorithms explored during the training processes. The column 'Abbreviation and references' gathers the abbreviations used in the text, corresponding to the name of the method in caret R package, and some articles using the methods in related fields.

Family	Specificities	HyperParameter	Abbreviation and references
Linear regression with variable selection	Forward Backward Stepwise	nvmax=1:30	leapForward leapBackward leapSeq
Linear regression with penalisation	Ridge Lasso ElasticNet	lambda = c(0.0001,0.001,0.005,0.01,0.05:0.05:1,2:1:10) fraction = c(0.0001,0.001,0.005,0.01,0.05:0.05:1) fraction = c(0.0001,0.001,0.005,0.01,0.05:0.05:1) lambda = c(0.0001,0.001,0.005,0.01,0.05:0.05:1,2:1:10)	ridge lasso enet
Principal Components Regression (PCR)	ncomp=1:15		pcr [67]
Partial Least Square Regression (PLS-R)	ncomp=1:15		pls [57] [38] [41] [24]
Support Vector Machine (SVM)	Linear kernel Polynomial kernel	cost = c(0,0.001,0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2,5, 10,20,30,50,100) degree = c(1,2,3) scale = 1 C = c(0,0.001,0.01, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2,5, 10,20,30,50,100)	svmLinear2 svmPoly [14] [72]
Relevance Vector Machine (RVM)	Linear kernel Polynomial kernel	/ degree = c(1,2,3) scale = 1	rvmLinear [73] rvmPoly

Table 2-2: Presentation of the machine learning algorithms explored in variable selection and training processes. The column 'Abbreviation and references' gathers the abbreviations used in the text, corresponding to the name of the method in caret R package, and some articles using the methods in related fields.

Family	Specificities	HyperParameter	Abbreviation and references
Generalised linear model (GLM)	gaussian Gamma poisson inverse.gaussian	link=c(identity,log,inverse) link=c(inverse, identity, log) link=c(log, identity, sqrt) link=(inverse, identity, log)	glm [17]
Generalised linear model with penalisation (glmnet)	gaussian poisson	alpha = c(0.0001,0.001,0.005,0.01,0.05,0.1:0.1:1) lambda = c(0.0001,0.001,0.005,0.01,0.05,0.1:0.1:10)	glmnet
Random forest (RF)		mtry=88	rf [14] [72] [74] [75]
Cubist		committees=c(1:10) neighbors=c(0:7)	cubist [76] [77]
Least Angle Regression (LARS)		fraction = c(0,0.1,0.25,0.5,0.75,1)	lars
Neural Network (nnet)		size=c(1:7) decay = c(0,0.0001,0.001,0.01,0.1:0.1:1)	nnet [39] [78] [72] [79] [80]
Multiple linear regression			lm [14] [78] [72] [9] [75]

2.6 Results

2.6.1 Description of the Datasets

The distribution of CSH from the pixellated and aggregated points of view is summarised for the calibration and validation dataset in Table 2-3, and the distribution for each day of acquisition in Figure 2-3. Those results showed that the CSH measurements were not normally distributed.

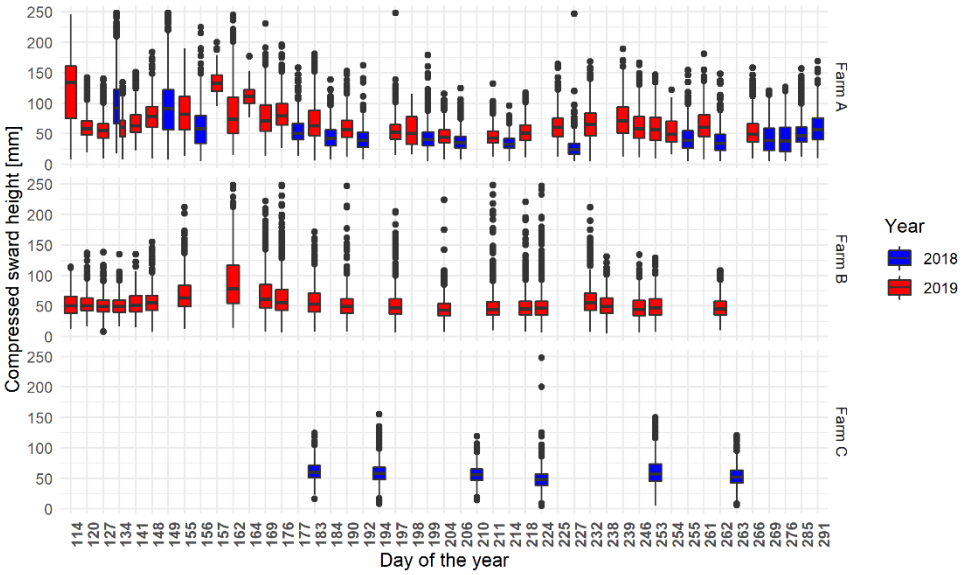


Figure 2-3: Distribution of the CSH acquired during the farm walks on each recorded farming area following the day and the year of acquisition, the day of the year equals one being the first of January. The error bars (whiskers) extend from the upper/lower hinge of the box to the largest/smallest value within 1.5 times the interquartile range. The theoretical frequency of acquisition was one acquisition a week. The data were aggregated through all the parcels of each farm.

Table 2-3: Summary of the compressed sward height [mm] of both training and validation datasets. The aggregated lines correspond to the mean compressed sward height (CSH) per parcel per acquisition date.

Dataset	N	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	sd
Whole	16,577	5.00	39.00	52.00	55.56	67.00	247.00	24.88
Training pixellated	11,625	5.00	39.00	53.00	56.66	69.50	247.00	26.69
Training aggregated	233	18.17	43.89	54.37	56.09	62.65	148.89	20.13
Validation pixellated	4,952	13.00	40.00	49.00	52.96	62.00	194.00	19.75
Validation aggregated	117	38.44	45.22	50.91	52.22	57.67	87.34	9.90

To reach 99% of cumulative percentage of variance, 41 principal components were considered. The distribution of GH for the training and validation records was summarised in Table 2-4. Records having a GH value higher than 5 represented less than 5% of each dataset: 150 (1 %) and 197 (4 %) in the training and validation sets, respectively. The higher number of outliers in the validation set is explained by the fact that the GH was calculated from the calibration set. Therefore, this means that a certain part of the variability in the data collected from Farm B was not totally covered. However, further investigations revealed that there was no apparent relationship between GH and CSH. It means that no subset of CSH occurs on a specific spot of the multidimensional space of components. A representation of the distribution of the GH per sampling date (data not shown) showed that there were more outliers for some recording dates. This could either reflect sampling issues or artifacts in the satellite data. It was chosen to keep these records, given that there is no certainty that these records are not linked to special meteorological events. Finally, one of the S2 surface coverage, corresponding to the classification "shadow of cloud", was represented only in the training dataset but no relationship between this classification and the GH stood out.

Table 2-4 : Summary of the standardised Mahalanobis distance (GH) of the training and validation datasets compared to the training dataset. Values above five are considered as potential outliers.

Dataset	Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
Training	0.11	0.31	0.48	1.00	0.85	283.13
Validation	0.12	0.47	0.69	2.35	1.38	1,035.03

Figure 2-4 shows the two most informative axes of the PCA. These components together explained 54.9% of the variability of the explanatory variables. Some faint grouping effects of variables could be seen. A group was positively related to the first component, mainly made up of raw and transformed S2 indices. Another group was related positively to both first and second components, mainly made of raw and transformed S2 bands. A third group was related negatively to the first and positively to the second component, mainly made of raw and transformed S2 bands. The large amount of S2 related variables in the datasets hid the position of the other variables. They spread well over the two-dimensional space of Figure 2-4 although they seem to be more correlated with other components given the length of their arrows. Based on these results, the integration of multiple data sources seemed relevant as they bring different information. Moreover, the variables with a transformation spread well on the two-dimensional space of the two first components of the PCA although the variables with an inverse transformation seemed to be mainly negatively correlated with the second axis

(data not shown). The integration of multiple variable transformations also appeared relevant. The proximity of some variables on the graph in Figure 2-4 implied a possibly non-negligible redundancy. Therefore, the selection of variables before the training of the models is critical.

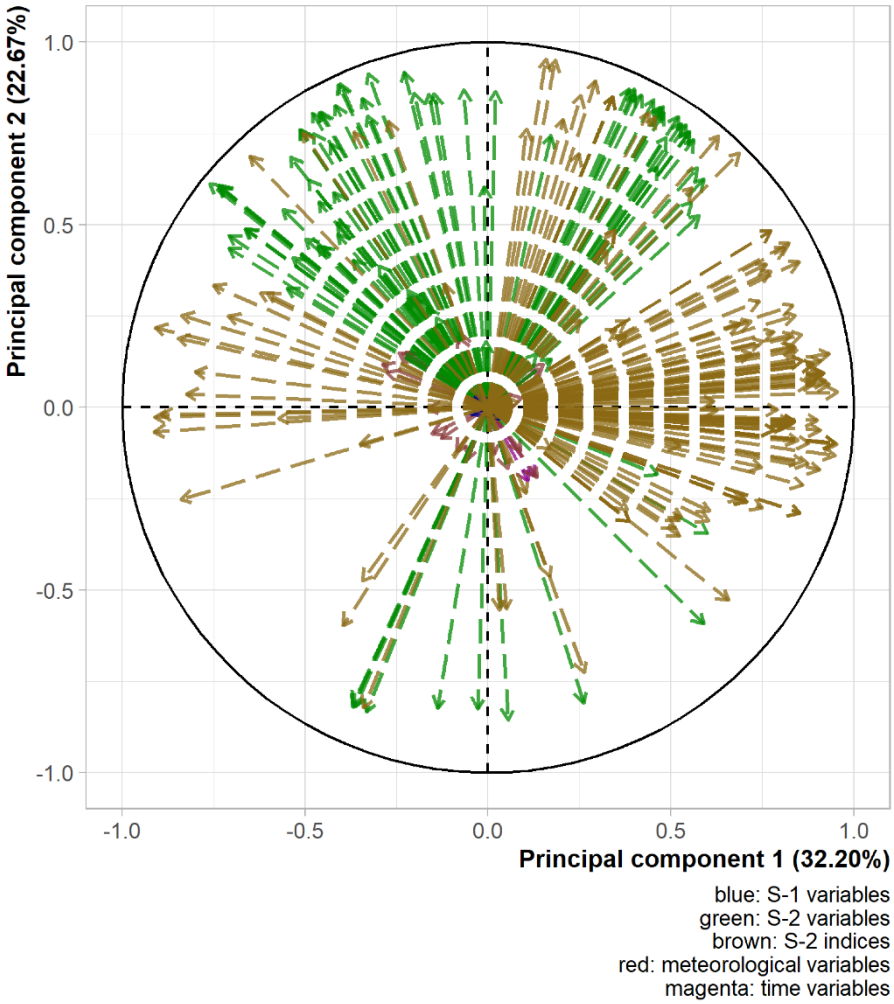


Figure 2-4: Variable position on the two first principal components for the calibration dataset. Transformed variables were pooled with their corresponding raw ones.

2.6.2 Variable Selection

The correlation plot of variable rankings between the 11 models used to select features is represented in Figure 2-5. Eleven models were represented, instead of 12, due to the inability of the rvm with a linear kernel to converge within a decent

time frame (one week). There were high correlation values between the ranking obtained by the models used, except for PLS.

The ranking scores for the 100 most-relevant variables were represented in Figure 2-6. The vast majority of those variables came from the index database (IDB [80]) family. The 10 most relevant were:

- S2T.B06.10exp which corresponds to the exponential transformation of band 06 of S2;
- S2T.B07.10exp which corresponds to the exponential transformation of band 07 of S2;
- S2T.B08.10exp which corresponds to the exponential transformation of band 08f of S2;
- IDB.032 known as Enhanced Vegetation Index;
- IDB.051 known as Hue;
- IDB.062 known as MCARI/MTVI2;
- IDB.071 known as mND680;
- IDB.221 known as Soil and Atmospherically Resistant Vegetation Index 2;
- CumT.DJ18.Last.10exp which corresponds to the exponential transformation of the degree-day with a basis of 18 ∘°C since the last CSH acquisition;
- S2T.B01.cube which corresponds to the cubic transformation of band 01 of S2.

The number behind the point in the "IDB.XXX" represents the index in Henrich et al. [80]. Band 03, 04 and 05 of S2 appeared in most indices. They corresponded respectively to green, red and near infra-red domains, which are known to be related to actual biomass (e.g., the NDVI ratio uses two of these bands). Another frequently appearing band was the first one, related to atmospheric correction and aerosol scattering. This suggested that there could be some residual effect of the pre-treatment concerning atmospheric condition.

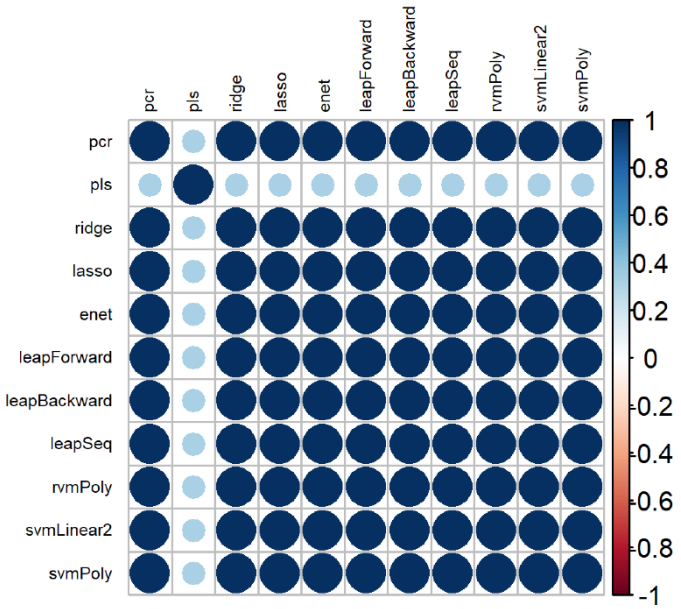


Figure 2-5: Correlation plot of the variable rankings between the different machine learning methods used.

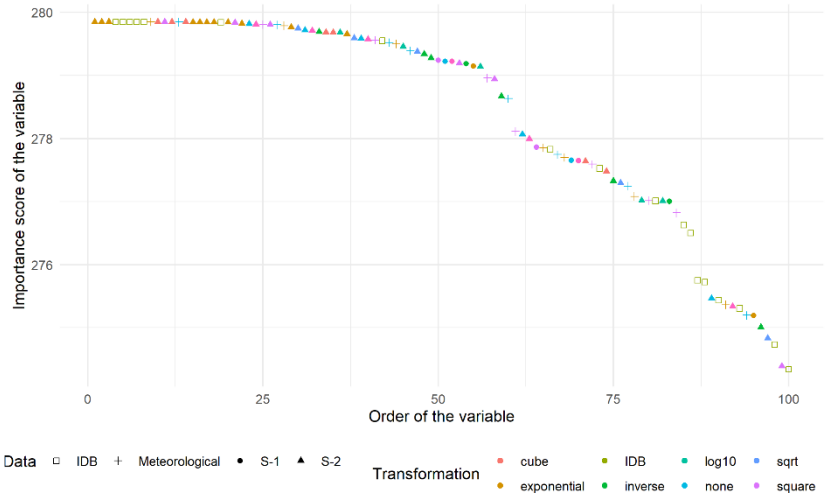


Figure 2-6: Ranking score of the 100 most relevant variables distinguished according to the type of data (shape) and the transformation applied (colour).

The R^2_{adj} curve related to the multivariate linear models built from an increasing number of variables (previously ranked on the basis of their score) is shown in Figure 2-7 with the breakpoints highlighted with vertical lines. The dot-dashed

breakpoints are the ones used in the analysis whilst the others (the plain lines) were left aside. The breakpoints taken into account in this study were related to linear models, including 7, 47, 111, 122 and 160 variables.

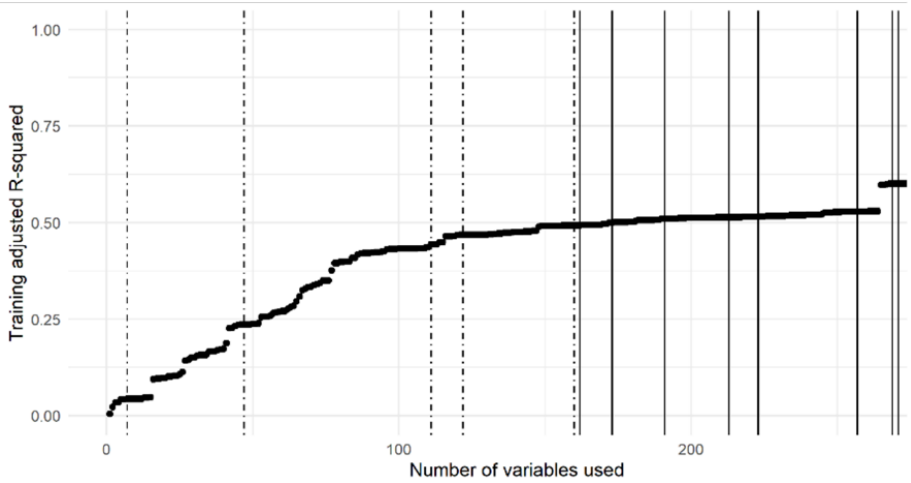


Figure 2-7 : Adjusted R-Squared curve trained on a cumulative number of variables sorted according to their ranking score. The breakpoints are represented with vertical lines. The dot-dashed lines correspond to the breakpoints taken into account, while the plain lines correspond to the other detected breakpoints.

2.6.3 Prediction of Compressed Sward Height (CSH)

From all subsets, approximately 50% of the 31 models run provided a RMSEcv lower than 100 mm of CSH (Table 2-5). Some models did not converge or provided extreme RMSEcv values. The models having RMSEcv values lower than 100 mm of CSH showed a tendency to decrease median RMSEcv values with an increase in the number of input variables. The minimum RMSEcv values were similar between subsets and lower than in the literature [25]. Figure 2-8 highlights the most powerful models based on their RMSEcv values. RF, cubist and enet models appeared to be highly repeatable between subsets. On the other hand, other models like the glm, lm, svmLinear2, leapSeq, ridge and svmPoly families showed a poor prediction ability.

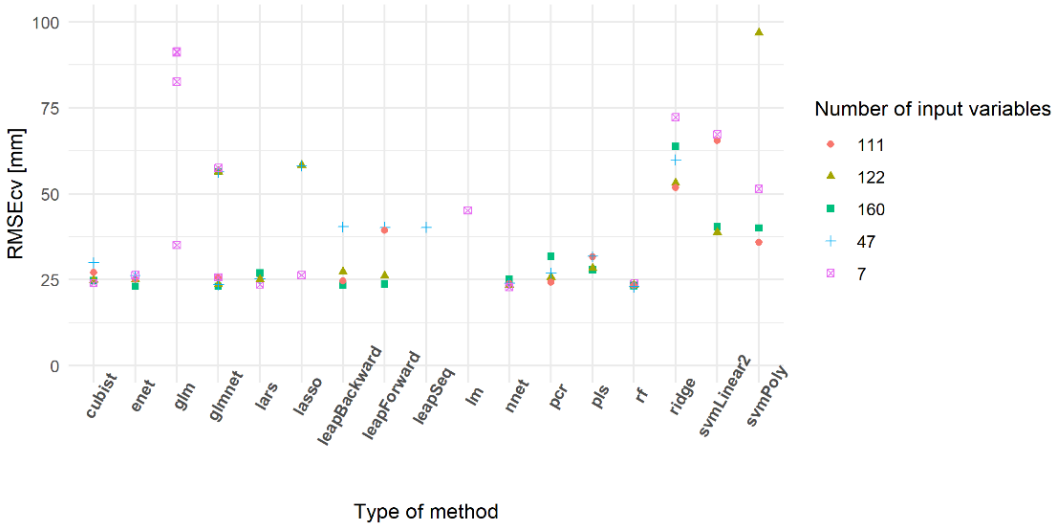


Figure 2-8: Cross-validation root mean square error (RMSEcv) of the best-predicting models.

Table 2-5: Cross-validation performances of machine learning models run on the selected subsets.

NbIn	NbInf	NbFail	NbSup100	NbOK	min	mean	median	max	sd
7	0	15	5	16	22.84	47.92	40.12	91.42	25.83
47	3	14	6	14	22.87	36.11	30.85	59.94	13.44
111	4	13	8	14	23.38	34.35	26.75	65.44	13.87
122	3	13	8	15	23.45	37.10	26.13	96.82	20.82
160	4	13	10	14	23.07	32.40	26.00	63.79	13.20

Note: NbIn: Number of input variables; NbInf: number of models having a non-finite mean RMSEcv; NbFail: number of models that failed to produce a usable expression; NbSup100: number of models having a mean RMSEcv above 100 mm of CSH; NbOK: number of models having a mean RMSEcv finite and below 100 mm of CSH

The prediction performances of the models were also assessed using an external validation dataset. As observed for the cross-validation (Table 2-5), some models did not provide realistic results (Table 2-6). The models with a validation RMSE lower than 100 mm showed a smaller range of variation than the one observed for the cross-validation. The minimum values of RMSE were also similar and confirm the potential interest of some developed models. Figure 2-9 shows the models having the lowest RMSEv. As observed in Figure 2-8, the rf, cubist and enet models seemed relevant as they had low RMSEv. Some models gained in terms of prediction capability, like rvmPoly that did not appear in Table 2-5 but well in Table 2-6. The distribution of RMSEv between subsets was relatively similar, except for the first subset. This could be related to a lack of information in the

inputs of the model. In other words, some models using seven variables could be under-fitted.

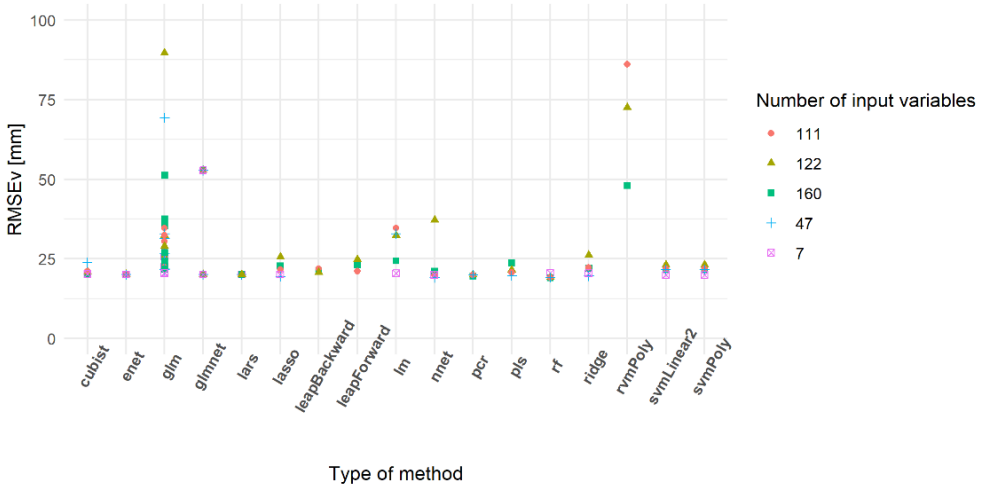


Figure 2-9: Validation root mean square error (RMSEv) of the best models predicting compressed sward height (mm).

The best models should have low RMSEv and RMSEcv values, as well as a low difference between them. The lower the difference, the better the model's robustness. Moreover, it is also important to have a distribution of the predicted values close to the one observed on the original dataset. Figure 2-10 developed the distribution of the actual CSH of the independent validation dataset and the CSH predicted by each model that provided realistic solutions on this dataset (i.e., NbOK model in Table 2-6). None of the models reproduced exactly the original distribution of the CSH. Two trends explained this differentiation: some models tended to group the prediction around the median of the distribution, and some models showed some saturation effect, resulting in a large amount of recording in the extreme classes. The cubist, nnet and rf methods and some linear regressions with variable selection were the methods that provided the best-fitting distribution curve. Some models provided extreme predictions.

The residuals were approximately centered on 0. Most models had a non-normal distribution of the residuals, given that their negative tail was often much larger than the positive one. The distribution of the residuals, according to the original CSH revealed that the most extreme residual values corresponded to CSH values near 50 mm of CSH for some glm models. Moreover, a higher absolute value of the residuals could be observed in the original CSH distribution extremes, meaning that all the information might not have been taken into account.

Table 2-6 : Validation performances of models which converged during the cross-validation for all selected subsets.

NbIn	NbInf	NbFail	NbSup100	NbOK	min	mean	median	max	sd
7	0	11	1	20	19.89	22.40	20.43	52.75	7.27
47	3	8	0	20	19.09	26.69	21.61	69.31	12.87
111	0	5	0	26	19.20	27.13	22.09	86.17	14.08
122	0	9	4	22	19.19	30.96	24.21	89.73	18.16
160	0	5	2	26	19.04	27.06	22.93	52.75	10.07

Note: NbIn: number of input variables; NbInf: number of models having a non-finite mean RMSEcv; NbFail: number of models that failed to produce a usable expression NbSup100: number of models having a RMSEcv above 100 mm of CSH; NbOK: number of models having a mean RMSEcv finite and below 100 mm of CSH.

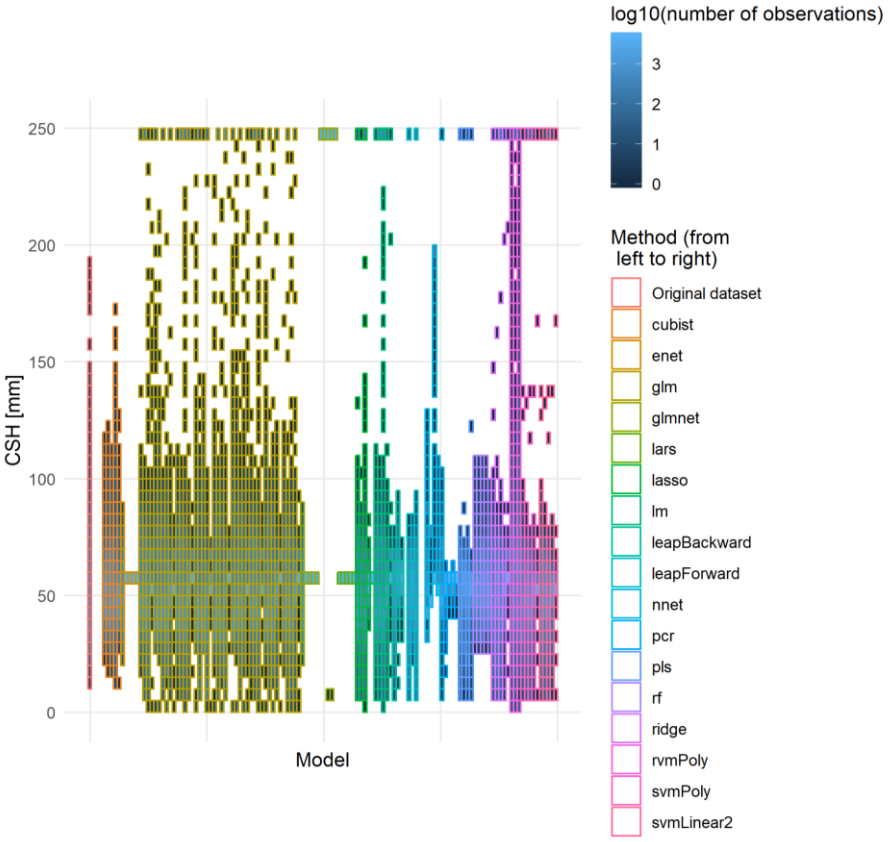


Figure 2-10: Distribution of the original and predicted CSH values. All models that managed to provide a usable expression were used to predict CSH (mm) on the independent validation dataset. The y-axis was divided into parts of 5 mm and for each part the number of records were counted. To ease the representation, the number of observations was transformed into its logarithm of base 10.

2.6.4 Effect of the Pixellation/Aggregation of Data

Beyond the pixel-base performances, the behavior of the models at the paddock scale was also assessed as the biomass that accounts for feed budgeting spread all over the parcel on which cattle graze. The RMSE of the training aggregated and of the validation aggregated dataset are summarised in Table 2-7. The lower threshold of around 19 mm of CSH that could be observed previously was completely blown up by the aggregation, with RMSE values reaching almost 4 mm of CSH in some cases. Unlike in the pixel analysis, the improvement of the quality indicators did not stand out for the validation dataset compared to the training one, although the RMSE was globally lower for the validation dataset. Two approaches were adopted in this table: on the one hand, the prediction was made solely on the parcel-scale aggregated data, and on the other hand, the prediction was made at the pixel level and then aggregated at the parcel-level. The first approach delivered better results for the calibration and the second for the independent validation dataset.

Table 2-7: Summary of the RMSE [mm] of the models on the CSH data aggregated at the parcel level. The difference between "Training aggregated" and "Aggregated training" was that the first case corresponded to a prediction on the aggregated training dataset and the second to an aggregation of the prediction made on the training dataset at a pixel scale. The same goes for the validation items.

Dataset	NbOK	NbFail	NbSup100	min	mean	median	max	sd
Training aggregated	121	3	0	9.16	17.68	16.17	55.79	8.70
Aggregated training	120	3	1	26.55	31.90	30.68	89.12	8.79
Validation aggregated	121	3	0	4.02	11.74	9.44	49.22	9.17
Aggregated validation	117	3	4	19.52	24.30	22.53	68.90	7.82

The RMSE decrease observed in Table 2-7 when the prediction was made on the pixel basis and then aggregated could be explained based on two different points of view. First, the variance of the reference datasets differed depending on the order of the prediction and aggregation. Secondly, a more materialistic explanation is the compensation of the sampling errors and the localization of the sampling: the GPS integrated to the RPM used did not have high precision, leading to positioning errors in the pixels. With the training data being comprised of medians of the records in each pixel, some records may have slipped from one pixel to another changing the value attributed to the pixel. The gathering of the information per parcel seemed to have erased a part of the error created. The effect of the difference in variance was taken into account by considering RPD.

A study of the prediction and the residuals similar to the one described above was performed on the validation dataset. The aggregation at the parcel-scale occurred after the prediction on the pixel. In general, the aggregation resulted in

deleting the most extreme values and a linearisation of the distribution of the predictions. However, multiple behaviours could be distinguished. First, some models narrowed the scope of the response, others drifted this range, and finally, a few broadened the values explored. The distribution of the residuals tended to be centered on 0, although some models drifted completely like before. No obvious relationship appeared between the residuals and the predicted CSH except for the models that already showed a tendency to drift.

2.6.5 Best Performing Models

The models that performed the best, i.e., having a low RMSE of independent validation are shown in Table 2-8. It summarises the main statistical parameters of the 20 models having the lowest RMSE of independent validation on the pixellated dataset: the family, the number of input variables, the RMSEcv and RMSEv, and residual prediction deviation on the two possible aggregations (pre- and post-prediction) for both training and validation datasets. A svmLinear reached the highest RPD for the training phase, an enet model for the independent validation, and a cubist model stood out regarding all the aggregations together. Concerning the effect of aggregation, the RPD was globally higher for the models aggregated pre-prediction on the training dataset. This, as well as the high number of negative differences of RPD for the training dataset, suggested that the aggregation at the parcel-scale before prediction was the most relevant method. The RPD difference related to the validation dataset implied the opposite, which matches the results of Table 2-7.

The correlation values between the predictions of the models on the validation pixellated dataset are represented in Figure 2-11. As Figure 2-10 suggested, the predictions of the various models were not similar, as correlation values were low. However, some models showed high correlations within the same method like rf, nnet and pcr or out of the family, such as rf's, nnet's, glmnet, and cubist. Conversely, some models of the same family with quite similar RPD values in Table 2-8 had correlations close to zero.

Table 2-8 : Description of the main features of the best-performing models from the RMSE of validation on pixellated independent validation point of view.

Method	NbVar	RPDcv	RPDiv	RPDta	RPDat	dRPDt	RPDva	RPDav	dRPDv
rf	160	0.88	0.98	0.78	0.92	0.14	1.00	0.88	-0.12
nnet	47	0.90	0.98	0.56	0.80	0.24	0.40	0.78	0.38
rf	47	0.86	0.99	0.99	0.86	-0.13	0.43	0.80	0.37
rf	122	0.88	0.99	1.21	0.84	-0.37	0.36	0.85	0.49
rf	111	0.88	0.99	1.04	0.95	-0.09	0.74	0.96	0.22
lasso	47	2.19	1.00	0.68	0.79	0.11	0.97	0.89	-0.08
pcr	160	1.20	1.01	1.18	0.00	-1.18	0.54	0.28	-0.26
ridge	47	2.26	1.01	1.02	1.00	-0.02	0.86	0.98	0.12
pls	47	1.20	1.01	1.06	0.85	-0.21	0.37	0.00	-0.37
nnet	111	0.90	1.02	0.45	0.83	0.38	0.41	0.81	0.40
svmPoly	7	1.94	1.02	1.23	0.95	-0.28	0.54	0.88	0.34
svmLinear2	7	2.54	1.02	1.32	0.85	-0.47	0.64	0.79	0.15
pcr	111	0.91	1.02	1.26	0.85	-0.41	0.34	0.86	0.52
pcr	122	0.97	1.02	1.09	0.86	-0.23	0.43	0.73	0.30
pcr	47	1.01	1.03	1.14	0.86	-0.28	0.39	0.80	0.41
nnet	7	0.86	1.03	0.69	0.71	0.02	0.39	0.00	-0.39
cubist	160	0.93	1.03	1.50	0.86	-0.64	0.74	0.96	0.22
glmnet	160	0.87	1.03	1.16	0.87	-0.29	0.51	0.83	0.32
enet	111	0.95	1.03	1.02	1.00	-0.02	0.86	0.98	0.12
enet	47	0.98	1.03	1.02	1.00	-0.02	0.86	0.98	0.12

Note: NbVar: total number of variables; RPDcv: RPD of the 19-fold cross-validation; RPDiv: RPD of validation on the independent validation dataset; RPDta: RPD of validation on the aggregated training dataset; RPDat: RPD of validation on the training dataset aggregated post-prediction; dRPDt: difference of RPD on the training dataset when an aggregation occurs; RPDva: RPD of validation on the aggregated validation dataset; RPDav: RPD of validation on the validation dataset aggregated post-prediction; dRPDv: difference of RPD on the validation dataset when an aggregation occurs

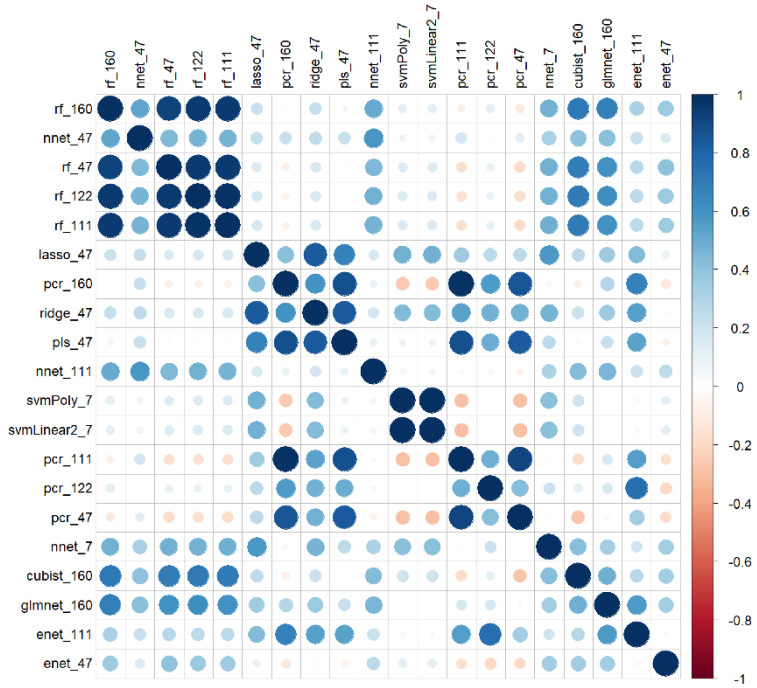


Figure 2-11: Correlations of the predictions of the 20 best models on the basis of the RMSE of independent validation.

Other quality indicators were also studied for some of these models: for each method, only the model with the lowest RMSE was kept. Figure 2-12 represents the distribution of the original independent pixellated validation dataset and the prediction made on this dataset. It appeared that most models did shrink the range of values adopted, although some extreme values appeared. From this graph, the cubist and rf models were the models most capable of reproducing the original distribution of predicted values, closely followed by a glmnet based on a gaussian family and nnet models.

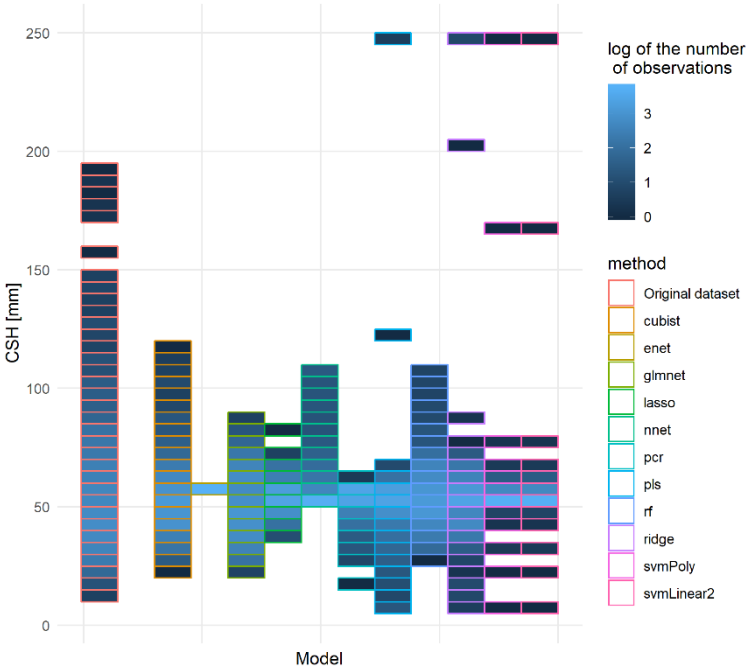


Figure 2-12: Distribution of the original and predicted CSH values of the independent pixellated validation dataset for the 11 best models, one per method. Each set of predictions of the models occupied a spot on the x-axis and the original pixellated validation dataset was on the left. The order of the models was the following: cubist, enet, glmnet, lasso, nnet, pcr, pls, rf, ridge, svmPoly, svmLinear2. The vertical axis corresponds to the CSH distribution: the axis was divided into parts of 5 mm of CSH and for each part the number of records was counted.

The erasure of the tails of the distribution of the predictions could be considered either as noise filtering or a loss of information. If this was a loss of information, it might complicate future detections of extreme cases. The subject noise filtering should be treated with caution. It could be interesting to recall that the parcels were divided into pixels of 10 m of resolution and the value associated with each pixel was the median of the values recorded in its range. A filtering step had thus already been applied. However, the amount of CSH recordings for each pixel was not constant. It implies that the filtering was not made equally. Some extreme values may have thus come through this filtering. Another factor in the noise filtering/erasure of the tails of the distribution of the prediction is the imbalance of the class used to train the model: the cross-validation occurred based on the date of acquisition, and no restriction had been set on the number of values in each sub-class of CSH which resulted in a training of the models that could have promoted central values. The application of some down-sampling techniques could be relevant to enlarge the scope of predictions.

Another point related to the distribution of the residuals was the trend to over- or under-estimate the actual value. The distribution of the residuals showed no clear trend: the center of gravity seemed to be positive, and the tail made of the negative values seemed to be more important than the one made of positive values of residuals. Table 2-9summarises the percentage of positive and negative values for each model. The models tended to produce positive residuals, which meant that the predictions were globally higher than the actual value, except for the glmnet and cubist models. This indicated that there was a global over-estimation. However, this trend was not that pronounced for cubist and pls models.

Table 2-9 : Percentage of positive (%>0) and negative (%<0) residuals values for each model tested.

Category	cubist	enet	glmnet	lasso	nnet	pcr	pls	rf	ridge	svmPoly	svmLinear2
%>0	52.6	65.5	49.3	60.4	64.6	60.5	55.9	58.1	56.7	57.4	57.4
%<0	47.4	34.5	50.7	39.6	35.4	39.5	44.1	41.9	43.3	42.6	42.6

The analysis of the indicators of the quality of prediction and residuals at a pixel-scale with a 10 m resolution led to the selection of four best models for predicting CSH: the cubist model using 160 variables, the glmnet model of the gaussian family based on 160 variables, the rf model based on 160 variables and the nnet model based on 47 variables.

2.7 Discussion

The main objective of this work was to predict grass biomass in pastures on the basis of merged heterogeneous data sources of information, multiple variable transformations and a broad scope of ML methods.

The first adaptation was the use of CSH measurements instead of real grass biomass. The acceptance of this proxy (CSH) in the scientific community [24,68,92,93] and the ease of acquiring these data were two of the reasons that led to this choice. Although it could have been possible to convert the CSH records to biomass with an equation provided for similar sward species and meteorological conditions by the manufacturer of the rising platemeter, the conversion was not made in order to prevent any interference with cut height. The experimental farm had records of biomass and linked differences of CSH before and after cutting since 2015. Linear regressions were trained on this basis with multiple combinations of scenarios: the fixation or not of the intercept, the determination of the model for each parcel sampled or all of them, the use of dry or raw biomass (data not shown). Some seasonal trends stood out but no consistency in the coefficient could be observed within this dataset (the values of the regression coefficient ranged from 50 to 1500 for CSH expressed in millimeters), nor with the equations provided by the seller (the pool of linear regression coefficient for the dry biomass was distributed around 150). Cudlin et al. [10] also faced a high variability in regression coefficient and intercept values. However, they had more observations of biomass and CSH combined—at least 24 for each regression compared to at best 10 records for each parcel for each period of the year in the case of the experimental farm. This means they could rely on their few significant relationships whilst the regression computed for the experimental farm was not resilient. Hakl et al. [17] also had results showing varying reliability and significance. The combination of these studies showed the sensitivity of the conversion from CSH to biomass, probably due to the floristic composition and, together with Ferraro et al. [94], their results underlined this sensibility to a seasonal effect whose expression might be more or less mitigated depending on the year and the location. Another aspect that could have been considered was the direct use of the biomass records mentioned before. However, they were acquired at a low frequency (approximately 49 days) and the bands mown to get this data were too small to occupy a full pixel scale. This means that the signal would have been noisy and blurred and the total amount of records would have been 239 since 2015, with some records acquired before the launch of S2 satellites. The use of CSH measurements as the observed variable was also motivated by the ease of

visualizing this data for farmers. Crémer [18] uses another meaningful example in his advice to farmers: the height of the grass compared to the ankle position on a boot.

Nakagami (2016) and Cimbelli and Vitale (2017) had similar trends in the distribution of the heights: namely, a non-normal distribution with a lower threshold and a peak near the small values, although this was less pronounced in their data. This might be related to pasture management practices in Wallonia. To allow for optimal quality in the consumed herbage grazing management should aim to maintain a tight sward where perennial species dominate and have no time to flower and reproduce in a vegetative manner. Applying such management obviously has an impact on the lower representation of flower and seed heads of higher height in the database.

Multiple sources of information were integrated into this work. The faint grouping effect of variables seen in the representation of the variables on the two first principal components (Figure 2-4) confirmed the non-redundancy of the information brought by the different groups of variables studied and their transformations. Another aspect that proved the relevance of integrating all these variables is the appearance of each group of variables, with or without transformation, at the beginning of the ranking of variables (Figure 2-6). S2 ranked first then meteorological data followed by S1 from the 50th onwards. Although a seasonal effect was observed in the attempts at conversion from CSH to actual biomass, the time markers inserted within the dataset did not seem to bring that much information given that the first appearance is in the 131st place. The use of one unique dataset of meteorological data could have led to the under-training of the models concerning these variables. The implementation of a spatialized meteorological dataset is considered. The use of the first band of S2 data and its rather higher position in the variable ranking indicate that there could be some kind of artifact/residual effect due to the pre-processing chain.

The sources of information mentioned above were integrated into their raw and transformed forms. The relevance of this integration was confirmed in the representation of the variables according to their decreasing ranking score (Figure 2-6). The “amplifying” transformations, i.e., cubic, exponential and square and the combinations of bands of S2 data were the most represented at the beginning of the ranking, although the other transformations also appeared, albeit less often. This means that the data transformations highlighted parts of the information.

Multiple ML methods were trained on the dataset based on the hypothesis that some features used in the modelling could handle collinearity differently and detect different parts of information. The methods that could handle collinearity

were used in the variable selection. It appeared that all of them, apart from the PLS, detected the same order of importance of the variables which was translated by the high Spearman correlation of the rankings. Although this high level of similarity could ensure the validity of the models, the dependence of the residuals on the original CSH slightly detected on the validation dataset indicated that all the information explaining CSH was not completely taken into account. The most promising models were a cubist, a glmnet of the gaussian family, a rf and a nnet models based on S1, S2 and meteorological data. While we could not find any article using glmnet in the literature, random forest, cubist and neural networks were related to spectral enhancement and prediction of other target variables, like leaf area index (LAI) [48,84,85,86,87,88,89,90,95]. However, they were never used to predict biomass or CSH on pastures with S1 and S2 datasets.

The hypothesis of pixellation of CSH data made at the beginning of this work was possible thanks to the geolocation linked to the measurements on the pastures. The objective of this fractionation of the study area was to increase the number of records that could be used during the training of the ML models, while limiting noise due to the allocation of different CSH for the same set of combinations of satellite variables. To enable posterior comparisons, the metrics of quality were computed both at the pixel scale and the parcel scale for the models developed at the pixel scale. This revealed that with and without aggregation, and whatever the aggregation method, the best-fitted models achieved better performances than the 29 mm RMSE shown by Cimbelli and Vitale [25]. They were the only ones to provide similar metrics in a similar context, although their height corresponded to the mean height computed on a picture instead of a CSH. The pixellation and aggregation processes acted as a partial filter concerning extreme values: the pixellation led to the erasure of the most extreme values for some pixels with the median filter, and the aggregation did the same with a mean filter.

One way to increase the size of the dataset could have been to create a continuum of CSH values within the parcels by using a kriging-like technique. However, this would mean that each record of CSH used would not have been completely independent from the others, given that a huge part of them would have been created as a combination of other measurements. Other types of gap-filling techniques were omitted due to their inherent complexity: sets of records on cloudy days were discarded because they made the use of S2 data meaningless. A way to avoid this decrease in the dataset size would have been to use tiles acquired on the nearest dates and interpolate the values for each pixel. However, this would have required that the pastures would have witnessed no change in their management or condition: no removal nor addition of cows, nor any other

operation like fertilization or mowing. Another obstacle to this process is the long-term archiving policy of ESA which complexified the acquisition of all relevant tiles, especially for the oldest ones. A similar problem of continuity in practices was dealt with using an approximation: the CSH data were not acquired exactly at the same time as S1 and S2 data. In this case, the closest tile in time was used, assuming that the frequency of satellite data acquisition is high enough to avoid major changes in pasture condition.

The future application of the models will be their integration within a DSS for the use of Walloon farmers. The objective will be to provide the farmers with information regarding the quantity of feed available on pastures. This means that the aggregated value of the predictions at the pixel-scale will probably be used. Filtering of extreme values would reveal useful for this assessment but the information about the refusals will be lost. The global over- and underestimation trend shown in Table 2-9 should be inserted in the DSS and treated with caution: an imprecision inherent to the sampling method using the rising platemeter has to be kept in mind. That is, it measures heights with a count of clicks on a ratchet and then converts this to CSH. The conversion is a source of error, and McSweeney et al. [34] warn that this type of platemeter often under-estimates actual CSH height. Another source of the unreliability of the developed models is the lack of diversity in the farms that were used and more importantly in the diversity of pasture botanical composition, soil types, fertility and management practices. The use of a completely independent validation dataset offsets this bias and imitated the future behaviour of the models on other farms.

2.8 Conclusions

The main objective of this work was to test the following hypothesis: the use of various sources of information and multiple transformations of these data as inputs for a broad range of machine learning methods may improve predictions of CSH in pastures. The combination of sources of information, data transformations, and multiple machine learning methods allowed the development of more precise models than previously described in the literature. Four models stood out: a cubist, a glmnet, a neural network and a random forest-based model. They were all based on Sentinel-1 sigma nought, Sentinel-2 reflectance and meteorological data. Their RMSE of independent validation was around 19 mm of CSH at the pixel-level. To better train the models, more records will be gathered in the years to come, in Wallonia and possibly internationally. The models trained through this framework will be used to establish a tool to help farmers in their daily decision making. It is also planned to enable prediction at a larger scale, include an extreme case detection in the predictions and sampling the places where the extreme cases did occur to increase the resilience of the models. Moreover, for both the DSS and large scale prediction, a combination of the outputs is considered, given that the similarity of distribution does not reflect the different repartitions of the predictions.

Author Contributions

Conceptualization, C.N. and H.S.; software, C.N., A.T. and B.T.; validation, A.T., I.D., F.L., B.T., Y.C., J.B. and H.S.; formal analysis, C.N., A.T. and H.S.; resources, I.D. and F.L.; data curation, C.N., A.T., I.D., F.L., B.T., Y.C., J.B. and H.S.; writing—original draft preparation, C.N.; writing—review and editing, A.T., I.D., F.L., B.T., Y.C., J.B. and H.S.; visualization, C.N. and H.S.; supervision, H.S.; project administration, H.S. All authors have read and agreed to the published version of the manuscript.

2.9 References

1. Hennessy, D.; Delaby, L.; van den Pol-van Dasselaar, A.; Shalloo, L. Increasing Grazing in Dairy Cow Milk Production Systems in Europe. *Sustainability* **2020**, *12*, 2443. [Google Scholar] [CrossRef][Green Version]
2. Elgersma, A. Grazing increases the unsaturated fatty acid concentration of milk from grass-fed cows: A review of the contributing factors, challenges and future perspectives. *Eur. J. Lipid Sci. Technol.* **2015**, *117*, 1345–1369. [Google Scholar] [CrossRef]
3. Lessire, F.; Jacquet, S.; Veselko, D.; Piraux, E.; Dufrasne, I. Evolution of grazing practices in Belgian dairy farms: Results of two surveys. *Sustainability* **2019**, *11*, 3997. [Google Scholar] [CrossRef][Green Version]
4. Cros, M.J.; Garcia, F.; Martin-Clouaire, R. SEPATOU: A Decision Support System for the Management of Rotational Grazing in a Dairy Production. In Proceedings of the 2nd European Conference on Information Technology in Agriculture, Bonn, Germany, 27–30 September 1999; pp. 549–557. [Google Scholar]
5. Amalero, E.G.; Ingua, G.L.; Erta, G.B.; Emanceau, P.L. A biophysical dairy farm model to evaluate rotational grazing management strategies. *Agronomie* **2003**, *23*, 407–418. [Google Scholar] [CrossRef]
6. Romera, A.J.; Beukes, P.; Clark, C.; Clark, D.; Levy, H.; Tait, A. Use of a pasture growth model to estimate herbage mass at a paddock scale and assist management on dairy farms. *Comput. Electron. Agric.* **2010**, *74*, 66–72. [Google Scholar] [CrossRef]
7. Romera, A.; Beukes, P.; Clark, D.; Clark, C.; Tait, A. Pasture growth model to assist management on dairy farms: Testing the concept with farmers. *Grassl. Sci.* **2013**, *59*, 20–29. [Google Scholar] [CrossRef]
8. Ruelle, E.; Hennessy, D.; Delaby, L. Development of the Moorepark St Gilles grass growth model (MoSt GG model): A predictive model for grass growth for pasture based systems. *Eur. J. Agron.* **2018**, *99*, 80–91. [Google Scholar] [CrossRef]
9. Moeckel, T.; Safari, H.; Reddersen, B.; Fricke, T.; Wachendorf, M. Fusion of ultrasonic and spectral sensor data for improving the estimation of biomass in grasslands with heterogeneous sward structure. *Remote Sens.* **2017**, *9*, 98. [Google Scholar] [CrossRef][Green Version]
10. Cudlín, O.; Hakl, J.; Hejcman, M.; Cudlín, P. The use of compressed height to estimate the yield of a differently fertilized meadow. *Plant Soil Environ.* **2018**, *64*, 76–81. [Google Scholar] [CrossRef][Green Version]
11. Lussem, U.; Bolten, A.; Menne, J.; Gnyp, M.L.; Schellberg, J.; Bareth, G. Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices. *J. Appl. Remote Sens.* **2019**, *13*, 034525. [Google Scholar] [CrossRef]
12. Laca, E.A.; Demment, M.W.; Winckel, J.; Kie, J.G. Comparison of weight estimate and rising-plate meter methods to measure herbage mass of a mountain meadow. *J. Range Manag.* **1989**, *42*, 71–75. [Google Scholar] [CrossRef][Green Version]
13. French, P.; O'Brien, B.; Shalloo, L. Development and adoption of new technologies to increase the efficiency and sustainability of pasture-based systems. *Anim. Prod. Sci.* **2015**, *55*, 931–935. [Google Scholar] [CrossRef]
14. MacAdam, J.; Hunt, S. *Using a Rising Plate Meter to Determine Paddock Size for Rotational Grazing*; Utah State University Extension Bulletin: Logan, UT, USA, 2015. [Google Scholar]
15. Gargiulo, J.; Clark, C.; Lyons, N.; de Veyrac, G.; Beale, P.; Garcia, S. Spatial and temporal pasture biomass estimation integrating electronic plate meter, planet cubesats and sentinel-2 satellite data. *Remote Sens.* **2020**, *12*, 3222. [Google Scholar] [CrossRef]
16. Wang, J.; Xiao, X.; Bajgain, R.; Starks, P.; Steiner, J.; Doughty, R.B.; Chang, Q. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 189–201. [Google Scholar] [CrossRef][Green Version]
17. Hakl, J.; Hrevušová, Z.; Hejcman, M.; Fuksa, P. The use of a rising plate meter to evaluate lucerne (Medicago sativa L.) height as an important agronomic trait enabling yield estimation. *Grass Forage Sci.* **2012**, *67*, 589–596. [Google Scholar] [CrossRef]
18. Crémer, S. *La Gestion des Prairies—Notes de Cours 2015–2016*; Fourrages-mieux, Marche-en-Famenne: Marloie, Belgium, 2015; pp. 1–142. [Google Scholar]
19. Nakagami, K. A method for approximate on-farm estimation of herbage mass by using two assessments per pasture. *Grass Forage Sci.* **2016**, *71*, 490–496. [Google Scholar] [CrossRef]
20. Fricke, T.; Wachendorf, M. Combining ultrasonic sward height and spectral signatures to assess the biomass of legume-grass swards. *Comput. Electron. Agric.* **2013**, *99*, 236–247. [Google Scholar] [CrossRef]

21. Bareth, G.; Schellberg, J. Replacing Manual Rising Plate Meter Measurements with Low-cost UAV-Derived Sward Height Data in Grasslands for Spatial Monitoring. *PFG J. Photogramm. Remote. Sens. Geoinf. Sci.* **2018**, *86*, 157–168. [Google Scholar] [CrossRef]
22. Legg, M.; Bradley, S. Ultrasonic Arrays for Remote Sensing of Pasture Biomass. *Remote Sens.* **2019**, *12*, 111. [Google Scholar] [CrossRef][Green Version]
23. Rayburn, E.B.; Lozier, J.D.; Sanderson, M.A.; Smith, B.D.; Shockey, W.L.; Seymore, D.A.; Fultz, S.W. Alternative Methods of Estimating Forage Height and Sward Capacitance in Pastures Can Be Cross Calibrated. *Forage Grazinglands* **2007**, *5*, 1–6. [Google Scholar] [CrossRef][Green Version]
24. Lopez diaz, J.; Gonzalez-rodriguez, A. Measuring grass yield by non-destructive methods. *J. Chem. Inf. Model.* **2013**, *53*, 1689–1699. [Google Scholar] [CrossRef]
25. Cimbelli, A.; Vitale, V. Grassland height assessment by satellite images. *Adv. Remote Sens.* **2017**, *6*, 40–53. [Google Scholar] [CrossRef][Green Version]
26. Ancin-Murguzur, F.J.; Taff, G.; Davids, C.; Tømmervik, H.; Mølmann, J.; Jørgensen, M. Yield estimates by a two-step approach using hyperspectral methods in grasslands at high latitudes. *Remote Sens.* **2019**, *11*, 400. [Google Scholar] [CrossRef][Green Version]
27. Zeng, L.; Chen, C. Using remote sensing to estimate forage biomass and nutrient contents at different growth stages. *Biomass Bioenergy* **2018**, *115*, 74–81. [Google Scholar] [CrossRef]
28. Tiscornia, G.; Baethgen, W.; Ruggia, A.; Do Carmo, M.; Ceccato, P. Can we Monitor Height of Native Grasslands in Uruguay with Earth Observation? *Remote Sens.* **2019**, *11*, 1801. [Google Scholar] [CrossRef][Green Version]
29. Michez, A.; Philippe, L.; David, K.; Sébastien, C.; Christian, D.; Bindelle, J. Can low-cost unmanned aerial systems describe the forage quality heterogeneity? Insight from a timothy pasture case study in southern Belgium. *Remote Sens.* **2020**, *12*, 1650. [Google Scholar] [CrossRef]
30. Borra-Serrano, I.; De Swaef, T.; Muylle, H.; Nuyttens, D.; Vangeyte, J.; Mertens, K.; Saeys, W.; Somers, B.; Roldán-Ruiz, I.; Lootens, P. Canopy height measurements and non-destructive biomass estimation of Lolium perenne swards using UAV imagery. *Grass Forage Sci.* **2019**, *74*, 356–369. [Google Scholar] [CrossRef]
31. Michez, A.; Lejeune, P.; Bauwens, S.; Lalaina Herinaina, A.A.; Blaise, Y.; Muñoz, E.C.; Lebeau, F.; Bindelle, J. Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system. *Remote Sens.* **2019**, *11*, 473. [Google Scholar] [CrossRef][Green Version]
32. Shalloo, L.; Donovan, M.O.; Leso, L.; Werner, J.; Ruelle, E.; Geoghegan, A.; Delaby, L.; Leary, N.O. Review: Grass-based dairy systems, data and precision technologies. *Animal* **2018**, *12*, S262–S271. [Google Scholar] [CrossRef][Green Version]
33. Eastwood, C.R.; Dela Rue, B.T.; Gray, D.I. Using a 'network of practice' approach to match grazing decision-support system design with farmer practice. *Anim. Prod. Sci.* **2017**, *57*, 1536–1542. [Google Scholar] [CrossRef]
34. McSweeney, D.; Coughlan, N.E.; Cuthbert, R.N.; Halton, P.; Ivanov, S. Micro-sonic sensor technology enables enhanced grass height measurement by a Rising Plate Meter. *Inf. Process. Agric.* **2019**, *6*, 279–284. [Google Scholar] [CrossRef]
35. Jouven, M.; Carrère, P.; Baumont, R. Model predicting dynamics of biomass, structure and digestibility of herbage in managed permanent pastures. 1. Model description. *Grass Forage Sci.* **2006**, *61*, 112–124. [Google Scholar] [CrossRef]
36. Thornley, J.H.M. *Grassland Dynamics: An Ecosystem Simulation Model*; CAB International: Wallingford, UK, 1998. [Google Scholar]
37. Sándor, R.; Ehrhardt, F.; Brilli, L.; Carozzi, M.; Recous, S.; Smith, P.; Snow, V.; Soussana, J.F.; Dorich, C.D.; Fuchs, K.; et al. The use of biogeochemical models to evaluate mitigation of greenhouse gas emissions from managed grasslands. *Sci. Total Environ.* **2018**, *642*, 292–306. [Google Scholar] [CrossRef] [PubMed][Green Version]
38. Brisson, N.; Mary, B.; Ripoche, D.; Jeuffroy, M.H.; Ruget, F.; Nicoullaud, B.; Gate, P.; Devienne-Barret, F.; Antonioletti, R.; Durr, C.; et al. STICS: A generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie* **1998**, *18*, 311–346. [Google Scholar] [CrossRef]
39. McDonnell, J.; Brophy, C.; Ruelle, E.; Shalloo, L.; Lambkin, K.; Hennessy, D. Weather forecasts to enhance an Irish grass growth model. *Eur. J. Agron.* **2019**, *105*, 168–175. [Google Scholar] [CrossRef]
40. Tamm, T.; Zalite, K.; Voormansik, K.; Talgre, L. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. *Remote Sens.* **2016**, *8*, 802. [Google Scholar] [CrossRef][Green Version]
41. Stendardi, L.; Karlsen, S.R.; Niedrist, G.; Gerdol, R.; Zebisch, M.; Rossi, M.; Notarnicola, C. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. *Remote Sens.* **2019**, *11*, 542. [Google Scholar] [CrossRef][Green Version]

42. Shoko, C.; Mutanga, O.; Dube, T. Determining optimal new generation satellite derived metrics for accurate C3 and C4 Grass species aboveground biomass estimation in South Africa. *Remote Sens.* **2018**, *10*, 564. [Google Scholar] [CrossRef][Green Version]
43. Kumar, L.; Sinha, P.; Taylor, S.; Alqurashi, A.F. Review of the use of remote sensing for biomass estimation to support renewable energy generation. *J. Appl. Remote Sens.* **2015**, *9*, 097696. [Google Scholar] [CrossRef]
44. Punalekar, S.M.; Verhoef, A.; Quaife, T.L.; Humphries, D.; Bermingham, L.; Reynolds, C.K. Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sens. Environ.* **2018**, *218*, 207–220. [Google Scholar] [CrossRef]
45. Mutanga, O.; Shoko, C. Monitoring the spatio-temporal variations of C3/C4 Grass species using multispectral satellite data. *IGARSS* **2018**, *2018*, 8988–8991. [Google Scholar] [CrossRef]
46. Darvishzadeh, R.; Wang, T.; Skidmore, A.; Vrieling, A.; O'Connor, B.; Gara, T.W.; Ens, B.J.; Paganini, M. Analysis of Sentinel-2 and rapidEye for retrieval of leaf area index in a saltmarsh using a radiative transfer model. *Remote Sens.* **2019**, *11*, 671. [Google Scholar] [CrossRef][Green Version]
47. Garioud, A.; Giordano, S.; Valero, S.; Mallet, C. Challenges in Grassland Mowing Event Detection with Multimodal Sentinel Images. *MultiTemp* **2019**, 1–4. [Google Scholar] [CrossRef][Green Version]
48. Alves, R.; Näsi, R.; Niemeläinen, O.; Nyholm, L.; Alhonoja, K.; Kaivosoja, J.; Jauhiainen, L.; Viljanen, N.; Nezami, S.; Markelin, L. Remote Sensing of Environment Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sens. Environ.* **2020**, *246*, 111830. [Google Scholar] [CrossRef]
49. Fernández-Delgado, M.; Sirsat, M.S.; Cernadas, E.; Alawadi, S.; Barro, S.; Febrero-Bande, M. An extensive experimental survey of regression methods. *Neural Netw.* **2019**, *111*, 11–34. [Google Scholar] [CrossRef]
50. García, R.; Aguilar, J.; Toro, M.; Pinto, A.; Rodríguez, P. A systematic literature review on the use of machine learning in precision livestock farming. *Comput. Electron. Agric.* **2020**, *179*, 105826. [Google Scholar] [CrossRef]
51. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. [Google Scholar]
52. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, Inc.: Boston, MA, USA, 2019. [Google Scholar]
53. Moot, D.J.; Scott, W.R.; Roy, A.M.; Nicholls, A.C.; Scott, W.R.; Roy, A.M.; Base, A.C.N. Base temperature and thermal time requirements for germination and emergence of temperate pasture species. *N. Z. J. Agric. Res.* **2010**, *8233*, 15–25. [Google Scholar] [CrossRef][Green Version]
54. Balocchi, O.; Alonso, M.; Keim, J.P. Water-Soluble Carbohydrate Recovery in Pastures of Perennial Ryegrass (*Lolium perenne* L.) and Pasture Brome (*Bromus valdivianus* Phil.) Under Two Defoliation Frequencies Determined by Thermal Time. *Agriculture* **2020**, *10*, 563. [Google Scholar] [CrossRef]
55. Anandhi, A. Growing Degree Days—Ecosystem Indicator for changing diurnal temperatures and their impact on corn growth stages in Kansas. *Ecol. Indic.* **2016**, *61*, 149–158. [Google Scholar] [CrossRef][Green Version]
56. Salvucci, M.E.; Osteryoung, K.W.; Crafts-brandner, S.J.; Vierling, E. Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation in Vitro and in Vivo 1. *Plant Physiol.* **2001**, *127*, 1053–1064. [Google Scholar] [CrossRef]
57. Greco, M.; Parry, A.J.; Andralojc, J.; Carmo-silva, A.E.; Alonso, H. Rubisco activity and regulation targets for crop in DNA In Posidonia oceanica cadmium as induces changes improvement methylation and chromatin patterning. *J. Exp. Bot.* **2013**, *64*, 717–730. [Google Scholar] [CrossRef][Green Version]
58. Copernicus. Open Access Hub. 2018. Available online: https://scihub.copernicus.eu/ (accessed on 7 december 2020).
59. European Space Agency. *Sen2Cor / STEP*; European Space Agency: Paris, France, 2018. [Google Scholar]
60. Schwalb-Willmann, J. getSpatialData: Get Different Kinds of Freely Available Spatial Datasets. R Package Version 0.0.4. 2018. Available online: https://rdrr.io/github/16EAGLE/getSpatialData/f/NEWS.md (accessed on 7 december 2020).
61. Filipponi, F. Sentinel-1 GRD Preprocessing Workflow. *Proceedings* **2019**, *18*, 11. [Google Scholar] [CrossRef][Green Version]
62. Filipponi, F. *Sentinel-1_GRD_Preprocessing_ Standard Workflow for the Preprocessing of Sentinel-1 GRD Satellite Data*; MDPI: Basel, Switzerland, 2020. [Google Scholar]
63. ESA. *SNAP—ESA Sentinel Application Platform v7.0.0*; ESA: Paris, France, 2020. [Google Scholar]
64. Rudant, J.P.; Frison, P.L. Télédéction radar: De l'image d'intensité initiale au choix du mode de calibation des coefficients de diffusion beta nought, sigma nought et gamma nought. *Revue Française de Photogrammétrie et Télédétection* **2019**, *219–220*, 19–28. [Google Scholar]

65. Ranghetti, L.; Busetto, L. Sen2r: Find, Download and Process Sentinel-2 data. R Package Version 1.2.1. 2019. Available online: **https://doi.org/10.5281/zenodo.1240384** (accessed on 7 december 2020).
66. Mueller-Wilm, U. *Sen2Cor Configuration and User Manual*; ESA: Paris, France, 2016. [**Google Scholar**]
67. GDAL/OGR Contributors. GDAL/OGR Geospatial Data Abstraction Software Library. Open Source Geospatial Foundation. 2020. Available online: **https://gdal.org/** (accessed on 7 december 2020).
68. Higgins, S.; Schellberg, J.; Bailey, J.S. Improving productivity and increasing the efficiency of soil nutrient management on grassland farms in the UK and Ireland using precision agriculture technology. *Eur. J. Agron.* **2019**, *106*, 67–74. [**Google Scholar**] [**CrossRef**]
69. Lugassi, R.; Zaady, E.; Goldshleger, N.; Shoshany, M.; Chudnovsky, A. Spatial and temporal monitoring of pasture ecological quality: Sentinel-2-based estimation of crude protein and neutral detergent fiber contents. *Remote Sens.* **2019**, *11*, 799. [**Google Scholar**] [**CrossRef**][**Green Version**]
70. Ruelle, E.; Delaby, L. Pertinence du Modèle Moorepark-St Gilles Grass Growth dans les conditions climatiques de l Ouest de la France; Description du modèle Moorepark-St Gilles Grass Growth a) b). 2017, pp. 158–159. Available online: **https://hal.archives-ouvertes.fr/hal-01595315/** (accessed on 7 december 2020).
71. Dowle, M.; Srinivasan, A. data.table: Extension of 'data.frame'. R Package Version 1.12.8. 2019. Available online: **https://cran.r-project.org/web/packages/data.table/index.html** (accessed on 7 december 2020).
72. Wickham, H.; François, R.; Henry, L.; Müller, K. dplyr: A Grammar of Data Manipulation; R Package Version 0.8.3. 2019. Available online: **https://dplyr.tidyverse.org/** (accessed on 7 december 2020).
73. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10*, 439–446. [**Google Scholar**] [**CrossRef**][**Green Version**]
74. Pebesma, E.J.; Bivand, R.S. Classes and methods for spatial data in R. *R News* **2005**, *5*, 9–13. [**Google Scholar**]
75. Bivand, R.S.; Pebesma, E.; Gomez-Rubio, V. *Applied Spatial Data Analysis with R*, 2nd ed.; Springer: New York, NY, USA, 2013. [**Google Scholar**]
76. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. R Package Version 3.0-7. 2019. Available online: **https://rdrr.io/cran/raster/** (accessed on 7 december 2020).
77. Bengtsson, H. Future: Unified Parallel and Distributed Processing in R for Everyone. R Package Version 1.16.0. 2020. Available online: **https://cran.r-project.org/web/packages/future/index.html** (accessed on 7 december 2020).
78. Bengtsson, H. Future.Apply: Apply Function to Elements in Parallel using Futures; R Package Version 1.4.0. 2020. Available online: **https://cran.r-project.org/web/packages/future.apply/index.html** (accessed on 7 december 2020).
79. Ali, I.; Cawkwell, F.; Dwyer, E.; Barrett, B.; Green, S. Satellite remote sensing of grasslands: From observation to management. *J. Plant Ecol.* **2016**, *9*, 649–671. [**Google Scholar**] [**CrossRef**][**Green Version**]
80. Henrich, V.; Götze, C.; Jung, A.; Sandow, C.; Thürkow, D.; Cornelia, G. Development of an online indices database: Motivation, concept and implementation. In Proceedings of the 6th EARSeL Imaging Spectroscopy SIG Workshop Innovative Tool for Scientific and Commercial Environment Applications, Tel Aviv, Israel, 16–18 march 2009. [**Google Scholar**]
81. IDB—Sensor_ Sentinel-2A. Available online: **https://www.indexdatabase.de/db/s-single.php?id=96** (accessed on 23 January 2021).
82. Lê, S.; Josse, J.; Husson, F. FactoMineR: A Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*, 1–18. [**Google Scholar**] [**CrossRef**][**Green Version**]
83. Kuhn, M. Caret: Classification and Regression Training. R Package Version 6.0-85. 2020. Available online: **http://topepo.github.io/caret/index.html** (accessed on 23 January 2021).
84. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. *Remote Sens.* **2015**, *7*, 16398–16421. [**Google Scholar**] [**CrossRef**][**Green Version**]
85. Houborg, R.; McCabe, M.F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 173–188. [**Google Scholar**] [**CrossRef**]
86. Houborg, R.; McCabe, M.F. High-Resolution NDVI from planet's constellation of earth observing nano-satellites: A new data source for precision agriculture. *Remote Sens.* **2016**, *8*, 768. [**Google Scholar**] [**CrossRef**][**Green Version**]
87. Houborg, R.; McCabe, M.F. A Cubesat enabled Spatio-Temporal Enhancement Method (CESTEM) utilizing Planet, Landsat and MODIS data. *Remote Sens. Environ.* **2018**, *209*, 211–226. [**Google Scholar**] [**CrossRef**]
88. Ali, I.; Cawkwell, F.; Green, S.; Dwyer, N. Application of statistical and machine learning models for grassland yield estimation based on a hypertemporal satellite remote sensing time series. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 5060–5063. [**Google Scholar**] [**CrossRef**]

89. Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3254–3264. [Google Scholar] [CrossRef]
90. Taravat, A.; Wagner, M.; Oppelt, N. Automatic Grassland Cutting Status Detection in the Context of Spatiotemporal Sentinel-1 Imagery Analysis and Artificial Neural Networks. *Remote Sens.* **2019**, *11*, 711. [Google Scholar] [CrossRef][Green Version]
91. Shi, Y.; Xiong, F.; Xiu, R.; Liu, Y. A comparative study of relevant vector machine and support vector machine in uncertainty analysis. In Proceedings of the 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2 mSE), Chengdu, China, 15–18 July 2013; pp. 469–472. [Google Scholar] [CrossRef]
92. Hanrahan, L.; Geoghegan, A.; O'Donovan, M.; Griffith, V.; Ruelle, E.; Wallace, M.; Shalloo, L. PastureBase Ireland: A grassland decision support system and national database. *Comput. Electron. Agric.* **2017**, *136*, 193–201. [Google Scholar] [CrossRef]
93. Murphy, D.J.; O' Brien, B.; Askari, M.S.; McCarthy, T.; Magee, A.; Burke, R.; Murphy, M.D. GrassQ—A holistic precision grass measurement and analysis system to optimize pasture based livestock production. In Proceedings of the 2019 ASABE Annual International Meeting, Boston, MA, USA, 7–10 July 2019. [Google Scholar] [CrossRef][Green Version]
94. Ferraro, F.P.; Nave, R.L.; Sulc, R.M.; Barker, D.J. Seasonal variation in the rising plate meter calibration for forage mass. *Agron. J.* **2012**, *104*, 1–6. [Google Scholar] [CrossRef]
95. Houborg, R.; McCabe, M.F. Daily retrieval of NDVI and LAI at 3 m resolution via the fusion of CubeSat, Landsat, and MODIS data. *Remote Sens.* **2018**, *10*, 890. [Google Scholar] [CrossRef][Green Version]

Chapter 3

Creation of a Walloon Pasture Monitoring Platform Based on Machine Learning Models and Remote Sensing

Adapted from: *Nickmilder, C.; Tedde, A.; Dufrasne, I.; Lessire, F.; Glesner, N.; Tychon, B.; Bindelle, J.; Soyeurt, H. Creation of a Walloon Pasture Monitoring Platform Based on Machine Learning Models and Remote Sensing. Remote Sens.* ***2023****, 15, https://doi.org/10.3390/rs15071890*

3.1 Outline

The main topic of this chapter is the creation of a platform predicting the compressed sward height at the scale of Wallonia with a sub-parcel resolution, i.e. based on a pixel division. The creation of the platform and the verification of the behaviour of the locally trained models once applied at a regional scale were detailed in the following article.

3.2 Abstract

The use of remote sensing data and the implementation of machine learning (ML) algorithms is growing in pasture management. In this study, ML models predicting the available compressed sward height (CSH) in Walloon pastures based on Sentinel-1, Sentinel-2, and meteorological data were developed to be integrated into a decision support system (DSS). Given the area covered (>4000 km^2 of pastures of 100 m^2 pixels), the consequent challenge of computation time and power requirements was overcome by the development of a platform predicting CSH throughout Wallonia. Four grazing seasons were covered in the current study (between April and October from 2018 to 2021, the mean predicted CSH per parcel per date ranged from 48.6 to 67.2 mm, and the coefficient of variation from 0 to 312%, suggesting a strong heterogeneity of variability of CSH between parcels. Further exploration included the number of predictions expected per grazing season and the search for temporal and spatial patterns and consistency. The second challenge tackled is the poor data availability for concurrent acquisition, which was overcome through the inclusion of up to 4-day-old data to fill data gaps up to the present time point. For this gap filling methodology, relevancy decreased as the time window width increased, although data with 4-day time lag values represented less than 4% of the total data. Overall, two models stood out, and further studies should either be based on the random forest model if they need prediction quality or on the cubist model if they need continuity. Further studies should focus on developing the DSS and on the conversion of CSH to actual forage allowance.

3.3 Keywords

pasture; decision support system; machine learning; remote sensing; Sentinel satellite; meteorological data

3.4 Introduction

The first working group of experts of the Intergovernmental Panel on Climate Change recently highlighted the importance of a greater understanding of greenhouse gas emissions in pastures [1]. Some experts have claimed that pastures are an unmissable opportunity to mitigate carbon dioxide (CO_2), methane (CH_4), and nitrous oxide (N_2 O) emissions [2]. Others have underlined the importance of grasslands in the assessment of biomass stocks and subsequent carbon storage [3] and asserted that carbon sequestration could be improved through better grazing management [4]. Besides these climatic considerations, [5] underlined a political interest as the European Union has deployed greening policies, including significant usage of pastures ([6–9]). These authors have also pointed out other key motivations, such as farmers' awareness of the preservation of landscapes and consumer perceptions ([5,10]). Moreover, grasslands represent a significant part of the global landmass (assessed as representing between 26% [11] and 40% [12]). In southern Belgium, 42.1% of the total cultivated area is dedicated to grassland, and more than 85% of these pastures are grazed [13]. All these reasons lead us to conclude that there is a strong positive economic advantage in grazing in appropriate climates (e.g., [14,15]), although some studies tend to be less affirmative, especially concerning grazing in the Netherlands [16] or Greece [17].

Unfortunately, as mentioned by [18,19], grass-based livestock ruminant production has not completely leveraged the advances in precision technologies to better understand and manage pasture, probably due to the constraints inherent in outdoor applications. [20] included remote sensing, global positioning systems (GPS), geographic information systems (GIS), and the Internet of Things (IoT), among the underlying technologies. A short, up-to-date review of the literature on the most relevant models to help the understanding and management of pastures highlighted some trends (Annex/SM 5.1). There is an increased use of remote sensing, as established in [21]. A progressive transition from mechanistic models to statistical/machine learning models is also observed with a diversification of their structure. Most papers using remotely sensed datasets stress the advantage of detecting the spatial heterogeneity of pastures, which, as underlined by [22], is a key component of grazing dynamics.

Given the multiple scientific, political, and economic interests of pastures and the technological opportunities available, there is an interest in developing a decision support system (DSS) from a prediction model. However, this has rarely been attempted. Only 11% of models mentioned in Annex/SM 5.1 are implemented into a DSS. This shortfall might either be located at the production level due to the poor performances of the prediction models, as suggested in [23], or at the user level. Sometimes the DSS design and the choice of information delivered by this application did not seem to match the needs of farmers fully [24]. Furthermore, there is a time lag induced by the information overload inherent in the integration of new data sources and data treatments, and that hinders the actual decision making [25]. Therefore, proper attention should be paid to data integration and transfer, as underlined in [24]. For the models that reach the DSS step, the underlying structure and resulting user application should be cheap, rapid, and provide relevant information to increase the interest and adoption rate by farmers.

Recently, to address the under-exploitation of the recent advances in sensor and machine learning algorithms, we have developed machine learning models to predict the compressed sward height (CSH) using cheap data available at a large scale, including Sentinel-1 (S1) and Sentinel-2 (S2) satellite images and meteorological data. Models with a prediction quality around 20 mm [26] present the advantage of producing pixel-based predictions, which enables the consideration of spatial heterogeneity as proposed by [22]. We intend to build a DSS based on these models. The use of CSH in the context of decision making was already included in decision making as an input feature in DSS, such as GrassQ and PastureBase Ireland. To meet the speed criterion and to decrease the computational power requirement, in this study, we focused on implementing and analyzing the usefulness of a platform that handles the prediction of CSH at the scale of Wallonia. Furthermore, the platform was designed to handle data acquired at different times. We assumed that predicting the available CSH at such a large scale with a fine temporal and spatial resolution would require greater time and computational power than would be acceptable for the application's end users. Therefore, this platform is intended to be the data provider for a future DSS that would handle the translation of the transmitted information into relevant metrics. Another major innovation is that, to our knowledge, other DSS primarily rely on mechanistic and empirical models and then try to integrate the remotely sensed data, while our goal is to implement integration at the core of the DSS. Moreover, we did not find any DSS or DSS data provider that could be easily and rapidly

adapted to changes either in the model structure or needs of the user, whereas this platform can offer this modularity as recommended by [27].

A last prospect highlighted by the process of implementing the prediction platform was the ability to study the behavior of the models that are trained and validated with relatively limited datasets once they are applied to huge databases. This led to a refinement of the selection process for the most relevant model to be used as the data provider for future DSS and other applications, depending on the most critical aspect: accuracy or temporal continuity.

3.5 Material and methods

3.5.1 Study area

The models predicting CSH were primarily developed with datasets covering areas located in the southern part of Belgium [26]. So, the same geographical region was used to test the prediction platform. To tackle different meteorological scenarios, the considered grazing periods (from April to October) ranged from 2018 to 2021.

3.5.2 Global design

The global architecture of the prediction platform, developed with both R v4.1 [28] and Python 3.6 [29], is summarized in Figure 1. For the sake of completeness, it should be noted that this platform was designed with a batch processing approach. This choice is in opposition to a streaming approach that is not suited for the combination of data with different acquisition frequencies. The R and python packages, as well as the other programs used, are referenced in Table 3-1. The Python scripts were used to configure and launch the R scripts in independent environments to avoid memory leaks that had happened when developing only in R.

Table 3-1 : List of software and package used

R			Python		Other		
Software/Package	Version	Reference	Software/Package	Reference	Software/Package	Version	Reference
R	4.1	[28]	Python v3.6	[37]	SNAP geoprocessing toolbox	8.0.0	[46]
sf	1.0–2	[29]	subprocess	[38]	7zip		[47]
data.table	1.14.0	[30]	os	[39]			
raster	3.4–13	[31]	time	[40]			
future	1.21.0	[32]	glob	[41]			
future.apply	1.7.0	[33]	datetime	[42]			
caret	6.0–88	[34]	re	[43]			
dplyr	1.0.7	[35]	sentinelsat	[44]			
e1071	1.7–8	[36]	pandas	[45]			

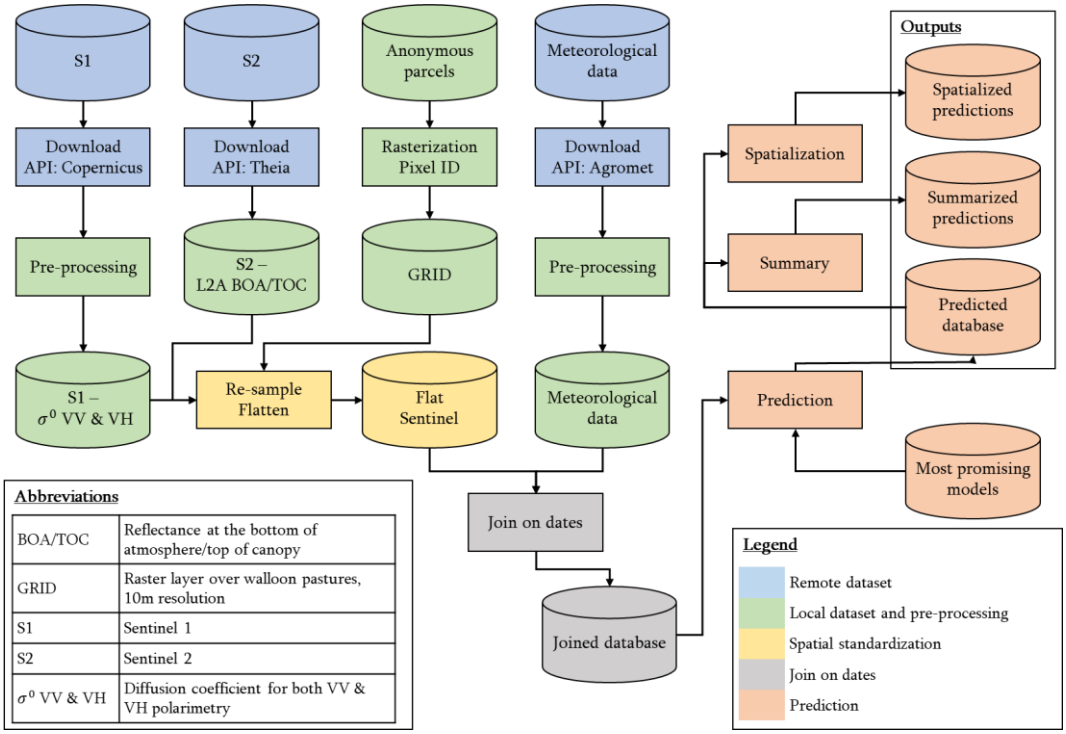


Figure 3-1: Design of the prediction platform

3.5.3 Data acquisition

The acquisition of remote data is illustrated in blue in Figure 3-1.

The S1 constellation is a set of two satellites, S1A and S1B, collecting space-borne synthetic-aperture radar data in the C-Band. Data was accessed from the European Space Agency's (ESA) API [48] for S1 with the help of the Sentinelsat Python package; the S1 data were retrieved in the form of GRD products. Both VV and VH polarization were used. The 37, 88, 110, and 161 relative orbits were used as they offer good coverage of the studied Walloon area. There are discussions in the scientific and remote sensing community about the speckle effect and the need to include the coherence product to get better and more consistent outputs. The drawback of this type of data handling is the need to consider the whole parcel, and the changing nature of these parcels makes this task challenging. In order to include improvements, the platform presented in this study was made to be modular regarding the models and the data treatment workflow. Regarding the consideration of the neighboring pixel values, [49] suggested that stacking convolution could help algorithms detect multi-scale effects. This would translate into a future evolution of the pretreatments, for the model training process and the platform, into a spatial convolution and the addition of features in the models.

The S2 constellation is a set of 2 satellites, S2A and S2B, collecting reflectance/optical data over 12 spectral bands. Theia's API [50] was accessed with the help of the theia_download.py script [51] and provided bottom-of-atmosphere/top-of-canopy reflectance products, also known as level-2A (L2A) products. The tiles covering the extent of Wallonia were: T31UDS, T31UER, T31UES, T31UFQ, T31UFR, T31UFS, T31UGR, and T31UGS. The main reason for the change in data provider was that the MAJA algorithm implemented behind Theia's API allows L2A products based on the exact computation of correction formulae, instead of using the lookup-table methodology underlying the L2A products acquired from the ESA API. Moreover, downloading past data was much easier using Theia's API compared to the difficulty generated by the long-term archiving policy of the ESA.

In our previous article, we used meteorological data from a meteorological station located on an experimental farm [26]. To be able to predict grass growth over the entirety of the Walloon Region, it was more relevant to consider meteorological data covering all of this geographical area. So, the Agromet platform [52] was used and provided data, related to the air, soil, and under leaf temperatures (°C), the wind speed at 2 meters above soil level (m/s), and its direction, solar radiation (J/cm^2), precipitation (mm), relative air humidity (%), and the potential evapotranspiration (mm/day), computed according to the FAO/Penmann-Monteith formula [53]. The data with the corresponding station identifier geographically localized were retrieved under .csv files from the day before the launch of the acquisition to ensure a complete recording of data within a day. The change in the data provider compared to the one used in [26] was motivated by a finer representation of Wallonia than was possible with one meteorological station, the standardization of the acquisition conditions, the near real-time availability of the information, and the convenience of retrieving the data through an API. Compared to [26], another change was made concerning the meteorological data. The choice between the previous computation of the degree-days, also used by [54], and the method currently proposed, also used by [55,56], was based on questioning the meaning of this variable. The 0 °C base temperature referred to the ability to grow in winter, and the 35 °C peak approximated the temperature of the diminishing activity of RuBisCO activase [57,58]. The lower threshold could be queried because some plants do not tolerate 0 °C, but this temperature was kept as it represents the water freezing point and, therefore, the limit of water availability.

3.5.4 Data pre-processing

The data pre-processing is illustrated in green in Figure 3-1. Before being usable by the models, the S1 GRD images need to be geocoded (i.e., properly projected from SAR geometry to "classical"/GiS usable geometry). The framework presented by [59] and based on the use of the Sentinel-1 Toolbox as a part of the SNAP software v8.0 [46] was applied to each tile. The S1 data was then converted using the backscatter coefficient (σ^0) and stored in .img raster files.

To ensure the quality of S2 data, some filters were applied: S2 tiles having more than 95% "NoData" values in the pixels, 95% saturated pixels, or 95% cloud-covered pixels were discarded to avoid biased information entering the data treatment chain. Furthermore, these "NoData", "Saturation", and "Cloud mask" filters were applied to the remaining tiles. Another filtering step was to remove the values inferior or equal to zero and superior to 1, given that the reflectance is supposed to be within the [0;1] range. The S2 data were then transformed into .tif raster files.

The meteorological data initially recorded on a minute basis were aggregated at the day level and recomputed to obtain the minimum, mean, and maximal temperature, the cumulative sum of the solar radiation, the cumulative sum of rain, wind speed, relative air humidity, and evapotranspiration. The degree days were computed on a 0 °C basis with an upper limit of 35 °C. This is translated in the following pseudo-code:

```
If(((Tmax+Tmin)/2)>Tbase):
        If (((Tmax+Tmin)/2)<Tup):
                DJ_00=((Tmax+Tmin)/2)-Tbase
        Else:
                DJ_00=Tup
Else:
        DJ_00=Tbase
```

where Tbase is the base temperature (here 0°C), Tmax corresponds to the maximum temperature of the day, Tmin is the minimum temperature of the day, and Tup is the upper limit temperature (here 35 °C). Moreover, the rolling sum of precipitation and degree-days over the previous 3, 7, and 15 days was also computed for each acquisition date. The meteorological data were incrementally added to a .csv file that acts as a database.

3.5.5 Spatial standardization

The manipulation of spatial data requires that special attention be paid to the format and referencing of data. To ease reference processing, the S1 and S2 datasets

were spatially standardized according to the same reference as illustrated in yellow in Figure 3-1. The spatial reference was obtained as follows. First, Walloon farmers have an obligation to report the parcels they use to the authority and their allocations. These data are then anonymized and available on the WalOnMap platform [60]. For this study, the parcel assignment declared for 2018 was used and was composed of 194,657 pasture parcels stored in a shapefile polygon file. A raster extending over the whole Walloon area with a resolution of 10 m was created using QGis v3.20.2 [61]. Parcels were identified and a unique identification number was attributed to each pixel and both were encoded as integers. Finally, this gridded version of the parcels was projected into EPSG 32631 and saved in a .tif file using integer encoding to avoid uncertainties on the high pixel values.

For this study, the spatial standardization was based on the Walloon parcel assignment realized in 2018. In theory, it would be optimal to use the parcels reported for the year of prediction. However, some constraints prevented this: parcel repartition is uploaded one year after the actual report, and some parcel IDs change. Thus, it might be relevant to discuss the use of the last repartition available. The substantial amount of permanently grazed swards, i.e., 89% of all the Walloon pastures were permanent [62] (for more details, see Annex/SM 3), reinforced the choice of ignoring this source of complexity. This implied that some land patches might either be predicted, although they are not supposed to be, or the opposite. The future DSS should therefore include a step to determine the parcels that farmers want to monitor. The inclusion of the pixel ID should ease this process.

The standardization step consisted in projecting and resampling each S1 and S2 tile independently into a copy of the reference spatial dataset. The resulting raster datasets were forced into a tabular dataset where each row represented a pixel, and the columns corresponded to the pixel ID, the parcel ID, and the S1 or S2 data. Each converted tile was saved in a .csv file

To cover the area of Wallonia, it was necessary to define how the meteorological stations impact each pixel. A simple assumption was chosen: the meteorological data for a pixel corresponds to the data acquired at the nearest station. Therefore, Voronoï polygons were drawn between the meteorological stations. To cover the entirety of the Walloon Region with these polygons, artificial stations were placed far away from the Walloon borders. After making sure that these stations did not appear inside the area of Wallonia, the polygon layer was cropped according to the actual limits of Wallonia with a 2 km buffer to include the pasture parcels that were shared between the 2 countries. The resulting spatial file was a polygon shapefile with the station ID. Then, an intersection with the parcel shapefile was

performed to generate a correspondence table that links the parcel IDs to the meteorological stations. In cases of multiple correspondences, only the first appearance was kept.

3.5.6 Merging datasets

The merging of datasets is illustrated in grey in Figure 3-1 and the precise workflow is shown in Figure 3-2. The first steps concern the joining of S1 and S2 data from the pixel identifier. First, the dates when the S2 tiles were acquired were identified. These dates were compared to those present in the joined database. Each date not yet treated was then processed one at a time. For the first date, all the S1 and S2 tiles acquired on the same day were fetched and joined based on the pixel identifier. The joining did not require an identifier match between the two datasets; therefore, parts of the joined dataset were filled with only S1 or S2 data. Then, the data acquired one day before were also retrieved and used to fill the remaining empty pixels. To mark the time lag between the datasets, a flag containing this time lag (dt) was included in the file. This filling continued until four days before the date of interest had been obtained. This methodology implies that there could be different dts for S1 and S2 records for one pixel. The joined dataset was then saved in the joined database, and the process was repeated for each date not yet gathered.

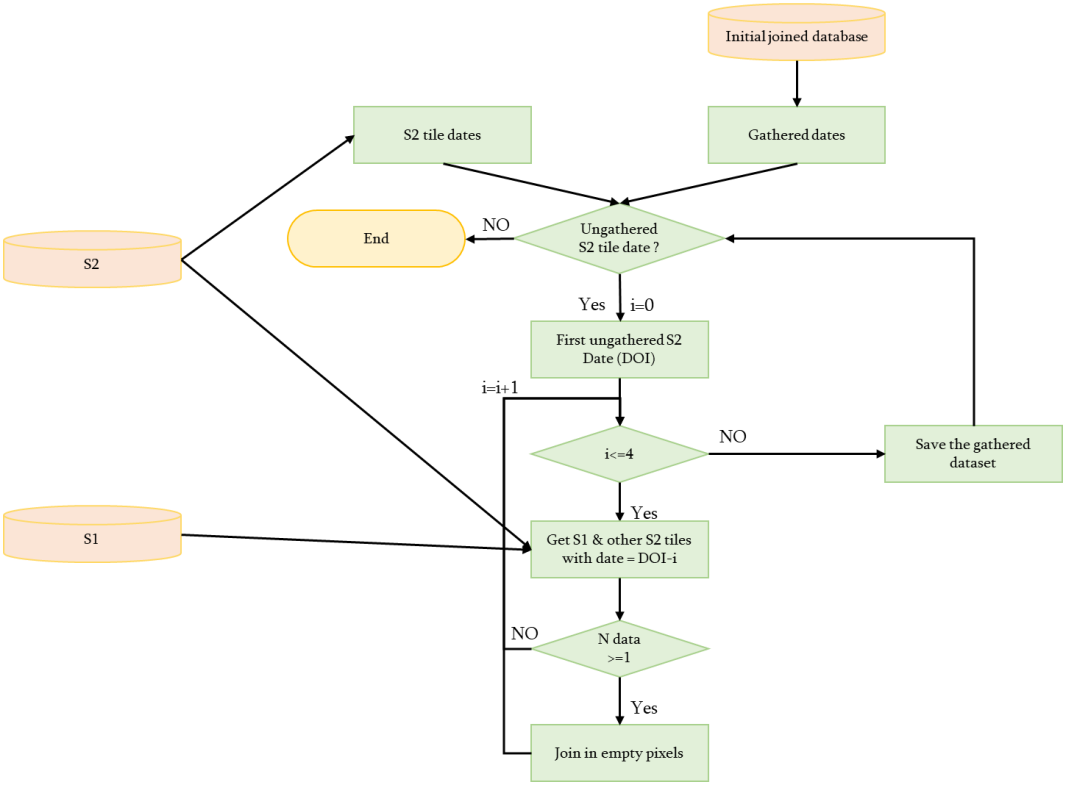


Figure 3-2: Workflow of the joining process

Once S1 and S2 datasets were joined, the addition of the meteorological data was realized from the original S2 date.

3.5.7 Model

Although we have already developed models predicting CSH [26] at the pixel level, new models were retrained due to the availability of more reference CSH values and some changes in the data treatment like the computation of degree days on a 0–35 °C basis, and the removal of the feature transformation related to band 01 of the S2 dataset due to the absence of this band in the dataset coming from Theia. Moreover, a new type of model was tested: an extreme gradient boosting variant (xgbTree). The feature selection process did not change from [26]. The hyperparameters used for every tested model are summarized in Table 3-2; the detail of the hyperparameters explored is presented with other prospects related to the model creation in Annex/SM 2. They were selected from the range of values published in [26] through the application of k-fold cross-validation based on the k dates of acquisition to avoid operator and meteorological data leakage. An alternative could be to perform a 10-fold stratified validation based on the dates to

increase the range of data on which the model was trained, but with the drawback of data leakage and the induced risk of over fitting. The accuracy of the models was assessed through computation of the root mean squared error (RMSE) for the hyperparameter tuning cross-validation, for the actual calibration with the best set of hyperparameters, and for the validation applied on an independent dataset, here consisting of data collected on a different farm. The validation ratio to performance deviation (RPD) was also computed. Moreover, the mean, standard deviation of the prediction, and percentage of underestimated values from the validation set were also used to assess the quality of the models. This percentage corresponds to the ratio of the number of underestimated values to the total amount of values. The calibration dataset was composed of 9376 records collected on 2 farms between 2018 and 2020. The validation dataset contained 871 records collected on another farm in 2019.

Table 3-2: Hyper-parameters tested and final values for the models used.

Model	Parameter	Final Hyper-Parameter Value
xgbTree	Nrounds; max_depth; eta; gamma; colsample_bytree; min_child_weight; subsample	200; 6; 0.1; 1; 0.5; 1; 1
Cubist	committees; neighbors	10; 0
Random Forest (RF)	mtry	88
glmnet	Alpha Lambda	1; 0.1
nnet	Size; decay	3; 0.01

3.5.8 Prediction

Once the whole dataset was gathered in the form of a table whose rows are the records representing one pixel and the columns are the features needed by the built model, the dataset was split into subsections of 100,000 rows to bypass the internal memory limitations of R. The following process was applied in parallel to these chunks: (1) check that all values are actual numbers (i.e., not logic); (2) apply all the required feature transformations; (3) perform the prediction. Due to the different handling of the missing values between the models, special attention had to be paid to the way those values were transferred or not. The resulting predictions were then gathered in a .csv file containing the parcel identification, pixel identifier, S1 and S2 data, as well as the meteorological data and the predictions. Each .csv file corresponds to one date.

3.5.9 Analysis of the predictions

The analysis of the prediction consisted of a visual check and statistical analysis.

The visual check required that the tabular data be transformed into raster data. This was achieved through a five-step process: (1) create a vector of "not a number" (NA) values with the number of cells in the Walloon pastures raster; (2) fill the places where there were predictions using the pixel ID; (3) convert the vector into a matrix respecting the fill order used in the raster; (4) stack this matrix with the original Walloon pastures raster; and (5) write the raster. The visual checks were performed with QGis and aimed to observe the global predicted CSH evolution and the occurrence of abrupt predicted CSH changes in parcels.

The quantitative approach consisted of (1) the study of the performances of the model in calibration, cross-validation, and validation; (2) the number of predictions at the parcel or pixel scale depending on the year and the restriction(s) applied in the joining process; (3) the variability of the CSH predictions performed at a large scale; and (4) a temporal analysis.

The analysis revealed criteria not used in [26] that could be relevant to assess the quality of the models and their appropriateness to the platform's ends. All the criteria were: the RMSE of validation, RPD of validation, the trend to over-/under-estimate, sensitivity to the time-lag inclusion, production of out-of-range values, temporal stability (changes in the values of the mean/standard deviation over the time and appearance of transient spikes), and the spatial heterogeneity of the predictions. Each model was given a ranking for each criterion, and the sum of the ranking gave the total ranking of the model. The exact relationship between the criterion and the ranking is that lower validation RMSE implies a lower ranking; idem for validation RPD; an over- or underestimation ratio closer to 50% implies a lower ranking; the less sensitivity to the time lag inclusion, the lower the ranking; the less production of out of plausible range values, the lower the ranking; the better the temporal stability, the lower the ranking; and the lower the spatial heterogeneity on known homogeneous parcels, the lower the ranking.

3.6 Results and Discussion

First, we will present the constraints related to the prediction process and the interest in the proposed solution to increase the frequency of data acquisition. Then, we will discuss the data choices and changes, the practical implications of the data fusion, the newly trained models' accuracy, and their prediction relevancy.

3.6.1 Practical constraints

3.6.1.1 Data availability

The interest in the developed prediction platform is strongly related to its ability to provide the farmer with reliable information on a routine basis. The prediction models proposed in this paper used three kinds of raw information: S1 and S2 satellite images and meteorological data. The data availability was critical as it directly impacts the potential prediction frequency. As the prediction model used daily aggregated meteorological data, the prediction of CSH can only be made after the test day. As the Agromet API provides meteorological records in real-time, if the day is not yet over, the aggregated value cannot reflect the entirety of the day. For this reason, we had to restrict the data download to the day before the launch. So, the platform should be launched after midnight to provide close to real-time CSH predictions. However, this is the easiest constraint to solve in terms of data availability.

3.6.1.2 Frequency of acquisition

The biggest problem is the acquisition of S1 and S2 data. The first part of this problem is the S2 revisit frequency. Theoretically, at the equator, each satellite has a revisit frequency of 10 days, leading to a total revisit frequency of 5 days when accounting for both satellites. Given the latitude of the Walloon Region, the satellite orbits partly cover each other. This leads to a part of Wallonia (the centre in this case) witnessing acquisitions more often than the rest. This asynchronous acquisition of S2 produced a requirement to account for spatially partial and asynchronous acquisitions. Furthermore, another complex factor was the acquisition frequency of S1 and the coverage of those tiles that were not synchronized with the S2 tiles. The last main constraint on the acquisition frequency was the presence of clouds that decreased the availability of S2 data. This led to an unbalanced dataset representation of S1 and S2 datasets, although the difference in the number of dates of acquisition is not in favour of one or the other dataset (Table 3-4). The huge increase in the number of S1 tiles used in 2021

is, in fact, a lack of tiles for the previous years; the long-term archiving policy of the ESA made the acquisition of old tiles difficult.

3.6.1.3 Hardware and time

Besides data availability, another constraint is the time needed for the prediction computation. Despite using up to 16 CPU cores and 250 Gb of RAM, the platform requires up to 5 hours to deliver predictions at the Walloon scale for one day if the data acquisition is particularly complex. This reinforces the need for launching the platform in the middle of the night. Moreover, consequent storage is needed: if no cleaning of the temporary files is performed, the storage required rapidly adds up. In the case of this study, for monitoring four years with missing dates, we used around 8 TB of disk space without accounting for the temporary files.

3.6.2 Model accuracy

The feature selection process led to a different number of features selected in this study compared to [26], where the nnet model was based on 47 variables, and the other most promising model on 160, while the newly trained model included either the 143 or 158 most promising features (Table 3-3). The order of importance of the features slightly changed and more details are provided in Annex 9.3. The main takeaway is that this variable importance ranking indicates a majority of meteorological and S2 variables in the most informative features, therefore, the choice of the S2 acquisition date as the key point for the gap filling seems relevant as a small shift in these variables might induce a higher change in the final values.

The RMSE performances observed during the hyper-parameter tuning cross-validation, calibration, and validation processes (Table 3-3) were similar to the values reported in [26]. Furthermore, the smaller calibration RMSE values observed for RF and xgbTree were due to their high capability to fit to the dataset, and the difference in the validation RMSE for these models indicates a potential overfit. Although the RMSE of calibration and validation seem high in this configuration, the high value of the RMSE of cross-validation and the inherent variability (standard deviation) show that we should expect a higher error than the error of validation when confronting the models to different conditions, and that a fine tuning of the hyper-parameters was required, especially for the Cubist model.

The validation RPD were similar between the models, although it seems to imply that the glmnet model could better reflect the variability of the validation dataset. The mean and standard deviation of the predicted values calculated on the validation set suggest a global under-estimation of the actual CSH compared to the original mean and standard deviation validation CSH values that were 57.1 ± 5.23 mm. This trend is also highlighted by the percentage of under-estimated

values that ranged from 57.52 to 78.07 %. It seems that the cubist model shows a better respect of the original distribution of the data whilst nnet under-predicts a lot. This trend to under-estimate is to compare to the impact of the rising platemeter used. As shown in [63], the ratchet-counter RPM used to calibrate our models tends to under-estimate the actual height on hard supports. Therefore, it should be expected the actual height is higher than the predicted one.

Table 3-3: Performances of the newly trained models with the new data and feature definition. (RMSEcv = cross-validation root mean squared error (RMSE); RMSEcal = calibration RMSE; RMSEval = validation RMSE; RPD = validation standard deviation (SD)/validation RMSE)

Model	N features	RMSEcv (mean +- SD) [mm CSH]	RMSE cal [mm CSH]	RMSE val [mm CSH]	RPD val	Mean+-SD prediction (validation set) [mm CSH]	Percentage of under-estimated values (validation set)
Cubist	158	23.77±16.20	17.61	17.91	0.85	53.69±9.95	57.52%
Glmnet (Gaussian)	158	22.48±7.00	21.59	15.15	1.01	53.24±6.85	63.15%
Nnet	158	24.08±7.26	23.11	18.67	0.82	46.71±6.36	78.07%
RF	143	22.39±5.55	7.08	17.68	0.86	50.56±8.67	66.70%
xgbTree	143	21.90±6.10	10.68	17.93	0.85	50.16±9.38	66.70%

The RMSE performances observed during the hyperparameter tuning cross-validation, calibration, and validation processes (Table 3-3) were similar to the values reported in [26]. Furthermore, the smaller calibration RMSE values observed for RF and xgbTree were due to their high capability to fit the dataset, and the difference in the validation RMSE for these models indicates a potential overfit. Although the RMSE of calibration and validation seems high in this configuration, the high value of the RMSE of cross-validation and the inherent variability (standard deviation) shows that we should expect a higher error than the error of validation when confronting the models to different conditions and that a fine tuning of the hyperparameters was required, especially for the Cubist model.

The validation RPD was similar between the models, although it seems to imply that the glmnet model could better reflect the variability of the validation dataset. The mean and standard deviation of the predicted values calculated on the validation set suggest a global underestimation of the actual CSH compared to the original mean and standard deviation validation CSH values that were 57.1 ± 5.23 mm. This trend is also highlighted by the percentage of underestimated values that ranged from 57.52 to 78.07 %. It seems that the cubist model shows better respect for the original distribution of the data, whilst nnet underpredicts a lot. This trend to underestimate is compared to the impact of the rising plate meter used. As

shown in [63], the ratchet-counter RPM used to calibrate our models tends to underestimate the actual height on hard supports. Therefore, it should be expected that the actual height will be higher than the predicted one.

3.6.3 Data fusion

3.6.3.1 dt tolerance

To improve the prediction frequency and the area covered, a time tolerance was applied when merging S1 and S2 data, consisting of considering data acquired within the previous four days. The reference data for this merging was the S2 acquisition date, hence the similarity between the number of dates covered and the number of dates of S2 acquisition in Table 3-4. To assess the improvement in terms of data acquisition (i.e., an increase in the quantity of data and dates covered), we compared the number of available dates with a delay of 0 days (dt0) and up to 4 days (all dts) (Table 3-4). This tolerance allowed the number of available dates to increase more than 2.5 times (from 104 without data augmentation to 276 with the gap filling) during the grazing period (214 days) if we consider the whole 4 years studied. The global frequency acquisition reached a mean value of 4 days. This improved temporal coverage can be explained not only by the merging tolerance used but also by the considered constraint for cloud presence. If we strengthened the condition of the cloud presence, described in the metadata description, to 25% to exclude a tile, 138 dates could be available for the whole studied period instead of 276 obtained when permitting up to 95% of cloudy/shadow area before excluding the tiles. This was possible because the prediction was made on a pixel basis and a strict application of the cloud mask to remove any biased (flagged as cloudy/shadowed) pixel. However, assuming a constant acquisition of 4 days is not true, as marked differences were observed between years (Table 3-4). Thanks to the merging tolerance, the number of available dates per year ranged from 49 to 86, and this acquisition was also not constant within a year (Figure 3-3). This is mainly explained by the meteorological conditions. Gaps in the prediction frequency correspond to drops in the mean daily solar radiation received on all the Agromet meteorological stations (Figure 3-4). The difference between the years matches reports from the Royal Meteorological Institute of Belgium [64], where some seasons were more humid than the average of the last 30 years, while others were significantly drier than the 4 years observed. The same is true for global temperatures.

Table 3-4: Descriptive statistics of the available dataset. (*) the number of available dates = number of dates for S2 data acquisition

Category	2018	2019	2020	2021
N dates without time delay	29	30	19	26
N dates*	86	71	70	49
N records/pixel	40.43	26.27	37.17	19.16
N records	1,611,879,463	1,047,054,399	1,481,945,618	764,039,165
Total amount of S1 tiles used	87	95	94	219
N dates acquisition S1	75	79	57	68
Total amount of S2 tiles used	352	276	298	164
N dates acquisition S2	86	71	70	49

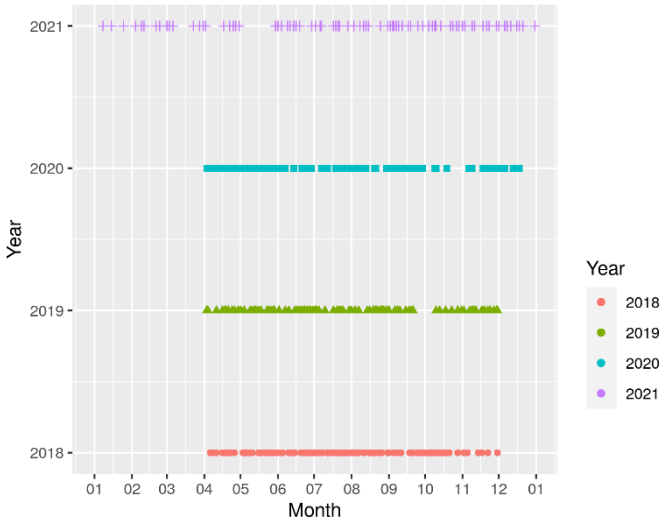


Figure 3-3: Repartition of the predicted dates throughout the years

In the raw prediction files, there was NA in the columns related to the time lag and the predictions. The number of records affected by at least one NA was significantly higher than the number of not affected records (e.g., in 2018, there were 3,118,009,467 records in total, compared to 1,611,879,463 full records). This mismatch was due to a non-exhaustive combination of the edgy pixel position (relative to the satellite orbits), poor weather, out-of-range/missing input values, or even absence of data acquisition, and part of the pixels filled with data from one dataset and not from the other. This led to partly incomplete databases. The amount of NA was smaller in the S1-related data, mainly due to the lower sensitivity to weather perturbations. Given the models currently implemented, the incomplete pixels could not be considered as inputs of the models to produce reliable predictions; hence they were excluded.

3.6.3.1 N Parcels and N Pixels

Even if the number of dates changed between the years studied, the numbers of parcels and pixels represented were equal (N = 194,364 and 39,866,540, respectively). This is expected as the covered area is the same, although the small number of not predicted parcels might indicate minor flaws in the coverage of the area of interest. However, the changes in the number of records per year implied different coverage within the year. The number of records per pixel ranged from 19.16 to 40.43. Although the number of tiles used for the last year was greater than for the other years, the number of records did not increase proportionally. This is due to the filling methodology applied that implied the discarding of past records if there was a temporally close record available.

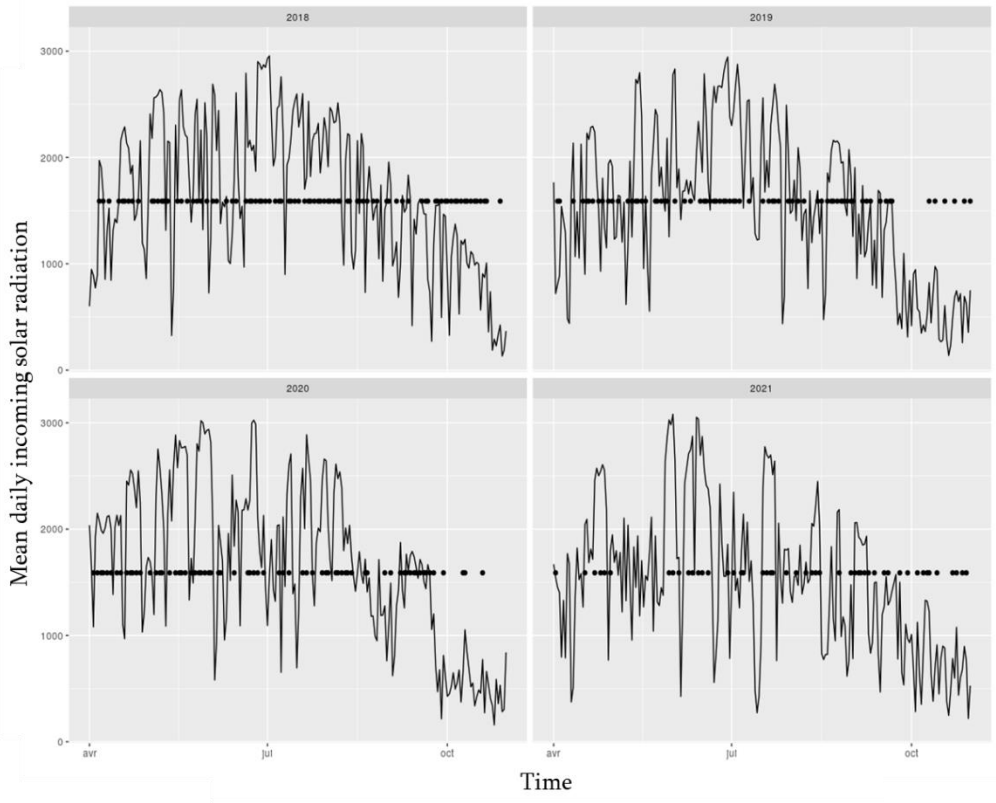


Figure 3-4: Data acquisition dates (points) and their link to the mean daily solar radiation (line and shown y values, in J/cm²) received on all the Agromet stations.

3.6.3.2 Impact of the merging tolerance

The goal of the gap filling is to produce maps that are as complete as possible. We decided to only work with past data to ensure that the platform could be

deployed on a real-time basis. The main idea was that the most reliable information about a parcel should be the last time data were acquired and with a revisit frequency between 3 and 5 days. A time window of 4 days was found to be the best option. The impact of the merging tolerance on the predicted CSH was dependent on the satellite platform and the year considered (as illustrated in Annex/SM 5), especially for higher time lag windows; the increase or decrease (depending on the model) of the range of predictions was more pronounced. This underlines the need for caution about the application of this gap filling methodology.

However, this does not mean that the predictions were completely biased. Using this gap filling method increased the available data between 2 and 3 times for the S1 data and between 2 and 1.5 times for S2. The breakdown of the percentage of data added by the increment of the time lag tolerance is shown in Table 3-5. It appears that most of the filled data were acquired between 0 and 2 days before the date of the merging (which corresponds to the most stable predictions), and the fourth-day consideration brings only a limited amount of information (maximum 4%) which corresponds to the most unreliable data. Therefore, there are two possible interpretations of the changed scope of prediction values observed at the four-day time lag repartition of prediction: either there are so few values that the entire scope of the values possible could not be represented, or the inclusion of the fourth day diminishes the reliability of the predictions—although this second hypothesis should be mitigated by the difference in impact between the models. The increase in the use of older information seen in 2020 and 2021 might be due to the poorer meteorological conditions of these 2 years.

Table 3-5: Percentage of data acquired within a given time lag (dt) from the computed date.

Time delay (days)	S1				S2			
	2018	2019	2020	2021	2018	2019	2020	2021
0	31.61%	38.35%	33.06%	43.72%	47.44%	50.17%	52.46%	59.84%
-1	36.17%	40.05%	26.17%	32.01%	23.31%	19.39%	20.86%	12.92%
-2	20.66%	14.83%	18.74%	20.87%	7.12%	6.48%	9.56%	5.26%
-3	7.71%	3.84%	18.01%	0.06%	21.98%	23.87%	17.05%	21.82%
-4	3.85%	2.93%	4.02%	3.34%	0.15%	0.09%	0.07%	0.16%

The number of predictions that can be made over each grazing season was increased by better temporal and spatial coverage. Indeed, using data in the past for S1 and S2 completion led to the retrieval of data on areas that were not covered at the specific date of acquisition. This type of partial recompletion is partly due to the satellite coverage of the Walloon extent. Indeed, in the case of S2, there are two acquisition orbits. Therefore, if the acquisition through these orbits were close

enough temporally speaking, the past data re-usage implies an increase in the level of spatial completion. However, it does not always happen, and another related phenomenon induces data augmentation; these two acquisition orbits partly overlap, leading to better coverage of these areas and, thus, of some parcels. The effect of this data augmentation can be seen in Figure 3-5, which illustrates the distribution of the number of times a parcel is represented based on the actual number of occurrences of each parcel per year.

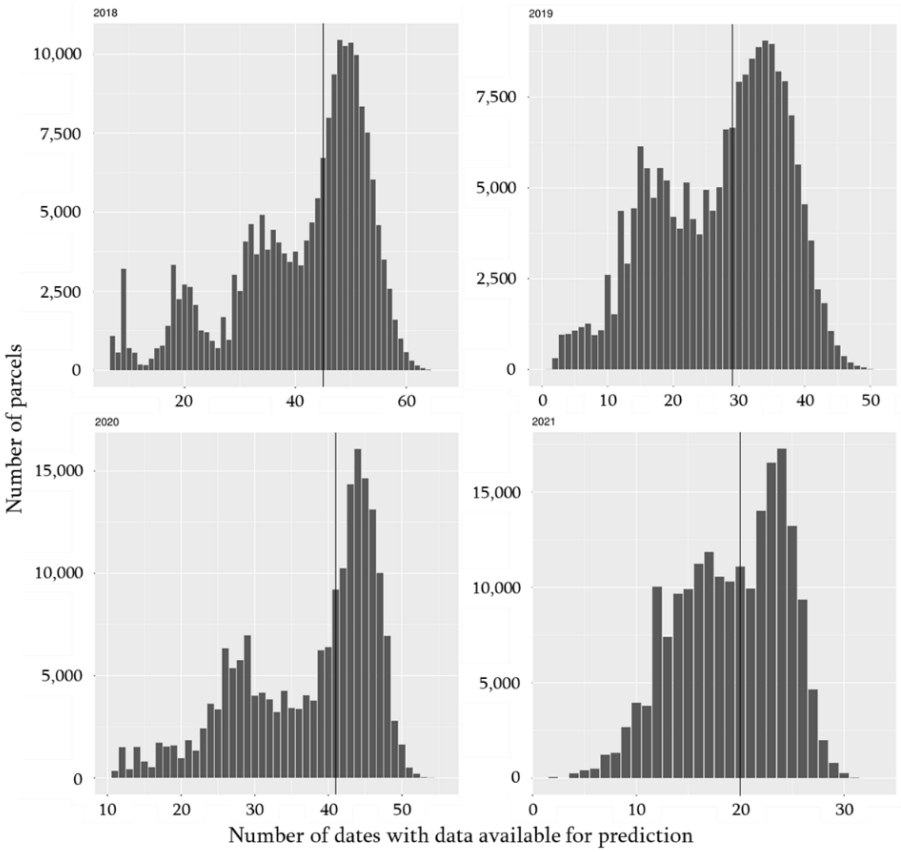


Figure 3-5: Number of parcels per number of dates with data available for prediction. This may be interpreted as the probability of getting x prediction a year. The vertical lines denote the median number of occurrences.

Multiple modes can be guessed in Figure 3-5. As the distribution is quite spread out, summarizing the information is complex. Nevertheless, the median seems to reflect well the majority of the information. The median number of occurrences per parcel was: 45, 29, 41, and 20 for the years 2018, 2019, 2020, and 2021, respectively. This means a median parcel coverage rate, with the proposed gap

filling methodology, between 41% (=20 days of the potentially 49 covered for the year 2021) and 59% (=45 days of the potentially 86 covered for the year 2018). As these percentage of odds to have a parcel covered during irregularly spaced time periods with varying numbers of recording dates across the said periods might seem hard to use for future users, a simpler yet more valuable approach could be the assessment of the probability of getting X prediction a year that is illustrated in Annex/SM 6.

3.6.4 Prediction relevancy

By studying the predictions made in the Walloon Region of Belgium during four annual grazing periods, the aim of this paper was mainly to check if the outputs of the prediction platform were believable and consistent. The amount of different topological and geopedological conditions studied was limited in our training and validation datasets compared to the variety of possibilities present in Wallonia (two agricultural areas were represented out of eleven). To increase the representativeness of the calibration and validation sets, the best method would be to increase the dataset size. This requires a consequent greater sampling effort. Even if this complex approach is needed, it might also be relevant to study the behaviour of the prediction at a large scale to see whether parts of the predictions were inconsistent.

Although CSH predicted values were mainly positive using all tested models, less than 1% of the CSH values predicted from the glmnet model were negative. This model also had a tendency for less than 1% of the obtained predictions to give values much greater than 250 mm, which is the maximum CSH measured by the rising plate meter [26]. For the other models, we observed positive values lower than 250 mm, even if the cubist model can sometimes present a maximum value greater than this threshold. It also appeared, following a model-by-model removal of the extreme values and iterative check of all the model predictions, that the records with an out-of-range prediction from the glmnet model did not correspond to the extreme values predicted with other models. Table 3-6 presents the annual descriptive statistics of the values predicted per pixel using the five studied models after removing the records with extreme values out of the 0–250 mm of CSH range (less than 2%), i.e., 346,465; 256,870; 447,156; and 528,830 records were removed in 2018, 2019, 2020, and 2021, respectively. The five models predicted 75% of the values below 75 mm. A direct interpretation of these values would be that all farmers are using their swards efficiently by maintaining the grass in a constant state of maximum growth. This interpretation should be counter-balanced with the accuracy of the models (most had an RMSE of

independent validation around 20 mm of CSH) and the trend to over-/underestimate the actual CSH, assessed through the percentage of over- or under-predicted values. In [26], the models developed tended to overestimate; whilst, in this new study, they tend to underestimate. More statistics are gathered in Annex/SM 7.

Table 3-6: Descriptive statistics of the cleaned dataset using the five studied models predicting the compressed sward height (mm). CV=coefficient of variation, the ratio of the standard deviation (SD) by the mean multiplied by 100

	2018 (N=1,137,991,583)	2019 (N=1,046,797,529)	2020 (N=1,481,945,618)	2021 (N=763,510,335)	Between year CV (%)
Model	Mean ± SD	Mean ± SD	Mean ± SD	Mean ± SD	
Cubist	56.01 ± 19.94	63.77 ± 20.11	60.07 ± 20.10	59.16 ± 18.48	5.34
Glmnet	48.62 ± 20.35	58.22 ± 21.24	54.97 ± 21.23	54.22 ± 17.95	7.39
Nnet	61.09 ± 21.85	66.48 ± 25.46	67.21 ± 25.44	61.26 ± 19.45	5.14
Rf	54.99 ± 20.63	65.51 ± 20.11	62.14 ± 20.11	61.27 ± 17.33	7.20
xgbTree	53.58 ± 21.16	64.11 ± 20.92	60.51 ± 20.92	60.61 ± 17.85	7.39
CV (%)	8.19	5.04	7.21	5.01	

Concerning the variability of the prediction, the first approach compares the mean and standard deviation values shown in Table 3-6. The small variability reflected by the coefficient of variation, computed as the ratio of the standard deviation by the mean multiplied by 100 (CV), values (below 10%) computed on the means indicate global consistency in the predictions–one year did not seem to be completely offset, nor groups appear to form, and no significant difference can be found for the model relevancy.

The CSH values observed during the model calibration and validation and for the predicted values on a larger scale were lower than the corresponding values observed in Norway [65], Germany [66], or England [67]. This difference is mainly seen because the goals of the pastures were not the same: those parcels were used to grow forage for harvest, while the parcels on which we trained our models and most pasture parcels in Wallonia are grazed. Therefore, the fact of obtaining a mean grass height per parcel ranging from ±30 cm to ±60 cm (Norway) or with more than 50% of values above 10.5 cm (England) would be considered a loss in our study, whereas it would be relevant for good forage yields. This difference is less pronounced when compared to the data from Germany (CSH ranged from 5.82 cm to 19.1 cm with a mean ranging from 7.27 to 14.87 cm).

The shape of the distribution of the prediction matches the "commonly used" descriptive distribution of pasture herbage mass as defined in [68], which is a log-normal/gamma distribution. [67] also reported a similar distribution of the sward height. Furthermore, a similar distribution pattern can be deduced from the comparison between the unmanned aerial vehicle (UAV)-derived sward height

and CSH in [69] or in the comparison between the light detection and ranging (LiDAR) sward height and the UAV-based sward height in [70]. The distribution of the predictions in this study included 75% of the predicted values in the [30–70 mm] range and a tail extended to the higher values.

To go further into the temporal analysis of the predictions, it might be relevant to study the within-year temporal variability of the predictions. In theory, there should be an increase during spring until the moment the cows return to pasture, and then, depending on the cattle load, the CSH should remain stable or even decrease. This kind of pattern is observed in Figure 3-6, where the mean CSH per date of each parcel is represented for each model. The glmnet model seems to react in the same way as modeled in [26]: the values seem to be globally lower for this model than for the others. The annual behavior of grass growth matches the results of [71]: an increase of the CSH (which is directly linked to the actual biomass) during the first part of the year and then a decrease throughout the grazing season. The variability during the year 2021 was more important than within the other years, and the relative decrease of the summer period trend was less pronounced than in the other years. Both those trends could be explained by the high amount of precipitation that occurred during that summer: fewer acquisitions were valid due to cloud cover, which decreased the acquisition frequency and the reliability quality, and the drought of the previous years was considered as a feed loss for breeders using pasture. The relative decrease of the summer compared to spring in 2018 and 2019 was more pronounced than during the 2 following years, and it correlates well with the drought periods.

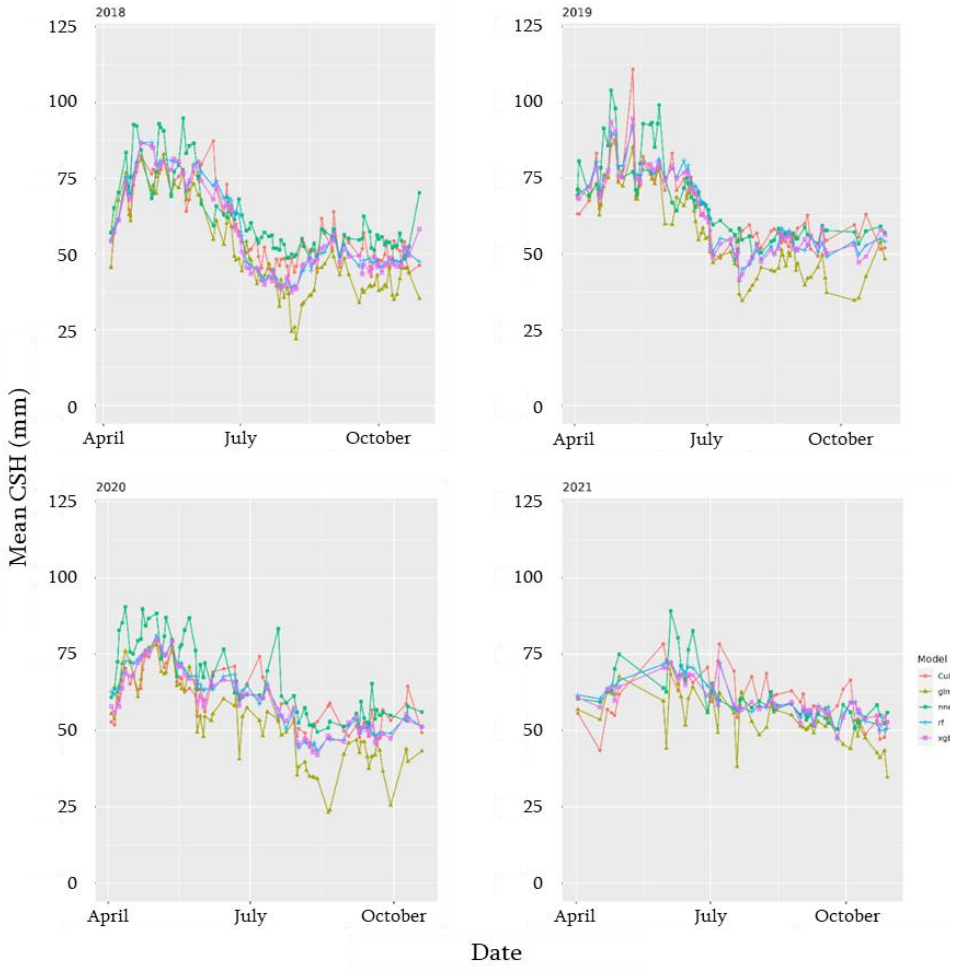


Figure 3-6: Mean per date of acquisition of the mean per parcel of the CSH predicted at the pixel level.

3.6.5 Spatial heterogeneity analysis

3.6.5.1 Per parcel per date variability distribution

Besides the study of temporal behavior, it is necessary to consider spatial heterogeneity as proposed by [22]. Thus, the variability of the predictions inside a parcel was studied. For each parcel and date, the coefficient of variation was computed. These statistics are shown in Table 3-7. Globally, the median values were always lower than 15%, which means there was high stability in the predicted values. The few extreme values are either due to extremely low standard deviation or extremely low means together with high standard deviation

coefficients. The presence of trees, bushes, or hedges inside or at the border of the parcels could induce a higher prediction variability. One way to discard this part of the problem is to erode the parcel file with a negative buffer that should decrease the impact of trees and bushes near the parcel edge. However, there are parcels with edge and solitary trees inside their boundaries. Thus, to further decrease the impact of woody vegetation, an additional step to detect and remove the woody vegetation area from the parcels could be considered in future developments of the platform. The globally lower CV values observed for the prediction compared to the observations of [72] are probably due to climatic differences (both climates are temperate, yet Uruguay has a drier period in winter that Wallonia does not). Moreover, the methodology of [72] to assess the grass height was based on a visual discrete-scale analysis that was then transformed into actual height measurements. This meant that close to extreme heights could have been poorly identified, meaning a higher dispersion of the results and, thus, a higher CV. Lastly, concerning the study by [72], a factor linked to the grass species might also explain the difference, although the composition of the pastures was not disclosed. The observed CV values in the present study fit more of the values than observed in [73]: with uncompressed sward height deduced from image analysis, they got coefficients of variation from 4.5% to 39.0%, which corresponds to 99% of the values observed in the current study.

Table 3-7: Descriptive statistics of the coefficient of variation of the CSH computed for each date for each parcel.

	Cubist	Glmnet	Nnet	RF	xgbTree
Minimum	0.00	0.00	0.00	0.00	0.00
1%	2.34	3.46	0.00	2.27	5.14
1st quartile	6.87	8.02	3.81	6.32	9.82
Median	10.18	10.69	11.24	9.18	12.59
Mean	11.55	11.85	13.43	10.30	13.56
Standard deviation	6.60	6.18	11.41	5.49	5.29
3rd quartile	14.72	14.19	21.06	13.12	16.26
99%	32.97	33.01	40.78	28.13	30.45
Maximum	128.85	330.39	68.65	73.39	79.85

The variability within a parcel can be easily visualized. Figure 3-7 represents predictions at the scale of the parcels of a known farm for the 21st April 2019 for a specific parcel. This date was chosen for the exact co-occurrence of S1 and S2 and for the absence of cloud cover. Checking for the presence of clusters of high/low values revealed a good repartition of the prediction, meaning that spatial over-fitting seemed to have been avoided. The specific zoom on this known farm revealed the ability to see patterns due to the topography and management.

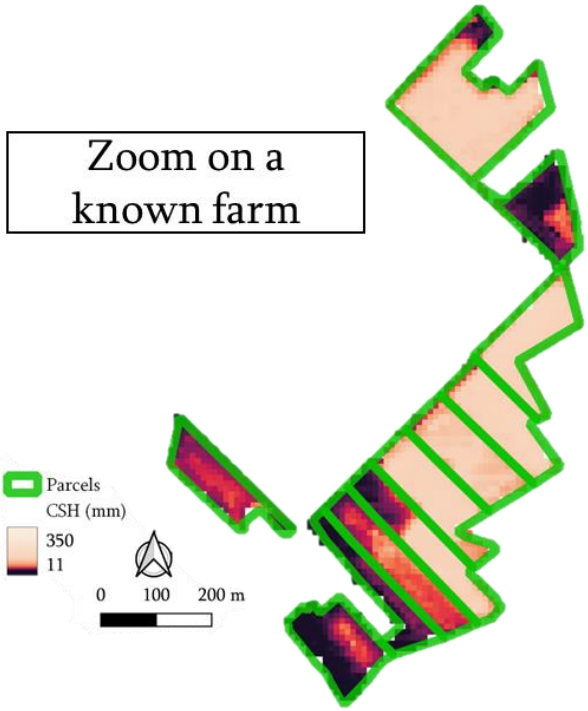


Figure 3-7: Geographical representation of the predictions for parcels known by our team.

Although a 10 m resolution might seem too fine given the area of the Walloon Region (16,901 km²), it is quite coarse when considering precision grazing, as underlined in [72] concerning the ability to reflect the internal variability of pastures. There could be improvement using the already available datasets: S1 tiles have a 5-m resolution. However, the gain in resolution would not be enormous for a huge increase in computation power needed, given that the current raster already had 15,413 rows, 26,006 columns, and 400,830,478 cells. If both datasets were to have a finer resolution, the increased computation cost could be more relevant. There are ways to create super-sampled datasets using algorithms, like the one created by [74], which generate space-borne optical data using the spectral resolution of the S2 satellites and the spatial resolution of the Planet satellites (around 2.5 m). This was not included in the first development of the platform, mainly to reduce complexity. It might be an improvement for future development.

Another spatial data improvement could be the inclusion of meteorological data spatialized at a finer level than the station. Although the Agromet platform [51] provides higher resolution data for air temperature and relative air humidity, a standard spatialization was chosen to avoid adding more complexity, but the

modularity of the platform offers this possibility as an evolution. Furthermore, using kriging techniques might lead to a better representation of the link between the meteorological data and the parcel. However, the inherent complexity and the need to account for local and geopedological/topographical variations made this choice irrelevant in this study. The modular conception of the platform allows for later integration of this type of spatial standardization.

3.6.6 Model selection

The use of this platform as a DSS data provider is currently hindered by the study of five models simultaneously, although the final user needs clear indications and thus should not have to choose between the models without knowing what lies behind them. Therefore, we suggest a complement to the analysis made in [26] to get the "best" model. Until now, the RMSE, RPD, and percentage of over-/underestimation were the major drivers to determine the models as "most promising" [26]. Here, criteria based on the application of the models at a large scale were added: sensitivity to the time lag inclusion, the trend to produce completely out-of-range possible values, temporal stability of the predictions (mean and standard deviation values of the prediction per year and the trend to witness abrupt changes during one year), and finally spatial heterogeneity. Furthermore, the capability to ignore missing data and substitute them could also be added as a criterion, although it might also be a problem given that the substitution is not controlled. The resulting ranking for the currently developed models is shown in Table 3-8 and suggests that the RF model is the best model for raw prediction performance, and Cubist is best for applications that require better temporal continuity.

Table 3-8: Ranking of the most promising models according to multiple criteria

Criterion	Cubist	Glmnet	Nnet	RF	xgbTree
RMSE (val)	3	1	5	2	4
RPD (val)	3	1	5	2	3
Over-/under-estimation (1)	1	2	5	3	3
Sensitivity to the time lag inclusion	1	4	5	1	1
Production of out-of-range values	4	5	1	1	1
Temporal stability (mean & SD)	2	5	1	3	4
Temporal stability (spikes)	3	5	4	1	1
Spatial heterogeneity	2	5	1	2	4
Cumulated ranking	19	28	27	15	21

3.6.7 On the choice of working with Compressed Sward Height

In this study, we worked with compressed sward height. It might be argued that it is more relevant to model actual biomass. The first point to highlight is that the spatial resolution of the pixels (10 m) is larger than traditional sampling quadrats: [71] used bands of 1 × 3.5 m for assessing biomass, [66] used bands of 1.5 × 3 m, and [75] used bands of 7 × 1.5 m. All those dimensions are below the actual resolution, which means that using the S1 and S2 data for the assessment of biomass is likely to introduce a lot of "not-exactly related" information and, thus, imprecision. Although we had datasets of biomass, the bands mown to get the biomass information were also smaller than the pixel resolution and were laid next to each other, which meant that the previous cut would influence the pixel value and therefore give inaccurate results. Besides this issue of coherence in the sampling, another problem was the small size of the dataset and the temporal variability of the relationship between the compressed sward height and the actual biomass, as illustrated in [76]. Furthermore, this variability was emphasized by the variability of the composition of pastures.

However, this platform was made to be modular. This means that if models were created using the same input variables names as we did, they could be implemented and used to perform predictions at a large scale. Therefore, this platform could be used to predict the biomass or the quality of the grass (e.g., protein/fiber content) from remote sensing data. Another approach that is currently being developed adds a conversion layer over the predicted CSH that would translate this value into the other aforementioned features, such as, in our case, the available dry biomass.

3.7 Conclusion

We assumed that predicting the CSH at the whole Walloon scale would require an intermediary platform to perform the prediction in order to decrease the computing load on the backend server of our future DSS. Given the time and resources required for the computation, this proved to be relevant. Despite issues underlined in the study, the platform is now up and running and is ready to serve as the data source for the future DSS. The main advantage of the models and data sources considered is their low cost. They are free to obtain, but a non-null cost is still considered as some processing is needed, and this requires hardware that has a certain cost. Another key advantage of the platform for the future DSS is that most of the computation will be performed beforehand, which means that the end-user application would not be computationally heavy and instead could be as reactive as the end-user hardware allows.

Although the platform is now operational, improvements are still underway. Concerning the prediction models, more CSH data are currently being sampled in order to increase the calibration and validation sets, and therefore, the robustness of the models that currently have a 20 mm of CSH RMSE at a pixel scale. Regarding the translation of CSH into available biomass, models are being tested to be used as a post-prediction layer. Some further developments might also be needed to implement the possibility of the user more precisely specifying their parcels with minimal additional computational cost on the user side, allowed by the use of a standardized sub-parcel/pixel reference spatialization. Furthermore, another important change will be the restriction of access to the data; if not enough restrictions are set in the DSS, there might be problems related to the European General Data Protection Regulation. Concerning the availability of the prediction, the current gap filling methodology restricts the prediction frequency to the acquisition of S2 information as this was the most informative data source (deduced from the relative variable importance in the models). Further research into the gap filling methodology and some paradigm changes would enable us to predict the CSH for every day of the year. Concerning the features considered, until now, we have focused on traditional feature resampling regarding the S1 and S2 datasets. Encompassing more pre-treatments is considered for the future.

Concerning the analysis of the models and the determination of the best model to use, the application of the models at a large scale revealed strengths and weaknesses for all models, resulting in the designation of the random forest model as the best model to predict CSH at our scale with our data. To expand the analysis of the spatial quality of CSH predictions, it might be relevant to account for the

relationship between the spatial and temporal behaviour and the topographical/geopedological properties.

Author Contributions: Conceptualization, C.N and H.S.; software, C.N. and A.T.; validation, A.T., I.D., F.L., N.G., B.T., J.B., and H.S.; formal analysis, C.N., A.T., and H.S.; resources, I.D., F.L., and N.G.; data curation, C.N., A.T., I.D., F.L., N.G., B.T., J.B., and H.S.; writing—original draft preparation, C.N.; writing—review and editing, A.T., I.D., F.L., N.G., B.T., J.B., and H.S.; visualization, C.N. and H.S.; supervision, H.S.; project administration, H.S. All authors have read and agreed to the published version of the manuscript.

3.8 References

1. Masson-Delmotte, V.; Zhai, P.; Pirani, A.; Connors, S.L.; Péan, C.; Berger, S.; Caud, N.; Chen, Y.; Goldfarb, L.; Gomis, M.I.; et al. *IPCC 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK, 2021. Available online: https://www.ipcc.ch/report/ar6/wg1 (accessed on 01/02/2023).
2. Roe, S.; Streck, C.; Obersteiner, M.; Frank, S.; Griscom, B.; Drouet, L.; Fricko, O.; Gusti, M.; Harris, N.; Hasegawa, T.; et al. Contribution of the land sector to a 1.5 °C world. *Nat. Clim. Chang.* **2019**, *9*, 817–828. https://doi.org/10.1038/s41558-019-0591-9.
3. Erb, K.-H.; Kastner, T.; Plutzar, C.; Bais, A.L.S.; Carvalhais, N.; Fetzel, T.; Gingrich, S.; Haberl, H.; Lauk, C.; Niedertscheider, M.; et al. Unexpectedly large impact of forest management and grazing on global vegetation biomass. *Nature* **2018**, *553*, 73–76. https://doi.org/10.1038/nature25138.
4. Henderson, B.B.; Gerber, P.J.; Hilinski, T.E.; Falcucci, A.; Ojima, D.S.; Salvatore, M.; Conant, R.T. Greenhouse gas mitigation potential of the world's grazing lands: Modeling soil carbon and nitrogen fluxes of mitigation practices. *Agric. Ecosyst. Environ.* **2015**, *207*, 91–100. https://doi.org/10.1016/j.agee.2015.03.029.
5. Lessire, F.; Jacquet, S.; Veselko, D.; Piraux, E.; Dufrasne, I. Evolution of Grazing Practices in Belgian Dairy Farms: Results of Two Surveys. *Sustainability* **2019**, *11*, 3997. https://doi.org/10.3390/su11153997.
6. European Commission. On the Implementation of the Ecological Focus Area Obligation under the Green Direct Payment Scheme. Brussels. 2017. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017DC0152&from=FR (accessed on 01/02/2023).
7. European Commission. The European Green Deal EN. Brussels. 2019. Available online: http://eur-lex.europa.eu/resource.html?uri=cellar:208111e4-414e-4da5-94c1-852f1c74f351.0004.02/DOC_1&format=PDF (accessed on 01/02/2023).
8. European Commission. List of Potential Agricultural Practices That Eco-Schemes Could Support. 2021.. Available online: https://ec.europa.eu/info/sites/default/files/food-farming-fisheries/key_policies/documents/factsheet-agri-practices-under-ecoscheme_en.pdf (accessed on 01/02/2023).
9. Tamm, T.; Zalite, K.; Voormansik, K.; Talgre, L. Relating Sentinel-1 Interferometric Coherence to Mowing Events on Grasslands. *Remote Sens.* **2016**, *8*, 802. https://doi.org/10.3390/rs8100802.
10. Michaud, A.; Plantureux, S.; Baumont, R.; Delaby, L. Les prairies, une richesse et un support d'innovation pour des élevages de ruminants plus durables et acceptables. *INRAE Prod. Anim.* **2020**, *33*, 153–172. https://doi.org/10.20870/productions-animales.2020.33.3.4543.
11. O'Mara, F.P. The role of grasslands in food security and climate change. *Ann. Bot.* **2012**, *110*, 1263–1270. https://doi.org/10.1093/aob/mcs209.
12. Taravat, A.; Wagner, M.P.; Oppelt, N. Automatic Grassland Cutting Status Detection in the Context of Spatiotemporal Sentinel-1 Imagery Analysis and Artificial Neural Networks. *Remote Sens.* **2019**, *11*, 711. https://doi.org/10.3390/rs11060711.
13. Biowallonie. Les Chiffres du BIO 2020. 2021. Namur. Available online: https://www.biowallonie.com/wp-content/uploads/2021/09/Biowallonie_ChiffresBio-2020-V2.pdf (accessed on 01/02/2023).
14. Peyraud, J.L.; Van den Pol-van Dasselaar, A.; Dillon, P.; Delaby, L. Producing milk from grazing to reconcile economic and environmental performances. *Grassl. Sci. Eur.* **2010**, *15*, 865–879. https://library.wur.nl/WebQuery/wurpubs/402666.
15. Murphy, D.; Murphy, M.; O'Brien, B.; O'Donovan, M. A Review of Precision Technologies for Optimising Pasture Measurement on Irish Grassland. *Agriculture* **2021**, *11*, 600. https://doi.org/10.3390/agriculture11070600.
16. Reijs, J.W.; Daatselaar, C.H.G.; Helming, J.F.M.; Jager, J.; Beldman, A.C.G. *Grazing Dairy Cows in North-West Europe: Economic Farm Performance and Future Developments with Emphasis on the Dutch Situation*; LEI Wageningen UR: Wageningen, The Netherlands, 2013. Available online: https://library.wur.nl/WebQuery/wurpubs/fulltext/265398 (accessed on 01/02/2023).
17. Papadopoulou, A.; Ragkos, A.; Theodoridis, A.; Skordos, D.; Parissi, Z.; Abraham, E. Evaluation of the Contribution of Pastures on the Economic Sustainability of Small Ruminant Farms in a Typical Greek Area. *Agronomy* **2020**, *11*, 63. https://doi.org/10.3390/agronomy11010063.

18. Shalloo, L.; Donovan, M.O.; Leso, L.; Werner, J.; Ruelle, E.; Geoghegan, A.; Delaby, L.; O'Leary, N. Review: Grass-based dairy systems, data and precision technologies. *Animal* **2018**, *12*, s262–s271. https://doi.org/10.1017/s175173111800246x.
19. Shalloo, L.; Byrne, T.; Leso, L.; Ruelle, E.; Starsmore, K.; Geoghegan, A.; Werner, J.; O'Leary, N. A review of precision technologies in pasture-based dairying systems. *Ir. J. Agric. Food Res.* **2021**, *59*, 279–291. https://doi.org/10.15212/ijafr-2020-0119.
20. Sishodia, R.P.; Ray, R.L.; Singh, S.K. Applications of Remote Sensing in Precision Agriculture: A Review. *Remote Sens.* **2020**, *12*, 3136. https://doi.org/10.3390/rs12193136.
21. Reinermann, S.; Asam, S.; Kuenzer, C. Remote Sensing of Grassland Production and Management—A Review. *Remote Sens.* **2020**, *12*, 1949. https://doi.org/10.3390/rs12121949.
22. Pontes-Prates, A.; Carvalho, P.C.D.F.; Laca, E.A. Mechanisms of Grazing Management in Heterogeneous Swards. *Sustainability* **2020**, *12*, 8676. https://doi.org/10.3390/su12208676.
23. Cockburn, M. Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms. *Animals* **2020**, *10*, 1690. https://doi.org/10.3390/ani10091690.
24. Eastwood, C.R.; Rue, B.T.D.; Gray, D.I. Using a 'network of practice' approach to match grazing decision-support system design with farmer practice. *Anim. Prod. Sci.* **2017**, *57*, 1536–1542. https://doi.org/10.1071/an16465.
25. Armstrong, L. *Improving Data Management and Decision Support Systems in Agriculture*; Burleigh Dodds Science Publishing: Cambridge, United Kingdom 2020. https://doi.org/10.19103/as.2020.0069.15.
26. 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.
27. Power, D.J. What are the characteristics of a Decision Support System. *DSS News*, 30 March 2003. Available online: http://dssresources.com/faq/pdf/13.pdf (accessed on 01/02/2023).
28. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019. Available online: https://www.R-project.org/ (accessed on 01/02/2023)..
29. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *R J.* **2018**, *10*, 439–446. https://doi.org/10.32614/rj-2018-009.
30. Dowle, M.; Srinivasan, A.; Gorecki, J.; Chirico, M.; Stetsenko, P.; Short, T.; Lianoglou, S.; Antonyan, E.; Bonsch, M.; Parsonage, H.; et al. Package 'Data. Table'. Extension of 'Data. Frame'. 2019. Available online: https://CRAN.R-project.org/package=data.table (accessed on 01/02/2023).
31. Hijmans, R.J. Raster: Geographic Data Analysis and Modeling. R Package Version 2.8-19. 2019. Available online: https://CRAN.R-project.org/package=raster (accessed on 01/02/2023).
32. Bengtsson, H. Future: Unified Parallel and Distributed Processing in R for Everyone. R Package. 2020. Available online: https://arxiv.org/abs/2008.00553 (accessed on 01/02/2023).
33. Bengtsson, H. Future Apply: Apply Function to Elements in Parallel Using Futures. R Package. 2021. Available online: https://cran.r-project.org/web/packages/future.apply/index.html (accessed on 01/02/2023).
34. Kuhn, M. Caret: Classification and Regression Training. R Package. 2021. Available online: https://CRAN.R-project.org/package=caret (accessed on 01/02/2023).
35. Wickham, H.; François, R.; Henry, L.; Müller, K. dplyr: A Grammar of Data Manipulation; R Package. 2019. Available online: https://CRAN.R-project.org/package=dplyr (accessed on 01/02/2023).
36. Meyer, D.; Dimitriadou, E.; Hornik, K.; Weingessel, A.; Leisch, F.; Chang, C.C.; Lin, C.C. *libsvm E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071)*; TU Wien: Vienna, Austria, 2020. Available online: https://CRAN.R-project.org/package=e1071 (accessed on 01/02/2023).
37. van Rossum, G. *Python Reference Manual*; Centrum voor Wiskunde en Informatica Amsterdam, 1995.
38. Python Software Foundation. Subprocess. 2021. Available online: https://docs.python.org/fr/3.6/library/subprocess.html (accessed on 01/02/2023).
39. Python Software Foundation. Os. 2021. Available online: https://docs.python.org/3.6/library/os.html (accessed on 01/02/2023).
40. Python Software Foundation. Time. 2021. Available online: https://docs.python.org/3.6/library/time.html (accessed on 01/02/2023).
41. Python Software Foundation. Glob. 2021. Available online: https://docs.python.org/fr/3.6/library/glob.html (accessed on 01/02/2023).

42. Python Software Foundation. Datetime. 2021. Available online: https://docs.python.org/fr/3.6/library/datetime.html (accessed on 01/02/2023).
43. Python Software Foundation. Re. 2021 Available online: https://docs.python.org/3.6/library/re.html (accessed on 01/02/2023).
44. Wille, M. Sentinelsat—Sentinelsat 1. 2016. Available online: https://sentinelsat.readthedocs.io/en/stable/index.html (accessed on 01/02/2023).
45. Mckinney, W. Pandas Documentation. 2011. Available online: https://pandas.pydata.org/docs/index.html (accessed on 01/02/2023).
46. European Space Agency. STEP—Science Toolbox Exploitation Platform. 2015. Available online: https://step.esa.int/main/ (accessed on 01/02/2023)
47. Pavlov, I. 7-Zip. 2012. Available online: https://www.7-zip.org/ (accessed on 01/02/2023).
48. European Space Agency. Copernicus Open Access Hub. 2022. Available online: https://scihub.copernicus.eu/ (accessed on 01/02/2023).
49. Sunil, C.K.; Jaidhar, C.D.; Patil, N. Empirical Study on Multi Convolutional Layer-based Convolutional Neural Network Classifier for Plant Leaf Disease Detection. In Proceedings of the 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), RUPNAGAR, India, 26–28 November 2020; pp. 460–465. https://doi.org/10.1109/iciis51140.2020.9342729.
50. Theia. Muscate—Atelier de Distribution. 2022. Available online: https://theia.cnes.fr/ (accessed on 01/02/2023).
51. Hagolle, O. 2022. Available online: https://github.com/olivierhagolle/theia_download (accessed on 01/02/2023).
52. CRA-W. Agromet. 2022. Available online: https://agromet.be/ (accessed on 01/02/2023).
53. Allen, R.G.; Pereira, L.; Raes, D.; Smith, M. *FAO Irrigation and Drainage Paper No. 56*; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998; Volume 56, pp. 26–40. Available online: http://www.kimberly.uidaho.edu/water/fao56/fao56.pdf (accessed on 01/02/2023).
54. Calvache, I.; Balocchi, O.; Alonso, M.; Keim, J.; López, I. Water-Soluble Carbohydrate Recovery in Pastures of Perennial Ryegrass (*Lolium perenne* L.) and Pasture Brome (*Bromus valdivianus* Phil.) under Two Defoliation Frequencies Determined by Thermal Time. *Agriculture* **2020**, *10*, 563. https://doi.org/10.3390/agriculture10110563.
55. Miller, P.; Lanier, W.; Brandt, S. *Using Growing Degree Days to Predict Plant Stages*; Ag/Extension Communications Coordinator, Communications Services, Montana State University-Bozeman: Bozeman, MO, USA, 2001; Volume 59717, pp. 994–2721. Available online: https://scholar.googleusercontent.com/scholar?q=cache:ZQoTYCYCD0YJ:scholar.google.com/&hl=fr&as_sdt=0,5 (accessed on 01/02/2023).
56. Anandhi, A. Growing degree days—Ecosystem indicator for changing diurnal temperatures and their impact on corn growth stages in Kansas. *Ecol. Indic.* **2016**, *61*, 149–158. https://doi.org/10.1016/j.ecolind.2015.08.023.
57. Salvucci, M.E.; Osteryoung, K.W.; Crafts-brandner, S.J.; Vierling, E. Exceptional Sensitivity of Rubisco Activase to Thermal Denaturation in Vitro and in Vivo. *Plant Physiol.* **2001**, *127*, 1053–1064. https://doi.org/10.1104/pp.010357.1.
58. Parry, M.A.J.; Andralojc, P.J.; Scales, J.C.; Salvucci, M.E.; Carmo-Silva, E.; Alonso, H.; Whitney, S. Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* **2013**, *64*, 717–730. https://doi.org/10.1093/jxb/ers336.
59. Filipponi, F. Sentinel-1 GRD Preprocessing Workflow. *Proceedings* **2019**, *18*, 11. https://doi.org/10.3390/ecrs-3-06201.
60. Région Wallonne. WalOnMap—Géoportail de la Wallonie. 2022. Available online: http://geoportail.wallonie.be/walonmap (accessed on 01/02/2023).
61. QGIS Development Team. QGIS Geographic Information System. QGIS Association. 2022. Available online: https://www.qgis.org (accessed on 01/02/2023).
62. SPW. Utilisation de L'espace Agricole. Namur. 2020. Available online: http://etat.environnement.wallonie.be/contents/indicatorsheets/AGRI 1.html (accessed on 01/02/2023).
63. McSweeney, D.; Coughlan, N.; Cuthbert, R.; Halton, P.; Ivanov, S. Micro-sonic sensor technology enables enhanced grass height measurement by a Rising Plate Meter. *Inf. Process. Agric.* **2018**, *6*, 279–284. https://doi.org/10.1016/j.inpa.2018.08.009.
64. Institut Royal de Météorologie. Météo en Belgique—IRM. 2019. Available online: https://www.meteo.be/fr/belgique (accessed on 01/02/2023).

65. Rueda-Ayala, V.; Peña, J.; Höglind, M.; Bengochea-Guevara, J.; Andújar, D. Comparing UAV-Based Technologies and RGB-D Reconstruction Methods for Plant Height and Biomass Monitoring on Grass Ley. *Sensors* **2019**, *19*, 535. https://doi.org/10.3390/s19030535.
66. Lussem, U.; Bolten, A.; Menne, J.; Gnyp, M.L.; Schellberg, J.; Bareth, G. Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices. *J. Appl. Remote Sens.* **2019**, *13*, 034525. https://doi.org/10.1117/1.jrs.13.034525.
67. Forsmoo, J.; Anderson, K.; MacLeod, C.J.A.; Wilkinson, M.E.; Brazier, R. Drone-based structure-from-motion photogrammetry captures grassland sward height variability. *J. Appl. Ecol.* **2018**, *94*, 237. https://doi.org/10.1111/1365-2664.13148.
68. Nakagami, K. A method for approximate on-farm estimation of herbage mass by using two assessments per pasture. *Grass Forage Sci.* **2015**, *71*, 490–496. https://doi.org/10.1111/gfs.12195.
69. Bareth, G.; Schellberg, J. Replacing Manual Rising Plate Meter Measurements with Low-cost UAV-Derived Sward Height Data in Grasslands for Spatial Monitoring. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2018**, *86*, 157–168. https://doi.org/10.1007/s41064-018-0055-2.
70. Michez, A.; Lejeune, P.; Bauwens, S.; Herinaina, A.A.L.; Blaise, Y.; Muñoz, E.C.; Lebeau, F.; Bindelle, J. Mapping and Monitoring of Biomass and Grazing in Pasture with an Unmanned Aerial System. *Remote Sens.* **2019**, *11*, 473. https://doi.org/10.3390/rs11050473.
71. Ali, I.; Barrett, B.; Cawkwell, F.; Green, S.; Dwyer, E.; Neumann, M. Application of Repeat-Pass TerraSAR-X Staring Spotlight Interferometric Coherence to Monitor Pasture Biophysical Parameters: Limitations and Sensitivity Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3225–3231. https://doi.org/10.1109/jstars.2017.2679761.
72. Tiscornia, G.; Baethgen, W.; Ruggia, A.; Carmo, M.D.; Ceccato, P. Can we Monitor Height of Native Grasslands in Uruguay with Earth Observation? *Remote Sens.* **2019**, *11*, 1801. https://doi.org/10.3390/rs11151801.
73. Cimbelli, A.; Vitale, V. Grassland Height Assessment by Satellite Images. *Adv. Remote Sens.* **2017**, *6*, 40–53. https://doi.org/10.4236/ars.2017.61003.
74. Latte, N.; Lejeune, P. PlanetScope Radiometric Normalization and Sentinel-2 Super-Resolution (2.5 m): A Straightforward Spectral-Spatial Fusion of Multi-Satellite Multi-Sensor Images Using Residual Convolutional Neural Networks. *Remote Sens.* **2020**, *12*, 2366. https://doi.org/10.3390/rs12152366.
75. Michez, A.; Philippe, L.; David, K.; Sébastien, C.; Christian, D.; Bindelle, J. Can Low-Cost Unmanned Aerial Systems Describe the Forage Quality Heterogeneity? Insight from a Timothy Pasture Case Study in Southern Belgium. *Remote Sens.* **2020**, *12*, 1650. https://doi.org/10.3390/rs12101650.
76. McSweeney, D.; Delaby, L.; O'Brien, B.; Ferard, A.; Byrne, N.; McDonagh, J.; Ivanov, S.; Coughlan, N.E. Dynamic algorithmic conversion of compressed sward height to dry matter yield by a rising plate meter. *Comput. Electron. Agric.* **2022**, *196*, 106919. https://doi.org/10.1016/j.compag.2022.106919.

3.9 Supplementary material

3.9.1 Overview of recent scientific papers related to the understanding of pastures

3.9.1.1 Method

A non-exhaustive review of the scientific literature was done and led to the retrieval of papers related to 93 models that aimed at providing a better understanding of the dynamic and parameter of grassland. The search was conducted on both Google Scholar and Scopus and the most used keywords were: “grass,” “grassland,” “sward,” “model,” “machine learning,” “remote sensing,” “management,” “decision support system.” Once the papers were downloaded and sorted per model they were related to, the first step of the analysis consisted in defining the research area(s), the input feature(s), the model(s) used, the objective variable(s)/output(s) and the integration of the model into a decision support system. To ease the analysis, the criteria were split in sub-categories with only a yes/no (1/0) answer possible. The detail of the sub-categories is represented in Table 3-9 and the references are classified per model in Table S2. The determination of the detailed criteria was made in three steps: first all the criteria that were thought to be possible were written, then criteria shown in previous reviews were integrated and finally, most new criteria encountered in the papers reviewed were implemented.

Table 3-9: Classification criteria and details for the models reviewed.

Classification criteria	Details		
Research area	empirical model		
	statistical model/ machine learning		
	cross-validation of methods		
	fusion/comparison of multiple data sources		
	mapping (sfm, upscaling, classification…)		
Feature inputs	Meteorological data (MD)	MD: Temperature	
		MD: Precipitations	
		MD: Solar radiation/ Sunshine hours	
		MD: atmospheric presure	
		MD: evapotranspiration	
		MD: wind (speed and orientation)	
		MD: Relative humidity	
	Time data (TD)	TD: Day of the year	
	Soil data (SD)	SD:Soil drainage	
		SD: geopedological conditions	
		SD: topography	
		SD: soil water content (actual and potential)	
	Pasture management (PM)	PM: NPK-input	
		PM: initial DM (kg/ha)	
		PM: initial LAI	
		PM: Grazing	
		PM: Mowing/ Cutting event	
	Grass height		
	Geographic location (lattitude,…)		
	Land cover/use classification		
	Grass properties		
	Remote Sensing (RS)	RS: Radar Satellite	RS: Rad: Sentinel-1
			RS: Rad: TerraSAR-X
		RS: Reflectance satelitte	RS: Refl: Spot (4)
			RS: Refl: MODIS
			RS: Refl: Landsat
			RS: Refl: Formosat
			RS: Refl: WorldView
			RS: Refl: Sentinel-2
			RS: Refl: Planet (formerly RapidEye)
		RS: UAV based	RS: UAV: Reflectance
			RS: UAV: multispectral/ hyperspectral

			RS: UAV: thermal
		RS: LiDAR	
	Proximal sensing (PS)	PS: laser	
		PS: TOF camera	
		PS: reflectance camera	
		PS: hyperspectral	
Type of model used	Inversion of the PROSAIL model and use of look up tables		
	Mechanistic model		
	Feature selection		
	Monte carlo simulation		
	Discriminant analysis (DA)		
	k nearest neighbour (k-NN)		
	Analysis of variance		
	Simple linear regression		
	Exponential linear regression		
	Multiple linear regression (MLR)		
	Logistic regression		
	Generalized linear model		
	Generalized additive model (GAM)		
	Principal component analysis/regression (PCR)		
	Partial least square regression (PLS-R)		
	Sparse Partial least square (S-PLSR)		
	Suppor vector machine/regression (SVM/SVR)		
	random forest (RF)		
	cubist		
	xgboost		
	Adaptative neuro fuzzy inference system (ANFIS)		
	Perceptron		
	Recurrent neural network (RNN)		
	U neural network (U-NET)		
	Convolutionnal neural network (CNN)		
	Long-short term memory neural network (LSTM)		
Outputs	Biomass/dry matter production [kg/area)		
	Grassland height (CSH/USH/…)		
	comparison/classification of data sources/ land cover		
	Leaf area index (LAI)		
	Quality of the grass (QG)	QG: Crude protein content	
		QG: Neutral detergent fiber content/digestibility	

		QG: Acid detergent fiber content/digestibility	
		QG: Ashes	
		QG: Structural composition of the grass	
		QG: Metabolizalbe energy	
	Soil/environmental conditions		
	Simulation of the effect of management		
	Mowing event detection		
	Spatial representation		
Integration into a DSS	DSS: Yes/NO		
	DSS: Name		

3.9.1.2 Use

Two main types of analysis could be made out of the table: the temporal evolution of the apparition of the criteria and the co-occurrence of criteria.

Concerning the concomitance of criteria, Figure 3-8 and Figure 3-9 illustrate the following trends: - The models focused on statistical/machine learning algorithms are mostly related to the fusion/comparison of multiple data sources, remote sensed data and the prediction of the biomass/dry matter production; - The models focusing on the fusion/comparison of multiple data sources mainly integrate remote sensed data; - The data acquired with UAV are related to the creation of spatial representation of the models through the use of structure from motion algorithms.

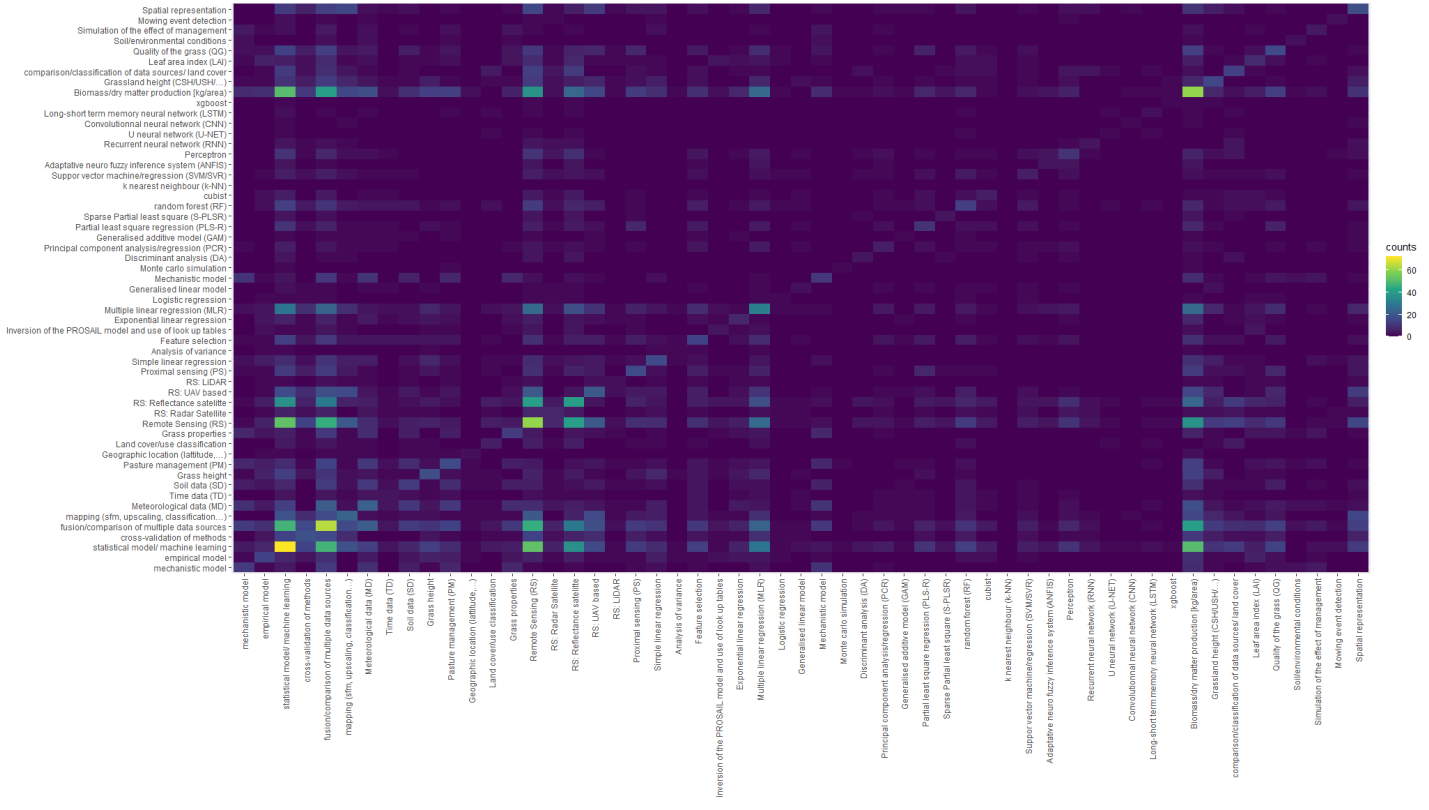


Figure 3-8: Concomitance of research areas and methods, with the raw values.

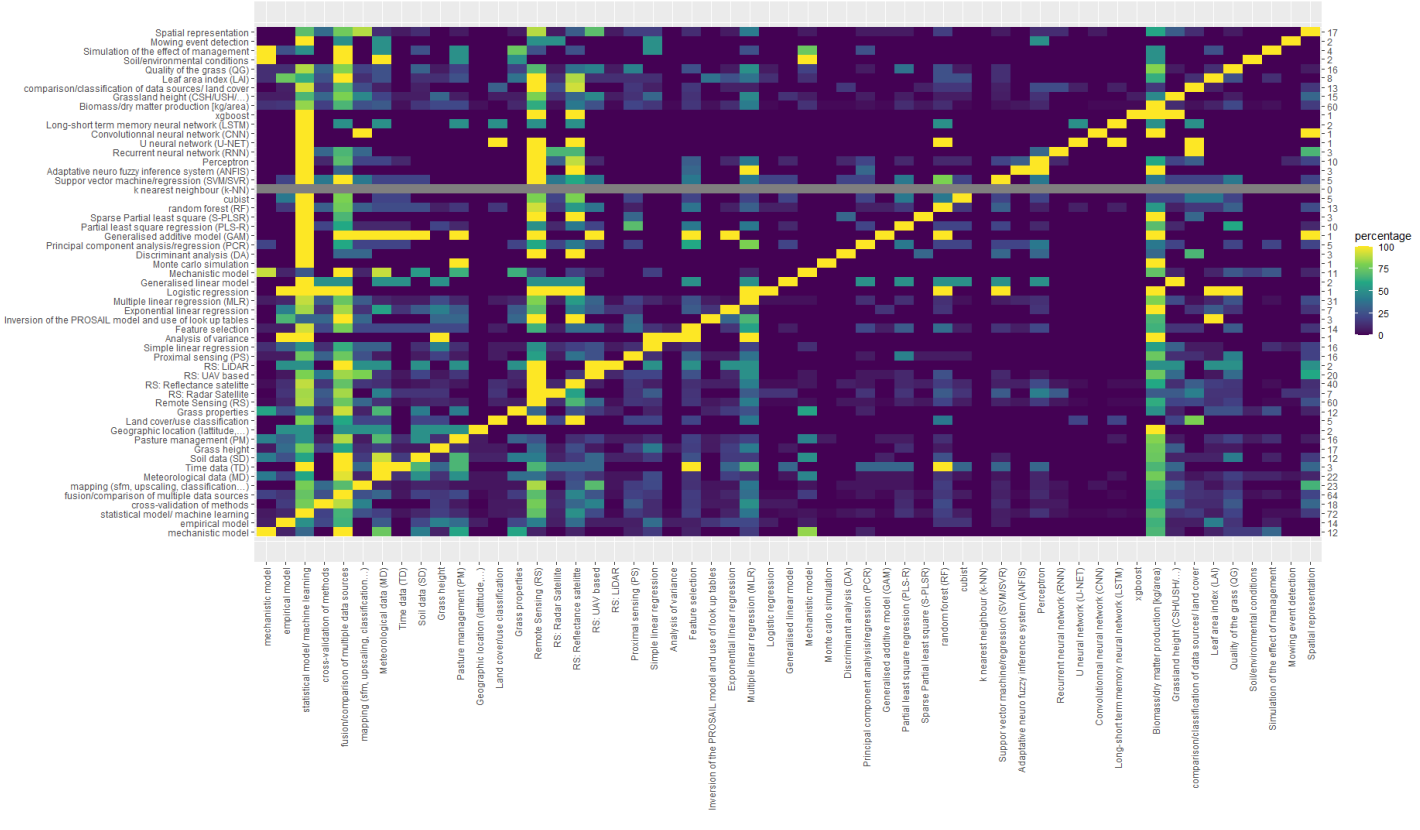


Figure 3-9: Concomitance of research areas and methods, with values relative to the amount of research paper concerning that field.

3.9.1.3 Associated references

The references on which the review is based are presented in Table 3-10. It includes the period of first publication and mention of the model, the geographic area covered by the said model, the model name, given by the authors or as the model was called in other papers. The last column is made of the references.

Table 3-10: References on which the short review is based

Period of first publication	Geographic Area	Name	
1983	England	"J&T"	Johnson, I.R.; Thornley, J.H.M. Vegetative crop growth model incorporating leaf area expansion and senescence, and applied to grass. *Plant Cell Environ.* **1983**, *6*, 721–729. Hurtado-Uria, C.; Hennessy, D.; Shalloo, L.; Schulte, R.P.; Delaby, L.; O'Connor, D. Evaluation of Three Grass Growth Models to Predict Grass Growth in Ireland. *J. Agric. Sci.* **2013**, *151*, 91–104.
1996	Ireland	"Br"	Brereton, A. J., Danielov,S. A.& Scott, D. Agrometeorology of Grass and Grasslands for Middle Latitudes. Technical Note No. 197. Geneva: World Meteorological Organisation **1996** Available online: **https://agris.fao.org/agris-search/search.do?recordID=XF9766146** Hurtado-Uria, C.; Hennessy, D.; Shalloo, L.; Schulte, R.P.; Delaby, L.; O'Connor, D. Evaluation of Three Grass Growth Models to Predict Grass Growth in Ireland. *J. Agric. Sci.* **2013**, *151*, 91–104.
1998	France	STICS	Brisson, N.; Mary, B.; Ripoche, D.; Jeuffroy, M.H.; Ruget, F.; Nicoullaud, B.; Gate, P.; Devienne-Barret, F.; Antonioletti, R.; Durr, C.; et al. STICS: A generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. *Agronomie* **1998**, *18*, 311–346. Brisson, N.; Ruget, F.; Gate, P.; Lorgeou, J.; Nicoullaud, B.; Tayot, X.; Plenet, D.; Jeuffroy, M.H.; Bouthier, A.; Ripoche, D.; et al. STICS: A generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. *Agronomie* **2002**, *22*, 69–92. Di Bella, C., Faivre, R., Ruget, F., Seguin, B., Guérif, M., Combal, B., Weiss, M. and Rebella, C., Use of SPOT4-VEGETATION satellite data to improve pasture production simulated by STICS included in the ISOP French system. *Agronomie* **2004**, *24*(6-7), pp.437-444. Courault, D.; Hadria, R.; Ruget, F.; Olioso, A.; Duchemin, B.; Hagolle, O.; Dedieu, G. Combined use of FORMOSAT-2 images with a crop model for biomass and water monitoring of permanent grassland in Mediterranean region. *Hydrol. Earth Syst. Sci. Discuss.* **2010**, *14*, 1731–1744.
1999	France	SEPATOU	Cros, M.J.; Garcia, F.; Martin-Clouaire, R. SEPATOU: A Decision Support System for the Management of Rotational Grazing in a Dairy Production. In Proceedings of the 2nd European Conference on Information Technology in Agriculture, Bonn, Germany, 27–30 September 1999; pp. 549–557 Amalero, E.G.; Ingua, G.L.; Erta, G.B.; Emanceau, P.L. A biophysical dairy farm model to evaluate rotational grazing management strategies. *Agronomie* **2003**, *23*, 407–418
2003	New-Zealand		McCall, D.G.; Bishop-Hurley, G.J. A pasture growth model for use in a whole-farm dairy production model. *Agric. Syst.* **2003**, *76*, 1183–1205.

2004	Australia	RIM model	Pannell, D.J.; Stewart, V.; Bennett, A.; Monjardino, M.; Schmidt, C.; Powles, S.B. RIM: A Bioeconomic Model for Integrated Weed Management of Lolium Rigidum in Western Australia. *Agric. Syst.* **2004**, *79*, 305–325. Torra, J.; Monjardino, M. Ryegrass Integrated Management (RIM)–Based Decision Support System. In *Decision Support Systems for Weed Management*; Chantre, G., González-Andújar, J.L., Eds.; Springer: Cham, Switzerland, 2020; pp. 249–278.
2005	Ireland	GrazeGro	Barrett, P.; Laidlaw, A.; Mayne, C. GrazeGro: A European Herbage Growth Model to Predict Pasture Production in Perennial Ryegrass Swards for Decision Support. *Eur. J. Agron.* **2005**, *23*, 37–56.
2006	France		Jouven, M.; Carrere, P.; Baumont, R. Model Predicting Dynamics of Biomass, Structure and Digestibility of Herbage in Managed Permanent Pastures. 1. Model Description. *Grass Forage Sci.* **2006**, *61*, 112–124. Hurtado-Uria, C.; Hennessy, D.; Shalloo, L.; Schulte, R.P.; Delaby, L.; O'Connor, D. Evaluation of Three Grass Growth Models to Predict Grass Growth in Ireland. *J. Agric. Sci.* **2013**, *151*, 91–104.
2007	USA		Rayburn, E.B.; Lozier, J.D.; Sanderson, M.A.; Smith, B.D.; Shockey, W.L.; Seymore, D.A.; Fultz, S.W. Alternative Methods of Estimating Forage Height and Sward Capacitance in Pastures Can Be Cross Calibrated. *Forage Grazinglands* **2007**, *5*, 1–6
2009	Australia		Bashari, H.; Smith, C.; Bosch, O.J.H. Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks. *Agric. Syst.* **2008**, *99*, 23–34.
2010	New-Zealand	PGSUS	Romera, A.J.; Beukes, P.; Clark, C.; Clark, D.; Levy, H.; Tait, A. Use of a pasture growth model to estimate herbage mass at a paddock scale and assist management on dairy farms. *Comput. Electron. Agric.* **2010**, *74*, 66–72. Romera, A.; Beukes, P.; Clark, D.; Clark, C.; Tait, A. Pasture growth model to assist management on dairy farms: Testing the concept with farmers. *Grassl. Sci.* **2013**, *59*, 20–29.
2012	Czech Republic		Hakl, J.; Hrevušová, Z.; Hejcman, M.; Fuksa, P. The use of a rising plate meter to evaluate lucerne (Medicago sativa L.) height as an important agronomic trait enabling yield estimation. *Grass Forage Sci.* **2012**, *67*, 589–596.
2012	USA Ohio		Ferraro, F.P.; Nave, R.L.; Sulc, R.M.; Barker, D.J. Seasonal variation in the rising plate meter calibration for forage mass. *Agron. J.***2012**, *104*, 1–6.
2013	Germany		Fricke, T.; Wachendorf, M. Combining ultrasonic sward height and spectral signatures to assess the biomass of legume-grass swards. *Comput. Electron. Agric.* **2013**, *99*, 236–247.
2013	Italy	C-Fix	Maselli, F.; Argenti, G.; Chiesi, M.; Angeli, L.; Papale, D. Simulation of grassland productivity by the combination of ground and satellite data. *Agric. Ecosyst. Environ.* **2013**, *165*, 163–172.
2013	Italy	BIOME-BGC	Maselli, F.; Argenti, G.; Chiesi, M.; Angeli, L.; Papale, D. Simulation of grassland productivity by the combination of ground and satellite data. *Agric. Ecosyst. Environ.* **2013**, *165*, 163–172.
2014	Ireland		Ali, I.; Cawkwell, F.; Green, S.; Dwyer, N. Application of statistical and machine learning models for grassland yield estimation based on a hypertemporal satellite remote sensing time series. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 5060–5063.

2014	Ireland France	MGGM (old version) MostGG	Paillette CA, Delaby L, Shalloo L, O'Connor D, Hennessy D. Developing a predictive model for grass growth in grass-based milk production systems. In18. Symposium of the european grassland federation 2015 Jun 15 (Vol. 20, p. np). Wageningen Academic Publishers. Ruelle, E.; Delaby, L. Pertinence du modèle Moorepark-St Gilles Grass Growth dans les conditions climatiques de l Ouest de la France; Description du modèle Moorepark-St Gilles Grass Growth a) b). **2017**; pp. 158–159. Ruelle, E.; Hennessy, D.; Delaby, L. Development of the Moorepark St Gilles grass growth model (MoSt GG model): A predictive model for grass growth for pasture based systems. *Eur. J. Agron.* **2018**, *99*, 80–91 McDonnell J, Lambkin K, Fealy R, Hennessy D, Shalloo L, Brophy C. Verification and bias correction of ECMWF forecasts for Irish weather stations to evaluate their potential usefulness in grass growth modelling. *Meteorological Applications*. **2018** Apr;25(2):292-301. McDonnell, J.; Brophy, C.; Ruelle, E.; Shalloo, L.; Lambkin, K.; Hennessy, D. Weather forecasts to enhance an Irish grass growth model. *Eur. J. Agron.* **2019**, *105*, 168–175.
2016	China		Li, F.; Zeng, Y.; Luo, J.; Ma, R.; Wu, B. Modeling grassland aboveground biomass using a pure vegetation index. *Ecol. Indic.* **2016**, *62*, 279–288.
2016	Estonia		Tamm, T.; Zalite, K.; Voormansik, K.; Talgre, L. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. *Remote Sens.* **2016**, *8*, 802.
2016	Ireland		Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3254–3264.
2016	Japan		Nakagami, K. A method for approximate on-farm estimation of herbage mass by using two assessments per pasture. *Grass Forage Sci.* **2016**, *71*, 490–496.
2016	Saudi Arabia		Houborg, R.; McCabe, M.F. High-Resolution NDVI from planet's constellation of earth observing nano-satellites: A new data source for precision agriculture. *Remote Sens.* **2016**, *8*, 768.
2016	South Africa		Sibanda, M.; Mutanga, O.; Rouget, M. Comparing the Spectral Settings of the New Generation Broad and Narrow Band Sensors in Estimating Biomass of Native Grasses Grown under Different Management Practices. *GISci. Remote Sens.* **2016**, *53*, 614–633.
2017	Germany		Moeckel, T.; Safari, H.; Reddersen, B.; Fricke, T.; Wachendorf, M. Fusion of ultrasonic and spectral sensor data for improving the estimation of biomass in grasslands with heterogeneous sward structure. *Remote Sens.* **2017**, *9*, 98.
2017	Ireland		Hanrahan, L.; Geoghegan, A.; O'Donovan, M.; Griffith, V.; Ruelle, E.; Wallace, M.; Shalloo, L. PastureBase Ireland: A grassland decision support system and national database. *Comput. Electron. Agric.* **2017**, *136*, 193–201.
2017	Ireland		Ali, I.; Cawkwell, F.; Dwyer, E.; Green, S. Modeling Managed Grassland Biomass Estimation by Using Multitemporal Remote Sensing Data-A Machine Learning Approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3254–3264.
2017	Ireland		Ali, I.; Barrett, B.; Cawkwell, F.; Green, S.; Dwyer, E.; Neumann, M. Application of Repeat-Pass TerraSAR-X Staring Spotlight Interferometric Coherence to Monitor Pasture Biophysical Parameters: Limitations and Sensitivity Analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3225–3231.

2017	Italy		Cimbelli, A.; Vitale, V. Grassland height assessment by satellite images. *Adv. Remote Sens.* **2017**, *6*, 40–53.
2017	Saudi Arabia		Houborg, R.; McCabe, M.F. A hybrid training approach for leaf area index estimation via Cubist and random forests machine-learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 173–188.
2018	Croatia		Gašparović, M.; Medak, D.; Pilaš, I.; Jurjević, L.; Balenović, I. Fusion of Sentinel-2 and PlanetScope Imagery for Vegetation Detection and Monitoring. In Proceedings of the Volumes ISPRS TC I Mid-term Symposium Innovative Sensing-From Sensors to Methods and Applications, Karlsruhe, Germany, 10–12 October 2018.
2018	Czech Republic		Cudlín, O.; Hakl, J.; Hejcman, M.; Cudlín, P. The use of compressed height to estimate the yield of a differently fertilized meadow. *Plant Soil Environ.* **2018**. *64*, 76–81.
2018	Germany		röhnert, M.; Anderson, R.; Bumberger, J.; Dietrich, P.; Harpole, W.S.; Maas, H.-G. Watching grass grow - a pilot study on the suitability of photogrammetric techniques for quantifying change in aboveground biomass in grassland experiments. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-2*, 539–542.
2018	Germany		Lussem, U.; Bolten, A.; Gnyp, M.; Jasper, J.; Bareth, G. Evaluation of RGB-based vegetation indices from UAV imagery to estimate forage yield in Grassland. *ISPRS-Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, 1215–1219.
2018	Germany		Bareth, G.; Schellberg, J. Replacing Manual Rising Plate Meter Measurements with Low-cost UAV-Derived Sward Height Data in Grasslands for Spatial Monitoring. *PFG J. Photogramm. Remote. Sens. Geoinf. Sci.* **2018**, *86*, 157–168.
2018	Italy		D'Urso, M.G.; Rotondi, A.; Gagliardini, M. UAV low-cost system for evaluating and monitoring the growth parameters of crops. In *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 2018, IV-5, Proceedings of ISPRS TC V Mid-term Symposium "Geospatial Technology—Pixel to People", Dehradun, India, 20–23 November 2018*; Kumar, A.S., Saran, S., Padalia, H., Eds.; ISPRS: Hannover, Germany, 2018; pp. 405–413.
2018	New-Zealand		Pullanagari, R.R.; Kereszturi, G.; Yule, I. Integrating Airborne Hyperspectral, Topographic, and Soil Data for Estimating Pasture Quality Using Recursive Feature Elimination with Random Forest Regression. *Remote Sens.* **2018**, *10*, 1117.
2018	Saudi Arabia		Houborg, R.; McCabe, M.F. Daily Retrieval of NDVI and LAI at 3 m Resolution via the Fusion of CubeSat, Landsat, and MODIS Data. *Remote Sens.* **2018**, *10*, 890.
2018	South Africa		Mutanga, O.; Shoko, C. Monitoring the spatio-temporal variations of C3/C4 grass species using multispectral satellite data. *IGARSS* **2018**, *2018*, 8988-8991.
2018	South Africa		Shoko, C.; Mutanga, O.; Dube, T. Determining optimal new generation satellite derived metrics for accurate C3 and C4 grass species aboveground biomass estimation in South Africa. *Remote Sens.* **2018**, *10*, 564.
2018	Switzerland	ModVege	KLOECKER D, CONTER G, DIRKSE A, FELTEN C. Méi Weed, Weideoptimierung durch die Anpassung der Weideführung an Witterung und Bodenverhältnisse. Internationale Weidetagung.:78.
2018	UK		Forsmoo, J.; Anderson, K.; Macleod, C.J.A.; Wilkinson, M.E.; Brazier, R. Drone-Based Structure-from-Motion Photogrammetry Captures Grassland Sward Height Variability. *J. Appl. Ecol.* **2018**, *55*, 2587–2599.

2018	UK		Punalekar, S.M.; Verhoef, A.; Quaife, T.L.; Humphries, D.; Bermingham, L.; Reynolds, C.K. Application of Sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. *Remote Sens. Environ.* **2018**, *218*, 207–220.
2018	USA Montana		Zeng, L.; Chen, C. Using remote sensing to estimate forage biomass and nutrient contents at different growth stages. *Biomass Bioenergy* **2018**, *115*, 74–81.
2019	Belgium		Michez, A.; Lejeune, P.; Bauwens, S.; Lalaina Herinaina, A.A.; Blaise, Y.; Muñoz, E.C.; Lebeau, F.; Bindelle, J. Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system. *Remote Sens.* **2019**, *11*, 473.
2019	Belgium		Borra-Serrano, I.; De Swaef, T.; Muylle, H.; Nuyttens, D.; Vangeyte, J.; Mertens, K.; Saeys, W.; Somers, B.; Roldán-Ruiz, I.; Lootens, P. Canopy height measurements and non-destructive biomass estimation of Lolium perenne swards using UAV imagery. *Grass Forage Sci.* **2019**, *74*, 356–369.
2019	Brasil		Schulte LG, Perez NB, de Pinho LB, Trentin G. Decision support system for precision livestock: Machine learning-based prediction module for stocking rate adjustment. InProceedings of the XV Brazilian Symposium on Information Systems 2019 May 20 (pp. 1-8).
2019	Brasil		Parente, L.; Taquary, E.; Silva, A.P.; Souza, C.; Ferreira, L. Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data. *Remote Sens.* **2019**, *11*, 2881.
2019	British Columbia Canada		Leach, N.; Coops, N.C.; Obrknezev, N. Normalization method for multi-sensor high spatial and temporal resolution satellite imagery with radiometric inconsistencies. *Comput. Electron. Agric.* **2019**, *164*, 104893.
2019	China		He, L.; Li, A.N.; Yin, G.F.; Nan, X.; Bian, J.H. Retrieval of Grassland Aboveground Biomass through Inversion of the PROSAIL Model with MODIS Imagery. *Remote Sens.* **2019**, *11*, 1597.
2019	France		Garioud, A.; Giordano, S.; Valero, S.; Mallet, C. Challenges in Grassland Mowing Event Detection with Multimodal Sentinel Images. *MultiTemp* **2019**, 1–4.
2019	France		Hubert-Moy, L.; Thibault, J.; Fabre, E.; Rozo, C.; Arvor, D.; Corpetti, T.; Rapinel, S. Mapping Grassland Frequency Using Decadal MODIS 250 m Time-Series: Towards a National Inventory of Semi-Natural Grasslands. *Remote Sens.* **2019**, *11*, 3041.
2019	Germany		Taravat, A.; Wagner, M.; Oppelt, N. Automatic Grassland Cutting Status Detection in the Context of Spatiotemporal Sentinel-1 Imagery Analysis and Artificial Neural Networks. *Remote Sens.* **2019**, *11*, 711.
2019	Germany		Lussem, U.; Bolten, A.; Menne, J.; Gnyp, M.L.; Schellberg, J.; Bareth, G. Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices. *J. Appl. Remote Sens.* **2019**, *13*, 034525.
2019	Germany		Lussem, U.; Bolten, A.; Menne, J.; Gnyp, M.; Bareth, G. Ultra-high spatial resolution UAV-based imagery to predict biomass in temperate grasslands. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *4213*, 443–447.
2019	Ireland	"GrassQ"	Murphy, D.J.; O' Brien, B.; Askari, M.S.; McCarthy, T.; Magee, A.; Burke, R.; Murphy, M.D. GrassQ—A holistic precision grass measurement and analysis system to optimize pasture based livestock production. In Proceedings of the 2019 ASABE Annual International Meeting, Boston, MA, USA, 7–10 July 2019.

2019	Ireland		McSweeney, D.; Coughlan, N.E.; Cuthbert, R.N.; Halton, P.; Ivanov, S. Micro-sonic sensor technology enables enhanced grass height measurement by a Rising Plate Meter. *Inf. Process. Agric.* **2019**, *6*, 279–284.
2019	Ireland		Askari, M.S.; McCarthy, T.; Magee, A.; Murphy, D.J. Evaluation of Grass Quality under Different Soil Management Scenarios Using Remote Sensing Techniques. *Remote Sens.* **2019**, *11*, 1835.
2019	Israel		Lugassi, R.; Zaady, E.; Goldshleger, N.; Shoshany, M.; Chudnovsky, A. Spatial and temporal monitoring of pasture ecological quality: Sentinel-2-based estimation of crude protein and neutral detergent fiber contents. *Remote Sens.* **2019**, *11*, 799.
2019	Italy		Stendardi, L.; Karlsen, S.R.; Niedrist, G.; Gerdol, R.; Zebisch, M.; Rossi, M.; Notarnicola, C. Exploiting time series of Sentinel-1 and Sentinel-2 imagery to detect meadow phenology in mountain regions. *Remote Sens.* **2019**, *11*, 542.
2019	New-Zealand		Legg, M.; Bradley, S. remote sensing Ultrasonic Proximal Sensing of Pasture Biomass. *Remote Sens.* **2019**, *11*, 2459 Legg, M.; Bradley, S. Ultrasonic Arrays for Remote Sensing of Pasture Biomass. *Remote Sens.* **2019**, *12*, 111.
2019	Norway		Ancin-Murguzur, F.J.; Taff, G.; Davids, C.; Tømmervik, H.; Mølmann, J.; Jørgensen, M. Yield estimates by a two-step approach using hyperspectral methods in grasslands at high latitudes. *Remote Sens.* **2019**, *11*, 400.
2019	Norway		Rueda-Ayala, V.P.; Peña, J.M.; Höglind, M.; Bengochea-Guevara, J.M.; Andújar, D. Comparing UAV-based technologies and RGB-D reconstruction methods for plant height and biomass monitoring on grass ley. *Sensors* **2019**, *19*, 535.
2019	The Netherlands		Darvishzadeh, R.; Wang, T.; Skidmore, A.; Vrieling, A.; O'Connor, B.; Gara, T.W.; Ens, B.J.; Paganini, M. Analysis of Sentinel-2 and rapidEye for retrieval of leaf area index in a saltmarsh using a radiative transfer model. *Remote Sens.* **2019**, *11*, 671.
2019	Uruguay		Tiscornia, G.; Baethgen, W.; Ruggia, A.; Do Carmo, M.; Ceccato, P. Can we Monitor Height of Native Grasslands in Uruguay with Earth Observation? *Remote Sens.* **2019**, *11*, 1801.
2019	USA Arizona		Gillan, J.K.; McClaran, M.P.; Swetnam, T.L.; Heilman, P. Estimating Forage Utilization with Drone-Based Photogrammetric Point Clouds. *Rangel. Ecol. Manag.* **2019**, *72*, 575–585.
2019	USA California		Liu, H.; Dahlgren, R.A.; Larsen, R.E.; Devine, S.M.; Roche, L.M.; O' Geen, A.T.; Wong, A.J.Y.; Covello, S.; Jin, Y. Estimating Rangeland Forage Production Using Remote Sensing Data from a Small Unmanned Aerial System (sUAS) and PlanetScope Satellite. *Remote Sens.* **2019**, *11*, 595.
2019	USA Michigan	SALUS MDP	Insua, J.R.; Utsumi, S.A.; Basso, B. Estimation of Spatial and Temporal Variability of Pasture Growth and Digestibility in Grazing Rotations Coupling Unmanned Aerial Vehicle (UAV) with Crop Simulation Models. *PLoS ONE* **2019**, *14*, e0212773.
2019	USA Oklahoma		Wang, J.; Xiao, X.; Bajgain, R.; Starks, P.; Steiner, J.; Doughty, R.B.; Chang, Q. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. *ISPRS J. Photogramm. Remote Sens.* **2019**, *154*, 189–201.
2020	Australia		Ara, I.; Harrison, M.T.; Whitehead, J.; Waldner, F.; Bridle, K.; Gilfedder, L.; Da Silva, J.M.; Marques, F.; Rawnsley, R. Modelling seasonal pasture growth and botanical composition at the paddock scale with satellite imagery. *Silico Plants* **2021**, *3*, 1–15.

2020	Australia		Gargiulo, J.; Clark, C.; Lyons, N.; de Veyrac, G.; Beale, P.; Garcia, S. Spatial and temporal pasture biomass estimation integrating electronic plate meter, planet cubesats and sentinel-2 satellite data. *Remote Sens.* **2020**, *12*, 3222.
2020	Australia		Smith, C.; Karunaratne, S.; Badenhorst, P.; Cogan, N.; Spangenberg, G.; Smith, K. Machine Learning Algorithms to Predict Forage Nutritive Value of In Situ Perennial Ryegrass Plants Using Hyperspectral Canopy Reflectance Data. *Remote Sens.* **2020**, *12*, 928.
2020	Australia The Netherlands		Togeiro de Alckmin, G.; Lucieer, A.; Roerink, G.; Rawnsley, R.; Hoving, I.; Kooistra, L. Retrieval of Crude Protein in Perennial Ryegrass Using Spectral Data at the Canopy Level. *Remote Sens.* **2020**, *12*, 2958.
2020	Belgium		Michez, A.; Philippe, L.; David, K.; Sébastien, C.; Christian, D.; Bindelle, J. Can low-cost unmanned aerial systems describe the forage quality heterogeneity? Insight from a timothy pasture case study in southern Belgium. *Remote Sens.* **2020**, *12*, 1650.
2020	Brasil		Dos Reis, A.A.; Werner, J.P.S.; Silva, B.C.; Figueiredo, G.K.D.A.; Antunes, J.F.G.; Esquerdo, J.C.D.M.; Coutinho, A.C.; Lamparelli, R.A.C.; Rocha, J.V.; Magalhães, P.S.G. Monitoring Pasture Aboveground Biomass and Canopy Height in an Integrated Crop–Livestock System Using Textural Information from PlanetScope Imagery. *Remote Sens.* **2020**, *12*, 2534.
2020	Brasil		Dos Reis AA, Silva BC, Werner JP, Silva YF, Rocha JV, Figueiredo GK, Antunes JF, Esquerdo JC, Coutinho AC, Lamparelli RA, Magalhães PS. Exploring the Potential of High-Resolution Planetscope Imagery for Pasture Biomass Estimation in an Integrated Crop–Livestock System. In2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS) 2020 Mar 22 (pp. 675-680). IEEE.
2020	Finland		Alves, R.; Näsi, R.; Niemeläinen, O.; Nyholm, L.; Alhonoja, K.; Kaivosoja, J.; Jauhiainen, L.; Viljanen, N.; Nezami, S.; Markelin, L. Remote Sensing of Environment Machine learning estimators for the quantity and quality of grass swards used for silage production using drone-based imaging spectrometry and photogrammetry. *Remote Sens. Environ.* **2020**, *246*, 111830.
2020	Ireland	GMOT	Murphy, D.J.; O' Brien, B.; Murphy, M.D. Development of a grass measurement optimisation tool to efficiently measure herbage mass on grazed pastures. *Comput. Electron. Agric.* **2020**, *178*, 105799.
2020	Spain		Campos-Taberner, M.; García-Haro, F.J.; Martínez, B.; Izquierdo-Verdiguier, E.; Atzberger, C.; Camps-Valls, G.; Gilabert, M.A. Understanding deep learning in land use classification based on Sentinel-2 time series. *Sci. Rep.* **2020**, *10*, 1–12.
2020	Zambia		Clementini, C.; Pomente, A.; Latini, D.; Kanamaru, H.; Vuolo, M.R.; Heureux, A.; Fujisawa, M.; Schiavon, G.; Del Frate, F. Long-Term Grass Biomass Estimation of Pastures from Satellite Data. *Remote Sens.* **2020**, *12*, 2160.
2021	Australia		De Rosa, D.; Basso, B.; Fasiolo, M.; Friedl, J.; Fulkerson, B.; Grace, P.R.; Rowlings, D.W. Predicting Pasture Biomass Using a Statistical Model and Machine Learning Algorithm Implemented with Remotely Sensed Imagery. *Comput. Electron. Agric.* **2021**, *180*, 105880.
2021	Australia		Chen, Y.; Guerschman, J.; Shendryk, Y.; Henry, D.; Harrison, M.T. Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learning. *Remote Sens.* **2021**, *13*, 603.

2021	Australia		Barnetson, J.; Phinn, S.; Scarth, P. Climate-Resilient Grazing in the Pastures of Queensland: An Integrated Remotely Piloted Aircraft System and Satellite-Based Deep-Learning Method for Estimating Pasture Yield. *AgriEngineering* **2021**, *3*, 681-702.
2021	Belgium		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.
2021	Belgium		Pranga, J.; Borra-Serrano, I.; Aper, J.; De Swaef, T.; Ghesquiere, A.; Quataert, P.; Roldán-Ruiz, I.; Janssens, I.A.; Ruysschaert, G.; Lootens, P. Improving Accuracy of Herbage Yield Predictions in Perennial Ryegrass with UAV-Based Structural and Spectral Data Fusion and Machine Learning. *Remote Sens.* **2021**, *13*, 3459.
2021	France	SenRVM	Garioud, A.; Valero, S.; Giordano, S.; Mallet, C. Recurrent-Based Regression of Sentinel Time Series for Continuous Vegetation Monitoring. *Remote Sens. Environ.* **2021**, *263*, 112419.
2021	Germany		Grüner, E.; Astor, T.; Wachendorf, M. Prediction of Biomass and N Fixation of Legume–Grass Mixtures Using Sensor Fusion. *Front. Plant Sci.* **2021**, *11*, 1–13.
2021	Iran		Ghayour, L.; Neshat, A.; Paryani, S.; Shahabi, H.; Shirzadi, A.; Chen, W.; Al-Ansari, N.; Geertsema, M.; Pourmehdi Amiri, M.; Gholamnia, M.; Dou, J.; Ahmad, A. Performance Evaluation of Sentinel-2 and Landsat 8 OLI Data for Land Cover/Use Classification Using a Comparison between Machine Learning Algorithms. *Remote Sens.* **2021**, *13*, 1349.
2021	Ireland		Murphy, D.J.; Shine, P.; O' Brien, B.; O' Donovan, M.; Murphy, M.D. Utilising grassland management and climate data for more accurate prediction of herbage mass using the rising plate meter. *Precis Agric.* **2021**.
2021	Malaysia		Muhammad Zumo, I.; Hashim, M.; Hassan, N.D. Mapping and estimation of above-ground grass biomass using Sentinel 2A Satellite Data. *Int. J. Built Environ. Sustain.* **2021**, *8*, 9–15.
2021	Portugal		Serrano, J.; Shahidian, S.; Paixão, L.; Marques da Silva, J.; Morais, T.; Teixeira, R.; Domingos, T. Spatiotemporal Patterns of Pasture Quality Based on NDVI Time-Series in Mediterranean Montado Ecosystem. *Remote Sens.* **2021**, *13*, 3820.
2021	USA	DeepPaSTL	Rangwala, M.; Liu, J.; Ahluwalia, K.S.; Ghajar, S.; Dhami, H.S.; Tracy, B.F.; Tokekar, P.; Williams, R.K. DeepPaSTL: Spatio-Temporal Deep Learning Methods for Predicting Long-Term Pasture Terrains Using Synthetic Datasets. *Agronomy* **2021**, *11*, 2245.
2021	USA Oregon		Jansen, V.; Kolden, C.; Schmalz, H.; Karl, J.; Taylor, R. Using Satellite-Based Vegetation Data for Short-Term Grazing Monitoring to Inform Adaptive Management. *Rangel. Ecol. Manag.* **2021**, *76*, 30–42.

3.9.2 Details on the model creation

The model creation process to produce the models implemented in the platform is similar to the one developed in our previous study [26]. To ease the understanding, the key points are summarized in this supplementary material section.

3.9.2.1 Reference data acquisition

As explained in the main body, the models aimed at predicting the compressed sward height. Recordings were acquired over 3 farms: the parcels of one of them were sampled regularly every year while the two others were regularly followed during a year, i.e. 2018 and 2019 respectively. The farm followed in 2019 was considered as the validation set to continue on the track of [26] and allow for cross-paper model comparison. Regarding the sampling strategy, CSH was recorded using a Jenquip EC20G rising platemeter (NZ Agriworks Ltd t/a Jenquip, New Zealand), and following a zigzag pattern across each parcel sampled.

3.9.2.2 Predictor variables

As explained in the main body, 3 "types" of predictor variables were considered: Sentinel-1 backscatter coefficient with the VV and VH polarization, Sentinel-2 multi-spectral information and meteorological data. The data went through feature transformations to compensate the possibly non-linear relationship between the CSH and the predictors. These feature transformations included: square, exponential, inverse, hyperbolic tangent, square root, logarithm of base 10 and composite features and indices.

3.9.2.3 Data association

The association of predictor features to the reference data required an intermediary step: defining a common reference. The strategy was: 1) define a grid (of 10m resolution) over each parcel then filling each pixel with the median value of the CSH records that fell within that pixel; 2) fill the pixels with the predictors variables, using an iterative approach to ensure a higher level of completion, with the iteration going from a time window equal to the exact date of acquisition to 4 days in the past and 4 in the future. This resulted in a training dataset made of 9,376 full records of 196 features and a validation dataset of 871 records of 196 features.

3.9.2.4 Feature selection

Given the increase in the number of features considered due to the feature transformation, there was a risk of too high dimensionality in the dataset compared to the number of records available. Therefore, a first cross-validation (CV) was performed with the algorithms considered as promising in [26] and one more that was not included: XGBTree, Cubist, Random forest, glmnet and nnet. This cross-validation was fold independent, with the 68 folds being the date of CSH sampling. It was performed on the whole training dataset across the set of hyper-parameters detailed in Table S3. Afterwards, the variable importance score were computed and summarized (mean divided by the standard deviation and

median divided by the interquartile range) across models. The summarized metrics were used to create a new variable importance ranking (illustrated in Annex 4). By integrating iteratively the most informative features into linear regressions, and recording the associated adjusted coefficient of correlation, it was possible to detect the inclusion of non-informative features and thus "critical" numbers of features the models should consider. The 5 first "breakpoints" in the number of features to consider were: 60, 97, 137, 143 and 158.

3.9.2.5 Hyper-parameter tuning

The hyper-parameters explored to tune the models are detailed in Table S3, altogether with the final value adopted. The hyper-parameter tuning was performed using a cross-validation with the 68 independent folds based on the date of CSH acquisition. For each algorithm tested, it was performed five times, the training dataset containing the 60, 97, 137, 143 and 158 most informative features. For each algorithm, the best performing models (lowest mean RMSE across the folds) were kept and the optimal parameters are shown, altogether with the number of features entering the best performing models, in Table 3-11.

Table 3-11: Hyper-parameter explored during the hyper-parameter tuning

Model	Parameter	Hyper-parameter values tested	Final number of features in the model	Final hyper-parameter value
xgbTree	Nrounds; max_depth; eta; gamma; colsample_bytree; min_child_weight; subsample	(200, 1000); (4, 6); (0.1, 0.25, 0.5); (0.1, 0.5, 1.0);(0.5, 1.0); (1, 3); (0.5, 1,0)	143	200; 6; 0.1; 1; 0.5; 1; 1
Cubist	committees; neighbors	(1:10); (0:7)	158	10; 0
Random Forest (RF)	mtry	88	143	88
Glmnet	Alpha Lambda	(0.0001,0.001,0.005,0.01,0.05,0.1:0.1:1); (0.0001,0.001,0.005,0.01,0.05,0.1:0.1:10)	158	1; 0.1
nnet	Size; decay	(1:7); (0,0.0001, 0.001, 0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)	158	3; 0.01

3.9.2.6 Model training and performances

Once the optimal number of features to consider and the optimal hyper-parameters have been determined for each modelling algorithm, it is time to create the production model. The performances in the fold-independent cross-validation

(that could be seen as the expected range of performances) and independent validation are summarized in Table 3-12.

Table 3-12: Performances of the models used for the platform creation

Model	Fold-independent cross-validation		Independent validation
	Mean RMSE (mm of CSH)	Standard deviation of the RMSE (mm of CSH)	RMSE
Cubist	23.8	6.2	17.9
Glmnet	22.5	7.0	15.1
Nnet	24.1	7.3	18.7
rf	22.4	5.6	17.7
xgbTree	21.9	6.1	17.9

3.9.3 Evolution of the Walloon grasslands

Using the parcel repartition available on WalOnMap (wallonne 2022) available for the years 2015 to 2020 at the time of writing, it is possible to draw the trends concerning the area and the number of Walloon parcels used as pastures. It is important to note that this term here encompasses the grazed pastures but also the mown ones and they could be either temporary or permanent. The number of pastures (totalPastureNumber) and the global area dedicated (totalPastureArea_m2) grew over these 6 years although it seems that the parcels are globally smaller given that the increase in the number of pastures was higher than the increase in area (Table 3-13). There were small differences between the sum of the computed areas and the sum of the declared areas, mainly due to rounding and delimitation of parcels not exactly fitting the border of the parcels. The rise in the percentage of pastures within the agricultural area attribution is due to both the increase in the area dedicated and the decrease of the total agricultural areas.

Table 3-13: Number of pastures and area covered between 2015 and 2020. The SAU corresponds to the global area dedicated to agricultural activities.

year	totalSAU_m2	totalPastureNumber	totalPastureArea_m2	percentPastureArea
2015	8,149,209,335	158,921	3,722,260,800	45.68
2016	8,145,473,197	161,959	3,738,523,652	45.90
2017	8,155,550,248	165,122	3,762,399,703	46.13
2018	8,153,854,595	194,657	3,988,696,285	48.92
2019	8,114,728,925	204,463	4,034,431,641	49.72
2020	8,094,418,326	209,907	4,037,578,677	49.88

The definition of permanent pastures is: area dedicated to pastures for at least five years. A way to check the number of permanent pastures in Wallonia is thus to check the occupation change, whether it concerns added or removed areas (Table 3-14). The maximum cumulative change over five years (addition and

removal of different pasture areas) was between 2016 and 2020 and was equal to 94,167.91 ha or 11.6% of the total agricultural area of 2020. This means that at least 88.4% of the Walloon pastures are permanent.

Table 3-14: Added and removed pasture areas between 2015 and 2020. Values were rounded to ha units

Reference year	Added areas					Removed areas				
	2016	2017	2018	2019	2020	2016	2017	2018	2019	2020
2015	15,879	25,924	64,927	61,766	65,818	14,249	21,895	38,228	30,464	34,193
2016	/	15,833	58,765	56,945	62,081	/	13,435	33,696	27,274	32,086
2017	/	/	50,453	50,482	57,607	/	/	27,782	23,209	30,010
2018	/	/	/	36,727	46,112	/	/	/	32,125	41,185
2019	/	/	/	/	17,269	/	/	/	/	16,945

3.9.4 Variable importance of the newly trained models.

The ranking of the 160 most important variable is shown in Table 3-15.The naming scheme is the following except for the time features:

- The variable origin: S1 for Sentinel-1, S2 for Sentinel-2, Met for meteorological data, IDB for Sentinel-2 based indices. A T is added when a transformation is applied
- The variable in itself: for S1 it is either the VV or VH or timelag; for S2 it is either the band of the timelag; for the meteorological data P represents the precipitations, ens for the solar radiation, ETP for the potential evapotranspiration, DJ00 for the degree-day on the 0 basis, HR for the relative humidity, WS for the wind speed; for IDB the indice in the online database
- The transformed nature: exp for exponential, inv for inverse, sqrt for the square root, log10 for the logarithm in base 10, tanh for the hyperbolic tangent, sq for the square, cube for the cube

Table 3-15: Variable ranking according to their importance, from the highest to the lowest.

VarName	Order	VarName	Order	VarName	Order	VarName	Order	VarName	Order
MetT.ens.exp	1	MetT.DJ00.3.sq	41	S2.B12	81	IDB.023	121	S2T.B08a.sq	161
Met.P.3	2	S2T.B04.sq	42	IDB.070	82	IDB.140	122	IDB.003	162
MetT.P.3.exp	3	S2.WVC	43	S2T.B12.sqrt	83	IDB.223	123	S2T.B06.inv	163
Met.P.7	4	S1T.VV.inv	44	S2T.B12.exp	84	IDB.001	124	IDB.110	164
MetT.P.7.exp	5	S1T.VV.exp	45	S2T.B12.log10	85	IDB.075	125	IDB.113	165
MetT.P.7.sq	6	Met.DJ00.7	46	IDB.040	86	IDB.135	126	S2T.B07.sq	166
MetT.P.3.sq	7	Met.WS	47	IDB.050	87	IDB.136	127	IDB.042	167
S2T.B02.cube	8	MetT.WS.exp	48	IDB.085	88	IDB.137	128	S2T.B08.sq	168
Met.ens	9	MetT.DJ00.3.exp	49	IDB.118	89	IDB.222	129	S2T.B08a.cube	169
S2.dTS2	10	MetT.WS.sq	50	IDB.095	90	IDB.130	130	IDB.182	170
MetT.ens.sq	11	MetT.DJ00.7.exp	51	S2T.B11.cube	91	IDB.036	131	S2T.B06.cube	171
S2T.B03.cube	12	MetT.DJ00.7.sq	52	IDB.046	92	IDB.084	132	IDB.002	172
MetT.ETP.sq	13	S2T.B05.sq	53	IDB.014	93	IDB.080	133	IDB.232	173
S2T.B04.cube	14	S2T.B03.exp	54	IDB.004	94	IDB.033	134	IDB.224	174
Met.P.15	15	S2T.B02.exp	55	IDB.091	95	IDB.234	135	IDB.101	175
MetT.P.15.exp	16	S2.B03	56	S2T.B11.sq	96	IDB.041	136	IDB.038	176
S2T.B02.sq	17	S2T.B03.tanh	57	IDB.152	97	IDB.074	137	S2T.B07.cube	177
MetT.ETP.exp	18	S2T.B03.sqrt	58	IDB.097	98	S2T.B11.sqrt	138	IDB.235	178
Met.ETP	19	S2T.B12.cube	59	IDB.229	99	S2T.B08a.sqrt	139	IDB.166	179
MetT.P.15.sq	20	S2.B02	60	IDB.061	100	S2T.B08a.log10	140	IDB.186	180
MetT.DJ00.sq	21	S2T.B02.tanh	61	S2T.B11.exp	101	S2T.B08a.tanh	141	IDB.068	181
MetT.HR.sq	22	S2T.B05.exp	62	IDB.086	102	S2T.B08.log10	142	S2T.B08.cube	182
MetT.HR.exp	23	MetT.DJ00.15.exp	63	IDB.090	103	S2.B08a	143	S2T.B11.log10	183
Met.HR	24	Met.DJ00.15	64	IDB.058	104	IDB.034	144	IDB.066	184
S2T.B03.sq	25	MetT.DJ00.15.sq	65	IDB.098	105	S2T.B07.sqrt	145	IDB.177	185
MetT.DJ00.exp	26	S2.B05	66	IDB.008	106	S2T.B08.sqrt	146	S2T.B06.sq	186
S2.AOT	27	S2T.B02.sqrt	67	IDB.115	107	S2T.B07.log10	147	S2T.B06.sqrt	187

S1T.VH.sq	28	S2T.B05.tanh	68	IDB.239	108	S2T.B07.tanh	148	S2T.B06.log10	188
Met.DJ00.3	29	S2T.B05.sqrt	69	IDB.191	109	S2T.B08.tanh	149	S2T.B06.tanh	189
S1.VH	30	S2T.B11.inv	70	S2.B11	110	S2T.B08.inv	150	IDGrid	/
S1T.VH.tanh	31	IDB.011	71	IDB.087	111	S2T.B08a.inv	151	Height_mm	/
S2T.B05.cube	32	S2T.B12.sq	72	IDB.233	112	S2.B07	152	Parcelle0	/
S1T.VH.inv	33	S2T.B12.inv	73	IDB.249	113	IDB.006	153	Site	/
S1T.VH.exp	34	S2T.B04.exp	74	IDB.017	114	S2T.B08a.exp	154	Date	/
Met.DJ00	35	DOY	75	S2T.B11.tanh	115	S2.B08	155	ModelFile	/
S1.dTS1	36	Month	76	IDB.069	116	IDB.031	156	Applicable	/
IDB.236	37	S2T.B04.tanh	77	IDB.081	117	IDB.226	157	method	/
S1T.VV.sq	38	S2.B04	78	IDB.121	118	S2T.B07.inv	158	S2T.B06.exp	/
S1.VV	39	S2T.B04.sqrt	79	IDB.231	119	S2T.B07.exp	159	S2.B06	/
S1T.VV.tanh	40	S2T.B12.tanh	80	IDB.067	120	S2T.B08.exp	160		

3.9.5 Distribution of the predicted CSH depending on the time lag tolerance considered

Impact of the merging tolerance, applied on the S2 dataset in Figure 3-11 and on the S1 dataset in Figure 3-10, on the prediction value. On each subplot, the boxplots groups are ordered from 0 day to 4 days time lag tolerance from left to right. The five models used are numerated from 1 to five for space convenience and correspond respectively to: the cubist model, the glmnet, the nnet, the rf and the xgbTree. The year and the dataset concerned by the splitting are represented in the upper right corner of each sub-plot.

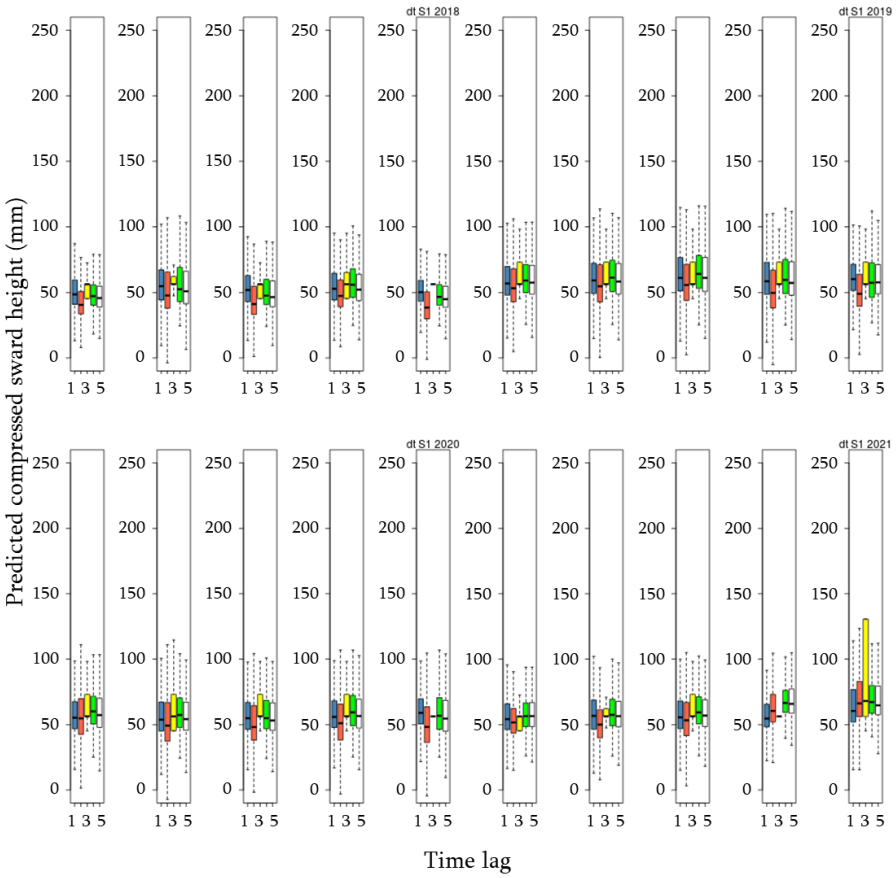


Figure 3-10: Impact of the merging tolerance (time lag), applied on the S1 dataset, on the prediction compressed sward height value.

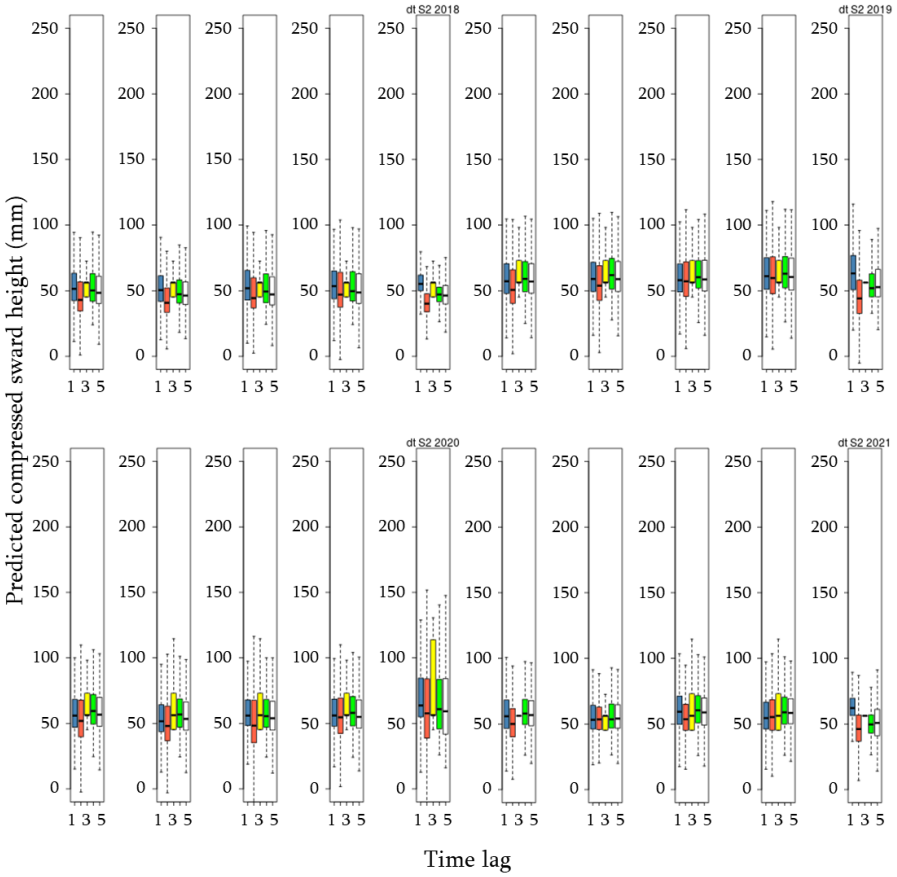


Figure 3-11: Impact of the merging tolerance (time lag), applied on the S2 dataset, on the prediction compressed sward height value.

3.9.6 Proposition of representation of the probability to get x predictions a year

For future users of the DSS, it might be relevant to assess the number of predictions they can expect per grazing period. An illustration of the probability curve is proposed in Figure 3-12. This representation was made on the basis of the whole dataset. Given that some areas were more covered by the chosen S1 and S2 tiles, it might be relevant in the future DSS to offer a finer spatial granularity. The proposed use of this representation is to start form a certain number of predictions the user wants per year and then check the range of probability to get this number of predictions.

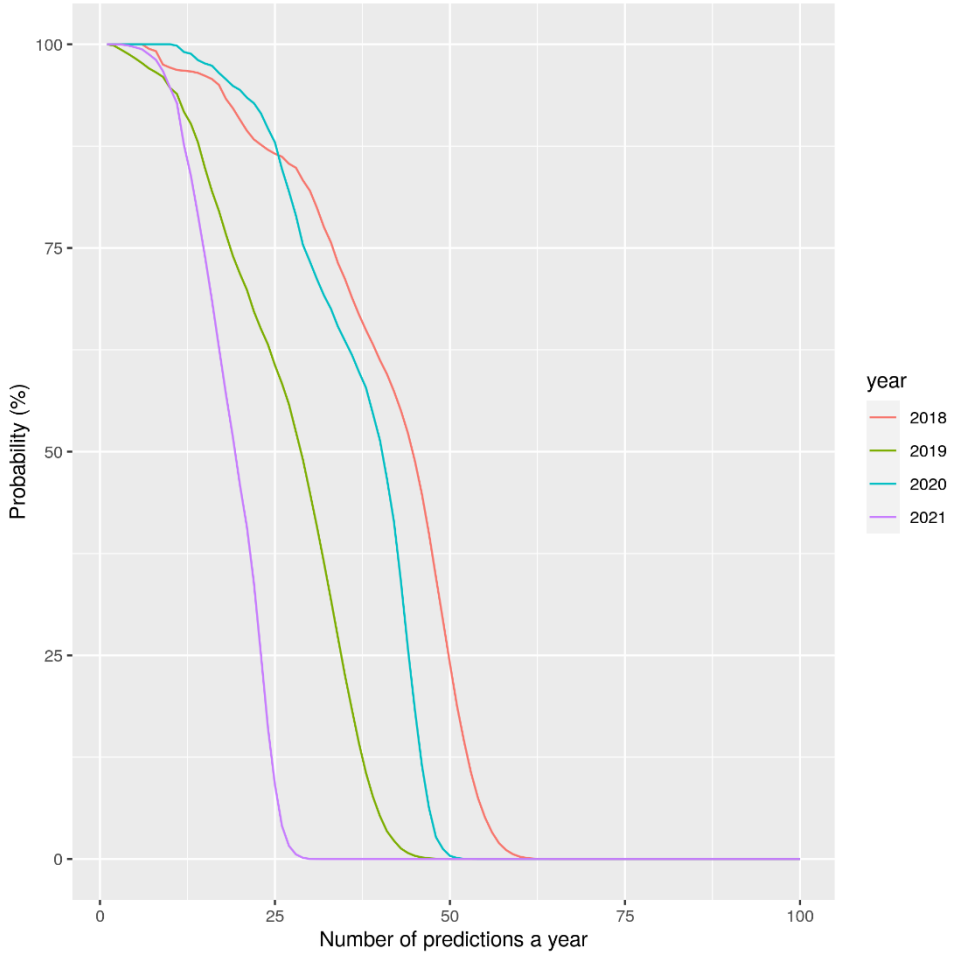


Figure 3-12: Probability of getting x number of prediction per year for each year studied.

3.9.7 Summary of the predicted CSH values at the pixel level

Summary of the predicted CSH values at the pixel level is represented in Table 3-16.

Table 3-16: Summary of the predicted CSH values at the pixel level. The left part relates to raw finite predictions (N= 1,426,156,171; 1,047,054,399; 1,482,392,774; 764,039,165 in 2018, 2019, 2020 and 2021) and the right one to data filtered to the]0:250]mm of CSH range (N= 1,425,852,569; 1,046,797,529; 1,481,945,618; 763,510,335 in 2018, 2019, 2020 and 2021). All values are in mm of CSH.

		Filtered on the finite values					Filtered on the range				
		Cubist	Glmnet	Nnet	Rf	xgbTree	Cubist	Glmnet	Nnet	Rf	xgbTree
2018	Minimum	5.9	-17,670,617.0	45.3	18.4	4.6	6.2	0.0	45.3	18.4	4.7
	1%	29.2	16.5	45.3	28.9	26.6	29.2	16.6	45.3	28.9	26.6
	1 quart.	42.7	35.0	45.3	41.6	40.0	42.7	35.0	45.3	41.6	40.0
	Median	51.7	43.0	56.2	49.3	47.8	51.7	43.3	56.2	49.3	47.8
	Mean	56.0	48.6	61.1	55.0	53.6	56.0	48.6	61.1	55.0	53.6
	SD	20.0	526.4	21.9	20.6	21.2	19.9	20.3	21.9	20.6	21.2
	3 quart.	63.4	58.0	56.2	62.2	60.4	63.3	57.6	56.2	62.2	60.4
	99%	132.9	114.0	130.6	130.4	134.4	132.8	114.0	130.6	130.4	134.4
	Maximum	348.6	4,272,111.0	130.7	208.0	230.0	250.0	250.0	130.7	208.0	230.0
2019	Minimum	0.8	-3.559.411.0	45.3	25.1	5.9	5.6	0.0	45.3	25.1	5.9
	1%	35.9	24.7	45.3	35.2	34.2	35.9	24.7	45.3	35.2	34.2
	1 quart.	49.0	43.0	56.2	50.4	48.9	49.0	42.6	56.2	50.4	48.9
	Median	58.6	54.0	56.2	60.6	58.3	58.6	53.9	56.2	60.6	58.3
	Mean	63.8	58.0	66.5	65.5	64.1	63.8	58.2	66.5	65.5	64.1
	SD	21.7	678.5	24.5	22.3	22.7	21.6	21.2	24.5	22.3	22.7
	3 quart.	71.9	70.0	73.1	73.7	72.2	71.9	69.6	73.1	73.7	72.2
	99%	144.8	120.1	130.6	146.1	148.8	144.7	120.0	130.6	146.1	148.8
	Maximum	353.1	11,662,154.0	130.6	217.3	239.0	250.0	249.9	130.6	217.3	239.0
2020	Minimum	0.4	-7,167,708.0	45.3	24.1	7.7	5.5	0.0	45.3	24.1	7.7
	1%	33.9	19.7	45.3	32.9	32.4	33.9	19.9	45.3	32.9	32.4
	1 quart.	46.5	39.0	56.2	48.4	46.5	46.5	39.0	56.2	48.4	46.5
	Median	55.1	51.3	56.2	58.3	55.5	55.1	51.3	56.2	58.3	55.4
	Mean	60.1	55.0	67.2	62.1	60.5	60.1	54.9	67.2	62.1	60.5
	SD	20.2	613.5	25.5	20.1	20.9	20.1	21.2	25.4	20.1	20.9
	3 quart.	67.5	67.2	73.1	70.8	68.6	67.5	67.2	73.1	70.8	68.6
	99%	134.3	116.0	130.6	135.1	137.9	134.1	116.0	130.6	135.1	137.9
	Maximum	352.2	13,335,017.0	130.6	204.8	226.5	249.9	249.9	130.6	204.8	226.5
2021	Minimum	3.1	-12,994,638.0	45.3	26.0	11.4	3.1	0.0	45.3	26.0	11.4
	1%	30.8	17.0	45.3	36.7	34.3	30.8	17.5	45.3	36.7	34.3
	1 quart.	46.6	42.0	56.2	49.2	48.7	46.6	42.3	56.2	49.2	48.7
	Median	55.4	52.0	56.2	57.8	56.9	55.4	51.9	56.2	57.8	56.9
	Mean	59.2	54.3	61.3	61.3	60.6	59.2	54.2	61.3	61.3	60.6
	SD	18.6	2265.8	19.4	17.3	17.9	18.5	17.9	19.4	17.3	17.9
	3 quart.	67.6	64.0	62.1	68.9	67.9	67.6	63.6	62.1	68.9	67.9
	99%	124.3	108.0	130.6	121.4	123.8	124.0	107.9	130.6	121.4	123.8
	Maximum	332.4	59,077,753.0	130.6	204.2	224.8	249.9	249.9	130.6	204.2	224.8

Chapter 4

Presentation of the retrospective analysis and complements addition to the work described until now:

Would I pack the same way?

4.1 Story

Dear reader, until now, you have been told what has been done. Now that we arrived at a tavern, it is time to check on our luggage and consider them regarding the partial conclusions of each chapter of this part of the journey. Indeed, taking a step-back will help us alleviate our bags load, adapt the tools we chose and bring complementary reflexions on the path choice we made along the way. Not all options could nor should be considered right away. In fact, there is enough points of discussion to drive another thesis.

4.2 Outline

This chapter is dedicated to a discussion about the choices made and the alternatives. Some options could be implemented right away, and some were actually used along the way. On the other hand, other changes would require consequent time, effort and/or computation resources to be implemented. This last part of the changes is to be considered as perspectives. Other potential evolution paths are also developed until they appear to be dead-end. This chapter is separated into 6 parts, with the 5 first corresponding to the different parts of the formula way of defining models in R as illustrated in Figure 4-1, related to:

- The objective and modelled feature;
- The data origin;
- The machine learning algorithms, on which the models are based, recall and detail;
- The data treatment to feed the models;
- The data augmentation through gap-filling;
- The use of the sub-products of this thesis beyond its scope.

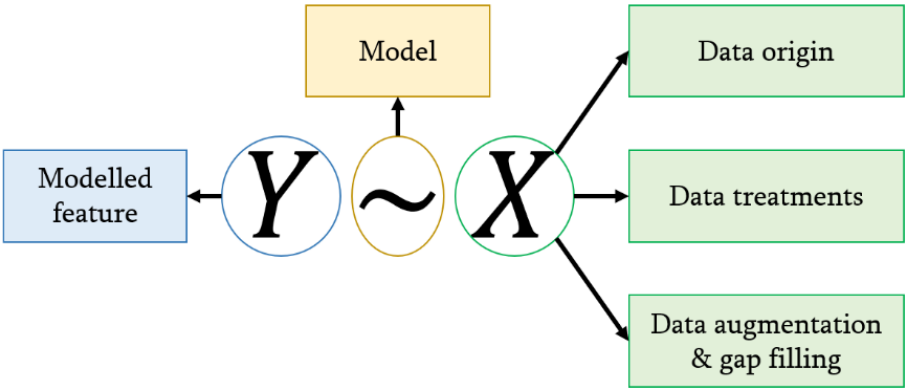


Figure 4-1: Illustration of the organisation of the discussion chapter in regard to the formula definition of the models in R.

4.3 Objective and modelled feature

As explained in the introduction, the feed available on pastures, and in particular the grass, can be represented under different proxies: surface height model, global height, CSH, LAI, (dry) biomass, crude protein content, acid and neutral detergent fibre, ashes, metabolizable energy. Choosing between these descriptors requires to make concessions between:

- The spatial representativity of the data. Apart from the surface height model, the afore mentioned metrics are not intrinsically spatially distributed. A positioning system like a GPS is needed for the discrete spatialisation of these data. This positioning could even be improved using corrections like Real Time kinematic (RTK) technologies, as in Michez *et al.* (2019) [ch-062]. More continuous spatial coverage requires spatial interpolation techniques like kriging;
- The ease of acquisition. Underlying this concept, one should consider the human time and the resources/material needed. E.g. the surface height model is often produced through an UAV acquisition and then surface model reconstruction through sfm on the images, as in Michez *et al.* (2020) [ch-018]. This means that, beside the human time, there is a need to consider the cost spreading of the acquisition platform and the computing power to perform the sfm and then handling these data. Other proxies like biomass require to dry the grass in the oven, protein and fibre content require chemical compounds.
- The understandability and meaningfulness of the metric for the application. From farmers' point of view, depending on the level of individual training, speaking of CSH, LAI, biomass, protein content and so on might or might not be as relevant as a visual assessment of the cover level on the sward. It is often opposed to a scientific' prospective based on using as much and as fine descriptors as possible to seize the whole reality and describe the most precisely the environment and therefore, in our case the feed available and the actual needs of the cattle.

In this thesis, it was chosen to work with CSH mainly for its ease of acquisition. It also opened up a rapid expansion possibility of the database exploited to develop the models as it is widely used in the scientific community. The spatial representativity of this data is connected to its ease of acquisition. Indeed, a RPM allows for a rapid acquisition and using a calibrated GPS enables a rapid spatialisation. Therefore, although this data is not intrinsically spatialised, the spatialisation can be done without work overload, especially given that recent

RPM are connected to a smartphone application relying on the integrated GPS data acquisition. As mentioned before, the acquisition of this metric of the available feed requires small investments for the material. On the other hand, it still requires time for wandering across the parcels although it is limited as, if the GPS and the RPM were calibrated properly, the data are available as soon as the field acquisition is done.

Concerning the representativity of this metric, it is more prone to discussions. CSH reflects a combination of both the actual height and the density of the forage. Therefore, this metric is hard to validate from a Walloon farmer's prospective as it might not be as meaningful as a casual sward height or an actual biomass value. From a more warned user/scientific prospective, the lack of information about the nutritional quality of the grass might also cripple this choice. There exist translation models that perform the conversion from the CSH to the actual biomass. However, these models have to be cherry picked as the reference acquisitions for the model training have to be very similar, notably regarding the structural organisation of the studied vegetation that have a significant impact according to Laca *et al.* (1989) [ch-042]. The main points of attention are the meaning of the CSH, and thus the RPM characteristics, and the biomass entering the conversion model. Comparing this choice between countries and study cases underlines the variability of the reference measurements: a Japanese study Nakagami and Itano (2014) [ch-020] relates whole CSH values to the above 5 cm aboveground dry biomass; another study by Hakl *et al.* (2012) [ch-021] focused on the lucerne, that is more often related to mowing management, and computed the relationship between the whole CSH to the above 4 cm aboveground dry biomass. In a context more related to the Walloon area, a popularisation paper Lefèvre *et al.* (2022) [ch-022] summed up the Walloon situation: the lower clipping threshold for biomass estimation varies between 3 cm to 5 cm and the CSH considered is often the whole CSH. The problem due to this change in reference does not only relate to the biomass identification but also to general properties of the grass: along the stem, there is a lignification gradient altogether with a biomass and chemical one. This translates for example in higher density for low CSH, hence the limits in Lefèvre *et al.* (2022) [ch-022] of 5 cm and 15 cm for the applicability range of their conversion equation. This change in composition along the stem height might seem trifling, yet in the case of mowing for silage, changing the minimum reference height might be one of the causes for a higher concentration in nitrate or the presence of soil residue that might lead to contamination of the silage and therefore of toxic gases as illustrated in Delforge and Sevrin (2022) [ch-023]. In

this thesis, we modelled the whole CSH, without a lower threshold. Future conversion layers should be adapted to fit this property.

As an alternative to the afore mentioned conversion models that should be handled with caution, especially regarding the input training data, it was also thought to use actual biomass values in the context of this thesis. This approach lost its relevancy when we fully seized the constraints of fiddling with datasets of various spatial resolution. As shown in Table 1-7, only a few satellites offer the possibility to get resolution finer than traditional sampling quadrats size. As a reminder, Ali *et al.* (2017) [ch-016] used bands of 1 m*3.5 m for assessing the biomass, Lussem *et al.* (2019) [ch-017] used bands of 1.5 m*3 m and Michez *et al.* (2020) [ch-018] bands of 7 m*1.5 m. Additionally, the "Centre des technologies agronomiques" (aka "centre for agronomical technologies", CTA) [ch-044] had spatial records of biomass on bands of ± 7 m*1 m with a different mowing timing for the 5 contiguous bands. As a result, no constellation with no cost presented a fine enough resolution to make sure that pixels fell only on the mown sub-parcels, without too much noise coming from the surrounding area of the quadrat. The option of going with costly satellite acquisition was discarded as we wanted to create a DSS available for all farmers in Wallonia and this would have meant huge costs for gains in spatial resolution with a possible loss of spectral resolution. Nowadays, models exist to compensate this loss in spectral resolution [ch-015] and [ch-024]. However, transferring the whole processing pipe on these data with a super-resolution (also called pansharpened) was assessed to fall out of the scope of this thesis, as it involved dealing with much larger datasets and thus increase the load on the hardware. This integration of pansharpened data is part of the prospective for a future development alike the integration of actual biomass direct prediction, although this integration would first need a re-definition of quadrats size and number.

As the alternatives for CSH have been discarded, the DSS drafts development still required a way to assess the fulfilment of the cattle needs. We included a conversion layer from per pixel CSH to per pixel biomass by 10,000 m² (ha) equivalent. The initial version of this conversion layer was the one defined by the RPM designer $dry\_biomass = CSHmm * 15 + 500$. As it appeared during the discussions that followed, this conversion is currently a hot topic as FourragesMieux and the CRA-W recently published an equation better suited for the Walloon pastures: $dry\_biomass = CSHmm\ above\ 5cm * \frac{215}{10}$ [ch-022]. Given that other teams are still working on this, like McSweeney *et al.* (2022) [ch-019] that also included the effect of the season and of the floristic composition, we thought during the creation of the second draft of the DSS to include it as a package

that could be updated/changed from within the GUI so the end-user could select which equation he wants to use.

A complementary approach, based on a **multi-scale modelling** paradigm was considered. It consisted in the re-training of the equations if the end-user provides local data. In the case the end-user wants a translation into an actual height, this should also be possible. As explained in the section dedicated to the future of the DSS, all these implementations are perspectives of this thesis.

4.4 Data origin

As expressed in the second article, some data providers had to be changed compared to what was described in the published paper in Chapter 2. For the sake of completeness, existing alternatives are also cited.

4.4.1 CSH data

The first models were developed using CSH data acquired with a EC-20 G from Jenquip RPM. Although there exists alternatives like the grasshopper used in Murphy *et al.* (2019) [ch-029], it was chosen to keep on using and acquiring data with the EC-20 G rpm to make sure the data used are the same across the whole dataset. Indeed, the change in measuring technology (ratchet count on the EC-20 G side versus the micro-sonic sensor on the grasshopper side) could induce a bias in values acquired as illustrated in McSweeney *et al.* (2019) [ch-030]. Another reason was practical, the GUI of the Grasshopper related smartphone application requires to define beforehand the parcels and the number of records intended. Once this number is reached, it stops recording and induces a need to setup the application again.

For the current discussion and step-back chapter, a new dataset was created. It consisted of data acquired at the CTA by the CTA team between 2018 and 2021, at various farms by FourragesMieux and at various farms by Killian Dichou from ULiège-GxABT in 2022. Due to errors partly located at the level of the smartphone application linking to the EC20-G and performing the GPS tracking, there were faulty recordings that had to be treated. First off, we had to discard a whole batch of data whose position was 0°N;0°E. This happened principally in the CTA dataset across all years and might be due to a loss of signal. The next filter was the removal of a batch of recording between the 23rd of July 2021 and beginning of May 2022, as the recordings were concentrated at specific points in and out of the parcels, denoting a poor implementation of the GPS use feature of the new version of the smartphone application, more precisely a too low sampling frequency. This hypothesis was proven true by the solution found by Killian Dichou which consisted in using another application with the GPS signal retrieval properly implemented to get a correct positioning. Other data treatments were also needed regarding the encoding of data: removal of special characters like accentuation marks; standardisation of the date encoding; and the rescue of longitude and latitude numbers that lost their decimal separator. The descriptive statistics of this dataset are shown in the supplementary material (Table 4-6) and illustrated per year without and with a distinction per parcel in Figure 4-2. The annual distributions seem to fit the gamma-distribution described in Nakagami (2016)

[ch-033] although some acquisitions campaigns seem to have gathered a very narrow range of data, hence the spike on the right part of Figure 4-2.

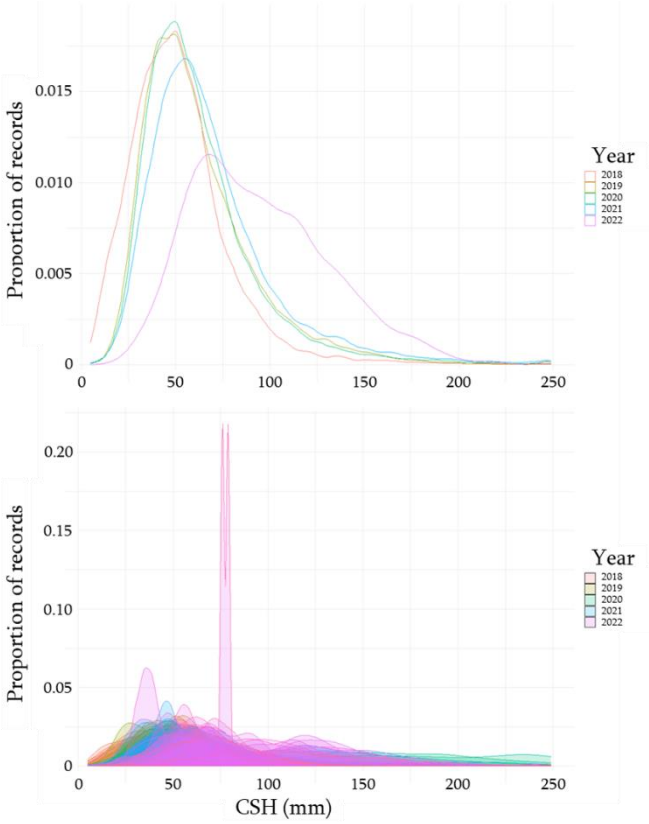


Figure 4-2: Density of the distribution of the CSH per year on the left. On the right each plot represents a different parcel-time of acquisition combination

4.4.2 Meteorological data

Instead of using data collected in one farm as in Nickmilder *et al.* (2021) [ch-000], the collection of meteorological data now relies on a distributed and standardised network of meteorological stations, called Agromet and operated by the CRA-W (2022) [ch-001]. It presented multiple advantages: a finer representation of Wallonia than with a single meteorological station, a standardisation of the condition of acquisition, a near-real time availability of the information and a handful and automatic way of retrieving data through an API. Recent advances in the Agromet project involved a finer representation of some features than the raw values located at the level of the stations. However, only some features received this kriging-like treatment. Implementing this dual-

resolution dataset would have complexified the handling of spatial data, especially regarding the choice of spatialisation. In fact, it would require retrieving the reference raster used for their kriging-like spatialisation, and associate to each pixel, that have, as far as I understood, a 1 km resolution, the Voronoï polygon, used until now, that covers it the most. This would mean that instead of one line per date per station, there would be one line per date per pixel. Therefore, the number of lines for each day of acquisition would largely increase. Hence, the computational cost of the integration of meteorological data would increase compared to the use of the station identifier until the last moment. As it might still be relevant in a near future, this option is a mid-term evolution perspective.

4.4.3 Sentinel-1

No changes were made regarding the data provider of S1 data. The way to acquire it has just been automated through the use of the sentinelsat python package Wille *et al.* (2016) [ch-028]. A technical precision is still needed: it had to be built locally on the server due to constraints on the permission rights. To fasten the acquisition and bypass the limitations on the number of tiles that can be asked to be retrieved from the long-term archiving, five accounts were created using all the old evolutions of the mail addresses of the university of Liège.

4.4.4 Sentinel-2

At first, the model created relied on data coming from the scihub-copernicus hub API [ch-006] from the ESA. When scaling up the data acquisition to the whole Walloon Region, the data provider for S2 tiles was changed to Theia [ch-007]. The main reason for this change of data provider was the conversion algorithm from top of atmosphere (S2-L1C) to bottom of atmosphere/top of canopy (S2-L2A) reflectance. It was made through look-up table in the sen2cor algorithm used for the ESA conversion [ch-008] whilst the MAJA algorithm, implemented in the Theia-CNES platform, relied on the exact computation of correction formula [ch-009]. The main implication of this difference is a significant change in the cloud detection as illustrated by Baetens *et al.* (2019) [ch-025] and Sanchez *et al.* (2020) [ch-026].

Another motivation for the change of data provider was the ease of access. Indeed, for retrospective studies, the ESA's long-term archiving policy slowed down a lot the data acquisition as the number of tiles that could be downloaded at once was limited. Furthermore, the acquisition process requires multiple identifications spaced of at least 1 hour.

4.4.5 Complementary data

The inclusion of other data was considered, especially regarding the topography and geo-pedological characteristics.

Theoretically, topography (altitude, slope and orientation) impact on S1 and S2 are already accounted for during the pre-treatments. However, it could somehow help the models to mitigate the meteorological data. Yet, this is not included in the data sources. The main reason was to minimize the introduction of bias through over-represented classes. Indeed, data were mainly collected in the eastern half of Wallonia and, given the topography of Wallonia, this would have induced a biased representation of the topology in the models and cripple the possibility to apply them at a larger scale.

Regarding the other considered data source, soil data were not included partly to also reduce bias as only a part of the Wallonia was covered. The other main reason was related to the encryption of the soil characteristics in the "Carte numérique des sols de Wallonie" (aka "Walloon numerical soil map", CNSW) that can be downloaded from the Service public de Wallonie (2023) [ch-045]. As stated in the layer description, there are more than 6,000 entries in the legend. These entries are the result of the combination of the substrate, texture, drainage, soil profile development, coarse elements load and other indicators. However, these values are not numbered, these are categorical letters. As a result, training the models on only a part of the range of the levels does not allow for extrapolation. There are ways to go back from level to actual number range and sub-products of this data source exist. However, the small range of training farm implies a small coverage of the total variability. Therefore, this data source was discarded, and its inclusion is a perspective to reconsider once the area, on which data are collected, will be enlarged. An alternative to the CNSW could be the coarser SoilGrid database (the finest resolution being 250 m) [ch-054]. This database includes pH, soil organic carbon content, bulk density, coarse fragments content, sand content, silt content, clay content, cation exchange capacity (CEC), total nitrogen as well as soil organic carbon density and soil organic carbon stock. This database integration is another perspective of this thesis.

Another data sources that could prove useful is the soil water content, as it is used in mechanistic modelling such as the Moorepark Saint-Gilles (MostGG) model [ch-056]. However, it is dependent on the soil and the meteorological history. As the soil description has been discarded, this data source is also discarded for now. Beside the soil related indicators, pasture management related indicators such as manure fertilizing as their importance is illustrate by their use in the afore mentioned study.

4.5 Machine learning modelling

ML models can be used to link the CSH to the other data sources. Five models were assessed to be the most promising in Chapter 2. In Chapter 3, it was assessed that two of them were even more relevant: a Cubist and a random forest. As the dataset size had substantially increased, it might be relevant to train the five most promising models anew. The modelling process will be globally similar to the one used in Chapter 2 and 3 (illustrated in Figure 4-3) but some points could be modified to be more efficient. This section aims to recall the algorithms used and to shed light on the strategical management of the dataset partitions for 1) guaranteeing the existence of a fully independent/ external validation dataset; 2) the assessment of the relevancy of the feature transformation; 3) the research of the optimal hyper-parameter; 4) the assessment of the range of possible performances of the models; 5) the assessment of the actual performances on the held-out external validation dataset. Furthermore, it also aims to demystify the larger validation prospect called "unsupervised validation".

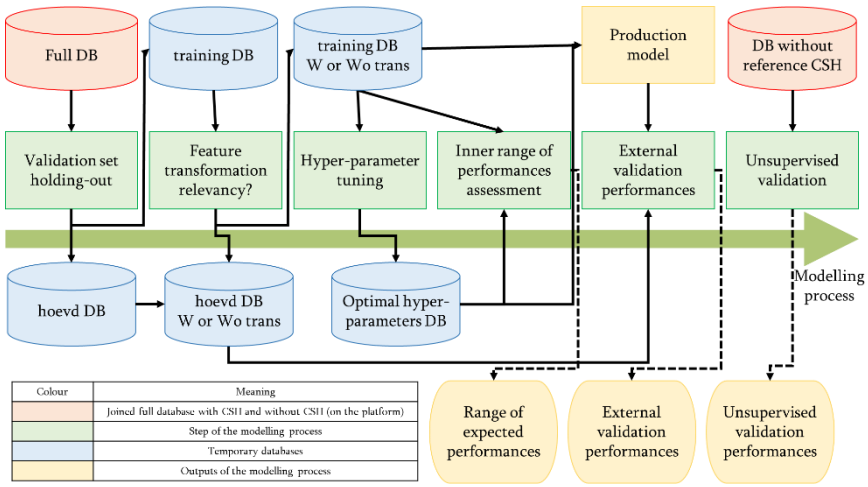


Figure 4-3: Modelling process workflow.

4.5.1 Technical foreword

To increase the computation speed and decrease the memory requirements of the cross-validation (CV) related sub-section, computation would be performed in python using float32 precision instead of the float64 format used in R. The decrease in precision could be considered as a potential advantage as it could lead

to a slight decrease in the over-fitting. The final/ "production" models should be trained in R using the best set of hyper-parameters as the platform was made to handle such models and spatial data handling is much easier and more reliable using R than python. Given that some hyper-parameters could or could not be accessed in each implementation, it required a particular attention regarding the package correspondence between the languages (as illustrated in Supplementary Material in Table 4-7).

4.5.2 Algorithm description

The first step of this analysis is the definition of the modelling algorithms that would be used. To ease the modelling and standardise the outputs, scikit-learn [ch-055] should be used as the main python library and the caret R package should be used. Regarding caret, there is a computational overhead due to the standardisation of the formats. It was assessed to be worth the time investment to avoid the hassle of debugging code. The correspondence of the hyper-parameters between the python and R implementation is detailed in Supplementary material in Table 4-7.

The models that would be explored again are:

- **Cubist**: regression algorithm generating rule-based predictive models. The python package that should be used, Cubist [ch-037], is a wrapper around the C-coded Cubist tool [ch-039] inspired from the work of Quinlan (1992) [ch-038]. It was designed to be compatible with the scikit-learn framework. Similarly, in R, the Cubist wrapper package is used [ch-040] and [ch-041].
- **Random forest (RF)**: meta-estimator fitting decision trees on various sub-samples of the dataset. The python version that should be used is the "RandomForestRegressor" from scikit-learn [ch-046]. On the other hand, the R "RandomForest" package should be used [ch-047]
- **Glmnet**: penalised regression based on the generalised linear model. In R, the glmnet package is used [ch-048]. In python, no package is as completely developed as in R. Therefore, a trick should be defined: a pipeline injecting a generalised linear model into a penalised regression. Scikit-learn pipeline [ch-049] should be used to inject the TweedieRegressor [ch-050] into the ElasticNet regression [ch-051]. The detail of the python functions arguments and default values is given with their original names. It should be noted that these names change when a pipeline is used. Given that it is part of the perspective of this thesis, all the prospects are not guaranteed to work, especially the warm-start ability (important for section 4.8.1) after getting through a pipe.

- **Xgboost**: Extreme gradient boosting is a library [ch-058] and [ch-059] based on the gradient boosting technique that consists in combining multiple "smaller" models. In this thesis, the boosted trees approach was adopted. It relies on a tree sub-model as the random forest, although it works in an association approach instead of an independent one. There is a R binding package [ch-060] and a python Scikit-Learn Wrapper interface for XGBoost [ch-061]. Given the high number of hyper-parameters, only the parameters easily accessible from the caret package are mentioned in Table 4-7.

Until now, only a basic approach of **neural network (NNet)** was adopted: multilayer perceptron with one hidden layer. However, more complex NNet such as the wide and deep convolutional NNet (a possible architecture is illustrated in Figure 4-4) would also be relevant to explore better the more complex relationships existing between the used features and to take profit of its ability for transfer learning. This is especially interesting to limit the computation cost during the modelling process and to speed the convergence.

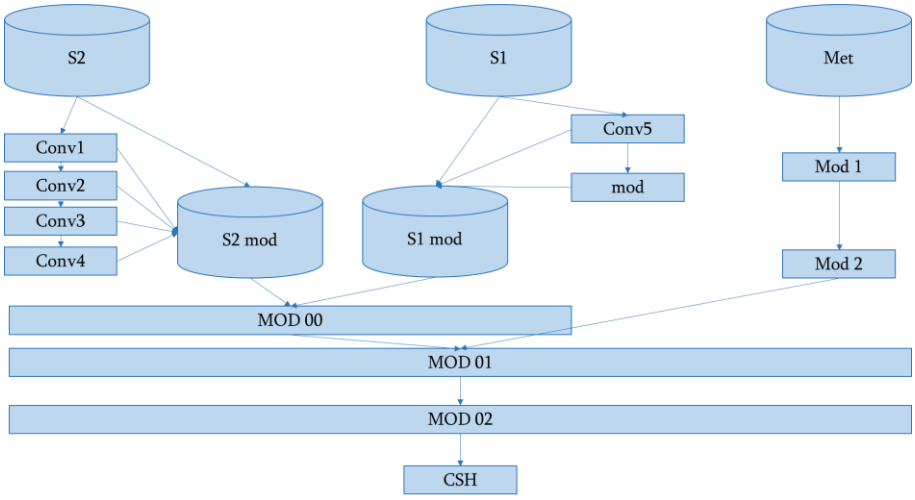


Figure 4-4: Example of wide & deep & convolutional NNet thought as a perspective model to explore. MOD means a transition layer with a transfer function, conv designates a convolution layers and CSH is the final layer that translates the values into actual CSH values.

4.5.1 Validation set holding-out

In Chapter 2, the independent validation set used for the external validation consisted in a farm whose data were acquired in 2019. To be consistent with this article the held-out external validation dataset (**hoevd**) for quality assessment of the production model will also consist in the recording related to this farm (and thus that single year of acquisition). Unless called the hoevd, all mention to a dataset will refer to the whole dataset without the hoevd.

4.5.2 Feature transformation

The inclusion of feature transformation was subject of many discussions, especially regarding potential dilution of the information. Therefore, a pre-analysis step is proposed. On the dataset covering the years 2018 and 2019 (to limit the number of records and thus the computational requirements), a tuning of the five most promising models recalled above will be performed using a 10-fold stratified CV based on the farm and on the date of acquisition. It should be performed twice, once on a dataset with and once without the feature transformation. Ideally, the performances of the different folds will form two-population showing a high skewness towards a low RMSE or a high R^2. Keeping the RMSE and applying a logarithmic transformation on these values should decrease their skewness. Therefore, a Student t-test could be applied to check whether the populations are different. If they are not, then there is no-need to transform the features. If they are and the transformed features show lower quality through its metrics, then there is also no point of keeping the feature transformation step. If they are and the transformed features show higher quality of the modelling through the associated metrics, it is worth the challenge and compute cost.

4.5.3 Tuning

To ease the automation of the search of optimal model hyper-parameters, the GridSearchCV tool from sklearn [ch-053] should be used to ease the automation of the search. To ensure the adequation of the hyper parameters to every type of conditions, the partitioning strategy for the fold definition of the 10-fold CV should be a stratified sampling. Compared to a fully random 10-fold CV, this stratification strategy allows to widen the number of cases on which the model is trained and tested. Therefore, it should increase the resiliency of the choice of the values of the hyper-parameters The performance metric to optimise should be the root mean square error (RMSE). Regarding the stratification feature, it should be a composite. Its composition and the inherent number of strata to sample requires a definition of the acceptance of missing values (NA) by the models. Considering

models that accept such NA and thus partial records (more details in section 4.7.2.1), combining the farms and the date of acquisition would lead to 110 strata, adding up the individual parcels would lead to 847 strata and including the S1 and S2 time-lag (dtS1 and dtS2), more details on these time-lag definition in section 4.7.1.1, would lead to 5811 strata to sample. To increase the resiliency of the models, they should be confronted to a maximum of conditions during training. Therefore, and given the drift with increasing time-lags seen for some models in Chapter 3, the combination of date, individual parcel (and thus farm), dtS1 and dtS2 should be considered. If the models don't handle well partial records, the meta-modelling approach developed in section 4.7.2.3 should be applied. It would lead to the three partitioning and strata consideration represented in Table 4-1. The conclusion is the same, the more strata considered and sampled, the more robust the choice of the hyper-parameters and thus this highest number of strata should be considered.

Table 4-1: Characterization of the stratification feature depending on the features considered.

	Stratification feature definition		
	Farm + acquisition date	Farm+ parcel + acquisition date	Farm + parcel + acquisition date + dtS1/dtS2
Full	66	487	1,919
Full without S1	66	487	706
Full without S2	110	847	2,339

The output of this tuning step would be a file with the optimal value for each hyper-parameter for each model.

4.5.4 Validation

4.5.4.1 Expected performances

Using the file with all the optimal hyper-parameters, these could be used to train and test the models using a fold independent CV to help assess the range of performances the final model should show once applied on an unmet-before set of conditions. Therefore, a fold-separation strategy needs to be defined. In this context, it is suggested to use one of the strata definition as a fold definition. To minimize run-time, the number of fold should be kept within reasonable values while offering an overview of the variability of the performances. The first solution would be to base the folds to be farm independent. This would imply to consider 7-folds. For a resiliency study, it might not be enough and instead considering the 70 parcels available in the dataset could lead to better information on the dispersion of the performances (e.g. standard deviation of the RMSE).

The output of this part of the modelling should be statistical descriptors (minimum, maximum, mean, standard deviation) and a histogram of the metric used to fit the models, i.e. the RMSE. There should also be a representation of the bias, the mean absolute error (MAE) and the coefficient of correlation (R^2). Regarding the use of the histograms, they should be checked to ensure the distribution is skewed towards better performance indication. Indeed, it would mean that most fits tend to be very similar with some deviation. These deviations would mean the inaptitude of the model to predict in certain conditions and could be due, in the particular context of this thesis, to poor conditions of acquisition.

4.5.4.2 Training and validation performances

Using the file with all the optimal hyper-parameters, these could be used to train the production model in python and in R. The reason for training the "production model" in both language is to ensure the translation of the hyper-parameter tuning was made properly by checking the similarity of the performances. Remaining in python would necessitate to re-code the Walloon prediction platform. Regarding the training modality, the dataset without the hoevd should be used for tuning and the hoevd serves as a basis to compute the actual metrics (RMSE, bias, MAE and R^2; the detail of these metrics is given in Table 4-2).

4.5.4.3 Unsupervised validation

During the realisation of this thesis, quality indicators of the models complementary to standard metrics were used to select the best trained models. The goal of this section is to go back through these indicators and their use, while completing the list. The first step is thus to re-define the indicators and provide associated measurements protocols (Table 4-2). To guide the model selection, a ranking can be associated per criterion, a lower value meaning a better ranking. Rankings for different criterion can be added up and the lower total sum indicates the best model.

Regarding the "unsupervised validation" mentioned in the section title, some criterion detailed in Table 4-2 are not directly related to a supervision of the training and validation process. Indeed, the sensitivity to the time-lag feature inclusion, the production of out-of-range values on the validation set, the temporal stability and the spatial heterogeneity are indicators not intended to be used directly on the training or the validation dataset. Instead, these indicators are intended to be evaluated on a larger database, ignoring reference measurement. Indeed, the point of these indicators is not to compare to actual values but instead to observe the behaviours of the prediction on a database. Therefore, as there is no comparison to an actual reference measurement, the "unsupervised" calling of the validation was chosen in this thesis, although it might trigger people given that it

relies on the prediction of a "Y-feature" and that it can be measured on the training and validation dataset altogether on larger databases (for which there is no reference measurement). This unsupervised approach of the model validation and selection is in fact relevant in other areas of research. For instance, applying such methodology on a milk-related mid-infrared spectra database allows to assess the relevancy of an equation to predict certain traits without having to perform loads of analysis as it might highlight points of failure. Therefore, focusing on these failure points might lead to hints to increase the robustness of the models.

Table 4-2: Summary of the criterion regarding the model selection, sd designates the standard deviation of the actual values, N the number of values, y the actual value, $\hat{y}$ the predicted value, and $\bar{y}$ the mean of the actual values.

Criterion	Explanation	Computation	Corresponding ranking
Root mean square error (RMSE) of validation	Metric indicating an expected error on a prediction	$RMSE = \sqrt{\frac{\sum(\hat{y} - y)^2}{N}}$	The lower the RMSE, the lower the ranking
Mean absolute error (MAE) of validation	Metric indicating an expected error on a prediction	$MAE = \frac{\sum\lvert\hat{y} - y\rvert}{N}$	The lower the MAE, the lower the ranking
R^2 of validation	Metric indicating the quality of fit of a prediction	$R^2 = \frac{\sum(\hat{y} - \bar{y})^2}{\sum(y - \bar{y})^2}$	The higher the R^2, the lower the ranking
Ratio of percent deviation (RPD) of validation	Ratio expressing the expected error compared to the actual variability of the dataset	$RPD = \frac{sd}{RMSE}$	The higher the RPD, the lower the ranking
Over-/under-estimation ratio	Ratio of the number of cases the prediction was over-/under-estimated, indicator not biased with extreme value prediction	$overestimation = \frac{N(\hat{y} > y)}{N\ total}$ $underestimation = \frac{N(\hat{y} < y)}{N\ total}$	The lower the over- and under-estimation ratio, the lower the ranking
Bias	Total deviation of the prediction	$Bias = \sum(\hat{y} - y)$	The lower the absolute value of the bias, the lower the ranking
Sensitivity to the time-lag inclusion	Ability of the model to run smoothly with the "backfill gap-filling methodology". A change in the distribution of the predictions with higher time lag values implies that the models are probably not reliable with the current gap-filling methodology.	Visual assessment of the boxplot of the prediction per category of time-lag	The more similar the boxplots, the lower the ranking
Extrapolation and prediction outside the training range	Ability to provide information in case the training range was a bit short compared to the amplitude that a feature could adopt	Sum of the number of values within a plausible range, out of the training range	The higher the sum, the lower the ranking

Production of out-of-range values	Occurrence of completely out-of-range prediction values (e.g. negative heights or sward heights >1 m)	Sum of the number of values out of logical bounds	The higher the number of out of range prediction, the higher the ranking
Temporal stability (regarding the metrics)	Aggregate the predictions per date (mean or standard deviation) and per land-use unit and see if the day-to-day variation is logical, e.g. the grass mown a certain day (and thus a global drop is seen) should take more than a week to be back at the pre-mowing level	Visual assessment	Appearance of artifacts imply a higher ranking
Temporal stability (regarding the spikes)	Aggregate the predictions per date (mean or standard deviation) and per land-use unit and see if there are spikes often appearing, that could hint a noise in the modelling	Visual assessment or computation of the number of inflexion points on a season	The less noisy the signal, the lower the ranking
Spatial heterogeneity	check if the representation of the predictions on known parcels makes sense and if the heterogeneity matches the known topographical properties, e.g. humid areas should appear different than drier ones.	Visual assessment	The better the spatial representation matches the known topographical properties, the lower the ranking

4.6 Data treatment

Linking the data presented in section 4.4 to the modelled feature detailed in section 4.3 through the machine learning models detailed in section 4.5 requires an adequation of the data format to the requirements of the models and thus some kind of treatments. The treatments of data throughout the model creation, the platform implementation and the DSS drafting were chosen at the time of realisation as they were assessed as the best solution at that moment. Taking a step back led to a reflexion on the choices that are detailed in this section.

4.6.1 CSH data

During the model creation, instead of the median per pixel used in Chapter 2, I would now rather use all the CSH values and the corresponding spatial data. First off, it is easier to work with points. The second main reason is that some models work in a decision-tree fashion. Therefore, more information could be found by keeping this variability. However, this assertion has to be mitigated by the fact that models tend to go for the mean value of the observed data when there are facing a similar dataset in prediction. Thus, it would expose the predictive capability to higher bias due the sensitivity of the mean to extreme values. The negativity of this impact of extreme values is alleviated by the overview of the performances of the most promising models once applied at a large scale: more than 75% of the predictions of CSH across all years were grouped below 75 mm (Chapter 3). This either reflects a very similar management across the whole region or a possible bias in the training datasets. Furthermore, increasing the variability and the number of configurations to which the models are confronted during training could help them improve their performances when confronted to extreme cases, hence section 4.7 that focuses on data augmentation through gap-filling.

4.6.2 Parcel definition

During the first iteration of the handling of spatial data, grids were drawn over each parcel and parcels were treated independently for the data extraction [ch-000]. It was rather inefficient concerning the access time to the files. Then, the second iteration (Chapter 3) included a way to handle the spatial data at the whole regional scale: the anonymous repartition of grassland parcels (PAA) was used for the year 2018 and transformed into a raster with 10 m resolution that became the reference for spatial data projection. All the area not marked as pastures were removed.

In the meantime, in the annex of the Chapter 3, 89% of the parcels could be called permanent pastures as these areas remained in the grassland definition for at least five years in a row during the period studied. This means that we probably miss 10% of the actual pastures by ignoring the non-grassland area of this anonymous grassland parcel repartition. Whether it is a problem or not is hard to tell: to provide data to the end user' application is the primarily goal of this platform. Therefore, not predicting over these area means a failure of the objective. However, given the high amount of storage needed per day of prediction, the addition of these areas should be taken with caution as any applied research in computer science should always keep in mind the need for sobriety.

To provide more insights on the storage requirements, for the most complete days, the file including the parcel and pixel identifiers, the corrected S1 and S2 and meteorological data and the predictions for the 5 most promising types of models of [ch-000] and trained again with more complete datasets weighs 14.6 GB. For future use, one might wonder why keep all the raw values and all the predictions. During discussions with a farmer that got out of bioengineering studies, it appeared that some want to mess themselves with the data. Therefore, the goal is to provide the end-user with a DSS that offers a high level of access to the data in order to match this wish of making their own equations/models.

Another prospect related to the parcel definition over the whole PAA is the futureproofing and inter-usability of the platform. Indeed, mapping the whole PAA with a rasterization and filling with data in the past could help scientist: it would provide an easily usable platform to retrieve the spatial data and use it as a base for future modelling prospects. This prospect remains open to discussion as there already exist platforms developed by tech giant like the Google earth engine, or by universities like Belcam [ch-057]. The problem I have with these two alternatives is that we do not have access to the source code, so we do not know what they are using to perform the computation and whether they are being done the right way for our application. Furthermore, especially regarding Google's tool, what would happen if they decided to have users pay for their download to pay off their servers. This would cripple the net benefit of going with freely available data. Another problem with this approach of using intermediary data providers is the definition of gap filling strategies. It is either hard or we have to download more data and post-process the data to have a chosen gap-filling like the one implemented in this thesis. Therefore, it loses its potential interest as compute resources consumption mitigator.

As mentioned in the chapter related to the platform creation, the resolution of 10 m might seem coarse especially regarding precision grazing, as underlined in

Tiscornia *et al.* (2019) [ch-014] and compared to the 2 m² used for the simulations in Ruelle *et al.* (2018) [ch-056]. To go with a finer resolution would require: 1) more computation power and storage, the current reference grid on which all the data are resampled has 15,413 rows, 26,006 columns and 400,830,478 cells and divide the resolution by 4 to have a resolution of 2.5 m would multiply the number of pixels by 16. 2) Go with a finer resolution offers a better spatialisation as we see finer combinations of S1 and S2 data. However, what would be the point of going that way with data that have a coarser resolution, for example the finer resolution of S2 being 10 m? Refining the global resolution would also require refining the input data resolution as Latte and Lejeune (2020) [ch-015] did by super-sampling of S2 tiles using data coming from another constellations of satellites.

Another prospect related to the complexity of the subject of the refinement of the resolution is inherent to coordinate reference systems (CRS). Three were used: EPSG 4326, EPSG 31370 and EPSG 32631. There exist translation equations from one to another CRS. However, it is based on statistical models and there is an error related to these models. This uncertainty is around 2 m for the translation between EPSG 31370 and EPSG 4326, as illustrated with the warning provided by QGis when manipulating layers in both CRS. This was confirmed in a personal communication with Philippe Lejeune in September 2021.

This error forms a complementary layer of uncertainty to data that have inherent uncertainty, especially regarding the positioning through GPS signals. Therefore, resolution refinement does not seem to be a priority and is a long-term perspective.

4.6.3 S1 data

4.6.3.1 Geocoding

The pre-processing of S1 products has witnessed only one change in the standardised workflow coming from Filipponi (2019) [ch-012] and Filipponi (2020) [ch-013] in the origin of the digital elevation model (DEM): we switched from the 3 seconds Shuttle Radar Topography mission data -SRTM 3sec- to a one second one -SRTM 1Sec HGT. The main reasons were that it should be more accurate given the better spatial resolution and it was not possible to access those data during the platform elaboration. To go further in the amelioration of the spatial resolution of the DEM, there is the possibility to use the LiDAR Walloon DEM [ch-031] that has a 1 m resolution. This path was explored to a certain point and the challenges encountered mainly relate to size and relevancy of the information.

The first challenge was to join all the DEM as they were heavy (the Walloon Brabant tile was around 3 GB whilst the 4 others weighed more than 12 GB). Once they were all joined and compared to the PAA, a problem appeared: some parcels

were out of the Walloon borders to which the DEM is limited. Therefore, there was a need for extending the extent of the merged raster. Furthermore, insufficient spatial coverage of the DEM compared to the S1 tile extent induced crashes in the SNAP software.

Hence, we faced a second challenge: upgrade the spatial extent of this fine DEM. At first, the SRTM 1Sec HGT was envisaged. Then Aster-GDEM was assessed as more fitted as it has a better resolution and is more recent. This DEM was downloaded from [ch-032], using a bounding box larger than Belgium to make sure that the area covered by the S1 tiles is in fact encompassed inside that DEM. Then this "basic" DEM is projected into another CRS (EPSG 31370) using QGis to have a resolution of 3 m (Figure 4-5). We did not go all the way to a 1 m resolution as we know we will have to project back to EPSG 4326 for the processing by the gpt command line of the SNAP software and as explained earlier there is an uncertainty about the spatial translation that makes irrelevant the will to use too fine resolutions. Projection to EPSG 31370 was performed to ease the replacement of the ASTER-GDEM values by the LiDAR ones.

The third challenge was related to the size of the data. The resampled and projected ASTER-GDEM weighted almost 64 GB and the raster merging operation largely outgrew the 32 GB of RAM of my workstation. Luckily, the on-disk work allowed to bypass this bottleneck. This, altogether with discussion with Cozmin Lucau (2022) led to the dismissal of the projected work (the remaining steps being detailed later). Indeed, the difference in type of signals emitted and received means that there could be inconsistencies between the layers. This was proven true as these layers are not similar and there are altitude differences between the two datasets, in the centre of the map and also on the edges. Given that we simply replace the centre, it could be only a minor problem as the reference would always be the same for all the operations and the relative changes are the focus of the modelling. However, the problem on the edge could cause extreme jumps in altitude values. Therefore, bias could be induced in the S1 correction and make irrelevant that merged product for the parcels located at the border of Wallonia although they were one of the primary concerns that led to the fusion of ASTER-GDEM and the Walloon LiDAR DEM. There exist ways to smooth out the transition that require more time to be implemented than what is left in this thesis. Indeed, we could imagine a convolution along the border and smooth the values with a simple mean or more complicated convolution windows. For this reason and the huge requirements on the system, this evolution path was discarded.

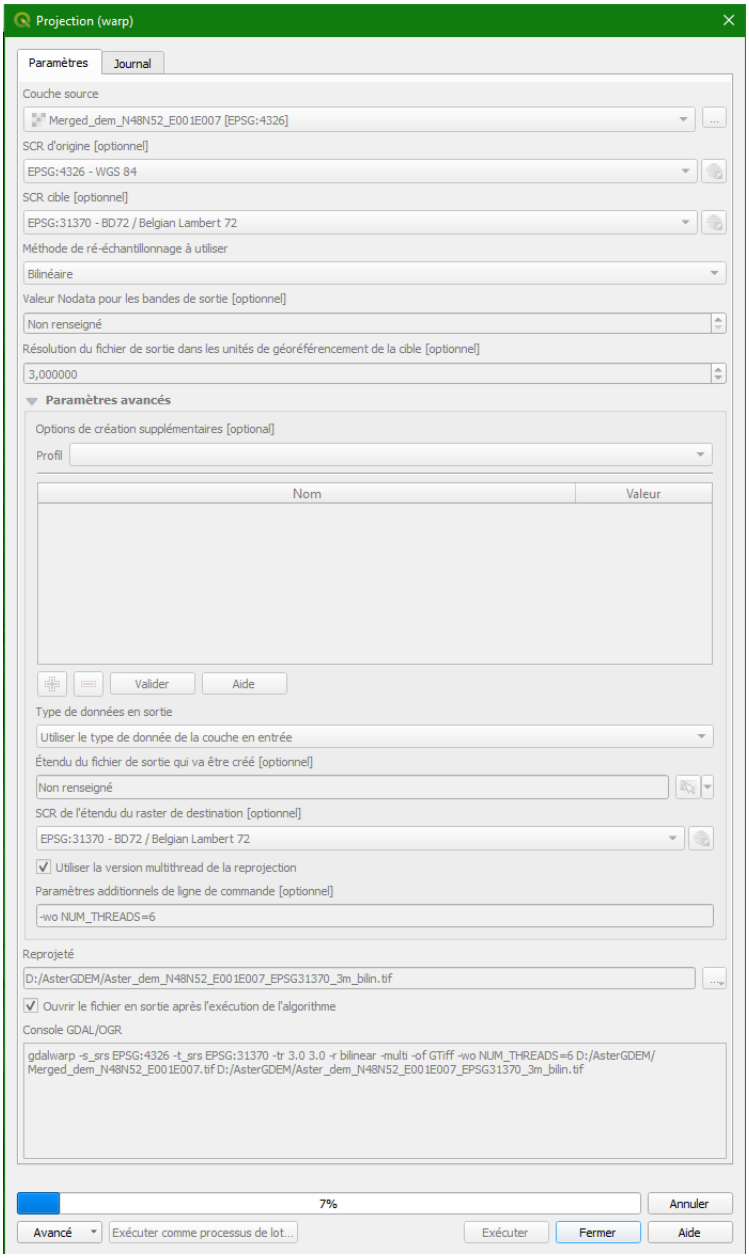


Figure 4-5: Parameters for the projection and resampling of the ASTER-GDEM to a 3 m resolution into the CRS EPSG 31370.

The rest of the theoretical path, notwithstanding the huge RAM and disk space consumption, should have been:

1) Merge all the Walloon DEM;
2) Resample this merged Walloon DEM according to the ASTER-GDEM, the aggregation function being the mean;
3) Create a 0/1 equivalent of the resampled layer where there is or isn't data in the resampled layer;
4) Multiply the ASTER-GDEM by this 0/1 layer;
5) Add the resampled Walloon raster layer to the multiplied ASTER-GDEM layer;
6) Project back to EPSG 4326 CRS;
7) Run the adapted SNAP gpt command line on any S1 Tile and monitor the memory (RAM) usage and the time taken.

4.6.3.2 Meteorological conditions

Beside the geocoding evolution, the precipitation effect inclusion might also be needed to further improve the performances of the models and reliability of the platform. As Tamm *et al.* (2016) [ch-010] and Zhao *et al.* (2021) [ch-011] suggest, there is an impact of precipitations on the wave retrieval. The fact that both in temperate grasslands [ch-010] and tropical conditions [ch-011] the impact is noticeable tends to mitigate the hypothesis formulated by a fellow scientist from Germany, met at the 29th European Grassland Federation general meeting of 2022: the impact of precipitations on remote sensing of grasslands might be due to the weight of the water laying down the grass.

4.6.3.3 Additional values

Beyond the conversion of the values already implemented in the pipeline, complementary information was assessed relevant in other studies and could be considered as perspectives of the databases enrichment. For instance, Garioud *et al.* (2021) [ch-005] integrated the "Ascending/Descending" characteristics of the orbit and the coherence products.

4.6.4 S2 data

The data provider for S2 was changed along the way, from the ESA Copernicus open access hub to the Theia hub (https://theia.cnes.fr). This translated in the change of some features transmitted as the tiles pre-treatment from L1C to L2A led to different bands in the final product. Hence, the bands entering the models were changed, all of them were included in the newly trained models.

During the elaboration of the platform, setting the restrictions on the proportion of clouds present in the tile to 25 % as in Nickmilder *et al.* (2021) [ch-000] was

rough and excluded too much data. Therefore, we changed the filter to exclude tiles that had more than 95% of their area covered with clouds or shadows and then used a stronger filter on the actual pixel extraction. It might be argued that this led to a higher permissiveness regarding the cloud cover. However, before arguing that way, one should keep in mind that Baetens *et al.* (2019) [ch-025] and Sanchez *et al.* (2020) [ch-026] showed that the cloud masking algorithm actually worked 7% better with the MAJA algorithm compared to the Sen2Cor one, used in backend for providing L2A products respectively on the Theia hub and the Copernicus one.

Synthetic indices were computed to highlight combinations of bands that could be related to plant characteristics (e.g. the NDVI is related to the greenness of the plant). All the formulas were retrieved from Henrich *et al.* (2023) [ch-043]. Related to the evolution of the models, instead of using these indices, convolutions across bands could be an alternative to take into account the interaction between the features. This evolution would fit perfectly in the context of the new NNet considered as perspectives.

4.6.5 Degree-day (DD) computation

Changes were made in the computation of the DD as expressed in the paper of the third chapter. Each way of computing deserves consideration as the first takes into account “standard temperature” and comes mainly from the construction conception and energy balance area and is used in vegetal physiology area by Calvache *et al.* (2020) [ch-002] while the second is mostly based on the physiological response of the plant to its environment. Indeed, the 0°C base temperature refers to the water freezing point and thus the water availability to the plant and thus the possible plant growth in winter and the 35°C approximated the temperature of the diminishing activity of the rubisco activase as expressed in Miller *et al.* (2001) [ch-003] and Anandhi (2016) [ch-004].

4.6.6 Joining the data

The join process starts at the level of the CSH recordings. As a quick reminder of an earlier statement, instead of using the median per grid pixel, now we spatially join the data using the position of the record point. To fasten the execution and to make sure that the datasets have the same characteristics between training, validation and production, the data were not directly sampled on the S1 Geocoded tiles nor on the S2 tiles. Instead, I took advantage of a temporary file created during the platform validation with the retrospective analysis. Indeed, all the S1 and S2 data were resampled according to the common grid. Therefore, the points were located on the grid to retrieve the specific pixel identifier that is used as a search

key in the S1 and S2 extracted file. Although it might be argued that it could decrease the accuracy of the data used as inputs of the models, it was assessed as the most relevant way, especially regarding the production context of this data extraction:

1) To deal with the amount of data to scrape. Indeed, for the years 2018 to 2021, it represents 540 S1 Tiles, each requiring half an hour, 80 GB of RAM and 16 CPU cores to be geocoded using the "SRTM 1Sec HGT" DEM, without accounting for the overhead of unzipping the files and the important number of read and write operations. The last point of time consumption is the loading and unloading of these datasets in RAM for a spatial search. Therefore, going with the already extracted files represent a huge time gain given that each extracted and geocoded file requires around 15 seconds to be read and 15 more seconds to retrieve the data of interest. From the S2 point of view, it translates into 1250 files to scrape with a lighter overhead as only the unzipping is needed before loading these datasets, applying the filters on shadows and clouds and extracting the data. Despite this lighter compute load, it does not match the speed of simply reading the already extracted files.
2) To deal intelligently with our hardware. The read and write time on the disk are a first limiting factor. The next one is the will to avoid performing redundant computation. Furthermore, the space available is finite and storing data in the extracted/resampled according to the grid form reduces the footprint. Indeed, before geocoding the S1 tiles weigh around 1 GB and after around 7.5 GB, while the resampled files weigh around 2 GB each. For S2, the compressed files are around 2 GB each and the resampled files could be heavier, up to 5 GB. This size increase is largely compensated by the speed of the data access and retrieval and the decreased strain on the storage support as it is only a read operation instead of a read/write/read sequence with zip files.
3) To standardise the data format between the training and validation to the one to which the models will be confronted in production.

Using exactly simultaneous acquisition leads to a dataset of 5,840 rows, representing less than 4% of the initial 155,099 CSH records. Beside the poor utilisation rate of the ground sampling, this exact co-occurrence of CSH sampling, S1 acquisition and S2 acquisition happened only at 3 dates on the whole period covered. As the DSS will rely only on S1 and S2 data (meteorological are ignored in the current demonstration, as they are always available), searching for exact co-occurrence of S1 and S2 and ignoring CSH sampling dates on the sampled area

leads to a dataset of 83,312 rows with 39 dates covered on the five years monitored. For a DSS, this temporal resolution is too low. Therefore, there is a need to perform data augmentation, for example using gap-filling.

4.7 Gap filling

The notion of gap-filling designates loads of different techniques and approaches. This section aims to provide insights into most prospects related to the object of this thesis: develop a DSS that would provide on a regular basis information to the farmers. The first notion to point out is the "gap". This designates values that are missing. This could be due to asynchronous acquisitions as illustrated in the section 4.6.6. Gaps are either a complete absence of values for a certain period of time or only partial information. If the information is partial, there are methods to complete it before or even during the modelling. If the values are completely missing, it requires to take a step back regarding the modelling paradigm chosen. The organisation of the methodologies is summarised in Figure 4-6.

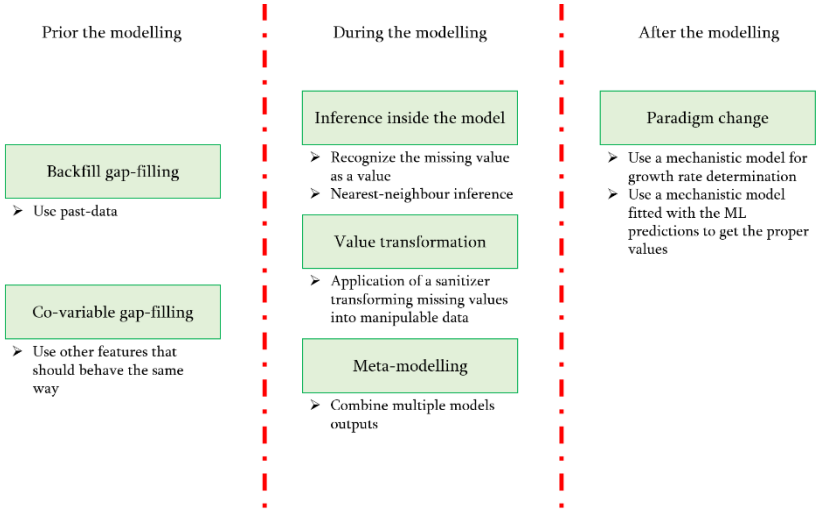


Figure 4-6: Summary of the gap-filling methodologies depending on their intervention location during the modelling process.

4.7.1 Prior the modelling

4.7.1.1 Backfill gap-filling

To moderate the temporal gaps, the closest values for each variable for each pixel in a time-window ranging from 5 days before to 5 days after the actual recording was used during the first iteration of the modelling of CSH growth (Chapter 2). The second iteration of the modelling process aimed at creating models that would be suitable for use in near real time. Therefore, there was an update in the joining algorithm, only closest data up to 4 days in the past were considered ("backfill gap-

filling"). During the evaluation of the performances of the platform (Chapter 3), this inclusion of past data implied a multiplication of almost 3 times of the number of dates covered compared to a join performed on exactly co-occurring data acquisition.

As illustrated in Chapter 3, there were serious instabilities in the prediction performances when predicting with some combinations of data far off in the past. To improve the reliability of the models when using data far off in the past, the algorithm to join data temporally should be improved following this methodology:

- For each record, the meteorological data corresponding to the closest meteorological station are added;
- For each record, a repetition of the record is performed for each S1 data available up to 7 days in the past (the 7th being excluded), including a NA in the range of data to add. This means that, if the S1 satellite were to fly over the area everyday of that period, that one record would be represented 8 times (7+1);
- For each of these new records, the same operation is performed with S2. In the same scenario, this means that for each of these new records there would be 8 new records and therefore, one original CSH acquisition could be matched with up to 64 combinations of S1 and S2 data, transforming one CSH acquisition into up to 64 records with different satellite combination.

To illustrate concretely the effects of this temporal gap filling, the 155,099 rows dataset described earlier went through the workflow and the resulting dataset was made of 2,481,158 rows. The descriptive statistics of this filled dataset are summed up in the Supplementary material (Table 4-8). Please note that due to the joining process, there has been duplication of some rows. The correct number of non-duplicated rows is 1,138,161. The descriptive statistics afore mentioned did not change much when removing the duplicated rows.

4.7.1.2 Alternatives

Another way to consider historical data and gap fill currently missing data could have been to use predictive models based on other features like Garioud *et al.* (2021) [ch-005] did with the prediction of NDVI using predictive models trained on S1 data. This reliance on another dataset did not seem that relevant during the algorithm update: the error propagation needs to consider both the error in the other dataset and its sensitivity towards changes in the grassland management.

4.7.2 During the modelling

Including partially NA in the prior-modelling gap-filling implies the need for a specific attention regarding the handling of the said NA by the algorithms. Some models will handle them gracefully whilst other will fail pitifully.

4.7.2.1 Inference inherent to the models

Extreme gradient boosting based algorithms and Cubist based models handle properly partial records. The Cubist case is very specific as it uses a nearest-neighbours algorithm to fill the NA. This option can be activated or not. If it is, it enables the inference of NA using the original dataset. However, this implies a storage of the training dataset and thus a consequent model size. The XGBoost ones learn the nature of NA by including them in the decision trees or by setting them to 0 depending on the type of algorithm. As we used trees, the NA are just learnt and not transformed. This translates in a lower storage cost for inference of NA.

4.7.2.2 Value transformation

A way to bypass the model NA understanding constraint is to use a "transformer"/ "sanitizer" model. This implies transforming the NA into an odd one that would be flagged as too extreme or that would not impact the modelling. The 0 attribution in the XGBoost algorithm detailed above corresponds to that kind of operation. Other values depend on the normal range: putting negative values as -9999 in only positive range could be a flag, especially for models working a decision tree fashion, or overflow the highest values could also constitute a flag.

This kind of negative sanitizer could be useful to improve the adequation of the data towards the algorithms requirements and is especially fit for models that take time to train, such as NNet.

4.7.2.3 Meta-modelling

The RF, the glmnet and the NNet do not handle well partial records. Beside sanitizing the data, there is another option to use data that are only partial: using multi-step modelling or meta-models (Figure 4-7). This implies training multiple sub-models that differ on the dataset on which they were trained. More precisely, in the context of this thesis, this would imply the creation of 3 sub-models per algorithm. The datasets would be: the full dataset (A) (including meteorological, S1 and S2 data, 487,999 rows) without any NA; a partial dataset (B) (including meteorological and S1 data, 628,076 rows) without any NA; a partial dataset (C) (including meteorological and S2 data, 657,555 rows) without any NA. Once each algorithm has been through the hyper-parameter tuning, range of performances assessment and the external validation for each dataset, this would result in three production models. Afterwards the meta-model is created using e.g. a decision-

tree fashion: are S1 data available? Are S2 data available? If the answer is yes-yes, then the model trained on the A subset is used, else if only S1 is available, the model trained on B is used and else if S2 is available, the model trained on the C subset is used. If neither S1 or S2 are available, then no prediction would be made, and post-modelling gap-filling is needed.

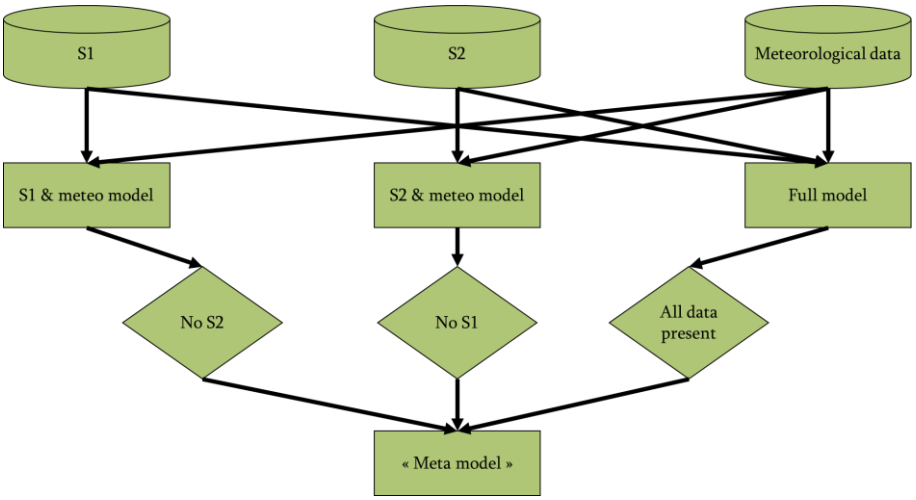


Figure 4-7: illustration of the meta-model concept

4.7.3 After the modelling

In some cases, despite all the efforts put into increasing the dataset size, sanitizing the data and meta-models usage, there could be pitfalls in the continuity of the predictions, with e.g. periods without S1 nor S2 acquisitions.

In this specific case where S1 and S2 data are both missing, the input features show little to no differences in adjacent farms, provided they are under the area of influence of the same meteorological station. Hence the prediction should not vary between nor within the parcels and thus diverge from the actual ground-truth. Therefore, to keep the possibility to deliver a correct information to the DSS and to enable the analysis of the feed cover variability, there is a need for a complementary method to come back at finely spatialised predictions. The first approach possible is to add the spatialised topographical and geopedological data mentioned in the complementary data sub-section (number 4.4.5) during the model training with the hope that the variability coverage is sufficient to enable specifically trained models to perform well once applied at the Walloon region

scale. In the meantime, in that same scenario of missing S1 and S2 data, the input features are in fact very similar to the "other side" of pasture modelling covering both areas of mechanistic and deterministic models. Consequently, it should be possible to go further than a make-do solution and reconcile both side of the feed availability estimation.

The first challenge is to find a model: suitable for the Walloon specific conditions, still under active development for a collaborative implementation and backed up by serious scientific resources. Based on the model paradigms and processes developed in Jouven *et al.* (2006) [ch-027], Prof Jérôme Bindelle and Essomandan Urbain Kokah are working on a deterministic model focusing on the assessment of the available feed based on meteorological, floristic, geopedological, pasture management data (detailed in Table 4-3) [ch-063].

Table 4-3: Input features of the considered mechanistic model for gap-filling, courtesy of J. Bindelle and E. U. Kokah , [ch-063]

Feature family	Examples of specific features
Meteorological data	Precipitations, potential evapotranspiration, Temperature, solar radiation (photosynthetically active radiation, PAR)
Floristic composition	Matrix of proportion of plants, including legume, grass, and other types of plants. Although in practice, only grass at the moment, with the 4 Groups of [i-105]
Geopedological data	Soil type: textural information (being developed at the time of writing)
Pasture management	Fertilisation, mowing events

This choice causes a second challenge to emerge: the need for corresponding outputs. As mentioned in the part 4.3 of this discussion section, a translation layer is needed to match the CSH output by the models developed in this thesis and the biomass output by this deterministic model.

As this model includes an historical prospective to account for the biomass evolution, a third challenge is the interfacing of the two models. Indeed, modelling in real time is not feasible given that it requires an overview of the past. Therefore, there are in fact two sub-challenges: run the deterministic model and use the results. As a result, multiple scenarios are considered and encompass both the hybridisation process and the DSS:

- A first scenario consists in running the models on the server for all the parameters combination encountered in Wallonia and then spatialise them. When the DSS asks for these data, they are sent over and the DSS fills the NA with these simulated biomass data. This presents the drawback of mixing data-sources and thus sources of error. Therefore, it could lead

breeders far astray of their actual forage allowance as the possible bias is not as systematic as it could be related to only one model.

- The second scenario also requires running the model on all the combination of parameters on the backend server. However, this is not the resulting biomass, output of the model, that presents an interest, it is the biomass growth/decrease rate once all the management and site-specific conditions are met. This rate could be applied on the last predicted data to get an idea of the current forage allowance status. The complexity of this method lies at the definition of the previous status.

These complementary modelling are not implemented in the platform nor in the DSS blueprints. These scenarios are written as perspectives and as a memo for the future roadmap.

4.8 Use beyond the scope of the thesis

All the work done to assess the available grass on Walloon pastures to feed a DSS to help farmers better manage their feed wedge could be redirected to other applications. Amidst the transfers, there are the change of the modelled feature, the change of the area the training dataset represents, the transfer of the prediction platform on another use-case, and the enlargement of the targeted market.

4.8.1 Models transfer

4.8.1.1 Transfer learning

As discussed earlier (section 4.3), the choice of the modelled feature is subject to discussions. If another choice was to be made, it would be interesting to decrease the dataset size needed to train the new models in order to decrease the inherent acquisition cost. A way to do this is transfer learning. It consists in using other data, that should be related to the one chosen, to prepare the models. The ways to do it depend on the type of model:

- For RF, the principle can be the addition of new trees, although it requires the same type of distribution and range of predicted values. This type of transfer learning relies on the warm-start ability of the model;
- For equations, the initial values of the coefficient can be set to the optimal values of the previously tuned production model and then apply the optimisation algorithm. This type of transfer learning relies on the warm-start ability of the model;
- For NNet, beyond using the warm-start ability, there is a complementary trick. Beheading the final conversion layer of a tuned network to replace it by a blank layer and then train the network with the new dataset should increase the convergence (represented in Figure 4-8, an evolution of the wide and deep and convolutional NNet proposed in Figure 4-4).

However, all models don't benefit from the warm-start ability, as it is incompatible with their structure, e.g. Cubist models. For these specific cases, meta-modelling could be a solution.

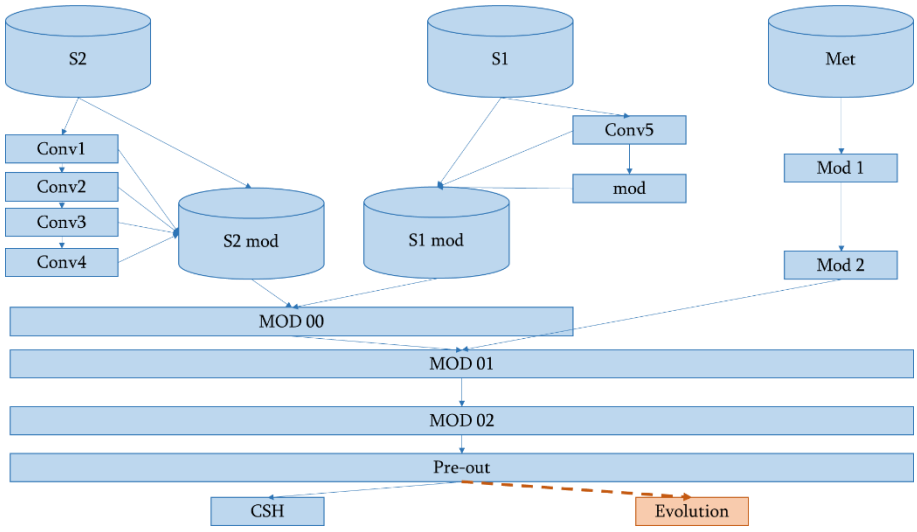


Figure 4-8: Example of wide & deep & convolutional neural networks thought as a perspective model to explore. MOD means a transition layer with a transfer function, conv designates a convolution layers, the pre-out layer could be a classical MOD and the output of the model is the CSH layer. The orange Evolution part represents the beheading process at the heart of transfer learning in NNet.

4.8.1.2 Multi-scale modelling

Another prospect related to using models in different context is the use of pre-trained models at a global scale to train and refine models on more specific local conditions. For instance, the models trained here account for different areas. If a farmer was to want models more specific to its own parcels, he could use the warm-start ability to re-train the models on data he would acquire on his own pastures.

4.8.1.3 Meta-modelling

Beside the use of the warm-start abilities and NNet beheading, meta-modelling is another way to transfer models to predict another feature that should behave the same way. In this context, meta-modelling still consists in training multiple models (or even meta-models) and then add a combination layer/model that would operate the translation of the previously modelled feature into the desired output. This is the case for the translation equation from CSH to biomass detailed in section 4.3.

4.8.2 Platform: SIMBA

The prediction platform was made to be modular and transposable to other server and use-cases. During the collaboration with the PDR SIMBA project,

funded by the national research fund (Brussels, Belgium) and the Luxembourg National Research Fund (Esch-sur-Alzette, Grand Duchy of Luxembourg), to detect grazing on pastures from Luxembourg, the transposability was assessed to be successful as it managed to deliver predictions, although it required to:

- Create a new reference raster for the area of interest;
- Redefine the S1 and S2 tiles of interest;
- Download and process those tiles.

The possible improvements description starts with the modularity of the platform, the use of "main" python scripts, to call R scripts in separated environments, reduces the number of times memory leaks were constated. However, to be more memory efficient and avoid having to retrain the model in R after all the selection, hyper-parameter tuning and estimation of performances process in python, it might be a good idea to translate the prediction script into python. This would not interfere with the rest of the pipe as the modularity enables to just swap the script called into the main python script to launch the computation. A prospect that might hinder this translation is the handling of big integers to keep the correct parcel identifiers. Regarding the automation, a crontab could be implemented to have the platform run every night at 01:00 am to update the database. Regarding the transposability of the platform, the scripts are on a Gitlab to be able to pull them at once. However, a perspective for transposability would be to package the scripts inside a Docker container. It would also guarantee that no general update of the server would break the scripts, as it happened once when the R version on the server was updated from R v3.6 to R v4.0.

4.8.3 Decision support system (DSS)

As cited in the introduction (Chapter 1) and in Chapter 2 and 3, a long-term objective of this thesis was to develop a decision support system (DSS) to help farmer manage their feed wedge. The development of the models in Chapter 2 and their implementation in the platform developed in Chapter 3 lead to an opportunity to focus on the DSS development. This opportunity became more concrete with the opportunity to present more precisely what we worked on during a popularisation event that occurred at GxABT at the end of May 2022 (Terra Innovation Day) and a second opportunity at "La Foire Agricole de Libramont" at the end of July 2022. As such DSS cannot be brought to the public without a proper comparison with the competing alternatives, a market definition was performed based on the work made by 3 master students (2nd year of master) studying in the agronomic section of bioengineering within the faculty ULiège-GxABT (Bughin Astrid, Lefèvre Théo and Pénasse Loïs).

4.8.3.1 Market definition

As stated earlier, this section is partly inspired from the work of Bughin Astrid, Lefèvre Théo and Pénasse Loïs. Their objective was to perform market research about a DSS related to the feed on pasture availability assessment. In this context, they created a survey. They sent it to every member of an association named "Fédération des jeunes agriculteurs" (**FJA**). This association aims to provide technical, logistical, and accounting support to its members. A total of 157 persons answered to the survey with various levels of completion. The respondent population had the following characteristics:

- ~50% of 155 respondent were actual **decision makers** in the farms;
- ~60% of 156 respondent were less than 25 **years old**, ~20% were between 25 and 30 years old, whilst the rest was more than 30 years old;
- All the respondents came from **Wallonia**;
- The main **activities** in the farms were dairy production, bovine meat production and cropping. It was mainly associated with secondary activities such as cropping and bovine meat production although some respondent did not have a secondary activity;
- Most respondents considered their farm to be small or medium **sized**;
- ~64% of 156 respondents were **interested in a DSS**, although only 7 respondents already used any form of DSS;
- 69 of 100 respondents emitted the will to be escorted by a **specialist**, and 74 of 101 respondents emitted the will to train on **tools**.

First and foremost, the term "decision support system" should be defined. [ch-064] suggest a fairly simple definition: "information system that provide decision relevant information and results". This definition is loose and, in the context of this thesis, implies the encompassment of products and techniques as basic as wandering across the pastures for a gross estimation altogether with highly technological solutions combining multiple data sources such as the DSS intended in this thesis.

The most basic solution - the tour inside the parcels - presents the disadvantage of being **time**-consuming and provide a gross estimation of the available feed. To refine the feed availability assessment without increasing the time spent on pastures, an alternative is to use a tool such as a RPM. Although it does not increase the time spent on pastures, one should account for one hour per 10 ha (depending on the topological conditions) and another major constraint: the need to repeat the measurement across the grazing season [ch-065]. The survey revealed that this time consumption was critical to respondents. Indeed, out of 101 responses, 53 were willing to allow less than 5 minutes a day to such DSS and 35 were willing

to give between 5 and 10 minutes a day to the use of a DSS. This means that there is a need for a rapid access to the relevant information.

This need for speed and **limited time-dedication** is also reflected by the high ratio of respondents (59/101) that wanted the DSS to use notifications to get warned when needed. However, the definition of "when needed" is subjective and the frequency that stood out in the survey –once a week (68/101 respondents)- illustrates the need to avoid overwhelming end-users with cumbersome loads of information.

To reduce the time the end-user should invest in the use of the DSS, a high level of **automation** is needed and as little computation as possible should be performed on the flight. This enforces our will to rationalise computation for CSH assessment by setting up a prediction platform that would act as a data provider for an end-user application, as requested by 58 out of 101 respondents.

This master students work also included an **analysis of existing DSS**, mostly based on Seuret al. (2014) [ch-066]. Additional DSS were analysed to highlight the strength and weaknesses of current DSS, with the principal highlights presented in Table 4-4 and meaning of the abbreviations in Table 4-5. Diving deeper in this analysis, one can highlight the important need to **enter data manually**. This means that the choice between the input data chosen for these DSS did not take into account the high level of automation mentioned earlier, although the use of remote sensing or meteorological data for some of these DSS could have led to such thought.

The second main highlight of this review of existing DSS is the **handling of the temporality** of the DSS. For simulation tools, it might seem trivial to consider if the output is continuous, but some DSS rely on models that do not provide continuously an output. Therefore, some DSS seem to have introduced ways to circumvent this inability to provide the end-user with an answer whenever he wants. However, this is far from the majority. In the case envisaged in this thesis, a "backfill" gap-filling methodology was included to increase the prediction frequency. Nonetheless, some periods witnessed meteorological conditions not fit for a prediction, hence the reflection about model hybridisation exposed in section 4.7.3. The other side of the temporality handling is the ability to provide indicators (and advices) for the future. Beside the interpretation of simulation DSS outputs, only a few DSS included future predictions. In their current state, the models underlying the targeted DSS are not able to provide future predictions. It would either require hybridising the models or include other types of algorithms such as the hidden Markov chains.

The third main highlight of this review of existing DSS is the modelling: most were based on dynamic models. Regarding the output on which it focused, it was mostly a simulation of the pasture management and the most used feed availability indicator was the biomass (mostly dried).

This review of existing DSS, altogether with the survey led to highlight points of attention regarding the conception of the DSS. Although, at the time of developing the first mock-ups, objectives were not expressed as clearly as of now, the need for speed and automation was already integrated.

Table 4-4: Highlights of the analysed DSS regarding the input data, the methodology, the output features, the level of user involvement, the fine spatial and temporal resolution, the intended public and the cost found. A"1" means that the DSS presents at least a partial response to the column. Other values in the cells are explained altogether with the abbreviations meaning in Table 4-5.

DSS	Input Data						Methodology			Outputs features					EM	SP	Temporal continuity			DE	Cost	
	MD	SD	HH	PM	GP	RS	McD	ML	DY	BM	HH	QG	SD	SM			TD	TP	TF		PV	PF
[ch-067]					1			1						1	N		1		1	E,C,S	NA	NA
[ch-068]	1			1				1		1					N		0	1		T	NA	NA
DIALOG				1	1									1	N					E,C	NA	NA
GMOT			1			1		1		1		1			N	1				T		
GRASSQ			1					1		1					N					T		
GRAZEIN							1													T		
GRAZEMORE	1	1		1	1		1			1					N		1			T		
Herb'Evol				1	1				1	1				1	N		1			C,E	NA	NA
Herb'ITCF, Herbo-LIS			1	1	1				1		1			1	N		1			C,E	NA	NA
Herb'Opti	1			1	1		1		1					1	N		1			C,E	NA	NA
Herb'sim	1	1		1	1				1	1				1	N		1			C,S	NA	NA
Herb'âge	1			1	1				1					1	N		1			C,E	NA	NA
Herb'Avenir			1	1	1				1						N		1			C,E	NA	NA
Herb'type				1	1				1			1		1	N		1			C,E	NA	NA
Pasture Base Ireland	1			1			1		1	1				1	A	1		1	1	T		
Pâtur'IN			1	1	1					1				1	N		1			S,C,E	NA	NA
Patur'Plan			1	1	1				1	1				1	N		1			C,E	NA	NA
PGSUS	1	1		1			1			1				1	N		1			T		
RIM					1		1		1					1	N			1		T		
SEPATOU				1	1		1					1			N	1			1	E,C,S	NA	NA

Table 4-5: Meaning of the abbreviations used in Table 4-4.

Abbreviation	Signification
DSS	Decision support system
MD	Meteorological data including temperature, precipitations, incident solar radiation, evapotranspiration, …
SD	Soil related data including topography, soil drainage, soil water content, …
HH	Any form of grass height, including the individual height of each plant, the CSH, the height averaged over a 3D reconstruction of the sward cover, …
PM	Any form of action linked to pasture management including fertilisation (NPK), grazing, mowing, …
GP	Grass properties like floristic composition, …
RS	Any remote sensing related data
McD	Dynamic mechanistic modelling
ML	Modelling based on probabilistic/Machine learning algorithms
DY	Dynamic modelling (and thus in near-real time)
BM	Biomass (often dry biomass) brought back to a certain reference area.
QG	Quality of the grass indicators such the protein content, ashes, fibres, available energy
SM	Modelling through simulation of the effects of pasture management
EM	Manual encoding: needed (N), lacking (A), or optional (O)
SP	Integration of the spatial prospect
TD	Sensitivity to external factors that could utterly block the outputting and thus provoke a temporal discontinuity in the prediction series
TP	Ability to provide an output temporal continuity, through the nature of the output or a combination of approaches
TF	Ability to predict the future or attempts were made to provide a prediction in the future
DE	Targeted audience: breeder (E), agricultural adviser (C), researcher/scientific personalities (S), public administrations (A), private company (such as an insurance) (A), all audience (T)
PV	Cost to use the DSS, often not found (NA)
PF	Payment frequency: annual (A), monthly (m), with use (U), free (G), other (O), not found (NA)

4.8.3.2 Hands-on approach

The development of a DSS aiming at providing Walloon farmers with assessments of the available feed is an ambitious project that requires to consider loads of perspectives. As the total workload might be overwhelming to face all the constraints and wishes of the stakeholders, an iterative approach was adopted: a mock-up was drafted then it was coded to be actually usable. Afterward it is used as a discussion starter to develop new mock-ups that will enter the loop again. The main iterations performed are described in supplementary material (1st draft in Supplementary material 4.10.4.1, 1st implementation in Supplementary material 4.10.4.2, 2nd draft in Supplementary material 4.10.4.3). The intermediary discussions occurred in July 2022 and involved actors from the scientific community and field advisors: Jérome Bindelle, Noémie Glesner, David Knoden, Essomandan Urbain Kokah, Charles Nickmilder, and Hélène Soyeurt. Out of these

round-table discussions, key points were highlighted that did not only relate to the DSS but also to the models and the platform developed in this thesis:

- The spatialisation of the parcels from the platform helps for parcels re-definition and management simulation. However, the 10-meter resolution could be too coarse for the fine representation of the spatial variability and repartition of the biomass for precision grazing;
- Depending on how deep users want do dive into the finer representation of the available feed, the final form under which the feed is expressed should either be the biomass for feed wedge management or sward height for variability management. This is illustrated through the expression "a cow eats biomass but sees a sward height". As work had been done concerning the conversion between the 2 metrics of feed, a conversion layer should be implemented. For the time being, it is a simple multiplicator. According to recent discussions with Françoise Lessire and Isabelle Dufrasne, research are being made in old databases in order to get better conversion equations or even switch the model type that converts CSH into biomass.
- Concerning the gaps due to poor satellite data acquisition conditions, a multi-model approach was suggested as it could retaliate against the gaps. The model that was thought of is the mechanistic models developed by the Urbain Kokah and Jérome Bindelle. This implementation would require a lot of collaboration and work on parameters definition (more details in section 4.7.3).
- An overview of the parcel history is a challenge in terms of storage management but could prove useful for long-term planning and land sell.

These hints pave the way towards the creation of the actual decision support system.

4.8.3.3 Complementary thoughts and perspectives

Regarding the actual development of the DSS, the drafts were coded in pyQt6 [ch-070]. However, this is not fit for a commercial use (due to licensing problems) and the implementation on the WalleSmart platform requires to split the application into a backend that would send .json packaged files to the graphical user interface (GUI) part. The GUI is made out of Javascript derivatives. A person was recruited on the WalleSmart project (Killian Dichou) to take over the development while being compliant with the platform constraints.

Regarding the potential users of the DSS, it was chosen to work with farmers as the main audience. However, the Walloon agricultural panorama is larger.

Farmers' advisers might also be interested in a DSS, hence the inclusion of a two-level of involvement GUI. Beside people directly concerned by the individual farms, insurances and public authorities have also to be accounted for. Indeed, in case of damages to the pastures, the current assessment of the compensation needed is either costly due to the need of experts or very generic due to regional generalisation. Our work could be beneficial for farmers as it would automate the attribution of subsidies and lead to a work alleviation. Insurers would also benefit from this paradigm shift as they would be guaranteed to pay the fair price in cases of damages. As public authorities also act as insurers in case of damages to the crops, our tool could therefore be of public interest as it would guarantee a fair constraint on the public finances. This public administration interest was confirmed in a personal communication with Véronique Dewasmes.

Regarding the nature of the output, the first part of this thesis focused on assessing if it was possible to assess the available feed on Walloon pastures using remotely sensed and meteorological data. We chose the CSH proxy as it offered the possibility to rapidly increase the database size and take advantage of recent advances in ML modelling. However, this is not a direct indicator a farmer could use. Converting to biomass would already prove more useful to assess the relevancy of the feed wedge. Yet, there are complementary information that could prove useful for decision making:

- Integrating the quantity of hay that could be harvested through mowing and how it would affect the feed availability in the near future, although our equations focus on the prediction of grass height on grazed pastures and may thus not be fit for use on mown pastures;
- Including the value of the hay that could be harvested and comparing it to the value of the fuel burnt to move the tractors and to the time of the farmer;
- Including a suggestion for complementary diet input;
- Including a comparison between the risk for the production value and the actual cost. For instance, some dairy cows farmer tend to supplement all the time to make sure they don't lose too much productivity when putting their cows on pastures.

These perspectives pave the way for the DSS elaboration and gives hint on how large this research area is and how much work there is still to do, even at the level of the available feed modelling.

4.9 Chapter bibliography

[ch-000] Nickmilder, C., Tedde, A., Dufrasne, I., Lessire, F., Tychon, B., Curnel, Y., Bindelle, J. and Soyeurt, H., 2021. 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 Sensing*, *13*(3), p.408. https://doi.org/10.3390/rs13030408

[ch-001] CRA-W. 2022, Agromet. https://agromet.be/ (accessed 17/02/2023)

[ch-002] Calvache, I., Balocchi, O., Alonso, M., Keim, J.P. and López, I., 2020. Water-soluble carbohydrate recovery in pastures of perennial ryegrass (Lolium perenne l.) and pasture brome (bromus valdivianus phil.) under two defoliation frequencies determined by thermal time. *Agriculture*, *10*(11), p.563. https://doi.org/10.3390/agriculture10110563

[ch-003] Miller, P., Lanier, W. and Brandt, S., 2001. Using growing degree days to predict plant stages. *Ag/Extension Communications Coordinator, Communications Services, Montana State University-Bozeman, Bozeman, MO*, *59717*(406), pp.994-2721. URL https://www.researchgate.net/profile/J-Tarafdar/post/Is_there_a_database_containing_growth_stage_phenophase_by_accumulated_growing_degree_days_by_agricultural_crops_available_in_the_public_domain/attachment/5c57fa923843b0544e63ec8e/AS%3A722503433453568%401549269650262/download/mt200103ag.pdf (accessed 17/02/2023)

[ch-004] Anandhi, A., 2016. Growing degree days–Ecosystem indicator for changing diurnal temperatures and their impact on corn growth stages in Kansas. *Ecological Indicators*, *61*, pp.149-158. https://doi.org/https://doi.org/10.1016/j.ecolind.2015.08.023

[ch-005] Garioud, A., Valero, S., Giordano, S. and Mallet, C., 2021. Recurrent-based regression of Sentinel time series for continuous vegetation monitoring. *Remote Sensing of Environment*, *263*, p.112419. https://doi.org/10.1016/j.rse.2021.112419

[ch-006] European Space Agency, 2022, Copernicus open access hub. URL https://scihub.copernicus.eu/ (accessed 17/02/2023)

[ch-007] theia. 2023, Pôle Thématique Surfaces Continentales URL https://theia.cnes.fr/ (accessed 17/02/2023)

[ch-008] Müller-Wilm, U., Devignot, O. and Pessiot, L., 2016. *Sen2Cor Configuration and User Manual, Ref.* S2-PDGS-MPC-L2A-SUM-V2. 3. European Space Agency. URL http://step.esa.int/thirdparties/sen2cor/2.3.0/[L2A-SUM]%20S2-PDGS-MPC-L2A-SUM%20[2.3.0].pdf (accessed 17/02/2023)

[ch-009] Hagolle, O., Huc, M., Desjardins, C., Auer, S., Richter, R., 2017, Algorithm Theoretical Basis Document (ATBD). URL https://www.theia-land.fr/wp-content-theia/uploads/sites/2/2018/12/atbd_maja_071217.pdf (accessed 17/02/2023)

[ch-010] Tamm, T., Zalite, K., Voormansik, K. and Talgre, L., 2016. Relating Sentinel-1 interferometric coherence to mowing events on grasslands. *Remote Sensing*, *8*(10), p.802. https://doi.org/10.3390/rs8100802

[ch-011] Zhao, X.B., Shao, W.Z., Zhao, L.B., Gao, Y., Hu, Y.Y. and Yuan, X.Z., 2021. Impact of rain on wave retrieval from Sentinel-1 synthetic aperture radar images in tropical cyclones. *Advances in Space Research*, *67*(10), pp.3072-3086. https://doi.org/10.1016/j.asr.2021.01.050

[ch-012]: Filipponi, F., 2019. Sentinel-1 GRD preprocessing workflow. In *International Electronic Conference on Remote Sensing* (p. 11). MDPI. https://doi.org/10.3390/ECRS-3-06201

[ch-013]: Filipponi, F., 2020. Sentinel-1_GRD_Preprocessing_ Standard Workflow for the Preprocessing of Sentinel-1 GRD Satellite Data. *MDPI: Basel, Switzerland*.Switzerland, 2020

[ch-014] Tiscornia, G., Baethgen, W., Ruggia, A., Do Carmo, M. and Ceccato, P., 2019. Can we Monitor Height of Native Grasslands in Uruguay with Earth Observation?. *Remote Sensing*, *11*(15), p.1801. https://doi.org/10.3390/rs11151801

[ch-015] Latte, N. and Lejeune, P., 2020. PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks. *Remote Sensing*, *12*(15), p.2366. https://doi.org/10.3390/RS12152366

[ch-016] Ali, I., Barrett, B., Cawkwell, F., Green, S., Dwyer, E. and Neumann, M., 2017. Application of repeat-pass TerraSAR-X staring spotlight interferometric coherence to monitor pasture biophysical parameters: limitations and sensitivity analysis. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(7), pp.3225-3231. https://doi.org/10.1109/JSTARS.2017.2679761

[ch-017] Lussem, U., Bolten, A., Menne, J., Gnyp, M.L., Schellberg, J. and Bareth, G., 2019. Estimating biomass in temperate grassland with high resolution canopy surface models from UAV-based RGB images and vegetation indices. *Journal of Applied Remote Sensing*, *13*(3), pp.034525-034525. https://doi.org/10.1117/1.JRS.13.034525

[ch-018] Michez, A., Philippe, L., David, K., Sébastien, C., Christian, D. and Bindelle, J., 2020. Can low-cost unmanned aerial systems describe the forage quality heterogeneity? Insight from a timothy pasture case study in southern Belgium. *Remote Sensing*, *12*(10), p.1650. https://doi.org/10.3390/rs12101650

[ch-019] McSweeney, D., Delaby, L., O'Brien, B., Ferard, A., Byrne, N., McDonagh, J., Ivanov, S. and Coughlan, N.E., 2022. Dynamic algorithmic conversion of compressed sward height to dry matter yield by a rising plate meter. *Computers and Electronics in Agriculture*, *196*, p.106919. https://doi.org/10.1016/j.compag.2022.106919

[ch-020] Nakagami, K. and Itano, S., 2014. Improving pooled calibration of a rising-plate meter for estimating herbage mass over a season in cool-season grass pasture. *Grass and Forage Science*, *69*(4), pp.717-723. https://doi.org/10.1111/gfs.12070

[ch-021] Hakl, J., Hrevušová, Z., Hejcman, M. and Fuksa, P., 2012. The use of a rising plate meter to evaluate lucerne (Medicago sativa L.) height as an important agronomic trait enabling yield estimation. *Grass and Forage Science*, *67*(4), pp.589-596. https://doi.org/10.1111/j.1365-2494.2012.00886.x

[ch-022] Lefèvre, A., Glesner, N., Mertens, A., Lorant, N., Curnel, Y., and Battheu-Noirfalise, C., 2022, Une première calibration pour adapter l'herbomètre aux conditions wallonnes, Sillon Belge, URL https://www.sillonbelge.be/9133/article/2022-05-05/une-premiere-calibration-pour-adapter-lherbometre-aux-conditions-wallonnes (accessed 17/02/2023)

[ch-023] Delforge, L., Sevrin, L., 2022, Formation de gaz nitreux en ensilage : peu courant mais ça existe !, Pleinchamp 22 septembre 2022, n°38, p9, URL https://centredemichamps.be/wp-content/uploads/2022/04/La-formation-de-gaz-nitreux-en-ensilage-2.pdf (accessed 17/02/2023)

[ch-024] Houborg, R. and McCabe, M.F., 2018. Daily Retrieval of NDVI and LAI at 3 m Resolution via the Fusion of CubeSat, Landsat, and MODIS Data. *Remote Sensing*, *10*(6), p.890. https://doi.org/10.3390/rs10060890

[ch-025] Baetens, L., Desjardins, C. and Hagolle, O., 2019. Validation of copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sensing*, *11*(4), p.433. https://doi.org/10.3390/rs11040433

[ch-026] Sanchez, A.H., Picoli, M.C.A., Camara, G., Andrade, P.R., Chaves, M.E.D., Lechler, S., Soares, A.R., Marujo, R.F., Simões, R.E.O., Ferreira, K.R. and Queiroz, G.R., 2020. Comparison of Cloud cover detection algorithms on sentinel–2 images of the amazon tropical forest. *Remote Sensing*, *12*(8), p.1284. https://doi.org/10.3390/rs12081284

[ch-027] Jouven, M., Carrère, P. and Baumont, R., 2006. Model predicting dynamics of biomass, structure and digestibility of herbage in managed permanent pastures. 1. Model description. *Grass and forage science*, *61*(2), pp.112-124. https://doi.org/10.1111/j.1365-2494.2006.00515.x

[ch-028] Wille, M., Clauss, K., Valgur, M., Solvsteen, J., 2016, sentinelsat, URL https://github.com/sentinelsat/sentinelsat (accessed 17/02/2023)

[ch-029] Murphy, D.J., O'Brien, B., Askari, M.S., McCarthy, T., Magee, A., Burke, R. and Murphy, M.D., 2019. GrassQ-A holistic precision grass measurement and analysis system to optimize pasture based livestock production. In *2019 ASABE Annual International Meeting* (p. 1). American Society of Agricultural and Biological Engineers. doi:10.13031/aim.201900769

[ch-030] McSweeney, D., Coughlan, N.E., Cuthbert, R.N., Halton, P. and Ivanov, S., 2019. Micro-sonic sensor technology enables enhanced grass height measurement by a Rising Plate Meter. *Information Processing in Agriculture*, *6*(2), pp.279-284. https://doi.org/10.1016/j.inpa.2018.08.009

[ch-031] Service public de Wallonie, 2015, Relief de la Wallonie - Modèle Numérique de Terrain (MNT) 2013-2014. URL https://geoportail.wallonie.be/catalogue/6029e738-f828-438b-b10a-85e67f77af92.html (accessed 17/02/2023)

[ch-032] NASA, 2023, data retrieval tool, URL https://search.earthdata.nasa.gov/search/ (accessed 17/02/2023)

[ch-033] Nakagami, K., 2016. A method for approximate on-farm estimation of herbage mass by using two assessments per pasture. *Grass and Forage Science*, *71*(3), pp.490-496. https://doi.org/10.1111/gfs.12195

[ch-034] Allen, R.G., Pereira, L.S., Raes, D. and Smith, M., 1998. FAO Irrigation and drainage paper No. 56. *Rome: Food and Agriculture Organization of the United Nations*, *56*(97), p.e156. URL http://www.kimberly.uidaho.edu/water/fao56/fao56.pdf (accessed 17/02/2023)

[ch-035] ESA, 2023, MultiSpectral Instrument (MSI) Overview, https://sentinels.copernicus.eu/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument (accessed 17/02/2023)

[ch-036] Theia, 2022, THEIA's L2A product format, URL https://labo.obs-mip.fr/multitemp/sentinel-2/theias-sentinel-2-l2a-product-format/ (accessed 17/02/2023)

[ch-037] Aselin, P., Pedersen, M., Morenon, H., Chaves, A., 2022, Python package for fitting Quinlan's Cubist regression model URL https://github.com/pjaselin/Cubist (accessed 17/02/2023)

[ch-038] Quinlan, J.R., 1992, November. Learning with continuous classes. In *5th Australian joint conference on artificial intelligence* (Vol. 92, pp. 343-348). https://doi.org/10.1142/9789814536271

[ch-039] Rulequest, 2019, Cubist homepage, URL https://www.rulequest.com/cubist-unix.html (accessed 17/02/2023)

[ch-040] Kuhn M., Weston S., Keefer C., Coulter N., Quinlan R., Rulequest Research Pty Ltd., 2023, Cubist: Rule- And Instance-Based Regression Modeling , URL https://cran.r-project.org/web/packages/Cubist/ (accessed 17/02/2023)

[ch-041] Kuhn, M., Quinlann, R., 2022, Fit a Cubist model, URL https://topepo.github.io/Cubist/reference/cubist.default.html (accessed 17/02/2023)

[ch-042] Laca, E.A., Demment, M.W., Winckel, J. and Kie, J.G., 1989. Comparison of weight estimate and rising-plate meter methods to measure herbage mass of a mountain meadow. *Rangeland Ecology & Management/Journal of Range Management Archives*, *42*(1), pp.71-75. https://doi.org/10.2307/3899662

[ch-043] Henrich, V., Krauss, G., Götze, C., and Sandow, C., 2023, Index DataBase - A database for remote sensing indices URL https://www.indexdatabase.de/db/s-single.php?id=96 (accessed 17/02/2023)

[ch-044] Centre des Technologies Agronomiques, 2022, Centre des Technologies Agronomiques, URL https://www.cta-stree.be/ (accessed 17/02/2023)

[ch-045] Service public de Wallonie (SPW), 2023, Carte Numérique des Sols de Wallonie , URL https://geoportail.wallonie.be/catalogue/38c2a87e-d38a-4359-9899-9d4a6b9f0c2a.html (accessed 17/02/2023)

[ch-046] sklearn, 2023, Random Forest regressor, URL https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html (accessed 17/02/2023)

[ch-047] Breiman, L., Cutler, A., Liaw, A., Wiener, M., 2022, randomForest: Breiman and Cutler's Random Forests for Classification and Regression, URL https://cran.r-project.org/web/packages/randomForest/randomForest.pdf (accessed 17/02/2023)

[ch-048] Friedman, J., Hastie, T., Tibshirani, r., Narasimhan, B., Tay, K., Simon, N., Qian, J., Yang, J., 2022, glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models, URLhttps://cran.r-project.org/package=glmnet (accessed 17/02/2023)

[ch-049] sklearn, 2023, Pipeline, URL https://scikit-learn.org/stable/modules/generated/sklearn.pipeline.Pipeline.html (accessed 17/02/2023)

[ch-050] sklearn, 2023, Tweedie Regressor, URL https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.TweedieRegressor.html#sklearn.linear_model.TweedieRegressor (accessed 17/02/2023)

[ch-051] sklearn, 2023, Elastic Net, URL https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.ElasticNet.html (accessed 17/02/2023)

[ch-052] Stackexchange, 2016, Correspondence for glmnet model, URL https://stats.stackexchange.com/questions/203816/logistic-regression-scikit-learn-vs-glmnet (accessed 17/02/2023)

[ch-053] sklearn, 2023, GridSearchCV, URL https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html (accessed 17/02/2023)

[ch-054] ISRIC, 2022, World Soil Information, URL https://www.isric.org/explore/soilgrids (accessed 17/02/2023)

[ch-055] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, pp.2825-2830. URL https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html (accessed 17/02/2023)

[ch-056] Ruelle, E., Hennessy, D. and 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*, pp.80-91. https://doi.org/10.1016/j.eja.2018.06.010

[ch-057] Belcam, 2020, Belgian collaborative agriculture monitoring, URL https://www.belcam.info/# (accessed 17/02/2023)

[ch-058] xgboost developers, 2022, XGBoost Documentation, URL https://xgboost.readthedocs.io/en/stable/index.html (accessed 17/02/2023)

[ch-059] Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

[ch-060] xgboost R developpers, 2023, xgboost: Extreme Gradient Boosting, URL https://cran.r-project.org/web/packages/xgboost/ (accessed 17/02/2023)

[ch-061] xgboost developers, 2022, Scikit-Learn Wrapper interface for XGBoost, URL https://xgboost.readthedocs.io/en/stable/index.html (accessed 17/02/2023)

[ch-062] Michez, A., Lejeune, P., Bauwens, S., Herinaina, A.A.L., Blaise, Y., Castro Muñoz, E., Lebeau, F. and Bindelle, J., 2019. Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system. *Remote Sensing*, *11*(5), p.473. https://doi.org/10.3390/rs11050473

[ch-063] Kokah, E.U., Knoden, D., Dumont, B., Bindelle, J., Modeling the biomass production of grasslands of Wallonia according to their functional type, 2023, International Congress, [under review]

[ch-064] Power, D. 2003. What are the characteristics of a Decision Support System" http://dssresources.com/faq/pdf/13.pdf (accessed 17/02/2023)

[ch-065] Scohy, D., 2020, L'herbomètre : un outil indispensable dans la gestion du pâturage ? URL https://www.web-agri.fr/herbe/article/168848/l-herbometre-est-il-un-outil-indispensable-dans-la-gestion-du-paturage (accessed 17/02/2023)

[ch-066] Seuret, J.M., Theau, J.P., Pottier, E., Pelletier, P., Piquet, M. and Delaby, L., 2014. Des outils d'aide à la gestion du pâturage pour mieux valoriser les prairies et renforcer la confiance des éleveurs. *Fourrages*, *218*(218), pp.191-201. URL https://hal.science/hal-01210866/document (accessed 17/02/2023)

[ch-067] Bashari, H., Smith, C. and Bosch, O.J.H., 2008. Developing decision support tools for rangeland management by combining state and transition models and Bayesian belief networks. *Agricultural Systems*, *99*(1), pp.23-34. https://doi.org/10.1016/j.agsy.2008.09.003

[ch-068] Schulte, L.G., Perez, N.B., de Pinho, L.B. and Trentin, G., 2019, May. Decision support system for precision livestock: Machine learning-based prediction module for stocking rate adjustment. In *Proceedings of the XV Brazilian Symposium on Information Systems* (pp. 1-8). https://doi.org/10.1145/3330204.3330222

[ch-069] FourragesMieux, Livret alimpentation de la vache laitière: Aliments, calculs de ration, indicateurs d'évaluation des déséquilibres de la ration et pathologies d'origine nutritionnelle URL https://www.fourragesmieux.be/Documents_telechargeables/Cuvelier_C_&_Dufrasne_I_Livret_alimentation_des_VL_2_Aliments_et_calculs.pdf (accessed 17/02/2023)

[ch-070] Riverbank Computing Limited, 2022, Python binding for the Qt application framework, URL https://www.riverbankcomputing.com/software/pyqt/ (accessed 17/02/2023)

4.10Supplementary material

4.10.1Detailed year per year descriptive statistics of the CSH dataset

Year per year descriptive statistics of the updated CSH dataset with the additional acquisitions.

Table 4-6: Year per year descriptive statistics of the CSH dataset used

Year	N	Min.	Mean	Max.	Sd.	Sk.	Kurt.	Quantiles						
								0.01	0.05	0.25	0.5	0.75	0.95	0.99
2018	22,502	5.00	52.18	248.00	27.73	1.89	7.55	9.00	17.00	34.00	49.00	64.00	98.00	155.00
2019	47,422	5.00	61.90	248.00	29.55	1.47	3.15	19.21	27.00	41.00	55.00	75.00	120.95	162.00
2020	58,476	5.00	62.71	249.00	30.71	1.90	5.77	20.00	28.00	42.00	56.00	75.00	120.00	180.00
2021	23,035	5.00	68.92	249.00	34.18	1.73	4.42	21.00	30.00	46.00	62.00	82.00	136.00	196.00
2022	3,664	20.00	94.66	246.00	36.83	0.63	0.09	31.63	44.00	66.00	90.00	118.00	163.00	191.00

Min. stands for minimum, Max. for maximum, Sd. for standard deviation, Sk. for skewness, Kurt. for kurtosis.

4.10.2 Correspondence table between the hyper-parameters of the models in R and python

In order to train models in both R and python the hyper-parameters should be tuned the same way. Therefore, the correspondence between different implementation was done in Table 4-7.

Table 4-7: Correspondence table between the hyper-parameters of the models in R and python (at the time of writing, i.e. beginning of 2023).

Algorithm	Hyper-parameter name in python	Hyper-parameter default python value	Hyper-parameter python range	Hyper-parameter name in R	Hyper-parameter default R value	Hyper-parameter R range	Hyper-parameter meaning
Cubist	n_rules	500	integer		100	integer	Limit of the number of rules Cubist will build
	n_committees	0	Integer		1	Integer	Number of rule-based model, beyond 1 means a try to correct the prediction errors of the prior constructed model (recommended = 5)
	neighbors	None	Integer within [1,9]		0	Integer within [1,9]	Number of instances used to correct the rule-based prediction through nearest neighbours consideration
	unbiased	False	Boolean		False	Boolean	Inclusion of a correction for classes of values over-represented in the total dataset
	composite	False	True/False/ “auto”		True	True/False	Combine the rule-based model and the nearest-neighbour corrections
	extrapolation	0.05	Float		100	Float within [0,100]	Level of adjustment of the rule predictions to be consistent with the training dataset

	sample	None	Float		0	Float within [0,99.9]	Percentage of the data set to be randomly selected for model building
	cv	None	Integer	/	/	/	Number of fold for the cross-validation
	random_state	Random integer	Integer within [0,4095]		Random integer	Integer within [0,4095]	Set the random seed for the C code
RF	n-estimators	100	integer	ntree	500	integer	Number of trees in the forest
	Criterion	”squared_error”	“squared_error”, “absolute_error”, “friedman_mse”, “poisson”	/	mse/mean-squared error	/	function to measure the quality of a split
	Max depth	None	Integer	/	/	/	maximum depth of the tree
	Min_samples_split	2	Integer or float	/	/	/	minimum number of samples required to split an internal node
	Min_samples_leaf	1	Integer or float	nodesize	5 (regression) 1 (classification)	integer	minimum number of samples required to be at a leaf node
	Min_weight_fraction_leaf	0.0	Float	/	/	/	minimum weighted fraction of the total of input weights required to be at a leaf node
	Max_features	1.0 (=all)	Sqrt(p), log2(p), float]0;1.0]	mtry	p/3 (regression) sqrt(p) (classification)	integer	Number of features to consider when looking for the best split
	Max_leaf_nodes	None	Integer	maxnodes	NULL	Integer	Maximum number of terminal nodes
	Min_impurity_decrease	0.0	float	/	/	/	Criterion to allow a split based on the impurity decrease
	Bootstrap	True	Boolean	Replace	True	boolean	Bootstrap samples used when building trees

	Oob_score	False	Boolean	/	/	/	Use out-of-bag samples to estimate the generalization score
	ccp_alpha	0.0	Float >=0	/	/	/	Complexity parameter used for Minimal Cost-Complexity Pruning
	max_samples	None	Integer or float	sampsize	Nrow or 632*nrow	integer	If bootstrap is True, the number of samples to draw from X to train each base estimator
	Warm_start	False	Boolean	"grow()" function + keep.forest	/	/	reuse the solution of the previous call to fit and add more estimators to the ensemble
GLMnet	Tw: Power Tw: link	0 'auto'	Float [0;1:3] 'identity' or 'log'	Family	/	Built-in or stats::glm() created	Family defining the relationship of the link between x and y
	/	/	/	Weights	1	Float	Observations weight
	Tw: alpha	1.0	float	/	/	/	Inner L2 regression, should be set to 0.0 to handle penalisation only with elasticnet
	Enet: alpha	1.0	Float]0;inf[	lambda			Regularisation parameter
	Enet: l1_ratio	0.5	Float]0.01;1] (<0.01 requires to provide the whole sequence)	alpha	1		Mixing parameter alpha=1 is the lasso penalty, and alpha=0 the ridge penalty
	Enet: positive	False	Boolean	Lower.limits	-Inf	Float <=0	Forces the coefficients to be positive (Enet) R: lower limit of the coefficients
				upper.limits			R: upper limits for each coefficient

	Tw: solver	'lbfgs'	'lbfgs' or 'newton-cholesky'	/	/	/	Algorithm to solve the optimisation problem R: type.log implies a change
	Tw: max_iter Enet: max_iter	100 1000	Positive float Positive float	maxit	10^5	Integer	Maximum number of iteration to fit the algorithm
	Tw: tol Enet: tol	10^-4 10^-4	Float	Thresh	10^-7		Stopping criterion for the gradient descent
	Warm start (Enet & Tw)	False		"update()"			Reuse the solution of the previous call to fit and add more estimators
	Enet: fit_intercept	True	Boolean				Whether the intercept should be regularized. ! solvers might cause problems [ch-052]
XGBoost	n_estimatos		int	nrounds			Number of boosting iterations
	max_depth		int	max_depth		6	Max tree depth
	Objective			Objective	reg:squarederror		Specify the learning task and the corresponding learning objective
	learning_rate			eta		0<eta<1	Learning rate
	booster	gbtree		booster	gbtree		Type of "sub-model" to boost
	gamma			gamma			Minimum loss reduction to create new split/shrinkage parameter
	min_child_weight			min_child_weight			minimum sum of instance weight needed in a child.
	subsample			subsample	1	0<subsample<=1	subsample ratio of the training instance

	colsample_bytree			colsample_bytree	1		subsample ratio of columns when constructing each tree
	lambda			lambda	1		L2 regularization term on weights
	alpha			alpha	0		L1 regularization term on weights
	monotone_constraints			monotone_constraints		Vector of -1/0/1	Constraint of variable monotonicity
	interaction_constraints			interaction_constraints			Constraints for interaction representing permitted interactions.

Note: Tw means tweedie regressor related parameter

4.10.3Descriptive statistics of the completed dataset.

Descriptive statistics of the completed dataset with the backfill gap-filling (Table 4-8). CSH is described again to show the preservation of the distribution. Please note that the two significant digits in the table are not always meaningful as some devices don't have such a precision

Table 4-8: Descriptive statistics of the CSH dataset with the meteorological, S1 and S2 data added.

Feature	NaN	Min.	Mean	Max.	Sd.	Sk.	Kurt.	Quantiles						
								0.01	0.05	0.25	0.5	0.75	0.95	0.99
CSH	/	5.00	64.16	249.00	34.51	1.68	4.02	15.00	25.00	41.00	56.00	77.00	134.00	189.00
tsa_min	/	-0.90	10.93	21.30	4.64	0.04	-0.32	1.40	2.30	7.90	10.90	14.40	18.50	21.30
tsa	/	1.60	16.54	25.90	4.87	0.07	-0.70	6.60	8.80	12.80	16.30	20.00	24.70	25.90
tsa_max	/	5.20	22.05	32.60	5.61	-0.13	-0.56	10.00	12.50	18.60	21.70	25.80	30.60	32.60
ens	785,202	254.20	1860.50	3110.00	698.71	-0.55	-0.63	254.20	559.10	1358.00	2053.10	2499.50	2751.00	2942.60
plu	/	0.00	0.67	18.70	2.03	5.23	35.05	0.00	0.00	0.00	0.00	0.10	3.90	8.50
vvt	/	0.40	1.92	5.70	0.93	0.92	0.89	0.50	0.92	1.20	1.80	2.60	3.80	4.80
hra	/	16.10	51.97	96.00	22.74	-0.05	-1.24	16.30	18.20	29.40	56.60	69.50	87.40	93.70
etp	785,202	0.60	3.40	6.10	1.38	-0.32	-1.05	0.80	0.90	2.40	3.60	4.70	5.20	5.50
DJ00	/	2.15	16.49	26.95	4.85	0.13	-0.59	7.05	9.00	12.60	16.25	20.00	25.30	26.95
Cum.P.3	/	0.00	3.73	41.70	6.06	3.13	10.89	0.00	0.00	0.00	0.20	1.70	18.50	22.60
Cum.P.7	/	0.00	6.07	93.10	10.55	3.54	19.68	0.00	0.00	0.00	1.20	10.50	22.10	50.20
Cum.P.15	/	0.00	14.03	104.50	18.45	2.58	8.32	0.00	0.00	1.30	7.40	21.30	50.00	94.30
Cum.DJ00.3	/	7.30	49.00	80.40	14.62	-0.24	-0.57	20.10	21.40	39.10	51.20	60.60	71.20	80.40
Cum.DJ00.7	/	25.10	114.00	185.00	30.48	-0.27	-0.19	45.80	55.80	90.80	116.20	136.60	164.70	185.00
Cum.DJ00.15	/	79.70	235.80	355.60	60.89	-0.48	-0.47	84.10	115.90	190.30	249.9	278.60	318.60	343.70
Sigma0_VH_db	1,240,579	-26.70	-19.00	-6.90	1.85	-0.03	0.28	-23.48	-22.08	-20.20	-19.00	-17.80	-16.01	-14.64
Sigma0_VV_db	1,240,579	-20.30	-13.00	5.50	1.79	0.13	0.60	-17.25	-15.89	-14.20	-13.00	-11.80	-10.06	-8.55
dtS1	1,240,579	-6.00	-3.00	0.00	2.01	0.03	-1.30	-6.00	-6.00	-5.00	-3.00	-1.00	0.00	0.00
S2.WVC	1,405,818	0.40	1.70	3.50	0.70	0.21	-0.47	0.45	0.55	1.20	1.60	2.20	2.90	3.36
S2.AOT	1,405,818	0.00	0.20	0.50	0.12	0.60	-0.21	0.01	0.03	0.10	0.20	0.30	0.45	0.49

S2.B02	1,405,818	0.00	0.00	0.10	0.01	0.58	1.72	0.00	0.00	0.01	0.01	0.01	0.05	0.06
S2.B03	1,405,818	0.00	0.10	0.20	0.02	0.31	2.00	0.02	0.03	0.04	0.05	0.06	0.08	0.10
S2.B04	1,405,818	0.00	0.00	0.20	0.03	0.94	2.14	0.00	0.00	0.00	0.00	0.07	0.08	0.11
S2.B05	1,405,818	0.00	0.10	0.30	0.02	0.33	1.57	0.05	0.06	0.1	0.1	0.1	0.14	0.16
S2.B06	1,405,818	0.10	0.30	0.60	0.05	0.08	0.04	0.18	0.22	0.30	0.30	0.30	0.39	0.43
S2.B07	1,405,818	0.10	0.40	0.80	0.08	0.37	-0.25	0.22	0.26	0.30	0.40	0.75	0.53	0.58
S2.B08	1,405,818	0.10	0.40	0.90	0.09	0.29	-0.20	0.22	0.28	0.30	0.40	0.50	0.55	0.60
S2.B08a	1,405,818	0.10	0.40	0.90	0.08	0.30	-0.15	0.26	0.31	0.40	0.40	0.50	0.57	0.62
S2.B11	1,405,818	0.00	0.20	0.50	0.05	0.42	0.29	0.14	0.16	0.20	0.20	0.30	0.31	0.35
S2.B12	1,405,818	0.00	0.10	0.30	0.03	0.66	0.32	0.06	0.07	0.10	0.10	0.10	0.17	0.20
dtS2	1,405,818	-6	-3	0	2.00	-0.12	-1.23	-6	-6	-5	-3	-1	0	0

Descriptive statistics abbreviation meaning: NaN stands for the number of no value records, Min. for minimum, Max. for maximum, Sd. for standard deviation, Sk. for skewness, Kurt. for kurtosis. The feature abbreviation meaning is explained in Table 4-9

The meaning of the feature abbreviations of Table 4-8 are presented in Table 4-9. For data related to the S2 satellites, the sensor calibration got them as close as feasible yet there are variations, so an approximate value is given for the central wavelength, detail available at [ch-035]. The correction multipliers are due to the storage process [ch-036] into integers to decrease the size of the data.

Table 4-9: Meaning of the feature abbreviation of Table 4-8.

Feature	Meaning
tsa_min	Daily minimum air temperature (°C)
tsa	Daily mean air temperature (°C)
tsa_max	Daily maximum air temperature (°C)
ens	Daily total solar radiation (J/cm²)
plu	Daily total precipitation (mm)
vvt	Daily mean wind speed at 2 meters above soil-level (m/s)
hra	Daily mean air relative humidity (%)
etp	Daily total potential evapotranspiration calculated through the FAO's Penmann-Monteith equation (mm/day) [ch-034]
DJ00	Daily degree-days computed on a 0°C-35°C basis (°C)
Cum.P.3	Cumulated precipitations over the past 3 days (mm)
Cum.P.7	Cumulated precipitations over the past 7 days (mm)
Cum.P.15	Cumulated precipitations over the past 15 days (mm)
Cum.DJ00.3	Cumulated degree-days over the past 3 days (°C)
Cum.DJ00.7	Cumulated degree-days over the past 7 days (°C)
Cum.DJ00.15	Cumulated degree-days over the past 15 days (°C)
Sigma0_VH_db	Backscatter coefficient for "vertical transmit-horizontal receive polarisation" (VH) (dB)
Sigma0_VV_db	Backscatter coefficient for "vertical transmit-vertical receive polarisation" (VV) (dB)
dtS1	Elapsed time / time lag between the CSH recording and the S1 acquisition (days)
S2.WVC	Value of the atmospheric water vapour content (g/cm²), with the correction multiplier applied Based on the "9th Sentinel-2 band" centred on 945nm wavelength with a 20nm bandwidth
S2.AOT	Value of the atmospheric aerosol optical thickness (), with the correction multiplier applied Based on the "1st Sentinel-2 band" centred on a 442.5nm wavelength with a 20nm bandwidth
S2.B02	Optical reflectance centred on the 492.5nm wavelength with a 65nm bandwidth (), with the correction multiplier applied
S2.B03	Optical reflectance centred on the 559.5nm wavelength with a 35nm bandwidth (), with the correction multiplier applied
S2.B04	Optical reflectance centred on the 664.7nm wavelength with a 30nm bandwidth (), with the correction multiplier applied
S2.B05	Optical reflectance centred on the 704nm wavelength with a 15nm bandwidth (), with the correction multiplier applied
S2.B06	Optical reflectance centred on the 740.3nm wavelength with a 14nm bandwidth (), with the correction multiplier applied
S2.B07	Optical reflectance centred on the 781.5nm wavelength with a 19nm bandwidth (), with the correction multiplier applied
S2.B08	Optical reflectance centred on the 832.9nm wavelength with a 105nm bandwidth (), with the correction multiplier applied
S2.B08a	Optical reflectance centred on the 864.3nm wavelength with a 21nm bandwidth (), with the correction multiplier applied
S2.B11	Optical reflectance centred on the 1612nm wavelength with a 92nm bandwidth (), with the correction multiplier applied
S2.B12	Optical reflectance centred on the 2195nm wavelength with a 180nm bandwidth (), with the correction multiplier applied
dtS2	Elapsed time / time lag between the CSH recording and the S2 acquisition (days)

4.10.4 Iterations of the DSS development

4.10.4.1 Schematic representation of the first draft

The first draft of the DSS was conceived the simplest way possible to favour the adoption rate. The mock-up is represented in Figure 4-9. It was thought to provide two interfaces: one showing a spatial repartition of the biomass on the parcels and one detailed the feed wedge. The parcel representation aimed at offering a way to the farmer for checking easily the validity of prediction by comparing the prediction repartition to their knowledge of their parcels. The feed wedge part would have been the "heart" of the DSS as it implements the advice to the farmer about the need for bringing complementary feed to the cattle on pastures.

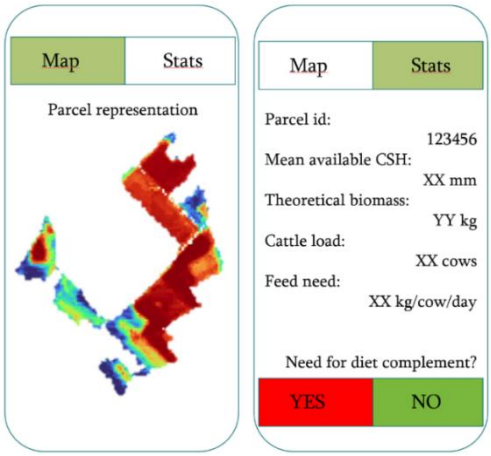


Figure 4-9: Mock-up of the first version of the decision support system

4.10.4.2 Actual implementation of the first iteration

The first draft of the application was developed in python v3.9.5 using PyQt6 [ch-070], a python binding to the Qt framework. Given that the sixth version was not implemented for long at the time of developing, there were bugs that required to also use the fifth version in a separate python environment for the parcel definition (the problem was located at the level of displaying the map on which the user input rely, that part had not been implemented yet). Another complexity faced during the development was the use of Spyder IDE that was also developed with PyQt. It led to some weird behaviours and crashes. Concerning the libraries, we used: os, matplotlib, sys, glob, re, numpy, pandas, json, rasterio, rasterio.mask, Fiona. Given the complexity of getting all the packages working properly and the need for understanding loads of thing to make it work, this application was packaged through pyinstaller to become a standalone ".exe" program able to work on any Windows machine without the need for handling the installation of any

python related environment/pieces of software. Relying in this packaging tool had one drawback: the geopandas package on which we rely had to be modified. Indeed, we had to suppress a style error: by default, it always charges the example data when loading and therefore relies on a side package that could not be hooked by pyinstaller. Therefore, we had to modify the main code to delete this automated import.

The mock-up was made mainly for show off and discussion triggering purpose. Therefore, it was assessed easier to create an app that could stand on its own. However, this also means that during the development, there will be a need to transit to a "web-app" compatible with the intended hosting platform (here WalleSmart, https://www.digitalwallonia.be/fr/publications/wallesmart/).

As suggested in the previous section, there was a will to create a GUI as **minimalist** as possible. This resulted in the creation of a 2 windows GUI (Figure 4-10), one for the sward height spatial repartition and the other for checking the actual numbers about the feed available and the corresponding available feed for the cattle. The key features were the date of interest definition (rolling menu), the set of parcels definition and, in the case of the feed wedge dedicated window, there was a rolling menu to refine the selection of the parcels based on the parcel identifiers from the PAA. Given the complexity of fitting everything in the restrained space of the windows without overwhelming the user with tons of parameters to encode, popup windows were implemented for the period of data retrieval on the server definition (Figure 4-11); the parcel drawing/definition (Figure 4-12); and the cattle need definition (Figure 4-13).

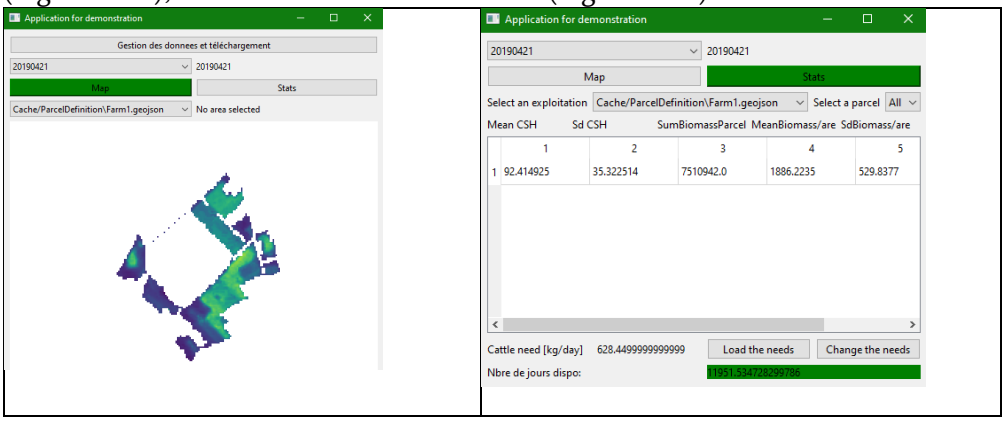


Figure 4-10: Overview of the mock-up of the two windows composing the first version of the GUI of the DSS. The left part relates to the spatial representation of the predictions and the right one to the definition of the constraints (cattle load) on the parcel to finally get a feed wedge.

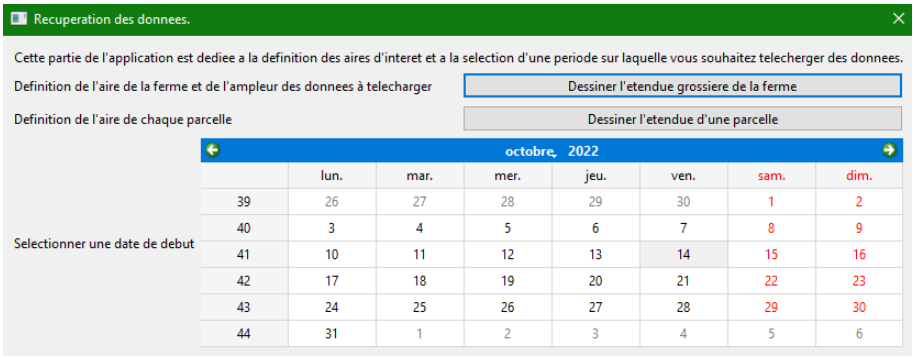


Figure 4-11: Pop-up window for the definition of the parameters on which the server data retrieval requests are based

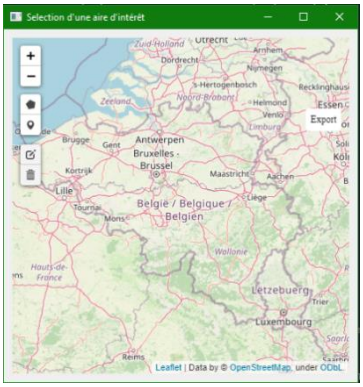


Figure 4-12: Pop-up window for parcel drawing/definition.

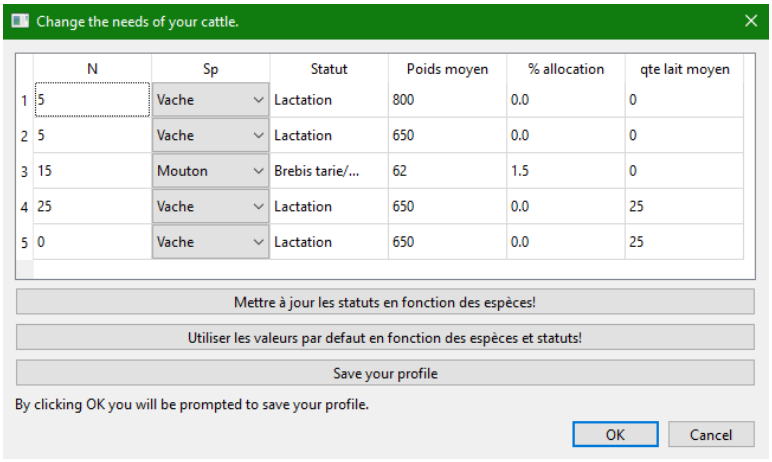


Figure 4-13: Pop-up window for the cattle need definition

Concerning these cattle needs, the algorithms implemented were based on the definition of the number of beasts per species per status and specific conditions (Figure 4-13). Despite the large dominancy of cows mentioned in the introduction chapter, cattle was enlarged to account for ovine. This scope widening was implemented to prepare the DSS concept to be exploited in non-planned scenarios. The equations and parameters (Table 4-10, inspired from [ch-069]) rely on an approximate animal approach and assume that properties of the grass like neutral/acid detergent fiber, protein content, ashes, etc. are quite similar across floristic composition. In case of multi-status grazing, the needs for each category are simply added, without any consideration for possible interactions between beasts of different species/status. Therefore, the shortcomings regarding the need computation and the grass composition were taken care of by implementing a modular approach using functions and datastores.

Table 4-10: Equations and parameters used to define the cattle need in the first mock-up of the GUI of the DSS

Species	Status	Equation	Default parameters
Cow	"A l'entretien"	$Need = N * 1.4 * \left(\frac{Weight}{100} + 2\right)$	Weight = 650 kg
	Lactating	$Need = N * \left(1.4 * \left(\frac{Weight}{100} + 2\right) + 0.3 * milk\right)$	Weight = 650 kg Milk = 25
	"Tarie en fin de gestation"	$Need = N * \left(1.4 * \left(\frac{Weight}{100} + 2\right) - 1.5\right)$	Weight = 650 kg
Sheep	"Brebis tarie/Adulte"	$Need = N * Weight * Allocat/100$	Weight = 62 kg Allocat = 1.5
	"Brebis 1/2 Gestation"		Weight = 65 kg Allocat =1.5
	"Brebis fin gestation"		Weight = 67 kg Allocat = 2.0
	"Brebis debut lactation"		Weight = 63 kg Allocat = 3.0
	"Brebis 1/2 lactation"		Weight = 62 kg Allocat = 2.5
	"Brebis fin lactation"		Weight = 62 kg Allocat = 2.0
	"Agneau allaitement"		Weight = 12 kg Allocat = 2.0
	"Agneau sevre"		Weight = 33 kg Allocat = 4.0

4.10.4.3 Schematic representation of the second draft

The second mock-up of the DSS relied on a two-fold approach: a minimalistic GUI dedicated to users needing only meaningful and summarised information e.g. casual farmers; a developed GUI including loads of parameters that could be tuned, and targeting users more prone to willing to tune their tool for their specific needs

e.g. agricultural technicians/advisers. The schematics for both side of the GUI running on the same "motor", i.e. ensemble of functions and ways of computations are represented in Figure 4-14.

It was chosen to display more information on the "simple interface" than before to give all the information a farmer would need in one go. A vertical structure was chosen so filling the constraints would be more intuitive. At the moment of re-designing the GUI, representing the spatial distribution of the CSH did not seem as relevant as giving a feed wedge for the simplest interface. The GUI includes the following key-features:

- a download button that would launch a pop-up window for the parcel definition and the period on which data should be downloaded. By default, it should use the last parcel valid definition and the last week of acquisition. Of course, this default behaviour is made modifiable by the user. This was implemented to reduce the time spent on defining the parameters.
- This version of the GUI relies on a two-level spatial definition: first the farm is to be defined and then the parcels of interest. This was to enable the use of this GUI by an agricultural technician that knows that all his parameters were correctly implemented and that would want a simple overview for one of the farms he is following.
- The choice of the parcel load/objective: number of animals and status. It also includes a mowing threshold definition.
- The big part of this GUI is the use of a feed wedge including a lower threshold limiting a no-go area in terms of forage allowance.

The "technician" GUI has an internal complexity represented by upper and inner tabulation/indexing. The upper one serves the purpose of better separating the prospective of feed and cattle management whilst inner ones might appear when one the said prospective is complex to summarize in a small window. The prospective judged relevant were:

- the spatial handling;
- the definition of the data wanted (only the prediction or also the pre-treated S1, S2 and meteorological data);
- the cattle load definition; the conversion from CSH to biomass with a function creation possibility;
- a management event recording and handling section (with a part related to mowing event impact on the available feed);
- a temporal representation of the feed on each pasture;

- the available feed and the resulting forage allowance, with subtleties as shown in the draft: the possibility to define the height to consider as not-usable for feed computation, variability analysis and concurrent/adjacent parcels behaviours.

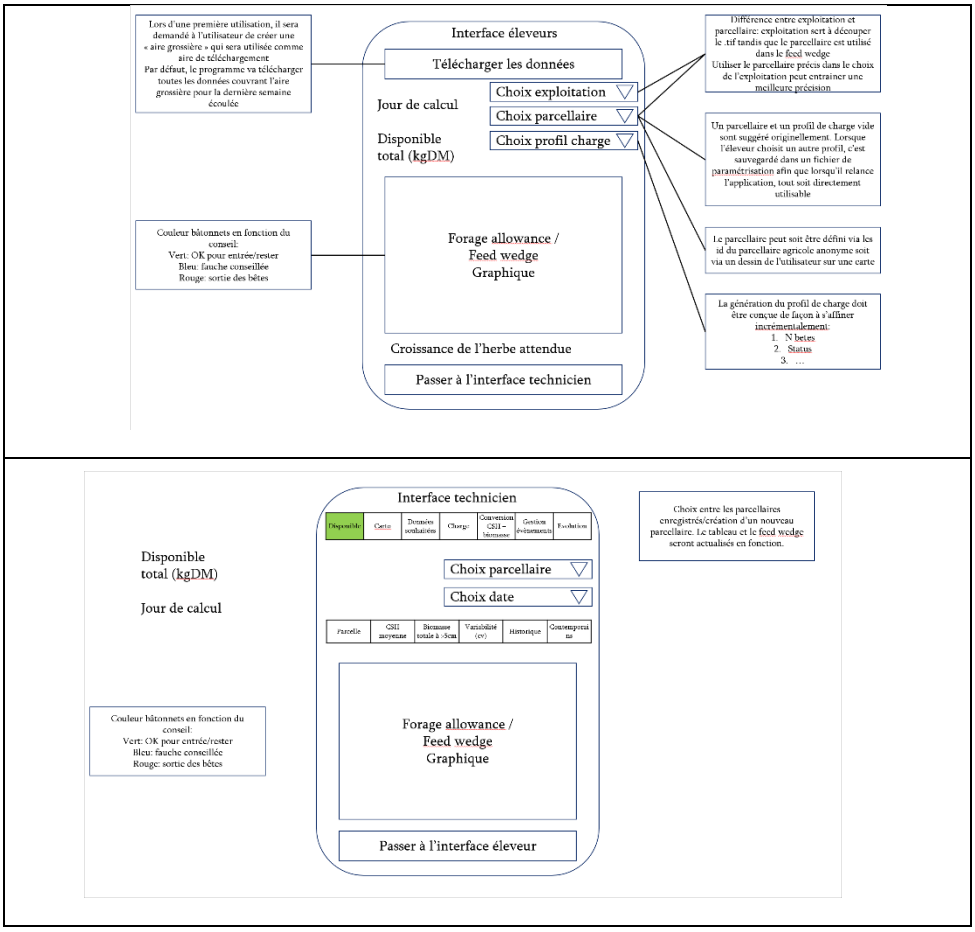


Figure 4-14 : Blueprints of the second version of the GUI of the DSS

Chapter 5

Conclusion and perspectives

5.1 Story

Dear reader,

As I said in the preamble this thesis was a demanding journey that revealed rich in emotions, encounters, and learnings. I enjoyed wandering it twice, once discovering pebbles and gravels on my own and a second time while relating you this tale. As you might want the answer to the question entitling this thesis, I will write it down in the next section. Afterwards, the main perspectives will be stated again. Then, I will give you my personal takeaways about this thesis. The redaction took some time but it was worth the effort as it helped me put words on feelings I developed during the thesis.

5.2 Answer to the question

After discussing the perspectives and points of amelioration of this thesis, it is time to give a conclusion to this work and summarize the answer to the research question stated in the introduction: "Is it possible to develop machine learning algorithms fed by meteorological and remotely sensed data to assess the available feed on pastures at the scale of Wallonia?" The short answer is yes. However, it is relevant to stress the beforehand steps that led to this answer:

- the choice of a metric of the available feed: the compressed sward height was chosen in this thesis as a proxy of the available feed, for its ease of acquisition, widespread scientific acceptance and ease of spatialisation;
- the choice of data sources. For cost, data quality and availability reasons, it was chosen to work with meteorological data coming from the Agromet project, C-band radar data acquired through the Sentinel-1 satellite constellation, and multispectral data acquired through the Sentinel-2 satellite constellation;
- the creation and selection of models. Multiple algorithms were tested to create models. Most converged towards the same type of error. Diverging algorithms were discarded, altogether with models requiring too much time and computation power to be trained. At this moment, the selected models are a cubist and a random forest with an error around 20 mm of CSH.
- the application of the models at the scale of interest. It required to shift the paradigm to obtain the prediction. Instead of predicting when we

need, it was more efficient, especially with a spatial approach, to predict every part of Wallonia and store the predictions. Therefore, the processing work of the DSS gets down to data scrapping and illustration construction. This larger scale application induced the introduction of the "unsupervised validation" designation.

- the creation of mock-ups of the DSS. At first, the mock-ups were made in python to ease their transferability. These drafts helped a lot with the discussions and the incremental construction of the bill of specifications the final DSS should fulfil;

Although the path is now better paved towards the creation of the DSS targeted in this thesis, there is still work to do. Indeed, the ML models developed until now have succeeded in providing feed assessments with a fine spatial resolution, an acceptable accuracy (especially considering the results averaged at the parcel level), and a rapid delivery of the information (also called inference). Yet, the synchronism with the S2 acquisition implies a sensibility towards the inherent S2 acquisition flaw: the meteorological conditions. Therefore, the ML models cannot be applied to in a way that would have pass the criterion of temporal regularity and availability of the information. To pass these criterions complementary modelling needs to be developed, using either hybridisation of models with the example of using a mechanistic model to fill the gaps, or using the "meta-models" that would involve building sub-models for specific conditions.

5.3 Perspectives summary

This perspective overview is related to the recall of the perspectives swept along Chapter 4:

- Regarding the relevancy of the modelled feature, multiple levels of perspectives were highlighted: 1) the choice of CSH was guided by practicality, in the short term, to increase the relevancy of this feature, a conversion to biomass can be implemented; 2) in the longer run, if models were to be created to predict other features using the same input data, the platform was conceived to be modular to ease the implementation of these models;
- Regarding the relevancy of the spatial representation, with the cost and computational constraints and the current options for data delivery, working at a 10 m resolution with S1 and S2 data was assessed to be the best compromise to provide a modular data source for the future DSS. Concerning the meteorological data, recent advances in the Agromet project involved a finer representation but the associated increase in computation cost relegated this for a mid-term evolution. Most remote sensing related evolutions could be implemented through an adaptation of the raster on which the platform relies to perform the spatial standardisation;
- Regarding complementary features implementation, in the mid-term it could be considered to implement a representation of the soils, whether using the CNSW or the SoilGrid database;
- Regarding the data treatments, in the short-term, the CSH value entering the models would not be the median per pixel anymore and rather each point would be associated to the combination of values;
- Regarding the backfilling, an evolution of the methodology for the training dataset elaboration is on its way to ensure a high enough representativity of each date in the past.
- Regarding the machine learning models, a whole part should be moved to python as 1) the transformation of spatial data to tabular data made in the platform avoids messing with the poorly implemented spatial manipulation tools of python, 2) it would avoid R's memory leaks and lack of efficiency;
- The training and validation strategy were stated in a clearer way for the short-term re-train of the models. Furthermore, the quality indicators

were stated more clearly to ensure their proper use for the next re-train of the models;

- To provide a better temporal continuity of the predictions, paradigm hybridisation and meta-modelling are considered as short-term evolutions to improve the databases and thus the data availability for the DSS.

Other perspectives of this work relate to the application of some parts in other context. Indeed, the prediction platform provides a data source that could be used for monitoring the general status of grasslands and thus serve for: an assessment of the evolution of the photosynthesis and the Carbon fluxes; the constitution of databases to improve the insurance to damages made on pastures; the detection of general management behaviours and categorisation of practices performances.

5.4 My takeaway

The main takeaway I get from this wandering throughout the arduous path that led from theoretical and field-remote science and knowledge to the tangible application can be resumed as follow: "As the definition of a model implies[2], whatever you do, it is never perfect. Therefore, instead of focusing again and again on improving your models, platforms, protocols and so on, what matters is to make things just good enough to have it working and articulate the different components of your work so that it might always evolve and improve itself latter."

The summary written above must be nuanced. Indeed, the work has to be made to include modularity. For instance, in the case of the platform, the models are not addressed by their names when they are called. Instead, all the models were located in a dedicated directory and all the models inside this directory are being called. The same way, multiple ways of joining data exist. Hence multiple scripts were written, and it is just a change in the call to the scripts that has to be done. The sub-script launched have been made to respect the same input and output dataflow. This modularity approach highlights two key-points I want to remember after this thesis: 1) any code related architecture cannot be done without a pencil and a sheet of paper. Drawing or even only writing the ideas before hands is vital to avoid messing up the codes. 2) code conception requires patience and brick by brick conception in order to get projects working.

As you might have guessed, this is only the first tome of this story. Hints were hidden throughout the paper about the second tome. You might have guessed that we intend to write it about the realisation of the decision support system and the integration of the collaborative/crowd-based science, the integration of multiple prospects like the near-real time pasture CSH assessment, the future production of the pastures, the translation from CSH to biomass…

[2] A model is an abstract representation of the reality