CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	National farm scale estimates of grass yield from satellite remote sensing
Authors	Marwaha, Richa
Publication date	2021-12-03
Original Citation	Marwaha, R. 2021. National farm scale estimates of grass yield from satellite remote sensing. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2021, Richa Marwaha. - https://creativecommons.org/licenses/by-nc-nd/4.0/
Download date	2026-11-15 17:38:56
Item downloaded from	https://hdl.handle.net/10468/12405

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork



NATIONAL FARM SCALE ESTIMATES OF GRASS YIELD FROM SATELLITE REMOTE SENSING

Thesis presented by

Richa Marwaha, B.Tech, M.Tech

for the degree of

Doctor of Philosophy

University College Cork

Department of Geography

Head of School/Department: Dr Kieran Hickey

Supervisors: Dr Fiona Cawkwell, Dr Stuart Green

2021







This research has been supported by the Teagasc Walsh Scholarship Program.

Contents

Contents ..... *i*
List of Tables ..... *v*
List of Figures ..... *vi*
Acronyms & Abbreviations ..... *ix*
Declaration ..... *xi*
Acknowledgements ..... *xii*
Abstract ..... *xiii*
Chapter 1 Introduction ***1***
1.1 General introduction *1*
1.2 Role of grasslands *7*
1.2.1 Food production *7*
1.2.2 Carbon sequestration *8*
1.2.3 Semi-natural habitats & biodiversity conservation *9*
1.2.4 Tourism *10*
1.3 Introduction to Irish grasslands *10*
1.3.1 Regional differences in farming enterprises *11*
1.3.2 Grass growth *13*
1.3.3 Factors affecting the grass growth rate *16*
1.4 Why measure grass? *16*
1.4.1 Fodder crises *17*
1.5 Grass measurement *19*
1.5.1 On-farm methods *19*
1.5.2 Biophysical simulation models *20*
1.5.3 Remote-sensing methods *21*
1.6 Rationale for this study *22*
1.7 Summary *23*
1.8 Chapter overview *25*
Chapter 2 Literature Review ***26***
2.1 Ground-based methods to estimate grass growth *27*
2.1.1 Non-remote sensing methods *27*
2.1.2 Proximal sensors *29*
2.2 Grass growth models *30*
2.2.1 Biophysical simulation models *30*
2.2.1.1 LINGRA *31*
2.2.1.2 GrazeGro *31*
2.2.1.3 APSIM *32*
2.2.1.4 MoSt GGM *33*
2.2.2 Biomass using radiative transfer models *33*
2.3 Biomass estimation using Earth observation data *34*
2.3.1 Unmanned aerial systems *35*
2.3.2 Synthetic Aperture RADAR *36*
2.3.3 Multi-sensor approach *37*
2.3.4 Optical sensors *38*

2.3.4.1 Statistical techniques ... *38*
2.3.4.2 Machine-learning models ... *39*
2.4 Conclusions ... *43*
Chapter 3 Study Area & Datasets ... ***44***
3.1 Introduction ... *44*
3.2 Geography & climate of Ireland ... *44*
3.3 Specific sites ... *45*
3.3.1 Moorepark Farm ... *45*
3.3.2 Teagasc farms ... *49*
3.3.3 Commercial Farms ... *51*
3.3.4 National Farms ... *54*
3.4 Datasets ... *56*
3.4.1 Landsat 8 data & processing ... *56*
3.4.2 Sentinel 2 data & processing ... *57*
3.4.3 Meteorological data ... *59*
3.4.4 PastureBase Ireland data ... *61*
Chapter 4 Determining the Influence of Growing Degree-Days on Grass Growth Rate in Ireland. ... ***69***
4.1 Introduction ... *69*
4.1.1 Need to monitor grass growth ... *69*
4.1.2 Factors affecting the grass growth rate ... *70*
4.1.3 Growing degree days ... *71*
4.2 Study area & datasets ... *76*
4.2.1 Study area ... *76*
4.2.2 Meteorological data ... *76*
4.2.2.1 Effect of SMD on GDD ... *79*
4.2.3 Validation data ... *83*
4.3 Methodology ... *83*
4.3.1 Modified GDD equation ... *83*
4.3.2 Statistical analysis ... *85*
4.4 Results & discussion ... *86*
4.4.1 Grass growth rate & GDD ... *86*
4.4.2 Correlation between meteorological variables & grass growth rate ... *89*
4.4.3 Ordinary least square regression (OLS) ... *93*
4.5 Discussion ... *97*
4.5.1 Effect of meteorological data on grass growth rate ... *97*
4.5.2 Regression models to predict grass growth rate ... *100*
4.6 Conclusions ... *101*
Chapter 5 Development of a Machine-learning Model for the Estimation of Grass Growth Rate in Ireland ... ***102***
5.1 Introduction ... *102*
5.1.1 Decision support system ... *102*
5.1.2 Biomass estimation using Earth observation ... *103*
5.1.3 Machine-learning for biomass estimation ... *105*
5.2 Materials & Methods ... *108*
5.2.1 Study area ... *108*
5.2.2 Datasets ... *108*

5.2.2.1 Vegetation indices *110*
5.2.2.2 Temporal matching of satellite data with ground data *111*
5.2.3 Brereton model *112*
5.2.4 Machine-learning algorithms *113*
5.2.4.1 Adaptive neuro-fuzzy inference system *113*
5.2.4.2 Random forest *114*
5.2.4.3 Feature selection *114*
5.2.4.4 Error metrics *115*
5.2.5 Methodology *117*
5.3 Results *119*
5.3.1 Brereton model *119*
5.3.1.1 Moorepark *119*
5.3.1.2 Johnstown Castle *122*
5.3.1.3 Athenry *124*
5.3.2 Machine-learning models *126*
5.3.2.1 Feature selection *126*
5.3.2.2 Models 1 & 2 (ANFIS Landsat 8) *130*
5.3.2.3 Models 3-6 (ANFIS Sentinel 2) *132*
5.3.2.4 Models 7 & 8 (RF Landsat 8) *135*
5.3.2.5 Models 9-12 (RF Sentinel 2) *137*
5.4 Discussion *141*
5.4.1 Limitations of ANFIS model *141*
5.4.2 Model performance *146*
5.4.3 Variable importance *149*
5.4.4 Sensor performance *150*
5.4.5 PBI as a source of ground truth data *150*
5.4.6 Grass growth curve *151*
5.4.7 Uncertainties & sources of error *154*
5.4.8 Applications & future developments *155*
5.5 Conclusion *156*
Chapter 6 A National Model for Biomass Estimation* *157
6.1 Introduction *157*
6.2 Materials & study area *159*
6.2.1 Study area *159*
6.2.2 Data acquisition and pre-processing *159*
6.2.2.1 Sentinel 2 ..... *159*
6.2.2.2 Meteorological data *161*
6.2.2.3 PastureBase data *161*
6.3 Random forest model development *161*
6.3.1 Feature selection *164*
6.3.2 Hyper-parameter tuning & data splitting *165*
6.3.3 Model evaluation *165*

6.4 Results *165*
6.4.1 Commercial dataset *165*
6.4.2 Agmodel- National dataset *167*
6.4.3 Agmodel 1-First part (January- June) *172*
6.4.4 Agmodel 2-Second part (July-Dec) *177*
6.5 Validation of final model 2020 data *181*
6.5.1 Agmodel- National model *181*
6.5.2 Agmodel 1- First part (January-June) *184*
6.5.3 Agmodel 2- Second part (July-December) *187*
6.6 Discussion *190*
6.6.1 National model limitations *191*
6.6.2 Model performance *192*
6.6.3 Variable importance *193*
6.6.4 PBI as a source of ground truth data *194*
6.6.5 Grass growth curve visually *195*
6.6.6 Variability in grass growth rate between individual paddocks *198*
6.6.7 Uncertainties & sources of model error *201*
6.6.8 Applications & future developments *202*
6.6.9 Applications & future developments *203*
6.7 Conclusions *204*
Chapter 7 Discussion ***205***
Chapter 8 Conclusion & Future Research ***213***
8.2 Future work *217*
8.2.1 Upcoming satellites *217*
8.2.2 Unmanned aerial vehicle (UAV) *219*
8.2.3 Proximal sensors *220*
8.2.4 Synthetic Aperture Radar *221*
8.2.5 Combination of biophysical models with machine-learning *222*
References ***225***
Appendices ***244***
5.1 Matching grass growth rate from PBI to image acquisition dates *244*
5.2 Script for Brereton model *245*
5.3 Script for ANFIS model *247*
5.4 Script for RF model *248*
5.5 RMSE for different values of 'k' for cross-validation *250*
5.6 Plot of OOB error with many trees with the 'mtry' tables *251*
5.7 Grass growth curve plots from model 10 (RF) (Chapter 5) *258*
6.1 Residuals of grass growth rate from PBI & model predictions (Chapter 6) *260*
6.2 Random Forest for national model *264*

List of Tables

Table 3-1 Grazing season length (in days) for Moorepark farm (2017 to 2020). *47*
Table 3-2 Soil type, elevation and bedrock type for Teagasc research farms. *49*
Table 3-3 Farm properties of the commercial farm cluster. *52*
Table 3-4 Soil type and bedrock for counties hosting the 179 farms used in Chapter 6 ..... *54*
Table 3-5 Landsat 8 OLI bands and their names, spatial resolution and wavelengths *56*
Table 3-6 Sentinel 2 bands and their names, spatial resolution and wavelengths *58*
Table 3-7 Specifications of Landsat 8 and Sentinel 2 satellites *59*
Table 4-1 Summary of the daily meteorological variables for Chapter 4. ..... *80*
Table 4-2 Correlations between grass growth and meteorological data. *90*
Table 4-3 Results for 2017 and 2018 showing R^2 and RMSE (kg DM $ha^{-1}day^{-1}$) *96*
Table 5-1 Input variables used for the ML regression models *109*
Table 5-2 Input variables used for the Brereton model *109*
Table 5-3 Number of cloud-free Landsat 8 and Sentinel 2 images (2017-2018) & available PBI data *111*
Table 5-4 Grass growth rate (kg DM $ha^{-1}day^{-1}$) output from the Brereton model (2017-2018). *120*
Table 5-5 Twelve models based on optimal variables from feature selection *129*
Table 5-6 Comparison of RF models 7 & 8. ..... *135*
Table 5-7 Error metrics for machine-learning models *140*
Table 5-8 Difference in actual & predicted growth for three farms (2017-2018). *146*
Table 6-1 Input variables from satellite data for the national data. *160*
Table 6-2 Accuracy metrics for the national test dataset *169*
Table 6-3 Accuracy metrics for test data for Agmodel 1 ..... *174*
Table 6-4 Accuracy metrics- R^2 for test data for Agmodel 2 ..... *179*
Table 6-5 Comparison of accuracy metrics for the national model *190*
Table 6-6 Farm summary for the 20 randomly selected farms from PBI for 2019. ..... *200*

List of Figures

Figure 1-1 Global distribution of pastures in the year 2000 ..... *4*
Figure 1-2 Pasture map of Europe in 2000 ..... *5*
Figure 1-3 The effect of improving grassland management on soil carbon storage. *9*
Figure 1-4 Number of the farm by category and region, 2016 (Source: CSO 2016). *12*
Figure 1-5 National grass growth curve with average growth rate values from PBI. *14*
Figure 1-6 Grass supply and demand curve (Snip, 2016). *15*
Figure 1-7 Poor fodder production in 2018 (Spatial Analysis Unit, Teagasc Ashtown). *18*
Figure 1-8 Thesis structure and chapter descriptions *24*
Figure 2-1 Ground-based methods to measure sward height *28*
Figure 2-2 An example of a sward stick. *28*
Figure 3-1 Moorepark farm in County Cork. *46*
Figure 3-2 Moorepark farm showing the location of 136 paddocks *47*
Figure 3-3 Monthly meteorological data for Moorepark farm (2010- 2020). *48*
Figure 3-4. Locations of Teagasc research farms with the outline of the farms. *50*
Figure 3-5 Locations of commercial farms in Counties Cork, Limerick & Tipperary *51*
Figure 3-6 Average monthly rainfall (in mm) for 9 commercial farms (2015-2020). *53*
Figure 3-7 Location of farms used in Chapter 6. ..... *55*
Figure 3-8 Irish meteorological stations. *60*
Figure 3-9 Grass wedge showing a scenario of the on-target wedge *62*
Figure 3-10 Grass wedge showing surplus condition on a farm. *63*
Figure 3-11 Grass wedge showing deficit condition on a farm. *64*
Figure 3-12 Grass wedge from PBI on 25th May 2020. ..... *65*
Figure 3-13 Screenshot of the information contained in PBI. *66*
Figure 3-14 A screenshot from PBI for Moorepark Farm. *67*
Figure 4-1 Average monthly accumulated GDD (1961-1990). *73*
Figure 4-2 Mean temperature for Moorepark Farm in 2017. ..... *77*
Figure 4-3 Mean temperature for Moorepark Farm in 2018. ..... *78*
Figure 4-4 Soil Moisture Deficit for 2017. ..... *81*
Figure 4-5 Soil Moisture Deficit for 2018. ..... *82*
Figure 4-6 Inter-annual comparison of SMD (mm) *83*
Figure 4-7 Grass growth rate for Moorepark farm (kg DM ha^{-1}day^{-1}). *86*
Figure 4-8 Comparison of the 7-day moving average of grass growth rate & GDD (2017). *87*

Figure 4-9 Comparison of the 7-day moving average of grass growth rate & GDD (2018). *88*
Figure 4-10 Grass growth rate vs. meteorological data (2017) *91*
Figure 4-11 Grass growth rate vs. meteorological data (2018) *92*
Figure 4-12 Actual vs. predicted grass growth rate at Moorepark (2017). *93*
Figure 4-13 Actual vs. predicted grass growth rate (June-Aug 2017). *94*
Figure 4-14 Actual vs. predicted grass growth rate at Moorepark (2017). *95*
Figure 4-15 Actual vs. predicted grass growth rate (Jun-Aug 2017). *96*
Figure 4-16 Mean monthly temperature at Moorepark Farm. *97*
Figure 4-17 Total monthly precipitation at Moorepark Farm. *98*
Figure 5-1 Architecture of Adaptive Neuro-Fuzzy Inference System (ANFIS). *113*
Figure 5-2 Methodology for Brereton, ANFIS and RF model *117*
Figure 5-3 Predicted grass growth at Moorepark using the Brereton model (2017-2018). *121*
Figure 5-4 Predicted growth at Johnstown Castle from Brereton model (2017-2018) *123*
Figure 5-5 Predicted grass growth at Athenry using Brereton model (2017-2018). *125*
Figure 5-6 Boruta plot for the Landsat 8 database. *126*
Figure 5-7 Landsat 8 VarImp output from the 'Caret' package *127*
Figure 5-8 Boruta plot for Sentinel 2 database. *128*
Figure 5-9 Sentinel 2 'Varimp' output plot showing the importance of all the variables. *128*
Figure 5-10 Actual vs. predicted grass growth for models 1-2. ..... *131*
Figure 5-11 Actual vs. predicted grass growth for models 3-4. ..... *133*
Figure 5-12 Actual vs. predicted grass growth for models 5-6. ..... *134*
Figure 5-13 Actual vs. predicted grass growth for models 7-8. ..... *136*
Figure 5-14 Actual vs. predicted grass growth for models 9-10. ..... *138*
Figure 5-15 Actual vs. predicted grass growth for models 11-12. ..... *139*
Figure 5-16 Actual vs. predicted grass growth for models 1-2. ..... *143*
Figure 5-17 Actual vs. predicted grass growth for models 3-4. ..... *144*
Figure 5-18 Actual vs. predicted grass growth for models 5-6. ..... *145*
Figure 5-19 RMSE for testing data of all the Brereton, ANFIS and RF models. *148*
Figure 5-20 Grass growth curve for Moorepark (2017-2018) using RF Model 10. ..... *152*
Figure 5-21 Grass growth curve for Johnstown Castle (2017-2018) using RF Model 10. *153*
Figure 5-22 Grass growth curve for Athenry farm (2017-2018) using RF Model 10. ..... *154*
Figure 6-1 Methodology flowchart *163*
Figure 6-2 Grass growth curve from Moorepark, Co. Cork (2017). *164*
Figure 6-3 Feature importance for Commercial dataset *166*
Figure 6-4 Actual vs. predicted growth on commercial farms. RF with all variables. *167*

Figure 6-5 Variable importance in national model. *168*
Figure 6-6 Actual vs. predicted grass growth in the national model *171*
Figure 6-7 Feature importance for Agmodel 1 ..... *173*
Figure 6-8 Actual vs. predicted grass growth for Agmodel 1 national model. *175*
Figure 6-9 Histogram of difference of days between Sentinel 2 and PBI. *176*
Figure 6-10 Variable importance for Agmodel 2 ..... *178*
Figure 6-11 Actual vs. predicted grass growth for Agmodel 1 national model. *180*
Figure 6-12 Actual vs. predicted grass growth rate for 2020 data using Agmodel *182*
Figure 6-13 Difference of acquisition days between Sentinel 2 & PBI (Agmodel) *182*
Figure 6-14 Residuals vs. difference of data acquisition days (Agmodel) *183*
Figure 6-15 Difference of acquisition days between Sentinel 2 & PBI vs. residuals *183*
Figure 6-16 Actual vs. predicted grass growth (2020) using Agmodel 1. ..... *185*
Figure 6-17 Residuals for Agmodel1 vs. difference of days between Sentinel 2 & PBI. *185*
Figure 6-18 Residual values (kg DM $ha^{-1}day^{-1}$) for Agmodel 1 vs. difference of days between Sentinel 2 & PBI. *186*
Figure 6-19 Histogram of difference of acquisition days between Sentinel & PBI (Agmodel 1) *186*
Figure 6-20 Actual vs. predicted grass growth rate (2020) using Agmodel 2. ..... *188*
Figure 6-21 Mean residuals for Agmodel1 vs. difference of days between Sentinel & PBI. *188*
Figure 6-22 Residual values from Agmodel 2 vs. difference of days between Sentinel 2 & PBI. *189*
Figure 6-23 Histogram of difference of acquisition days between Sentinel 2 & PBI (Agmodel 2) *189*
Figure 6-24 Grass growth curve for 2020 using Agmodel. *195*
Figure 6-25 Grass growth curve for 2020 using Agmodel 1. ..... *196*
Figure 6-26 Grass growth curve for 2020 for Agmodel 2. ..... *197*
Figure 6-27 Grass growth variability in 20 farms selected from PBI. *198*

Acronyms & Abbreviations

AGB - Above-Ground Biomass

ANFIS - Adaptive Neuro-Fuzzy Inference System

ANN - Artificial Neural Network

CV - Cross-Validation

DOY - Day Of Year

DSS - Decision Support System

EO - Earth Observation

EU - European Union

EVI - Enhanced Vegetation Index

FC - Field Capacity

FPAR - Fraction of Absorbed Photosynthetically Active Radiation

GDD - General Data Protection Regulation

GHG - Greenhouse Gas

GPP - Gross Primary Production

HNV - High Nature Value

IoT - Internet of Things

kg DM/ha - Kilograms of dry matter per hectare

LAI - Leaf Area Index

LiDAR - Light Detection and Ranging

ML - Machine-learning

MLR - Multi-Linear Regression

MODIS - Moderate Resolution Imaging Spectroradiometer

MoSt GGM - Moorepark St. Gilles Grass Growth Model

MSE - Mean Squared Error

MSPE - Mean Squared Prediction Error

NDVI - Normalised Difference Vegetation Index

NDRE - Normalised Difference Red-Edge index

OLI - Operational Land Imager

OLS - Ordinary Linear Regression

OOB - Out-Of-Bag

PA - Precision Agriculture

PAR - Photosynthetically Active Radiation

PBI - PastureBase Ireland

PE - Potential evpotranspiration

RF- Random Forest

RPM - Rising Plate Meter

RMSE - Root Mean Squared Error

SAR - Synthetic aperture RADAR

SMAPE - Symmetric Mean Absolute Percentage Error

SMD - Soil Moisture Deficit

SPOT - Satellite Pour l'Observation de la Terre

SVM - Support Vector Machine

SWIR - Short-Wave Infrared

UAS - Unmanned Aerial Systems

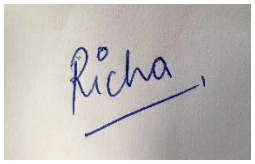
VI - Vegetation Index

VV - Vertically transmitted and vertically received SAR signal

VH - Vertically transmitted and horizontally received SAR signal

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism and intellectual property.



Richa Marwaha

Acknowledgements

I would like to thank my supervisors Dr Fiona Cawkwell (Department of Geography, University College Cork) and Dr Stuart Green (Spatial Analysis Unit, Teagasc), for their guidance, expertise and support throughout the PhD research. Thanks to Dr Rob O'Hara for the constructive feedback and invaluable advice for the thesis, without which I would not have reached this stage. Also, to Dr Jesko Zimmerman for providing the data and valuable suggestions. Thank you to Dr Carlos Alvarez for sharing his knowledge and experience, which helped me immensely during my research. I would like to thank all my departmental colleagues in Teagasc Rural Economy & Development Programme, particularly the Spatial Analysis Unit. I am grateful to Teagasc for funding this PhD under the Walsh Scholarship Programme.

I am grateful to all my friends in India and Ireland for their constant support and friendship. Finally, very special thanks to my family for their blessings and love.

Abstract

Globally, grasslands are an important source of food for livestock and provide additional ecosystem services such as greenhouse gas (GHG) mitigation through carbon sequestration, habitats for biodiversity, and recreational amenities. Grass is the cheapest source of fodder providing Irish farmers with an economic benefit against international competitors. Hence, to maintain profitability, farmers have to maximize the proportion of grazed grass in cow's diet or save it as silage.

The overall objective of the current research project was to build a machine-learning model to estimate grass growth nationally using earth observation imagery from the Sentinel 2 satellite constellation and ancillary meteorological data, which are known to influence grass growth. Firstly, the impact of meteorological data and Growing Degree Days (GDD) was assessed for Teagasc Moorepark experimental farm (Fermoy, Co Cork, Ireland). GDD was modified to include Soil Moisture Deficit (SMD), which included the impact of summer drought conditions in 2018. Results demonstrated the importance of GDD for grass growth estimation using ordinary linear regression (OLS). The potential evapotranspiration (PE) 0.65 (r=0.65) and evaporation (r=0.65) were equally significant variables in 2017, while in 2018 the solar radiation had the highest correlation (r=0.43), followed by potential evapotranspiration and evaporation with r of 0.42. The standard and modified GDD were equally significant variables with r of 0.65 in 2017, but both had a reduced correlation in 2018 with modified GDD (0.38, $p<0.01$) performing slightly better than the standard GDD (0.26, $p<0.01$) calculation. These models only explained 53% (RMSE of 18.90 kg DM $ha^{-1}day^{-1}$) and 36% (RMSE of 27.02 kg DM $ha^{-1}day^{-1}$) of variability in grass growth for 2017 and 2018, respectively.

Considering the importance of meteorological data, an empirical grass model called the Brereton model, previously used for Irish grass growing conditions were tested. Since this model lacks a spatial element, we compared the Brereton model with the previously used machine-learning model ANFIS and Random Forest (RF) with the combination of satellite data and meteorological data for eight Teagasc farms.

Overall, the machine-learning algorithms (R^2= 0.32 to 0.73 and RMSE=14.65 to 24.76 kg DM $ha^{-1}day^{-1}$ for the test data) performed better than the Brereton model (range of R^2=0.03 to 0.33 and RMSE=41.68 to 82.29 kg DM $ha^{-1}day^{-1}$). The RF model (with all the variables except rainfall) had the highest accuracy for predicting grass growth rate, with (R^2= 0.55, RMSE = 14.65 kg DM $ha^{-1}day^{-1}$, MSE= 214.79 kg DM $ha^{-1}day^{-1}$ versus ANFIS with R^2 = 0.47, RMSE = 15.95 kg DM $ha^{-1}day^{-1}$, MSE= 254.40 kg DM $ha^{-1}day^{-1}$).

When developing a national model, meteorological data were missing (except precipitation). A different approach was followed, whereby the grass growing season was subdivided (January-June Agmodel 1 and July–December Agmodel 2). Phenologically, the peak grass growth in Ireland typically occurs in May, with a slow decline in subsequent months. Spring is the most important season for grassland management, where growing conditions can impact the grass supply for the whole year. The national models were developed using Sentinel 2 band metrics, spectral indices (NDVI and NDRE), and rainfall for 179 farms. Data from 2017-2019 was divided into training and testing data (70:30 split), with 2020 data used for independent validation of the final trained model. Test accuracy was higher for Agmodel 1 (R^2 = 0.74, RMSE= 15.52 kg DM $ha^{-1}day^{-1}$) versus Agmodel 2 (R^2 = 0.58, RMSE= 13.74 kg DM $ha^{-1}day^{-1}$). This trained model was used on validation data from 2020, and the results were similar with better performance for Agmodel1 (R^2 =0.70) versus Agmodel2 (R^2=0.36). The improved spatial resolution of Sentinel 2 and the availability of red-edge bands showed improved results compared with previous work based on coarse resolution satellite imagery.

1

Chapter 1 Introduction

1.1 General introduction

The global human population is growing rapidly and is projected to reach 9.8 billion by 2050 (Ma et al., 2020). Ensuring sufficient food and raw materials to meet demand is a massive societal challenge (Timsina, 2018). In conjunction with continuing population growth, food producers globally must also deal with other factors that can threaten food production. Those factors are, for example, climate change (Gomez-Zavaglia et al., 2020), increasing urbanisation on agricultural lands (Arntzen et al., 2017), as well as a range of biotic stresses, for example, pests and plant diseases (van Zonneveld et al., 2020) and abiotic stresses, for example, water deficit and drought (Dormatey et al., 2020). At the same time, there is an increasing public awareness around the need for food production to be more sustainable, have a less environmental impact, and use fewer resources to produce more and better quality food (Yue et al., 2020).

The concept of sustainable food production is reflected in the current European Union (EU) Farm to Fork Strategy, which aims to transform agricultural production across the Union, to reduce greenhouse gas (GHG) emissions, consumption of natural resources and loss of biodiversity ensure access to sufficient, safe, nutritious and sustainable food (Commission, 2020). Additionally, EU Commission Agricultural Policy reform is putting pressure on European farmers to ensure that food production is both economically and environmentally sustainable (Hennessy et al., 2020).

Grasslands are a major ecosystem that makes a significant contribution to the global food supply (Zarei et al., 2020), supplying fodder for ruminants such as cattle, sheep, goats and horses, which in turn provide meat, milk, and other raw materials (for example, wool and leather) for human consumption (Nelson et al., 2017). For example, the volume of cow milk produced worldwide has risen from ~497 million tons in 2015 to ~532 million tons in 2020 (Statista, 2021). Grasslands cover approximately 40% of the terrestrial area. They are broadly characterized by grasses and herbaceous vegetation with little or no tree cover (White and Rohweder, 2000). They occur across various geographical areas and climatic conditions, including tropical and subtropical grasslands located in semi-arid to semi-humid climate regions in Africa, Australia, and South America. Temperate grasslands are found in semi-arid to semi-humid climates in Europe, North & South America and New Zealand.

Grasslands are used worldwide to feed livestock for meat and dairy production, although the level of management and expected yield vary considerably between regions. Figures 1-1 indicates that large areas of the world are under some form of pasture. A more detailed map of European pastures is illustrated in Figure 1-2. In Europe, a total agricultural area, nearly 21% is grassland (EUROSTAT, 2015), which includes a spectrum of management levels, from natural grasslands with minimum human interference, to extensive (semi-natural) low yield grasslands to high yielding, intensively managed grasslands with high inputs, and regular grazing/silage cutting (Velthof et al., 2014).

The EU is a major dairy producer producing over 157 million tons of milk in 2020 from a dairy herd of ~22 million dairy cows (EUROSTAT, 2021b). Intensive, pasture-based grazing systems dominate in Western Europe, with the Republic of Ireland leading the field in the proportion of area (93.4%) devoted to grazing (van den Pol-van Dasselaar et al., 2020, Eurostat, 2021a). Cheap, abundant grass and widespread grazing underpin the competitiveness of Ireland's domestic beef and dairy livestock industries. Recent surveys (Teagasc National Farm Survey 2018) (Dillon et al., 2018c) reported Ireland's national herd was approximately 6.5 million

cattle, which produced 8,075 million litres of milk and 617000 tons of meat in 2017. From 2015 to 2019, the number of dairy cows nationally increased by 17% (EUROSTAT, 2021b), resulting in a 20% increase in milk production (FAO, 2021). These figures are in the context of a program of agricultural intensification in the past decade, including Food Harvest 2020 and Food Wise 2025 strategies that put ambitious targets on primary production.

As well as food production, grasslands also provide additional ecosystem services, such as habitats for biodiversity (Bengtsson et al., 2018) or as places of tourism/aesthetical value (Schirpke et al., 2019). A critical function of grasslands is supporting biodiversity. Extensively managed grasslands such as semi-natural in Ireland are low-output and high diversity grasslands supporting pollination and biological control. Carbon sequestration means carbon storage over time, including above-ground and below-ground biomass, soil, and dead organic matter in terrestrial ecosystems. Grasslands contribute to ~34% of the global carbon stock (João Pedro Silva Justin Toland and O'Hara, 2008). Grasslands are associated with aesthetic value and are used for recreation activities such as bird watching and hiking (Schirpke et al., 2019). A critical ecosystem service, in the context of global climate change, is their role in carbon sequestration, the storage of carbon over time in both above-ground and below-ground biomass, as well as soil and dead organic matter in terrestrial ecosystems (João Pedro Silva Justin Toland and O'Hara, 2008). However, grasslands in Ireland are a net carbon source because of the drainage of organic soils (Renou-Wilson et al., 2015).These additional services are also important for improving the sustainability of grassland production in an economically viable way that is resilient to future change (OECD, 2003).

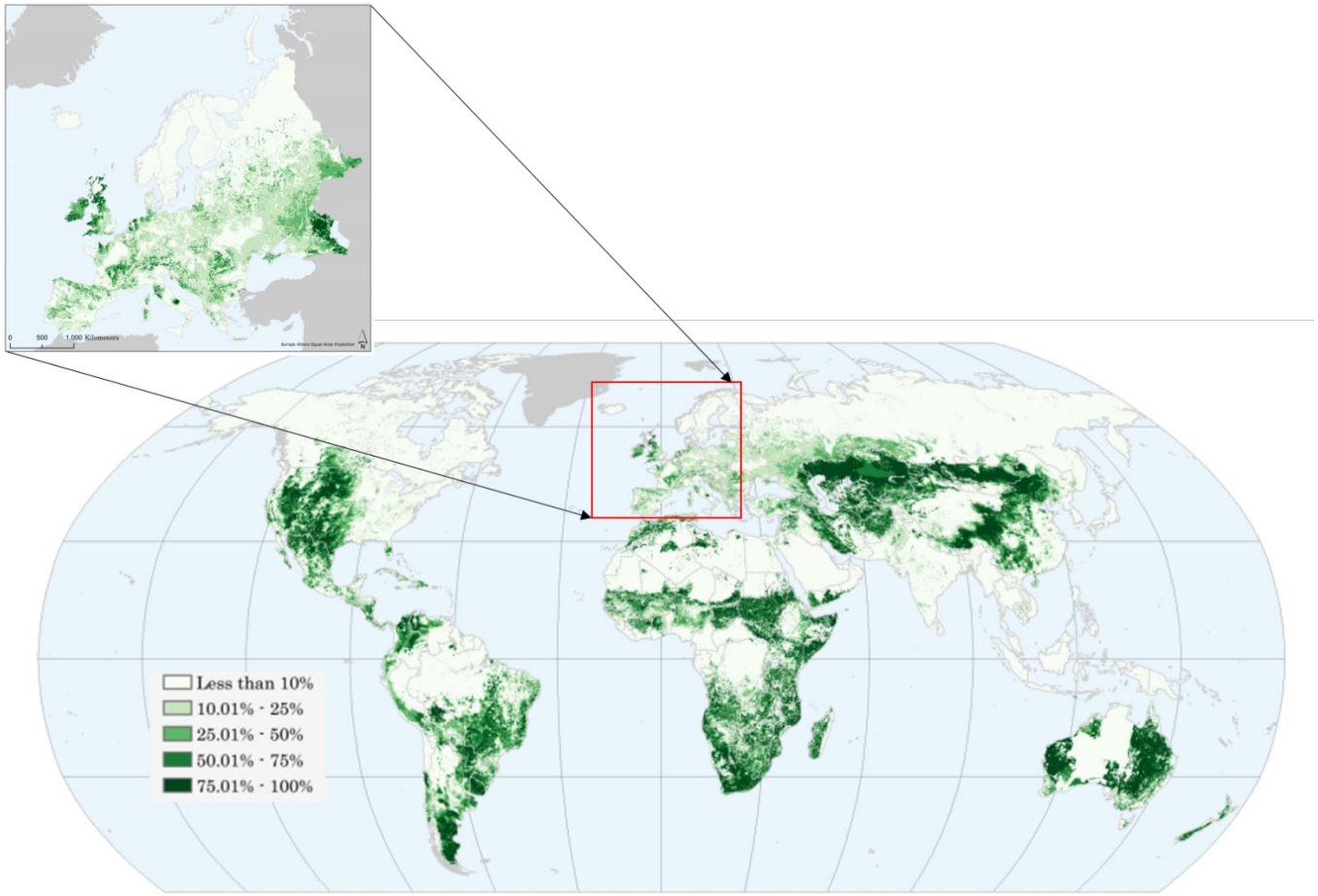


Figure 1-1 Global distribution of pastures in the year 2000

(Source: NASA Socioeconomic Data and Applications Center (SEDAC) Pasture map (Ramankutty et al., 2008). Darker green shades depict higher proportion of pastureland. These maps were derived from Moderate Resolution Imaging Spectroradiometer (MODIS) and Satellite Pour l'Observation de la Terre (SPOT) vegetation-based Global Land Cover 2000 data and combined with agricultural inventory data from the UN Food and Agriculture Organisation (FAO).

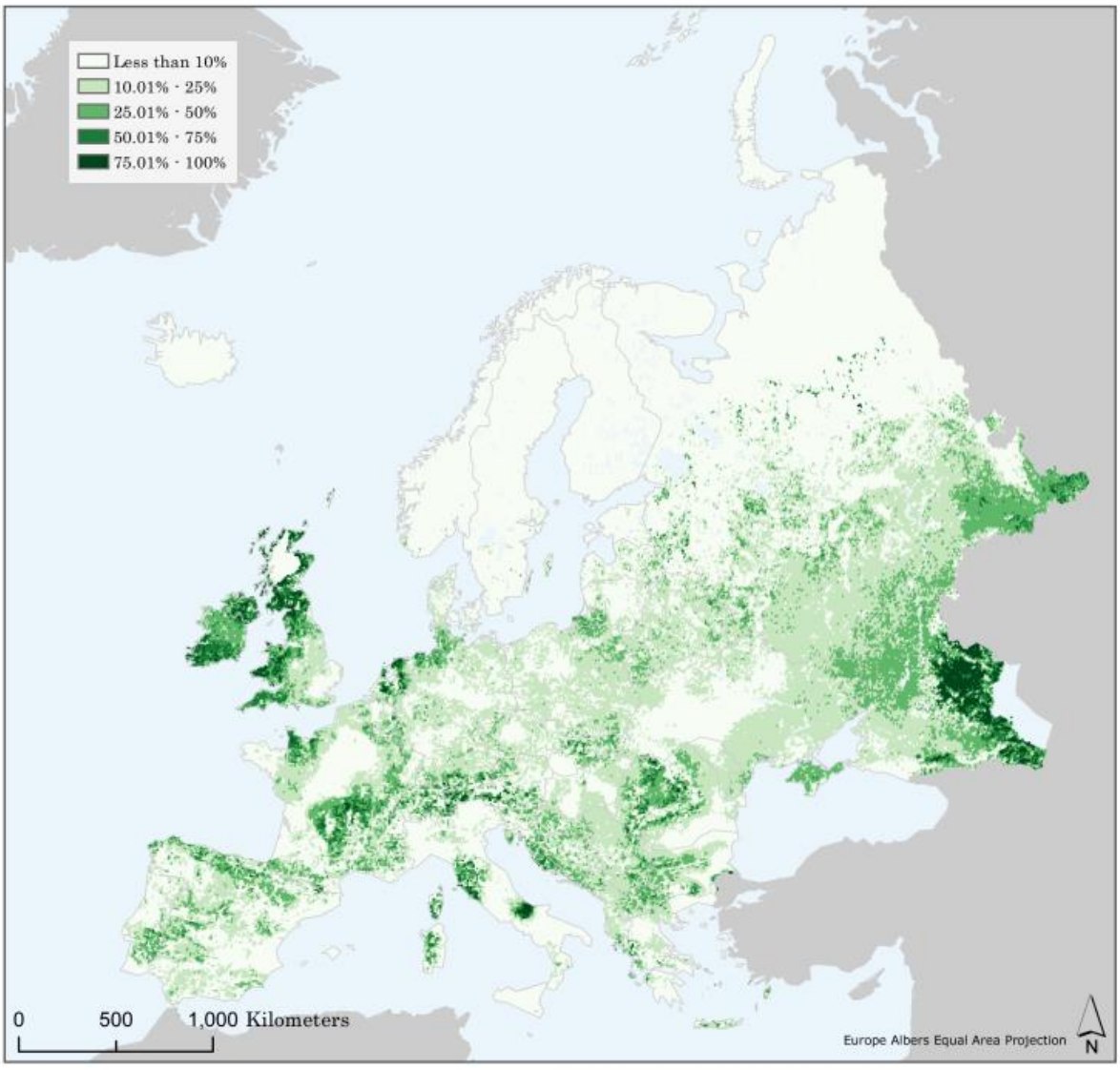


Figure 1-2 Pasture map of Europe in 2000

The darker the green shade, the higher is the proportion of area under pasture. The map was created using 1km resolution MODIS Global Land Cover Product and SPOT Vegetation Global Land Cover 2000 dataset (GLC2000) (Ramankutty et al., 2008).

For intensive agricultural grassland systems, such as those found in the Republic of Ireland, a crucial step in achieving sustainable production is increasing the proportion of fodder in the livestock diet. This can be done by growing and utilising more grass and minimizing the purchase of supplemental feedstuffs (O'Donovan et al., 2020), maximizing production while reducing environmental impacts such as increased emissions and nutrient leaching (Shalloo et al., 2021). Improved grazing management is a key step for sustainable grass production, which includes getting reliable data to facilitate targeted and timely management decisions. Using data to

target underperforming areas of a farm is the foundation of Precision Agriculture (PA). Increasingly, PA technologies are being adopted by livestock farmers (Michels et al., 2019). These include a vast array of sensors from the Internet of Things (IoT), biosensors and Earth observation satellites. One management intervention essential to all grass growers is measuring grass to understand growth rates and better manage surpluses or deficits.

Grass measurement is a common task for grassland farms, where farmers try to match their stocking rate (the number of livestock units they have per unit of area) with grass availability from spring to autumn while saving enough to feed the herd during winter when animals are housed indoors. There are management support systems available to farmers in several countries to help them better understand their grass availability and plan their grazing rotation (Vinogradovs et al., 2020, Sturm et al., 2018). In Ireland, a web- and smartphone-based grassland decision support system called PastureBase Ireland (PBI) was launched in 2013 (Hanrahan et al., 2017, Dasselaar et al., 2017). The input to PBI is typically in-field measurements by the farmers. These may be destructive methods, for example, where a small amount of grass must be cut and dried to gain accurate measurements of dry matter content (the cut and dry approach), or they may be non-destructive methods. These non-destructive techniques include a range of methods, from low-technology, point-based approaches, for example, visual assessment, simple measuring sticks or more technological approaches using rising plate meters (RPM, see Section 1.5.1 below for further details). These approaches are highly accurate but not spatially representative.

There are increasingly technical approaches to estimating grass biomass using biophysical models or Earth observation satellite data. Biophysical simulation models rely on physiological growth processes to predict yields and simulate yield responses to changing environmental conditions (for example, climate change) (Zhang et al., 2021a). The computer-simulated biophysical models lack a spatial component and can be constrained by the accuracy of the input datasets (Bellocchi et al., 2010). Earth observation methods offer a solution to the spatial limitations of

field-based and biophysical simulation methods by collecting data continuously over large geographical footprints at high spatial and temporal resolution. Satellite estimates of biomass have been used previously in Ireland. Coarse-resolution biomass estimates were made for two independent sites in Moorpark and Grange using Adaptive Neuro Fuzzy Inference Systems-ANFIS (Ali, 2016). The 8-day composite MODIS image time series was used, along with *in-situ* data. The results showed that machine-learning could retrieve grassland biomass with high performance (R^2 = 0.86 for Moorepark and R^2=0.76 for Grange). The study demonstrated the potential of using remote sensing and weather data to predict grassland biomass using machine-learning algorithms. The major limitation of the work was the coarse resolution satellite dataset, which cannot detect the variations between the fields at the farm scale. However, this issue will be resolved in terms of spatial resolution by the Sentinel 2 data in this work. In this research, a national model for Ireland will be developed to estimate the grass growth rate. The machine-learning models, .i.e. Random Forest (RF) and ANFIS, will be used with Sentinel 2 and meteorological data as input variables which will be validated against *in-situ* data from PBI.

1.2 Role of grasslands

From the introductory section above, it is clear that grasslands provide an essential service in food production and have other roles in carbon sequestration, biodiversity, and other areas. These applications will be explored in greater detail in the following sections. As noted previously, grasslands provide several services, including food supply, carbon storage, habitats for biodiversity and amenity/recreation value.

1.2.1 Food production

Grasslands play a significant role in food production and global food security. According to the Food and agriculture Organisation (FAO), food security, having access to adequate and nourishing food which meets daily dietary requirements, is comprised of four elements: food availability, food access, utilization and stability (Schmidhuber and Tubiello, 2007). Food availability focuses on producing enough food to meet the increasing global demand. As the population grows, there is

increasing pressure on agriculture to meet global needs sustainably. Factors such as weather and economical price fluctuations can affect the availability of food over time (Charlton, 2016). In conjunction with population growth, the improvement in incomes has seen a shift in consumption of meat (beef and poultry) and dairy products (milk and cheese) in the recent past (O'Mara, 2012, Baldi and Gottardo, 2017), particularly in low to middle-income countries (Drewnowski, 2018). These products primarily come from grass- or fodder-crop fed ruminants. As grass forms an abundant and cheap feedstuff, proper grasslands management to increase yield and fodder availability is critical for food production (O'Mara, 2012).

1.2.2 Carbon sequestration

Carbon sequestration is the process of capturing and storing atmospheric carbon dioxide (CO_2) in soils or vegetation, which can reduce global climate change (Jain et al., 2012). It is increasingly recognized that grasslands may have an essential role in carbon sequestration by acting as carbon sinks (Bossio et al., 2020), particularly in more intensively managed grasslands, where grazing and cutting can help re-stock soil carbon and prevent carbon loss from soils (Conant, 2010). As Figure 1-3 illustrates, excessive grazing on poorly managed grasslands does not allow grass to grow, leading to degradation. Simultaneously, under-grazing can also affect carbon storage, as insufficient organic content is added to the soil. Improved grazing practices are encouraged to maintain the carbon cycle in the grassland system (Abberton et al., 2009). The grassland management approaches such as cattle slurry application, grazing, fertilization, grass species diversity (mixed with clover) and planting hedgerows can supplement soil carbon stocks, thus sequestering atmospheric carbon in the soils.

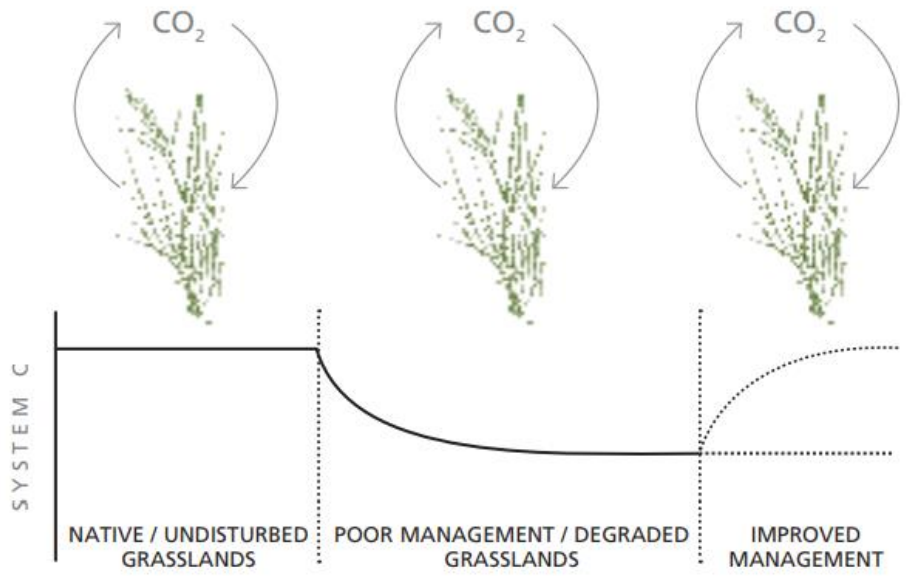


Figure 1-3 The effect of improving grassland management on soil carbon storage. By improving poorly managed or degraded grasslands, more atmospheric C can be sequestered and stored long-term in soils (Conant, 2010).

1.2.3 Semi-natural habitats & biodiversity conservation

Extensive grasslands have had centuries of low-intensity management, which has allowed them to support large amounts of vertebrate and invertebrate species. Intensification of grassland production includes frequent grazing, fertilization and pesticide use and artificial drainage, which impacts the richness and diversity of flora and fauna. Many of Europe's grassland ecosystems are now under threat from the intensification of agriculture or the abandonment of marginal lands (Gaujour et al., 2011). Grasslands can vary from monocultures to multi-species vegetation with varying management practices. Semi-natural grasslands are traditionally managed grasslands, with low inputs and infrequent interventions that are low-producing grasslands but very often-rich and very attractive to pollinators such as bees and butterflies. Semi-natural grasslands have declined between 2007 and 2012 due to conversion to woodlands or by the intensification of management in Ireland (Devaney et al., 2013).

EU policies such as the Commnon Agricultural policy (CAP) (Environnement, 2019) and EU Green Deal (Hart et al., 2020) support the conservation of semi-natural grasslands and High Nature Value (HNV) farmland. The farmers are advised to use fewer pesticides and fertilizers or use them more responsibly to protect the threatened habitats. According to proposed CAP reforms (Environnement, 2019), practices that conserve permanent areas, maintain species diversity or establish ecological focus areas will be compensated. Farmers are advised to keep some space on their farm for habitats such as trees, hedges, and ponds to promote biodiversity. Such habitats attract butterflies, bees, and birds, which perform different ecological services affecting other organisms' diversity, such as pollination.

1.2.4 Tourism

Many grasslands have an aesthetic attractiveness, combining landscape and species biodiversity that can promote well-being or have tourism potential (Parente and Bovolents, 2012). These activities can help farmers actively conserve biodiversity on their farms to enhance such an environment. Tourism has been used as a strategy for farms to find supplementary incomes by supplying accommodation to tourists or facilitating on-farm activities.

1.3 Introduction to Irish grasslands

Ireland’s temperate climate and fertile soils are well adapted for growing grass. Grasslands have a hugely important role in the Irish economy. As noted previously, Ireland has the highest proportion of total land area under grass within the EU. This abundance of grass provides a cheap source of nutrition for cattle and sheep and gives Ireland’s agro-food industries an advantage over international competitors (Hennessy et al., 2020). The Irish agri-food industry is export-orientated, with ~90% of the food produced exported to over 180 countries worldwide (DAFM, 2020). The largest export destination is the United Kingdom (UK), with 38% (€5.5 billion) of total exports by value in 2019, followed by the United States and the Netherlands, both importing over €1 billion of Irish agri-food products (DAFM, 2020). Brexit will have a negative impact on Irish export value, impacting sterling value and high trade costs between the two countries (Matthews, 2017).

1.3.1 Regional differences in farming enterprises

There are regional differences in the kind of farm enterprise carried out in Ireland (Walsh and Horner, 1984). There are physical reasons for these, such as soil type, climate, topography and geology. At NUTS Level 2, Ireland is divided into two parts, the Border, Midland and Western region (BMW) and the South and Eastern region (SE). The BMW Region consists of 13 counties in the Border (Louth, Monaghan, Cavan, Leitrim, Sligo, and Donegal), Midland (Longford, Westmeath, Offaly and Laois) and Western (Mayo, Galway, and Roscommon). Southern and Eastern region is composed of Dublin (Dublin City, Dun Laoghaire, Fingal, and South Dublin), Mid-East (Meath, Kildare, and Wicklow), South-East (Carlow, Kilkenny, Tipperary South, Waterford, and Wexford), Mid-West (Clare, Limerick, and Tipperary North), South-West (Cork and Kerry).

On average, dairy farms in the SE region are more extensive in terms of land area than in the BMW region, with an average farm size of 38.3 hectares compared to 27.1 hectares. Most dairy farms are located in the SE region of well-drained soil and long growing season (Dillon et al., 2018c). The West of Ireland has heavy clay and poorly drained soils with high rainfall. High rainfall can lead to excessive losses through drainage, affecting soil trafficability (Shalloo et al., 2004).

Trafficability is soil moisture conditions to support agricultural traffic such as planting and cultivating, and grazing without affecting the soils (Müller et al., 2011). In poorly drained soils, the growing season is shorter, which can affect dairy production. As a result, the milk production is reduced, and since there is no grass during the winter, the cows have to be supplemented using other feedstuff, which leads to an increase in production cost.

According to the recent data by CSO Farm Structures Survey (CSO, 2012), there were 137,500 farms in Ireland in 2016, 52.7% of which were located in the BMW region and the rest of the farms in the SE area (Figure 1-4). 43,200 of these farms were located in the Border, Midland and Western region. Specialist dairy farms (12,900) were located in the Southern and Eastern region.

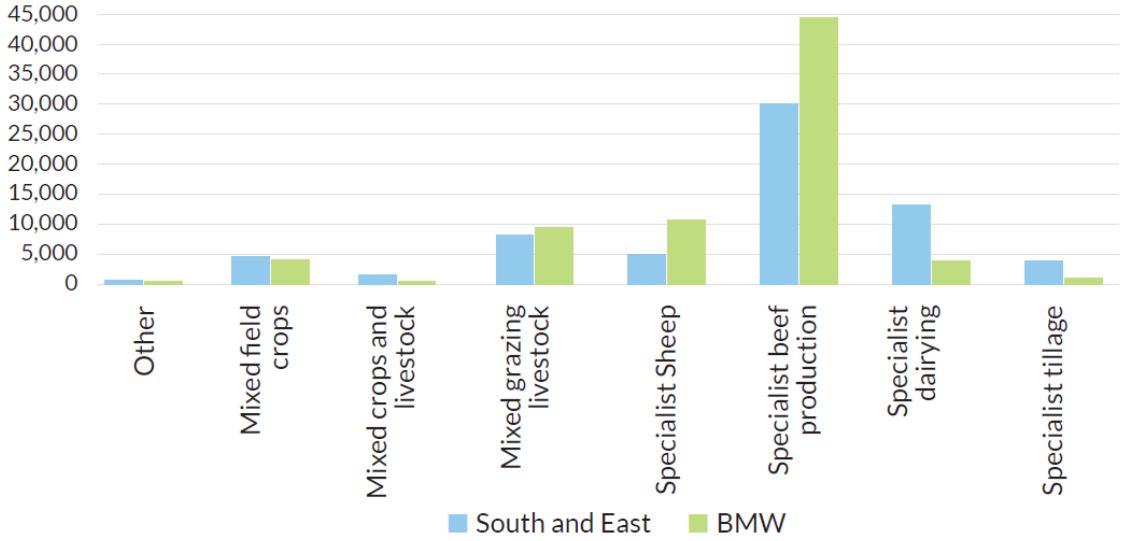


Figure 1-4 Number of the farm by category and region, 2016 (Source: CSO 2016). SE regions are represented by blue bars, and the BMW area is shown as green bars, where the y-axis is the number of farms and the x-axis is the category of farms.

1.3.2 Grass growth

The minimum temperature required for grass to grow in Ireland is 5 ºC. In Ireland, perennial ryegrass (*Lolium perenne)* is the dominant grass species. Ryegrass can support three leaves at one time. When the fourth leaf emerges, the leaf at the bottom dies. The rate at which a new leaf appears depends on the time of the year and season. For example, in early spring, a new leaf emerges every 30-40 days, which means the grazing rotation is long. However, during peak season in May and June, a new leaf is produced every 7-8 days so that the growing season cycle can take 18-21 days. If this grass is not grazed within 21 days, then the grass-grown is lost. It is because of such variation in grass growth rate that grass budgeting is essential. A farmer needs to match the grazing rotation to the time at which a new leaf appears. The farm is divided into a small number of paddocks. Each paddock is grazed until 3-4 cm residual height in rotation, known as rotational grazing.

The Irish grassland systems is a grass-based system focused on spring-calving. In grass-based feeding systems, cattle are fed on grass outdoors during the warmer season and are housed for winter (Dillon et al., 1995). During spring, the proportion of grazed grass is more than during winter, and as the proportion of concentrates increases in the feeding system, the milk prices increase, and production decreases. Ireland has one of the lowest milk production costs (Dillon et al., 2008). In spring-calving systems, cows calve early in the spring and start grazing to be fed by a grazed-grass diet. The grass growth curve for Ireland for the period of this study is illustrated in Figure 1-5. The beginning of the grass growing system and growth over the season is affected by several factors, including management factors such as fertilizer grazing of animals and environmental factors such as meteorological conditions and soil type.

It is crucial to increase grazed grass in a cow's diet to achieve high farm output. Proper grassland management is essential to achieve high grass production sustainably. When the grazed grass does not meet the cattle's feed requirement, then the proportion of other feed in the diet increases. Calving date and stocking rate are essential factors in a grass-based system and affect milk production. Late calving can lead to more concentrates and silage in the diet leading to the under-utilization of grass. The stocking rate is defined as the number of animals per hectare.

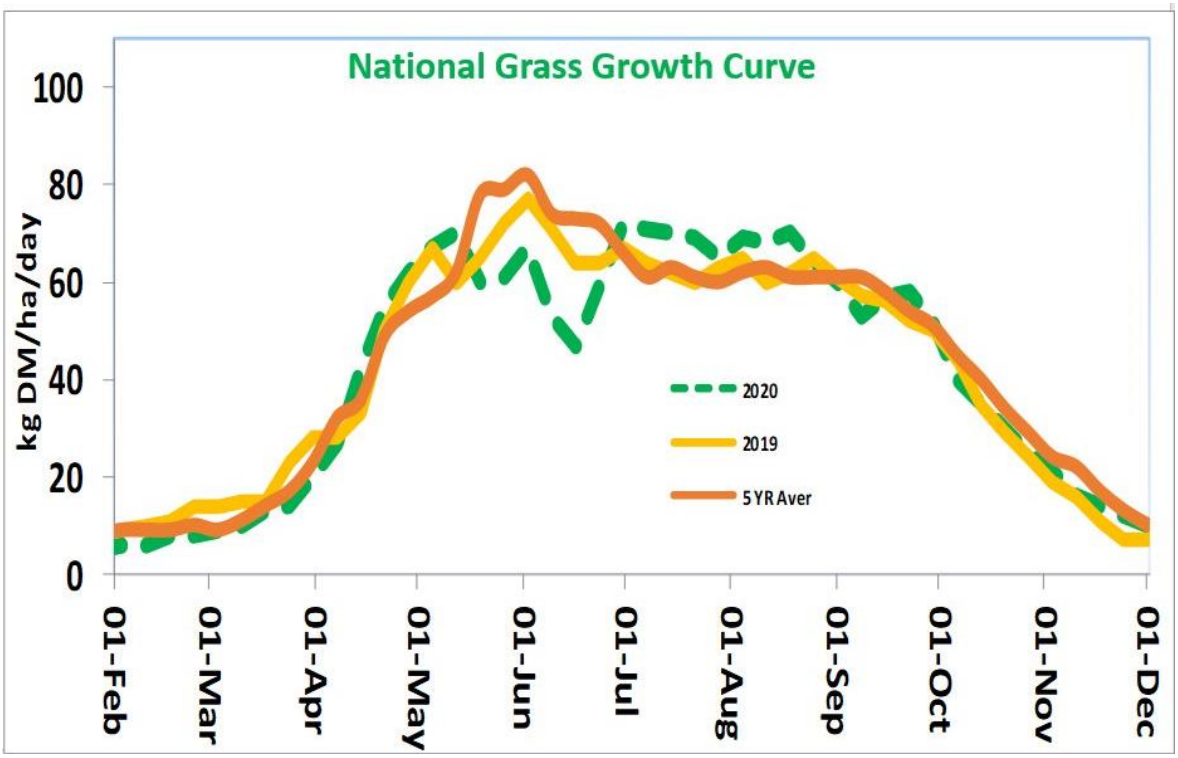


Figure 1-5 National grass growth curve with average growth rate values from PBI. The orange curve shows the five-year average, and the values from 2019 and 2020 are represented with a yellow and green curve, respectively.

The national average grass growth curve with average growth rate values from PBI consists of a peak in May and early season growth is higher than late season growth (Figure 1-5) (Hanrahan et al., 2017). In grassland, feed budgeting is important, i.e. matching grass growth with the grass demand on the farm to plan and decide on the farm (Figure 1-6). The grass budgeting might be challenging due to the uncertainty of pasture availability in the future (Barrett et al., 2004). The calving begins in February for spring calving animals, and the requirement for fresh grass is high.

Grass growth starts in February, and therefore, there is a deficit during the beginning of the season. During autumn, the grass growth starts to decline again. The cows are in the lactation period, and therefore the demand is more than the supply of grass. There is a surplus amount of grass during peak season, and the surplus can be used as hay or silage and kept for the deficit season during autumn (Snip, 2016). If there is an imbalance between demand and grass supply, there might be excess grass leading to wastage or deficit leading to under-feeding of animals and reducing the lactation period.

There is poor synchronisation between grass supply and feed demand on farms, as shown in Figure 1-6. The green curve is the grass growth curve with a primary peak in May. A blue curve represents the demand curve. There is a surplus of grass from mid-April to mid-August and a deficit for the rest of the year. Usually, the surplus grass during May and June is used as silage or hay to use during the winter when there is a fodder deficit.

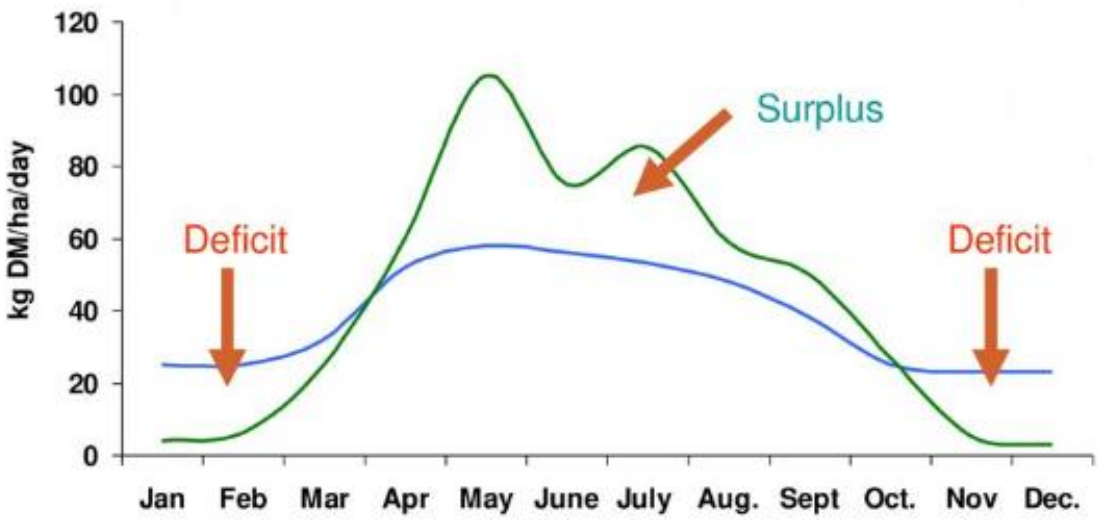


Figure 1-6 Grass supply and demand curve (Snip, 2016).

The green curve is the grass growth curve with the primary peak in May and the secondary peak in July. A blue curve represents the demand curve. From April till August, a surplus of grass can be saved as silage to be used during winter when the cattle are housed. There is a deficit of grass before the beginning of spring and in winter.

1.3.3 Factors affecting the grass growth rate

The grass growth rate is affected by climatic conditions, which are not controlled by a farmer and the grass management, which depends on the farmer. Management factors include fertilisation application rate, grazing or silage cutting, and grazing rotation. Climatic conditions such as rainfall, solar radiation, soil moisture deficit, and temperature affect the grass growth rate (Hurtado-Uria et al., 2014). The grass growth rate increases as the average air temperature increase up to 30 ºC, and when the temperature is below 5 ºC, the grass growth stops. Growing degree days (GDD) gives the amount of heat accumulated required for grass growth as species mature only after attaining a certain number of degree days (Fealy and Fealy, 2008). Extreme weather affects the quality and quantity of grass produced and, in turn, affecting livestock production. It is crucial to reduce the dependency of animals on indoor feeding during winters to decrease the production cost. Climate change affects agricultural production (Hopkins and Prado, 2007). It can lead to extreme weather conditions such as drought, flooding, and storm. The grazing season can be affected by soil and weather conditions (Lapple et al., 2012).

1.4 Why measure grass?

Ireland has an advantage over its international competitors for milk and meat production because of cheap and abundantly available grass. Grass growth is highly seasonal and can vary significantly both within and between farms. Such variation in growth makes grass budgeting challenging. It becomes essential for farmers to identify any surplus or deficit on their farm, helping them make management decisions. It was estimated that, on average, 8t grass DM ha^{-1} $year^{-1}$ of grass is utilised nationally on dairy farms using data from the commercial and research farms (O'Donovan, 2017). In 2017, a campaign was launched by Teagasc known as Grass10, and the main aim was to increase grass utilisation on the farms. Number 10 means to utilize 10 tonnes of grass DM/ha/year with ten grazings per paddock on farms by improving grazing management practices (Maher et al., 2019). This promoted regular measurement of grass grown on the farm and the maximum utilisation of grass.

1.4.1 Fodder crises

Grass growth is highly dependent on meteorological conditions such as rainfall, soil moisture, solar radiation, and temperature. In late 2012 and early 2013, Ireland was affected by the fodder crises (Dermody, 2013). The summer of 2012 had below-average temperatures and higher than average rainfall resulting in an insufficient quantity of fodder for the winter. The autumn was cold and led to low grass growth, and the farmers used up their available fodder rapidly. The spring of 2013 was colder and wetter than average, which affected the grass growth leading to a fodder shortage in 2013. Grass growth productivity is not always the result of a single, extreme event but the cumulative impact of small events such as long winter (White et al., 2020).

The cold and wet weather in the summer of 2017 led to low grass growth, which affected the winter fodder. Storm Emma was one of the most significant snowfall events of recent years that hit Ireland in March 2018, delaying the spring (Falzoi et al., 2019). A heatwave followed with low on-farm soil moisture with drought-like conditions in the country. In the summer, delayed spring and drought affected the grass growth rate during summer, with some regions producing 5-10% less grass than average. According to Dillon et al. (2018c), the drought-affected dairy farms the most, and their concentrate expenditure increased by 42% and the unfavourable weather conditions along with an increase in production cost led to a decline in farmer's income by one-third on average to €61,273.

The extent of the impact of the fodder crisis of 2018 was illustrated in a map produced by the Spatial Analysis Unit, Teagasc (Figure 1-7). The 10-day composite NDVI MODIS images were produced each month in 2018 and summed up from Feb to October, compared with average February-October growth for the ten years 2002-2012. The grass growth rate was below average for Eastern region farms and above normal for North-West regions. Poorly-drained soils performed better than well-drained soil as the poorly-drained soils can hold moisture for much longer and benefit from high temperature. The fodder crises in 2013 and 2018 highlighted that the national grass growth rate is essential, which means more measurement of grass

on the farm. If there is a surplus growth, it can be used for silage for winter. If there is a deficit, then farmers have to plan. In the next section, methods to measure grass growth are discussed.

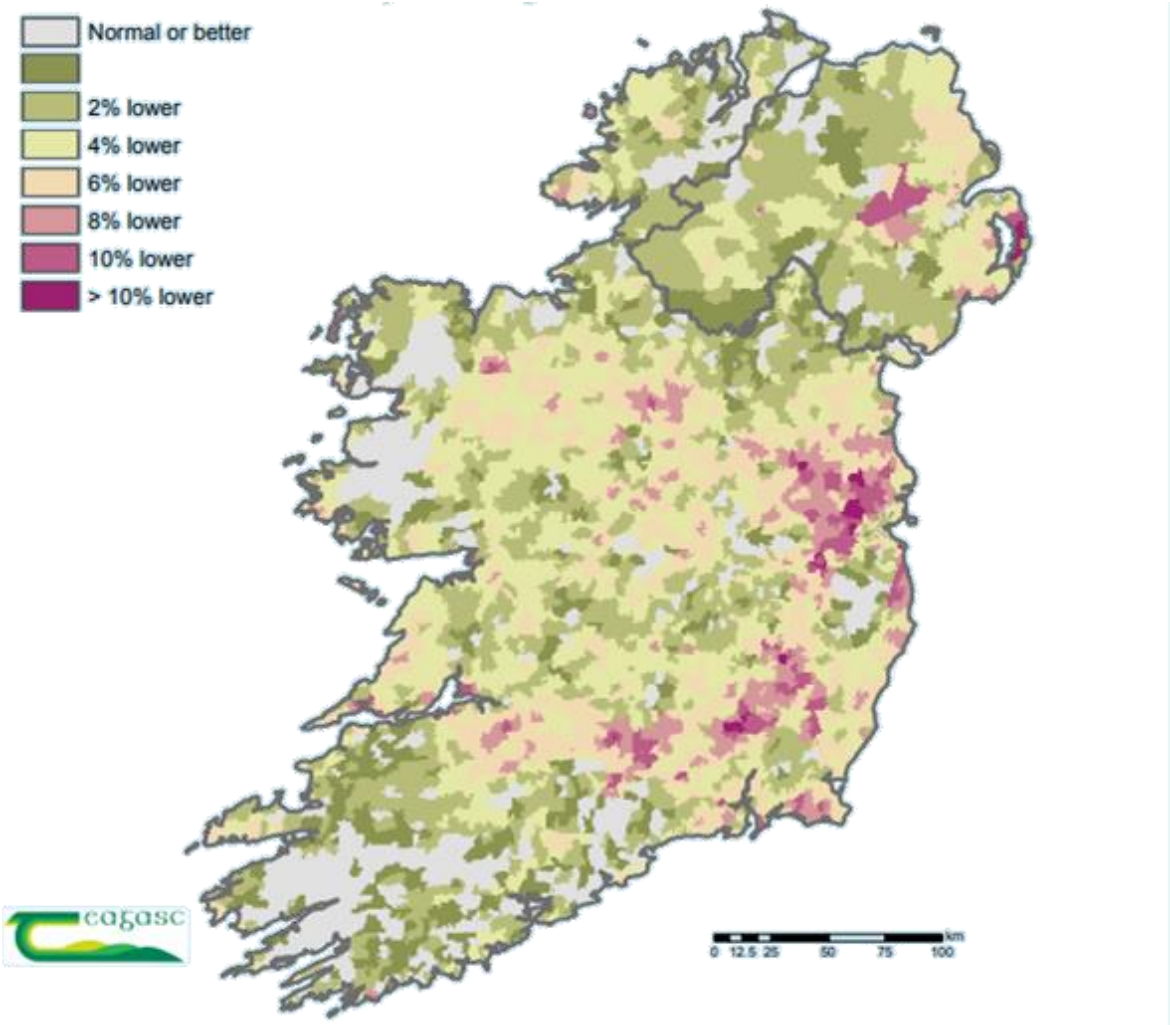


Figure 1-7 Poor fodder production in 2018 (Spatial Analysis Unit, Teagasc Ashtown). The map was created using NASA's 10-day composite NDVI MODIS images produced each month in 2018 and summed up from February to October, compared with average Feb-Oct growth for the ten years 2002-2012. The eastern regions produced between 5 and 10% less grass than normal grass. The poorly drained soils in the north and west fared best, producing more biomass than usual.

1.5 Grass measurement

Grass growth is highly spatially variable across pasture farms for various environmental and management specific factors (see Sections 1.3.2 above). As discussed in Section 1.3, grass measurement is a critical part of the grass budgeting process, helping farmers maximize utilization and increase profit margins. This section outlines several methods currently available for measuring grass, including on-farm methods, biophysical methods and remote sensing methods.

1.5.1 On-farm methods

On-farm methods of estimating grass growth rate can be divided into two categories- destructive and non-destructive methods. A standard method is a destructive cut and dry method where the grass is cut, dried and weighed to estimate grass biomass. Cut and dry is an accurate method but time-consuming. Another standard method is a non-destructive visual known as the "eye-ball" method, where a farmer estimates grass biomass by visually checking the biomass (O'Donovan et al., 2002). This method is time-effective but is inaccurate.

Biomass can also be estimated using grass height. A rising plate meter (RPM) is a farmer's tool to walk the farm and calculate the grass height (McSweeney et al., 2019). It consists of a one-meter shaft and a plate. The shaft can freely move through the plate. When the shaft is pressed downwards, the plate rises. The distance the plate moves is recorded relative to the one-meter shaft, which gives the grass height. If multiple samples are taken throughout the farm, then the average distance is calculated.

A similar device is a GPS-enabled Grasshopper™ which calculates grass height using a micro-sonic beam. The measured grass height is sent via Bluetooth to a mobile application. Also, GPS helps to map the paddocks on the farm. This device also gives an estimation of yield. Both the RPM and Grasshopper™ are non-destructive but time-consuming. Grasshopper™ is more accurate than RPM for calculating grass height/biomass and is less prone to measurement errors

(McSweeney et al., 2019). Both RPM and Grasshopper™ can get a more spatially representative assessment of grass height across the field. They do require farmers to measure grass, which can be time-consuming and repetitious actively. They also require farmers to spend money on additional equipment. As noted above, farmers can input grass growth into a decision support system such as PBI and include other factors such as stocking rate and several paddocks to see where grass surpluses/deficits exist on their farm. PBI is explained in greater detail in a subsequent chapter (See Section 3.4.4).

1.5.2 Biophysical simulation models

A model is a simulation of reality without performing actual experiments. Biophysical simulation models predict the predict crop yields under different management strategies (Rossiter, 2009). The major limitation of field-based methods is that they can give biomass estimate for few paddocks on a farm but will not give the whole farm performance. At the same time, biophysical models give predictions for the whole site, region or country. These growth models are parametrized and categorized into empirical and mechanistic models (Marshall et al., 2018).

- Empirical models estimate biomass by forming a statistical relationship between ground biomass and vegetation indices calculated using satellite images. The model obtained using these techniques are usually site-specific and need recalibration for every site.
- Mechanistic models involve equations based on crop physiology and the response of crop to environmental changes, that needs proper calibration before using them (Estes et al., 2013). Such models are input dependent, which can be difficult to acquire.

Many grass growth models have been developed for grasslands, such as DairyMod, EcoMod (Johnson et al., 2008) and LINGRA model (Schapendonk et al., 1998). Some of the grass growth models used for Irish conditions are Moorepark St Gilles (MoSt) grass growth model (Ruelle et al., 2018b), Jouven and Johnson & Thornley, which are mechanistic models, and the Brereton model is empirical (Hurtado-Uria et al., 2012). These models have been developed for perennial ryegrass in a temperate climate. These models can be used at a local scale (farm or catchment) or for a small

region with the same agro-meteorological conditions. They are data and parameters dependent, which must be calibrated for the site. The limitation of these models is that they are heavily dependent on the local parameters, field measured parameters, and cannot be used for any other site where they have not been calibrated (Higgins et al., 2019).

1.5.3 Remote-sensing methods

As discussed in Sections 1.5.1 and 1.5.2, farmers' on-farm measurement methods are often destructive and time-consuming, while the biophysical models must be parametrized and might apply to specific conditions. Remote sensing is an alternative method of grass growth estimation. Remote sensing is the science of observing and recording phenomena from a distance to obtain information about it. Earth observation (EO) is a branch of remote sensing that uses satellite platforms to measure reflected or emitted electromagnetic (EM) radiation. Depending upon the electromagnetic spectrum's wavelength, they can be optical sensors (433-2300 nm) such as Landsat 8, Sentinel 2, and microwave sensors (1 mm-1 m) and an example for it is Sentinel 1.

Satellite remote sensing has been used in agriculture for various applications such as biomass estimation (Ali et al., 2017b, Shoko et al., 2018), management intensity mapping (Griffiths et al., 2020, De Vroey et al., 2021), fertilizer application (Hollberg and Schellberg, 2017), identifying artificial drainage on grasslands (Hara et al., 2020) and classification (Barrett et al., 2014). Multispectral imagery from optical sensors can derive vegetation indices such as the Normalised vegetation index (NDVI) and Enhanced vegetation index (EVI). Vegetation indices have been used for biomass estimation over the years, such as using statistical models, e.g. multiple linear regression models (Shammi and Meng, 2021). The drawback of such approaches is that they assume that data have a normal distribution. It assumes a straight-line relationship between the dependent and independent variables, which is hardly true in real-world data.

The research has been shifting towards machine-learning algorithms. These are powerful algorithms that are non-parametric and do not depend on the data distribution. These models are not limited by any specific site or scale of application. The machine-learning algorithms have been used for crop biomass estimation, such as soy bean (Schwalbert et al., 2020), sugarcane (Shendryk et al., 2021), and rice (Guo et al., 2021). The use of remote sensing for Irish grasslands biomass estimation is relatively new, and it needs to be explored more due to the dynamic nature of these grasslands. For two sites in Ireland, a machine-learning algorithm was used to estimate grass growth rate with MODIS data (Ali et al., 2017b).

1.6 Rationale for this study

In the previous sections, it has been demonstrated that grasslands are essential for food security, biodiversity, carbon storage and tourism. The grass growth rate is variable over the seasons and can depend on meteorological and management factors. Extreme climatic factors such as drought and floods can affect the grass growth rate. A farmer needs to understand the spatial variation of grass on the farm to manage the feed budget. Conventional methods to estimate grass biomass rely on either visual methods or empirical/mechanistic grass models. The problem with visual methods is that they depend on a person's experience, which requires much training, as discussed in Section 1.5.1 (O'Donovan et al., 2002). The grass growth models can vary from a straightforward empirical model to complex mechanistic models. The problem with mechanistic/empirical models is that they depend on the site where they are applied and are highly dependent on parameters' field measurements (Hurtado-Uria et al., 2012). They cannot be applied over large geographical areas and cannot be used at a national level.

EO offers a promising alternative to conventional methods of grass biomass estimation. Satellite data is used widely for grassland biomass estimation but is still developing in Ireland. The potential was demonstrated in a previous study that used coarse resolution, multispectral data from the MODIS satellite at two locations (Ali et al., 2017b). It is not possible to detect the variations within the fields using coarse

resolution. This study builds on this earlier study by developing a national model with higher spatial and shorter temporal resolution imagery. A new dataset, Sentinel 2 and a new algorithm, i.e. Random Forest will be used in this work. An accurate model of national grass growth can help farmers optimize grass production and utilization.

1.7 Summary

In this chapter, the importance of grasslands globally has been introduced. The Irish grazing system has been explored, and the need for accurate grass measurement has been outlined. ML and EO's use for retrieving biomass in an intensively managed grassland is investigated in the remainder of this thesis. In the next chapter, grassland biomass estimation's current status is investigated, focusing on current biophysical and EO-based methods and a more in-depth discussion of machine-learning for grassland monitoring. The thesis structure is shown in Figure 1-8.

The objectives of this research are to:

i. Evaluate the role of growing degree-days (GDD) for estimating and understanding grass growth rate at a farm-scale for one farm- Moorepark (Chapter 4);

ii. Compare the performance of the Brereton model estimating grass growth rate at 8 farm locations against RF and ANFIS regression models (Chapter 5); and

iii. Transfer the machine-learning model developed in (ii) above to a broader national scale using 179 representative farms which contribute grass measurements to PBI (Chapter 6).

This chapter has introduced grasslands and grassland production in Ireland. The Irish grazing system has been explored, and the need for accurate grass measurement has been outlined. In the next chapter, the current state of the art in grassland biomass estimation's is discussed, focusing on recent studies using biophysical simulation and satellite remote sensing methods.

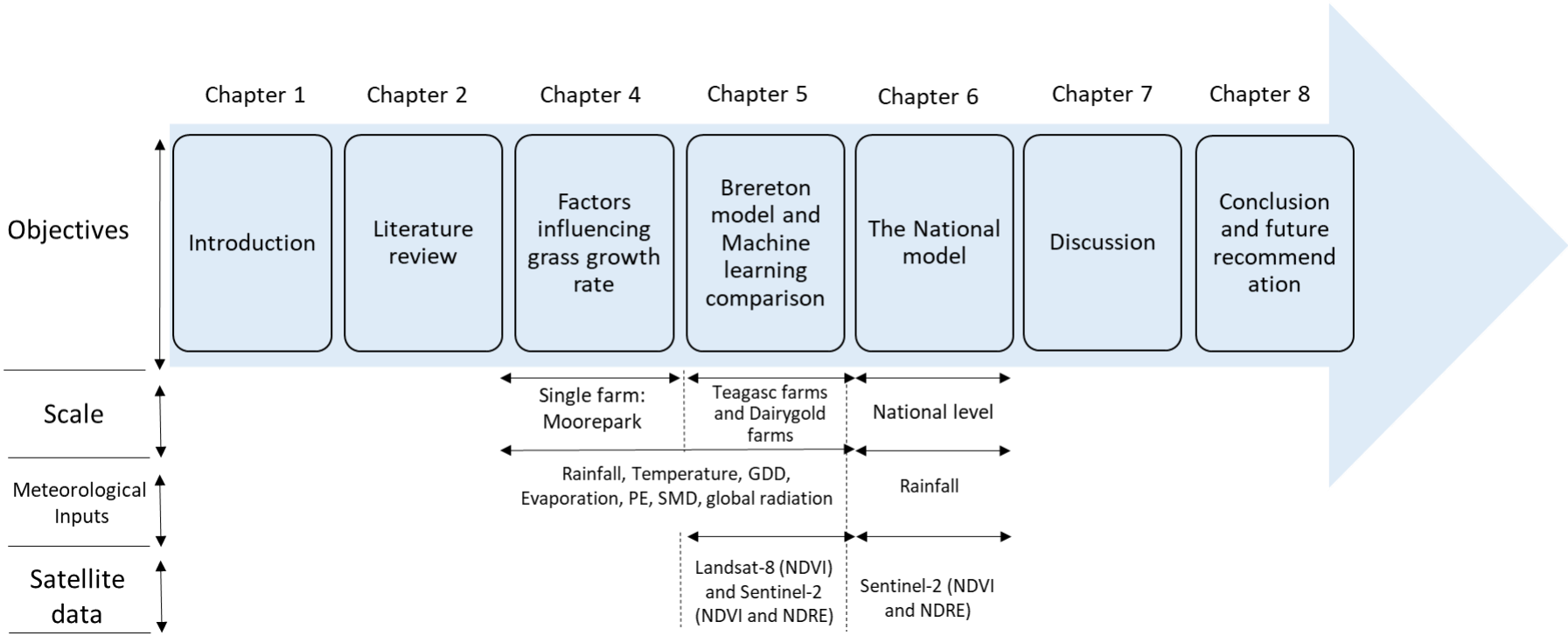


Figure 1-8 Thesis structure and chapter descriptions

1.8 Chapter overview

The structure of the remainder of this thesis is as follows:

- Chapter 1 presents an introduction to grasslands, their importance, and their relevance. It also defines research objectives and outlines all the chapters.
- Chapter 2 gives a detailed literature review of the methods to estimate grassland biomass. The methods using field-based, biophysical and remote-sensing methods will be discussed.
- Chapter 3 introduces the study area and the dataset used, such as input data and the ground data from PastureBase Ireland.
- Chapter 4 presents the factors influencing the grass growth rate in Ireland. The importance of growing degree days is presented.
- Chapter 5 compares the biophysical simulation model, the Brereton model, and machine-learning models: ANFIS and RF. Both models were evaluated, and the method to perform better was selected for further use. This chapter develops a machine-learning approach for grass growth rate estimation using remote sensing. Two machine-learning models were also compared.
- Chapter 6 extends the model developed in Chapter 5 to a national level. It includes 179 representative farms in Ireland.
- Finally, Chapters 7 and 8 discuss the overall outcomes and conclusions for all the chapters and present the recommendations for future work.

2

Chapter 2 Literature Review

This review chapter aims to place the research presented hereafter into a broader research context, examining the need for more accurate grass growth models, particularly using satellite data and machine-learning. The review will demonstrate a gap in current literature regarding available grass growth models for Western European temperate grasslands. The role of Earth observation (EO) in biomass modelling is examined, focusing on pasture biomass. This chapter is arranged as follows: Section 2.1 presents field-based, destructive methods (the "cut and dry" approach) as well as non-destructive methods (including visual estimation, proximal sensors, and rising plate meters (RPM), including "Grasshopper™". Section 2.2 discusses empirical and mechanistic grass growth models, presenting and their advantages and disadvantages. Section 2.3 analyses the use of remote-sensing methods of grass biomass estimation, including RADAR and optical satellites and unmanned aerial systems (UAS).

EO-based estimation of biomass is a very active area of research. Ali et al. (2016) presented a review of the current status of grassland monitoring/observation methods and applications based on satellite remote sensing data, and related technological and methodological developments, to retrieve grassland information. They noted that the retrieval of grassland biophysical parameters moved from standard regression analysis to more robust mechanistic modelling approaches, primarily driven by the increasing volume of satellite imagery. Ali et al. (2017) also explored the use of machine-learning and EO data for biomass estimation as part of an Irish study on biomass estimation using MODIS (Ali et al., 2017b). In their study, a model was developed for two Teagasc farms in the Republic of Ireland, Moorepark, Co. Cork

and Grange, Co. Meath. The model used 8-day composite imagery collected over 12 years. A more recent review paper by Reinermann et al. (2020) presented optical and SAR data applications in grassland use intensity and management, including monitoring grazing, mowing, irrigation and fertilizer applications. The paper highlighted the most commonly used spectral indices, which were - NDVI (62% of studies reviewed), EVI (15%), SAVI (9%) and 8% using leaf area index (LAI). The authors also drew attention to the impact grassland heterogeneity has on biomass estimation. Since the launch of Sentinel 2A in 2015 followed by Sentinel 2B in 2017, the researchers have been able to use higher spatial resolution data when estimating biomass.

2.1 Ground-based methods to estimate grass growth

2.1.1 Non-remote sensing methods

Grass biomass is positively correlated with sward height (Hakl et al., 2012). Sward height can be measured in the field using an RPM or a more straightforward "sward stick" that can be linearly related to canopy biomass by regression equations. The "cut and dry" approach is the most accurate and direct method to estimate grass biomass but is time-consuming and destructive, limiting its applications (Harmoney et al., 1997). The "cut and dry" methods are typically used as a benchmark for calibrating and validating the other ground-based methods. The RPM consists of a shaft and movable disc plate (Figure 2-1 (a)). When the shaft is placed on the grass, the plate moves up to the canopy to measure compressed sward height. There are disadvantages to using RPM; for example, it is labour intensive to record sufficient measurements to get an accurate canopy height value (Wachendorf et al., 2018). As a result, another version of an RPM, known as the 'Grasshopper™', uses an ultrasonic sensor to measure the sward height emitting ultrasonic waves to the plate and measuring the return time. It also has an integrated to record the measurement location (Figure 2-1 (b)) (McSweeney et al., 2019).



(a) *(b)*

Figure 2-1 Ground-based methods to measure sward height

(a) RPM with manual shaft and disc (b) The Grasshopper™, which automatically measures grass height location coordinates. Source: (Teagasc, 2017)



Figure 2-2 An example of a sward stick.

This measure has markings for grass height and corresponding grass yield. The colour-coded markings offer advice based on the sward height and estimates of yield in kg/DM/ha

A sward stick is a method comprising a ruler with corresponding biomass values in kg DM/ha (Figure 2-2). The red markings are for low yield with less than 4 cm height. Between 150-1150 kg DM/ha, the grass is let grow. Between 1150-1900 kg DM/ha grass is available for grazing, while cover greater than 1900 kg DM/ha is considered excessively high. RPM and sward sticks are affordable, quick estimators of yields and are in everyday use among farmers. In a comparative study, (O'Donovan et al., 2002) reported visual estimation by the farmer and RPM had had

an R^2 of 0.95 and 0.94 against actual yields, and for the sward stick, R^2 was 0.72. Sanderson et al. (2001) found that RPM measurements calibrated for ryegrass and clover swards in New Zealand yielded high error when used on pastures in the USA with R^2 of 0.31 and RMSE 653 kg/ha.

Serrano et al. (2020) used a capacitance probe known as Grassmaster II™ to estimate pasture biomass. The capacitance probe produces an electric field, which changes when brought close to the pasture. The capacitance of a material is proportional to the dielectric of the pasture, which changes with the water content and can be correlated with the biomass. The experimental data using the capacitance probe from 2007-2018 was used for calibration, and the data from 2019 was used as validation data. The probe is sensitive to the moisture content and, therefore, to the wet biomass and does not work well during the summer as there is a low moisture content in pastures, which affects the estimation accuracy. The ground-based methods are costly and difficult to implement over a large geographical area or an area that is difficult to access and therefore does not give the spatial distribution of biomass.

2.1.2 Proximal sensors

Proximal sensors are field-based sensors that can provide accurate measurements and help identify temporal and spatial variability within paddocks. Legg and Bradley (2019) used an array of low-frequency ultrasonic sensors mounted on a moving vehicle by sending ultrasonic echoes through the vertical depth of the pasture to get the height. A regression model was developed using measured height and biomass from the cut and dry method. The R^2 was in the range of 0.70 to 0.80. The ultrasonic sensors used were low frequencies, which might not be helpful with increased driving speeds and could result in inaccurate measurements. Nguyen et al. (2020) used a vehicular Light Detection and Ranging (LiDAR) sensor and a Real-Time Kinematic positioning system called DairyBioBot to measure grass volume to estimate biomass. The ground truth data were measured using a mower and weighing the samples by the cut and dry method. There were 160 experimental plots with three rows per plot. There was a high correlation between grass volume and ground

biomass with R^2 of 0.71 at row level and 0.73 at plot level. The authors observed that in the reproductive phase of grass, the biomass also includes the flower heads, creating complexity in LiDAR measurements. Overall, the proximal sensors are a promising technology for grass biomass estimation; it can be easily used by farmers, and can provide results instantly.

2.2 Grass growth models

2.2.1 Biophysical simulation models

Biophysical simulation models provide an understanding of the complex biophysical interactions between site-specific soil, weather and management conditions that affect grass growth rate. They can also simulate future climatic conditions and the impact of climate change on grass growth. For example, Holden and Brereton (2002) used a grass growth model to predict the impact of climate change on grass growth rate in Ireland for over 100 years for blocks of 30 years- 1961-1990, 2041-2070 and from 2061-2090.

Grass growth models are typically divided into two types- empirical models and mechanistic models. Empirical models describe the relationship between the local climatic variable and pasture biomass. Empirical models are relatively straightforward to implement as they have a small number of input variables. The Brereton model is an example of an empirical grass model developed at Teagasc Research Centre in Co. Wexford, Ireland (Brereton and Keane, 1992). Empirical models are site-specific and are further limited by not considering processes such as photosynthesis and respiration.

Mechanistic models overcome these limitations by including crop physiological information and variables such as fertilizer application and grazing dates. Mechanistic models are more complex and require additional parameters but can be used for various soils and climates (White and Snow, 2012). Dairymod (Johnson et al., 2003a), EcoMod (Johnson et al., 2008) and Sustainable Grazing Systems- SGS Pasture Model are collectively known as GrazeMod (Johnson et al., 2003b), are

mechanistic models to estimate pasture growth. They are based on Johnson and Thornley (1983) model. The models are a daily step model considering different soil types, management, irrigation, and fertiliser in Australia and New Zealand. GrazeMod was applied to pasture in Australia and New Zealand under a range of climatic, soil and management conditions for temperate and tropical pastures. The model predicted the pasture yield with an R^2 of 0.73 (Cullen et al., 2008). Their model could not take into account the variability of yield in paddocks due to grazing.

2.2.1.1 LINGRA

LINGRA (Light INTerception and UtiLization simulator- LINTUL-GRASS) is another example of a mechanistic grass growth model developed by Schapendonk et al. (1998). The model is based on the energy conversion from radiation to carbon stored in the grass, depending on light efficiency and leaf area indexes. The authors used the LINGRA model, calibrated using the experimental data from 35 sites across 16 European countries with different meteorological and soil conditions from 1982-1986. The model was tested on a subset of farms from Southern and Northern Europe (the exact number was not mentioned), divided into two categories- irrigated and non-irrigated to assess the models potential to work in irrigated and in water-limited conditions. The fields were irrigated when the soil-water deficit exceeded 12.5 mm. The average normalised error between actual and predicted grass yield was 14% for irrigated farms with over-prediction and 19% for non-irrigated farms with under-prediction. Non-irrigated farms were water-limited, and the impact of deficit soil moisture led to under-estimation.

2.2.1.2 GrazeGro

GrazeGro, proposed by Barrett et al. (2005), is based on the LINGRA model described above and was used to predict pasture yield for sites in Spain, France, Norway, Northern Ireland and the Netherlands. The inputs to the model were photosynthetically active radiation (PAR), mean air temperature, rainfall and the rate of nitrogen fertilizer application. The mean normalised error for the accumulated yield varied from 1.9% to 15.3%, and the R^2 varied from 0.27 to 0.85. A data limitation, such as the soil type and rate of nitrogen fertilizer application data, was

not available, and the assumptions had to be made for each site. The growth data were questionable as some sites had unusually high grass growth rates.

Barrett et al. (2004) compared the Brereton model, LINGRA, Johnson & Thornley and a version of LINGRA, which considers the seasonal growth effects adapted by the authors. The models were tested for two sites with data for 17 years for site 1 and 11 years of data for site 2. For site 1, the Brereton model performed the best with an R^2 of 0.67 and mean squared prediction error (MSPE) of 355.3 kg DM/ha, and the LINGRA model performed the worst with an R^2 of 0.27 and MSP of 1194.4 kg DM/ha. The second site modified LINGRA model performed the best with R^2 of 0.65 and MSPE of 333.8 kg DM/ha, and the Brereton model performed the worst with R^2 of 0.20 and MSPE of 975.7 kg DM/ha. The two sites were different because site 1 was affected by drought leading to low grass growth rates during summer and the Brereton model (34.5 kg DM $ha^{-1}day^{-1}$) predicted the values close to the actual values (41.0 kg DM $ha^{-1}day^{-1}$).

2.2.1.3 APSIM

Agriculture Production System Simulator (APSIM) has a pasture growth prediction module called AgPasture. It was used to predict the pasture growth for 27 sites in New Zealand over 37 years (Li et al., 2011). The measured and simulated annual pasture biomass accumulation was 7277 and 6833 kg DM/ha, respectively, with an R^2 of 0.83. The model did not take into account pasture phenology and management, such as grazing and fertilizer. There was heterogeneity on the sites with multi-species pastures and different grazing management, affecting the biomass.

2.2.1.4 MoSt GGM

The Moorepark St. Gilles grass growth model (MoSt GGM) is a mechanistic model which is a modified form of Jouven et al. (2006). It uses soil, weather, and N content and management data to predict perennial ryegrass growth at a paddock scale (Ruelle et al., 2018b). The model was modified to include the meteorological forecasts, which performed better than the model with climatological data (McDonnell et al., 2019). The model can predict grass growth up to 6 days in advance. To calculate the farm grass growth rate, the model has to run for each paddock and average the results, which is time-consuming. Another limitation of the model is that it could not predict extreme grass growth values.

In a comparative study of grass growth models used under Irish conditions, Hurtado-Uria et al. (2012) compared one empirical model - Brereton and two mechanistic models - the Johnson and Thornley model (Johnson and Thornley (1983)) and the Jouven Model (Jouven et al. (2006)). The Johnson and Thornley model used the meteorological data and initial conditions such as dry matter and leaf area index (LAI). The initial conditions are the values needed at the beginning of the simulation period. The Jouven model uses meteorological data, N index and soil water holding capacity. The three models were tested at Moorepark, County Cork, using data from 2005-2009. The study concluded that no one model was ideal for grass growing conditions in Ireland. The average annual grass growth rate for 5 years was 50.3 kg DM $ha^{-1}day^{-1}$. The Johnson and Thornley and the Brereton model over-predicted the growth by 88.9 and 8.3 kg DM $ha^{-1}day^{-1}$ respectively, while the Jouven model under-predicted the growth rate by 13.9 kg DM $ha^{-1}day^{-1}$. The site-specific parameters required for the empirical model limits their applicability for the diverse range of conditions on farms nationally.

2.2.2 Biomass using radiative transfer models

Radiative transfer models are used to measure grass biophysical parameters such as LAI, which can be used to estimate AGB. Grass AGB is positively correlated with sward height (Andersson et al., 2017) and LAI (Crabbe et al., 2019). Schwieder et al. (2020) compared random forest (RF) with a radiative transfer model called soil–leaf-

canopy (SLC) for estimating grass biomass using Sentinel 2 imagery. The RMSE was normalised (NRMSE) according to the validation data or easier comparison. RF performed better with NRMSE of 11% for biomass estimation than the SLC model with NRMSE of 47%. The grasslands were heterogeneous in this study area, and therefore SLC modelled biomass with high variability. For RF, many field samples were needed from all grass species types to estimate biomass with high accuracy.

Punalekar et al. (2018) used LAI to estimate grass biomass for three sites in the UK. LAI was derived from an inversion of radiative transfer model PROSAIL, NDVI from the proximal hyperspectral sensor and Sentinel 2 imagery using an exponential equation. Linear regression between the LAI using an instrument called a ceptometer (a device with an array of light interception sensors to measure photosynthetically active radiation and LAI) and grass biomass using RPM gave an equation to estimate grass biomass. The obtained equation was used to estimate biomass from LAI derived from Sentinel 2 and the inversion model. The R^2 ranged from 0.22 to 0.76 for PROSAIL derived LAI and from 0.16 to 0.73 for NDVI derived LAI for three different dates. The limitation of PROSAIL is that it assumes vegetation as homogenous and therefore does not work well with multispecies grass. The equation used to derive AGB from LAI was developed specifically for these sites, and to use it for new locations, new regression equations would be required. The major drawback is that they are site-specific and incapable of capturing the non-linear and complex patterns in data from other locations and cannot be applied generically across diverse pasture ecotypes with dissimilar management practices.

2.3 Biomass estimation using Earth observation data

As noted above, ground-based methods are time-consuming and labour-intensive, while grass growth models are complex and do not consider spatial element. Consequently, there has been a shift toward EO data for grass biomass estimation. EO-based methods can overcome some of the spatial disadvantages of the previous methods. The satellite can record data synoptically over large geographical areas, including areas that may be difficult to access.

2.3.1 Unmanned aerial systems

Unmanned Aerial Systems (UAS) can be mounted with various sensors to measure biomass, including standard RGB cameras, as well as multispectral or hyperspectral sensors. UASs have an advantage over satellite data as they can be deployed on-demand and capture very high spatial and temporal resolution data. Moeckel et al. (2018) used UAS RGB imagery to estimate minimum, mean, and standard deviation grass height as input variables for RF and SVM models to predict sward height. RF (R^2 = 0.89 - 0.97) performed slightly better than SVM (R^2 = 0.87 - 0.91) for three crop types (eggplant, tomato, and cabbage). The algorithms were able to work efficiently with crops of different structure and heights. Predicted crop heights were then used to estimate biomass (RF with R^2 = 0.88-0.95).

Barnetson et al. (2020) used UAS RGB imagery with random sample consensus (RANSAC) and a - decision tree-based pipeline optimisation tool (TPOT) to estimate pasture biomass. The site had a mix of arid and sparse grasslands with woodland. Grass height from UAS (9-10 mm pixel size) using photogrammetry was used as an input to the two models. The field-based biomass collected using RPM was used as reference data. The R^2 between the heights estimated using UAS and field height using RPM was 0.44. Both approaches under-estimated the actual yield of 1840 kg/ha, and the RANSAC model performed slightly better with an average yield of 1230 kg/ha versus 960 kg/ha for TPOT. There was a limitation of using UAS for height measurement - occlusion of grass by woodland and shrub affecting the height calculations as the site had sparsely located grass and woodland.

De Rosa et al. (2021) used a multispectral camera on UAS to calculate NDVI for two sites in Australia at < 1 m spatial resolution. They reported an exponential relationship between ABG and NDVI using pre-and post- grazing biomass, rainfall, temperature, fertilization, season and soil type. A generalised additive model (GAM) and RF was used to predict pre-grazing pasture biomass using UAS data. RF performed better with 27.7% error than GAM with 22.9% error. UAS are suited to biomass estimation and can provide timely data at farm and paddock scale; however, they are limited in providing regional or national scale estimates.

2.3.2 Synthetic Aperture RADAR

Synthetic aperture RADAR (SAR) sensors work in the microwave region of the electromagnetic spectrum. They have an advantage over optical satellite data owing to the fact that microwaves can work for any atmospheric conditions because of their ability to penetrate clouds. SAR can also capture data at night and are not sensitive to high values of biomass. The usefulness of SAR sensors depends on the sensor specifications such as polarisation, soil moisture and surface roughness, which can affect biomass estimation

Crabbe et al. (2019) correlated ESA satellite Sentinel 1 data with pasture leaf area index, height and above-ground biomass (AGB). The study assessed VV (vertically polarised waves transmitted and vertically polarised waves received) and VH (vertically polarised waves transmitted and horizontally polarised waves received) polarisation, scattering entropy and anisotropy, and mean scattering angle. The VH channel was significantly related to AGB with an R^2 of 0.71. In contrast, the VV channel had an R^2 of 0.35. A generalised additive model (GAM) was used to estimate AGB (R^2=0.66; RMSE=391.93 kg /ha). In a second study, Crabbe et al. (2020) used the VH polarisation to identify grass clumps (the clusters of residual grass after grazing) due to grazing because VH is sensitive to volume scattering and grass canopy.

In intensively managed grasslands, there are multiple grazing and silage cuttings in a year. It is vital to monitor such management events for accurate grass biomass estimation to get accurate estimates of total harvest yield. Taravat et al. (2019) used an artificial neural network (ANN) to detect mowing events on grasslands under different grazing and moisture conditions. The model used 50 Sentinel 2 images in 2016. The ANN inputs were VV, VH polarisation backscattering intensity and texture metrics such as homogeneity, entropy, contrast and dissimilarity. The methodology worked well, with an overall accuracy of 85.71%. The study did not consider the impact of each variable on the detection capability.

In a similar study, Sentinel 1 data were used to detect grazing and mowing events by De Vroey et al. (2021). After any grazing or mowing event, the coherence values increase due to the reduction of biomass. The summer grazing events were detected with 71% accuracy as the grass takes time to re-grow and have similar coherence values until complete re-growth. Such a method can be affected by mixed pixels due to the presence of trees or shrubs nearby, leading to false detection.

2.3.3 Multi-sensor approach

There have been studies into multi-sensor approaches, combining optical and SAR data to benefit from the specific capabilities. Wang et al. (2019) combined Landsat 8 with Sentinel 2 and Sentinel 1 to predict grass biomass using SVM and RF. The combination of Landsat 8 and Sentinel 2 provided a finer spatial and temporal resolution to capture the seasonality of pastures, such as the start and end of the season. However, there can be issues combining Landsat and Sentinel because whilst designed to be comparable, the optical sensors have different specifications and performance characteristics such as different bandwidths and band centres. For biomass above 0.50 kg/m^2, the model using Sentinel 1 had the lowest RMSE and the model with optical data- Landsat 8 and Sentinel 2 had the highest RMSE. Overall, the model with the combination of Sentinel 1, Sentinel 2 and Landsat 8 was the best, with an R^2 of 0.78 and the lowest RMSE of 0.00011 kg/m^2.

A European project, GrassQ, developed by Teagasc and Maynooth University, is a decision support system to estimate grass yield (Murphy et al., 2019). GrassQ is a multi-sensor approach with ground-based data collected using RPM, providing the data such as grass height (mm) and yield (kg DM/ha). The spectral indices were calculated using the data from Sentinel 2 and UAV. The grass biomass was estimated with Sentinel 2 ($R^2 > 0.7$) and UAV ($R^2 > 0.8$) using stepwise multi-linear regression (MLR). The SWIR bands (band 11 with 1.610 µm and band 12 with 2.190 µm) from Sentinel 2 were the most important in grass biomass estimation.

2.3.4 Optical sensors

Optical satellite data provide coarse to medium spatial resolution from visible to infrared wavelength region and are free of cost with a large geographical area for biomass estimation. A common technique to estimate grass biomass is to establish a regression equation between ground biomass and EO data, typically a vegetation index. Variables used to estimate grass biomass are such as vegetation indices, grass height (Yang et al., 2018), LAI (Punalekar et al., 2018) and Fraction of Absorbed Photosynthetically Active Radiation (FPAR) (Wu and Fu, 2018). In such studies, these variables are used as the predictor variable and, as such, can be used to train machine-learning regression models. Vegetation indices, such as NDVI, SAVI, and EVI, correlate well with grassland biomass (Yin et al., 2018) and can be calculated with satellite data across different spatial resolutions. The vegetation index and satellite data choice depend on the study area, vegetation type and structure, and application. For example, for the grass biomass estimates at a paddock-scale, high temporal and spatial resolution satellite data are required, whereas low spatial resolution data will suffice for global biomass modelling.

2.3.4.1 Statistical techniques

MODIS NDVI has been widely used to describe the phenology of grasslands and pastures due to high-frequency observations (Green et al., 2018, Ali et al., 2017b). However, the spatial resolution of the sensor is often larger than the parcel size of pastures being observed, so it will include surrounding land cover types within each pixel (known as mixed pixels).

Clementini et al. (2020) used - linear, power and exponential regression models to estimate grass biomass. Actual biomass was estimated using an RPM. EO-estimates of biomass were based on Sentinel 2 NDVI. The power function model performed best on calibration and validation dataset with RMSE of 572.29 kg/ha. The Sentinel 2 data coincided with the ground data collection date. The developed model was subsequently used to estimate biomass using MODIS, SPOT and Landsat 5 NDVI in the 21 years (1996-2017). The authors concluded that the model developed on specific vegetation characteristics and climatology can be used on another

geographical area with the same characteristics and climatology, especially where there is insufficient data availability.

Zumo and Hashim (2020) used linear regression to estimate pasture AGB. Vegetation indices from Sentinel 2 from 2017 to 2018 and temperature and rainfall data were used as input data. Correlation analysis was done for each variable with a vegetation index number (VIN) as the lowest RMSE of 0.00175 kg/m^2. During the wet season (April to September in Nigeria), the rainfall was positively related with AGB with R^2 of 0.86, while the temperature was negatively correlated with AGB with R^2 of -0.86. Regression models were developed separately for each variable, whereas there was no combined model with all the variables.

Amies et al. (2021) developed a regression model using 4 years (2015–2020) of Sentinel 2 NDVI imagery to estimate annual pasture yield (kg/ha/year). Measured yield data were collected from a literature review for 21 sites in New Zealand. A linear model was used to predict pasture yield national at field-scale by fitting median NDVI from Sentinel 2 and the measured data from 21 sites. The fitted model was used to create a national pasture yield map using median values of NDVI. The standard error was 2200 kg/ha/year. The field-based pasture yield for 21 sites ranged from 4800 t/ha/y to 17200 t/ha/y, and any values outside this range, the model will not be accurate.

2.3.4.2 Machine-learning models

Most of the studies above assume a linear relationship between features of the input data in the model and use statistical techniques to examine these relationships, such as multiple regression analysis and least-square methods. This assumption might lead to the underestimated results showing an insignificant relationship between the variables (Rasoolimanesh et al., 2018). Machine-learning algorithms can model non-linear problems (Xia et al., 2018). ML models such as ANN and SVM contains a non-linear mathematical function that can help model input-output relationships (Zekić-Sušac et al., 2014). See section 5.2.4 for discussion on ML algorithm and how they work.

In Zeng et al. (2019), NDVI and EVI, along with factors such as topographical and meteorological data, were used as an input into the RF model, which was effective in handling the non-linearity in the data (R^2 = 0.85 and RMSE=46.57 g/m^2) and was able to capture spatial heterogeneity in the grassland. However, the vegetation index was calculated using MODIS data with 250 m resolution, a coarse resolution dataset. In several studies, RF has proven to be robust, which can model complex interactions among the variables in grass biomass modelling (Gao et al., 2020a). RF performed better (R^2=0.95 and RMSE=208.88 kg DW/ha) than SVM and MLR in a study for biomass modelling with NDVI, SAVI and EVI along with topological variables (Liang et al., 2019). In a similar study by Meng et al. (2018), although RF performed the best among all the models (R^2=0.78 and RMSE=10.84%), Back propagation-ANN (BP-ANN) was used as a final model as its stability was the highest with standard deviation in R^2 of 0.062. There can be some temporal and spatial mismatch between ground data collection and the satellite data leading to inaccuracy and error in the model. For example, the use of NDVI3g data with 8 km resolution (Xia et al., 2018) and MODIS data with 250 m and 500 m resolution might not capture the phenological changes in the grass growth curve because the data is 16-day composite (Gao et al., 2020b). The higher resolution satellite images than MODIS such as L-8 (30 m) (Wang et al., 2017), S-2 (10 m) and World-View-3 (1.24 m) (Naidoo et al., 2019) can lead to high accuracy in modelling biomass in more complex grasslands with varying grassland types.

Compared to Landsat 8 data, WorldView-2 (WV2) images provide a potential for estimating biophysical parameters with higher spatial and spectral resolution and red-edge band. For example, (Zhu et al., 2017) suggested the red-edge band was the most appropriate band for estimation of LAI using SVM, RF and ANN with reasonable accuracy improvement (3.79%, 2.70% and 4.47% for ANN, SVM and RF). Apart from optical data, microwave data can also be used for biomass modelling. The microwave data gives textural information about the grass, and optical data gives the reflectance data. The complementary information from both the sensors can be combined into one model to give better accuracy than the individual models. For example, RF with Sentinel 1 had an R^2 of 0.56, and with

Sentinel 2, it was 0.60, but when both datasets were combined in the same model, R^2 increased to 0.63 (Naidoo et al., 2019).

Grass height is an indicator of biomass, especially in the early growth stages. AGB can be indirectly estimated using grass height, in which grass height can be estimated using an RF model with vegetation indices, topographical and meteorological variables as input (R^2=0.51 and RMSE=6.15 cm) (Yin et al., 2020). RF and support vector machine (SVM) algorithms could handle high correlation between the variables (Moeckel et al., 2018). RF can also be used to estimate forest height. L-band SAR data were used as an independent variable, and the reference data used was Lidar height as an input to the RF (Urbazaev et al., 2018). An important observation was that as RF averages the predictions from each tree, it could lead to under or overestimation of the values. For example, the height of small and tall trees was over-estimated and under-estimated, respectively. Machine-learning such as RF worked well in grassland system with LiDAR used for data collection. Jansen et al. (2019) observed that the canopy height and intensity data were the most accurate model with an R^2 of 0.64 as the intensity of the return wave are more for the green crop than the bare soil. Biomass can be affected by grazing intensity by the livestock. As the pixel size increases, it is difficult to observe the impact of grazing on the spatial heterogeneity of the grassland. A finer resolution is required to quantify the impact of grazing on grasslands.

Another application of machine-learning algorithms in grasslands is evaluating grassland degradation using biomass and net primary production (NPP). (Lyu et al., 2020) employed a multi-layer perception neural network (MLPNN) for grassland degradation monitoring using Hyperion data with varying degree of degradation levels. The indices used for the study were AGB, species information, vegetation coverage and soil information. The authors presented a detailed study combining ground data with satellite data, which can potentially be used as the reference method for other studies. Grassland degradation can also be indicated by the changes in nutrients of the soil such as carbon, phosphorus and nitrogen. SVM was

implemented to predict these nutrients to indicate the severity of grassland degradation (Li et al., 2017).

A comparison of parametric and non-parametric models was made for the estimation of sawgrass biomass. The non-parametric models used were SVM, RF, k-nearest neighbour (k-NN), ANN, and the parametric model used was Multiple Linear Regression (MLR) (Zhang et al., 2018). Non-parametric models performed better than parametric with R^2 of 0.58 for live biomass and 0.47 for total biomass. ANN and SVM performed better among non-parametric models with R^2 0.91 and 0.75 for total biomass receptively. Chen et al. (2021) used a multilayer perceptron (MLP) neural network to predict pasture biomass for two years, 2017 and 2018. The model was developed for 5 dairy farms in Australia using Sentinel 2, field biomass and climate variables such as precipitation and minimum temperature, solar radiation and the vapour pressure deficits as input data. Two models were developed- one with all the bands and NDVI from Sentinel 2 and the other one using Sentinel 2 and climatic data. The best model used all the variables- Sentinel 2 and climate data with R^2 of 0.60 and RMSE of 356 kg DM/ha. There were missing factors from the model such as management and soil information.

2.4 Conclusions

It was seen that none of the conventional models, empirical and mechanistic, is ideal for Irish grassland conditions, and some modifications such as adapting to a range of management and soil might be needed to predict the grass growth rate accurately (Ruelle et al., 2018b). EO methods are increasing, a good amount of data is now available. Previous work by Ali et al. (2017b) showed a promising machine-learning model which worked well for Moorepark and Grange but was limited by spatial resolution as it used the MODIS data (250 m).

The choice of data collection methods can affect the scale, complexity, accuracy and time to estimate the biomass. The methods to estimate grass growth rate are moving towards new technology such as UAS and proximal sensors. These technologies work well at a paddock scale but are limited to large geographical areas. Several studies have shown how synoptic, wide-area EO imagery can play an important role in biomass estimation on managed grasslands. There is still a need and an opportunity for improved biomass estimation using Sentinel 2 data. The model in this work will try to improve upon the developed model by bringing in more farms with a higher resolution dataset and a new algorithm.

The most important lesson from the literature on empirical, mechanistic and machine-learning models is that there is an enormous variety in grasslands in terms of management and enterprise across regions at all scales and between seasons. Such factors make the performance of algorithms developed in one site to another very variable, and that differences in seasonal performance can outmatch the model range. The EO data can capture such variability, which can provide spatial and temporal data to model national grass growth.

3

Chapter 3 Study Area & Datasets

3.1 Introduction

In Chapter 1, it was established how a grass growth model at the farm scale could have several applications, including on-farm day-to-day management and developing or supporting policies relating to food production, carbon budgeting, and climate change mitigation. This chapter presents information on the study areas used in this research and introduces the datasets used. The general study area, the Republic of Ireland, is introduced and several specific locations, including Teagasc research farms and several large commercial farms.

3.2 Geography & climate of Ireland

The Republic of Ireland is part of an island, approximately 70,273 km^2 in area, located in North-West Europe. The central portion of Ireland comprises lowlands with a ring of coastal uplands reaching 1,038 m above sea level (a.s.l.). Ireland has a maritime oceanic climate, which means mild summers and cool, but not cold, winters. Mean annual temperatures for Ireland range between 9°C and 10°C and average annual rainfall for Ireland is approximately 1230 mm (Éireann, 2012). Ireland receives between 1100 and 1600 sunshine hours, and May and June are typically the sunniest months (Éireann, 2012).

Chapter 4 focuses on just one study site, i.e. Moorepark Farm, a Teagasc research farm in County Cork, with additional sites, both research and commercial farms from PastureBase Ireland (PBI), added to it in Chapter 5. A national model was developed using 179 farms that are registered with PBI in Chapter 6.

3.3 Specific sites

3.3.1 Moorepark Farm

The Teagasc Animal and Grassland Research and Innovation Centre, Moorepark, Fermoy, Co. Cork (hereafter Moorepark Farm) (50°7 N 8°16 W; see Figures 3-1) comprises 136 paddocks (as shown in Figure 3-2) covering approximately 220.77 ha and is used for grazing and scientific experiments. A typical grazing season lasts from February to November, but the grazing season length varies according to the management and weather conditions, as shown in Table 3-1. Moorepark is an experimental farm, and the grazing season length may have become shorter from 2017 to 2020 because there are different experiments over different paddocks and the size and grass of each paddock could change every year. 10-year variations in maximum, mean and minimum temperature between 2010 and 2020 are illustrated in Figure 3-3. These data are taken from a Met Éireann weather station located on-site, which provides hourly, daily, and monthly data on rainfall, air temperature, wind speed, sunshine duration, evaporation/potential evapotranspiration and soil moisture deficit (SMD).

Soil type at Moorepark Farm ranges from sandy loamy to loamy and free-draining soils. The bedrock at this location is pale grey and fine gain carbonate limestone of the Waulsortian Formation (GSI, 2021) (Murray and Henry, 2018). A mean elevation across the farm of 40m above sea level (a.s.l.) was calculated based on the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) with a resolution of 90 m.

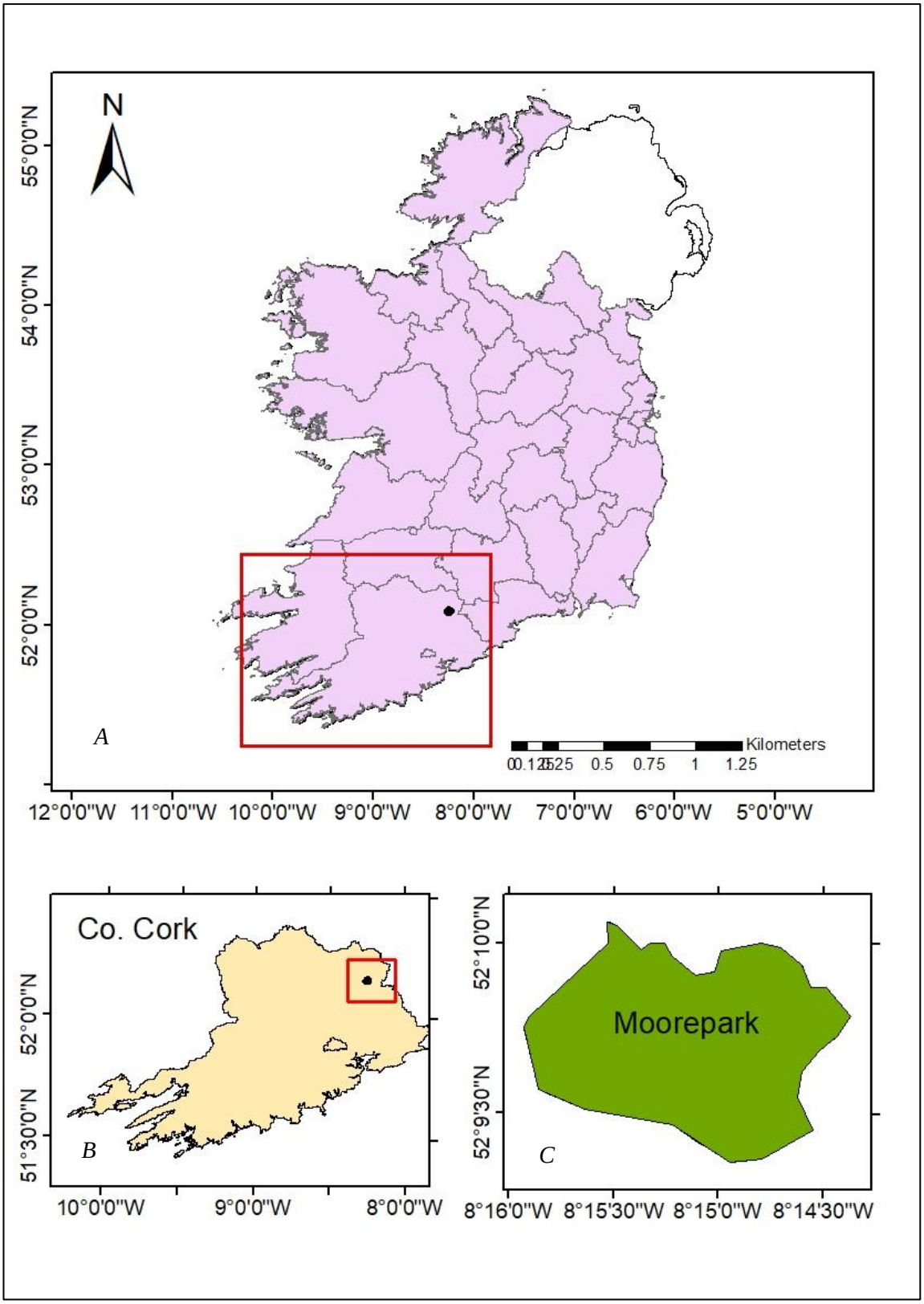


Figure 3-1 Moorepark farm in County Cork.

(A) Location in Republic of Ireland (B) Location in County Cork (C) Moorepark farm boundary

Table 3-1 Grazing season length (in days) for Moorepark farm (2017 to 2020). Grazing season length can vary depending upon weather and management conditions.

Year	Start grazing	End grazing	Grazing season length
2017	18th January (18)	9th December (343)	325
2018	3rd February (34)	6th December (340)	306
2019	2nd February (32)	3rd December (337)	305
2020	16th February (47)	26th November (331)	284

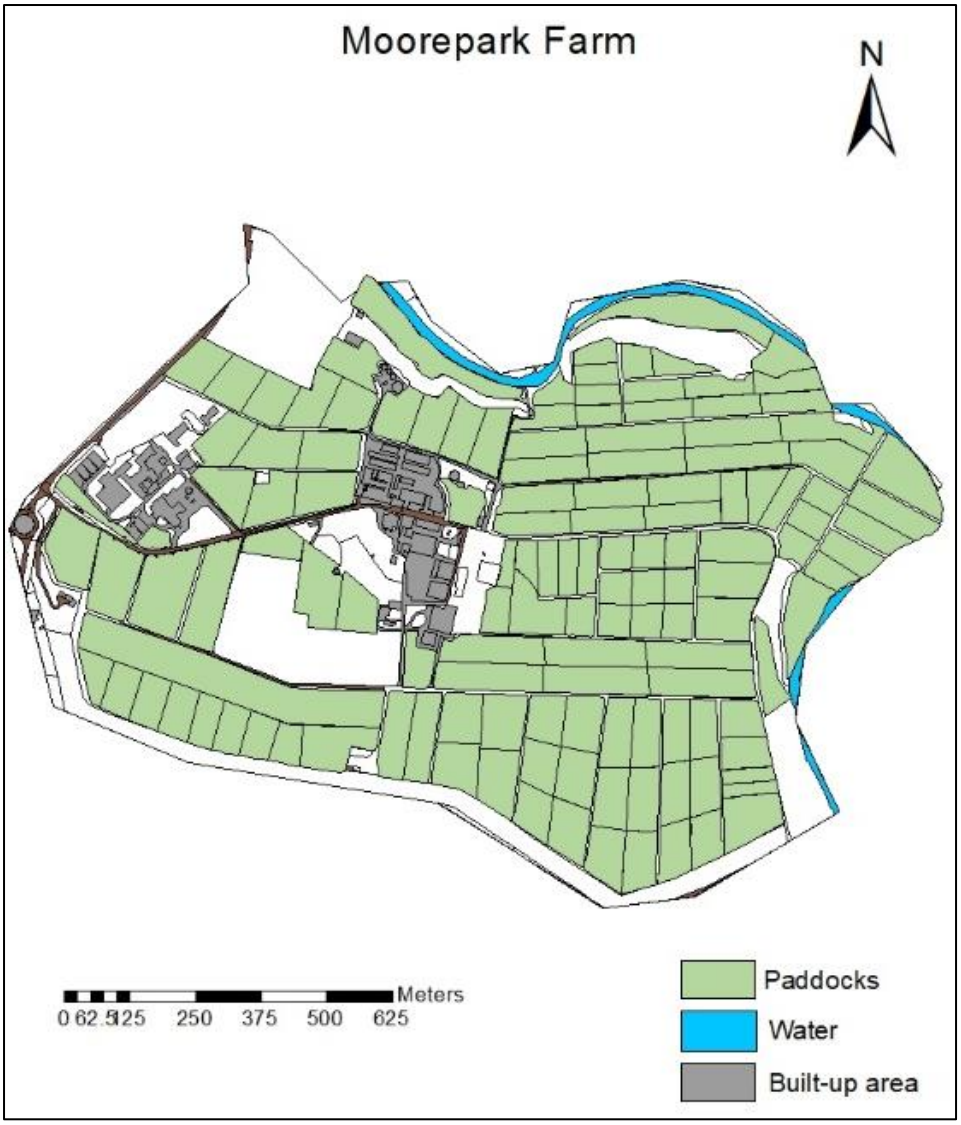


Figure 3-2 Moorepark farm showing the location of 136 paddocks

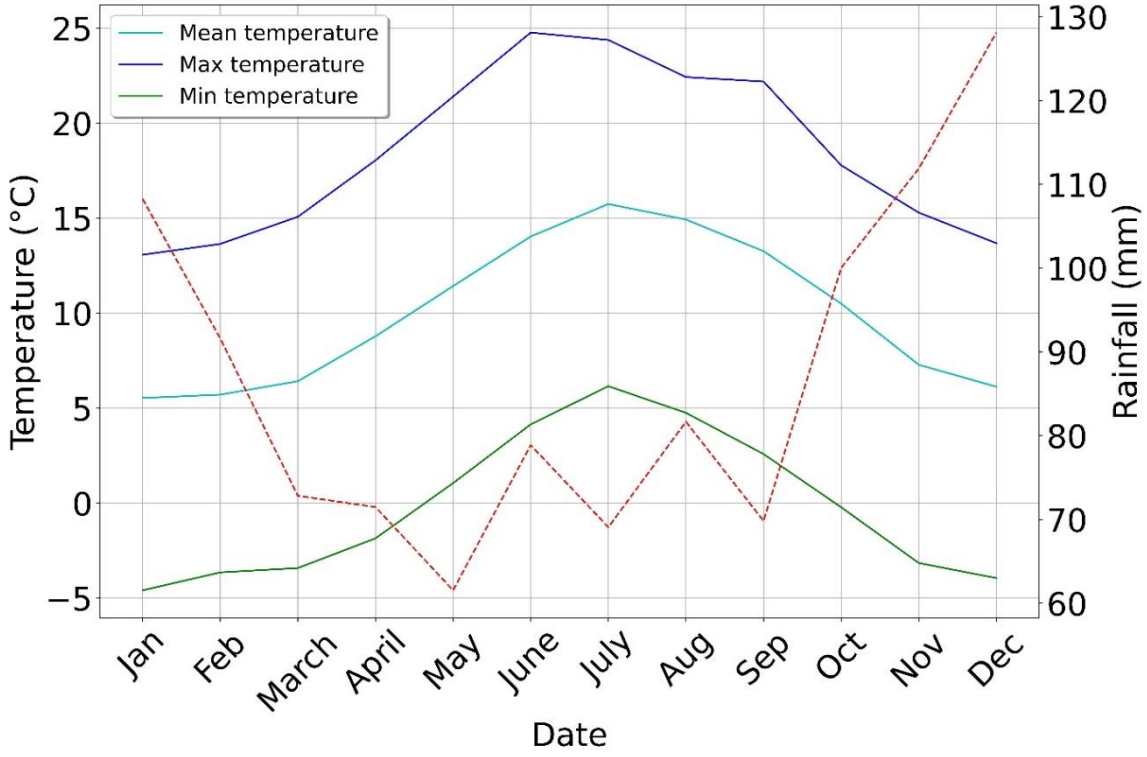


Figure 3-3 Monthly meteorological data for Moorepark farm (2010- 2020).

The average monthly mean temperature is shown in a cyan curve, the minimum temperature in green, the maximum temperature in blue and the average monthly rainfall in a red dotted curve.

3.3.2 Teagasc farms

The study area for Chapter 5 was extended to include seven additional Teagasc research farms distributed across the South and West of Ireland (see Figure 3-4). These farms are in three clusters, with three farms in Galway, three farms in Fermoy, County Cork and two farms in Johnstown Castle, Wexford. Incorporating farms nationwide is more representative of the different farming systems, farm sizes, soil types and variable weather conditions found across the country. The soil properties, elevation and bedrock type for each of these locations are presented in Table 3-2. The farms have contrasting soil and bedrock properties, with the farms in Galway and Wexford dominated by fine loamy soils, with fine and coarse loamy soil in Cork.

Bedrock was a mixture of sedimentary limestone and greywacke in Galway, pale grey limestone in Cork, and Greywacke slate is found in Wexford (extracted from 100K geological maps from the Geological Survey of Ireland (GSI, 2021). Soil association types were taken from the Irish Soil Information System (Irish SIS) project.

Table 3-2 Soil type, elevation and bedrock type for Teagasc research farms.

Farm	Soil	Elevation	Bedrock
Galway: Farm 1	Fine loamy	31.40 m	Limestone
Galway: Farm 2			
Galway: Farm 3			
Cork: Farm 1	Fine loamy	51.77 m	Limestone
Cork: Farm 2	Coarse loamy	39.25 m	
Cork: Farm 3		45.56 m	Limestone with shale
Wexford: Farm 1	Fine loamy	51.88 m	Grey-green meta-greywacke & slate
Wexford: Farm 2		57.28 m	

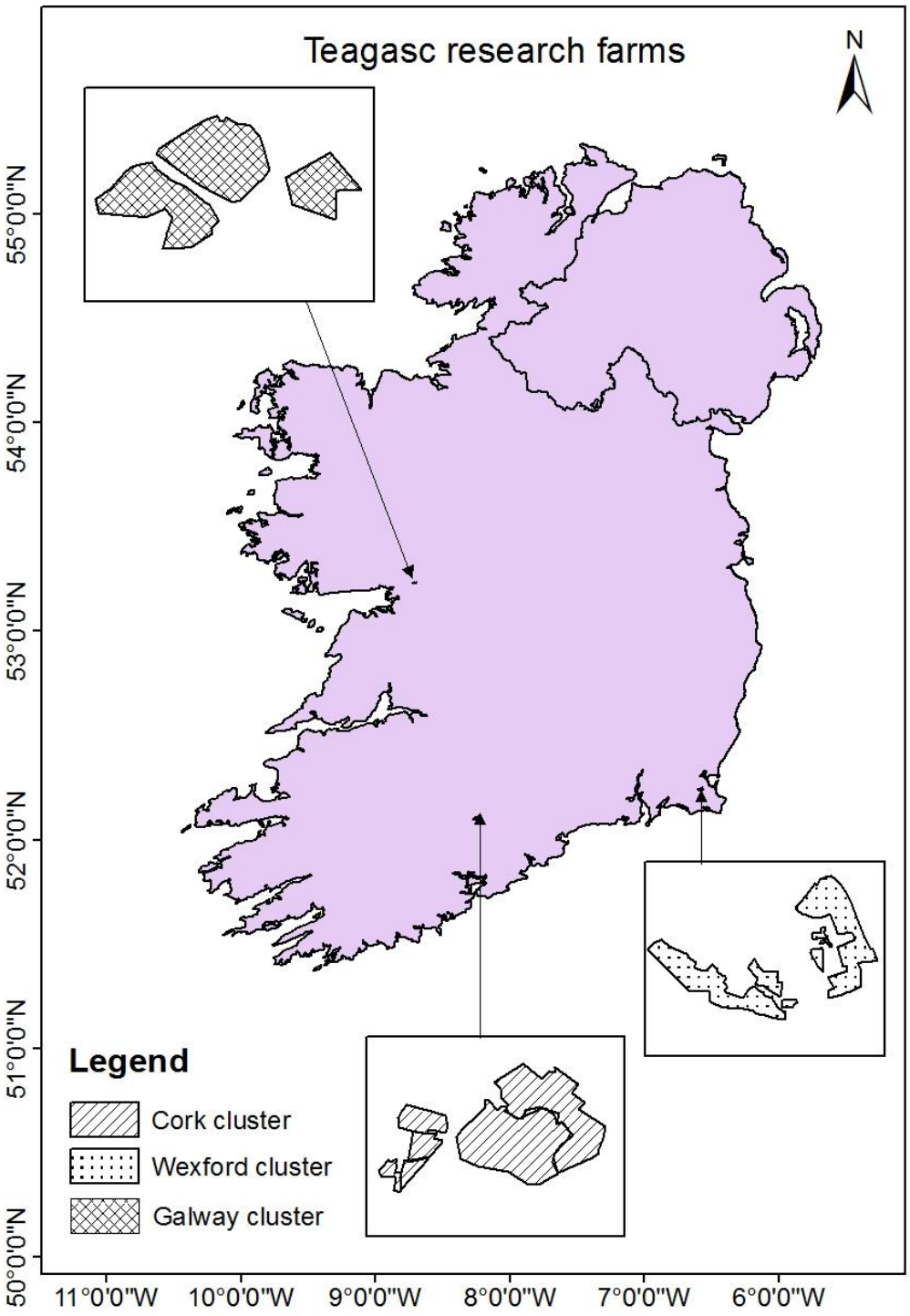


Figure 3-4. Locations of Teagasc research farms with the outline of the farms. The map shows three clusters of farms. The first cluster is in Galway in the West (crosshatched farms), containing three farms. The second cluster is in County Cork with three farms (diagonal lines), and the third cluster is in County Wexford (dotted) with two farms.

3.3.3 Commercial Farms

The study area for Chapter 6 also includes 10 commercial farms distributed across Counties Limerick, Tipperary and Cork (Figure 3-5). Table 3-3 describes these locations in elevation, proximal rain station, soil properties and bedrock type.

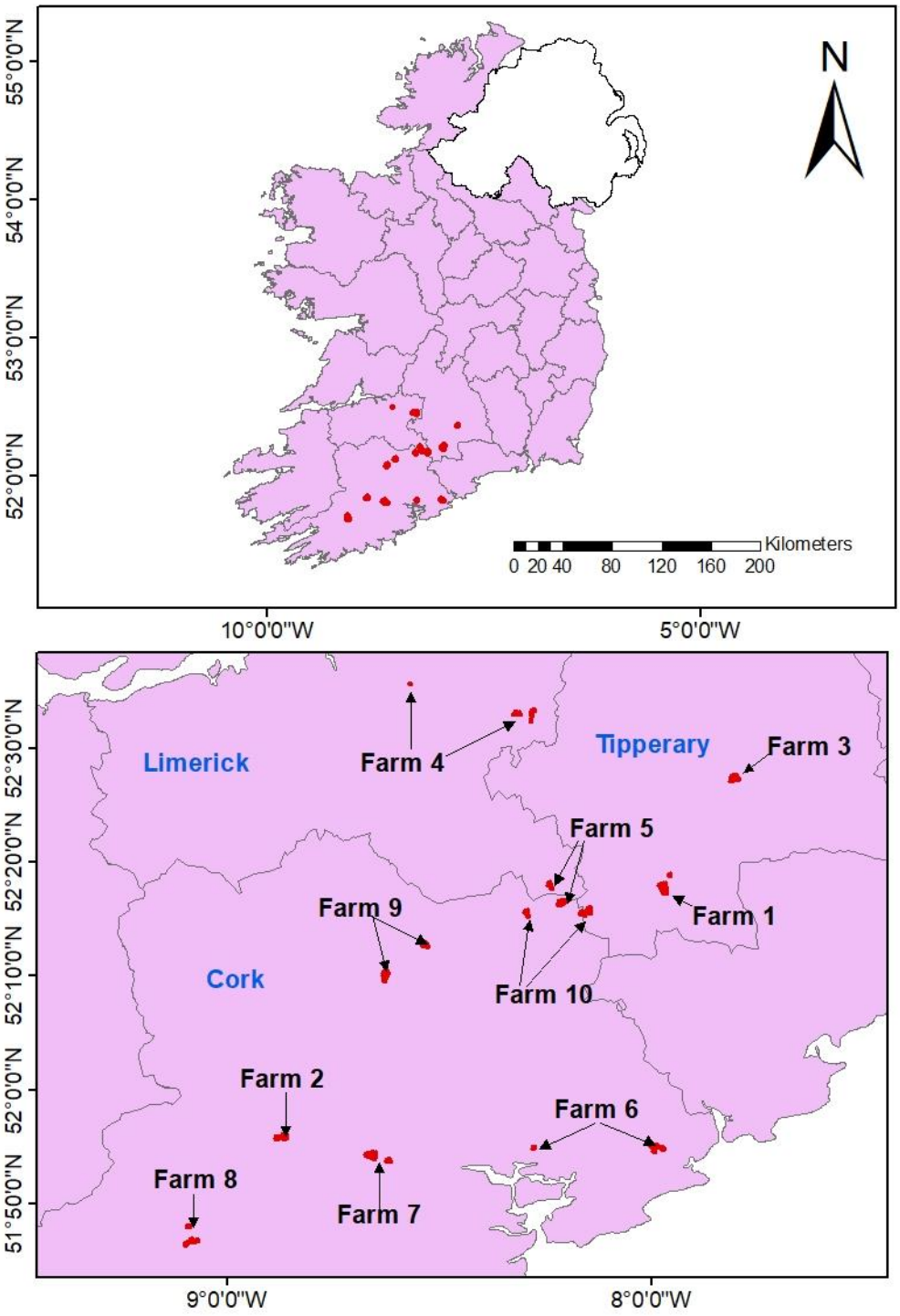


Figure 3-5 Locations of commercial farms in Counties Cork, Limerick & Tipperary A farm may consist of several parcels, such as Farm 2 and 7 in Cork has two parcels.

Table 3-3 Farm properties of the commercial farm cluster.

Table lists county, elevation, nearby rain station, soil properties and bedrock type.

Farm	Parcel	Location	Rain station/ elevation	Distance to station (km)	Bedrock	Soils
Farm 1	1	Tipperary	Ardfinnan, 46 m	5.6	Pale-grey limestone	Fine loamy
	2			5.8		
	3			6.2		
Farm 2	1	Cork	Carrigadrohid, 65 m	5.2	Dark grey argillaceous limestone	Coarse loamy
	2			4.8	Shale and sandstone	Fine loamy
Farm 3	1	Tipperary	Cashel, 123 m	0.6	Purple mudstone and sandstone	
	2			2.0		Coarse loamy
Farm 4	1	Limerick	Cooga, 88 m	6.3	Limestone	Fine loamy
	2			6.3	Basaltic lava flows	
	3			4.6	Waulsortian soil	Clayey
	4			21.2		
Farm 5	1	Cork/ Tipperary	Skeheenarinky, 335 m	6.0	Grey, fine-grained limestone	Coarse loamy
	2	Tipperary		5.2	Pale-grey limestone	
	3	Limerick		5.4	Dark-grey clean to muddy limestone	
	4			5.5		
Farm 7	1	Cork	Inishcarra, 24 m	1.7	Limestone	Fine loamy
	2			0.6	Basaltic lava flows	
Farm 8	1		M.Inchigeelagh, 299 m	4.1	Pebbly sandstones with purple mudrock	Coarse loamy
	2			5.4	Pale to dark grey limestone	
	3			5.7	Limestone	River alluvium
	4			4.6		
	5			3.4		
Farm 9	1		Mallow, 61m	3.0	Mudstone and siltstone	Coarse loamy
	2			4.3	Purple mudstone and sandstone	Fine loamy
	3			10.5		
Farm 10	1		Mitchelstown, 168 m	4.6	Purple siltstone and fine-grained sandstone	Peat
	2			4.8		
	3			5.6	Purple and green medium to coarse-grained and sandstones	Coarse loamy
	4			4.9		
	5			5.0		
	6			5.5		

Only rainfall data are available for these farms as there are no climatological stations nearby to provide temperature, global radiation, SMD, and evapotranspiration values. The average total monthly rainfall from 2015-2020 (mm) for the rainfall station closest to each farm is shown in Figure 3-6. Farm 6 was excluded from the plot as it had only 2018 rainfall data. There are two types of grass growth rate data available on PBI for this work- paddock and farm average data. The average farm grass growth rate is used for this work.

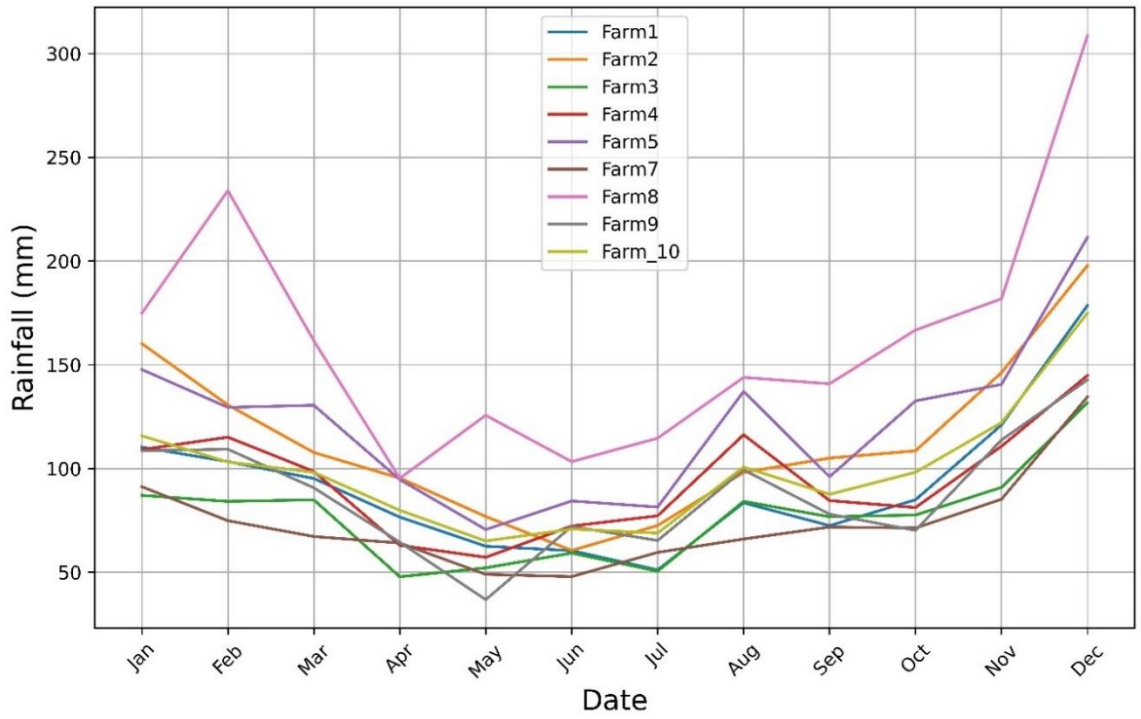


Figure 3-6 Average monthly rainfall (in mm) for 9 commercial farms (2015-2020). Rainfall stations closest to the commercial farms (excluding Farm 6)

3.3.4 National Farms

These 179 farms were located in Counties Donegal, Galway, Sligo and Cork (Figure 3-7). They were selected because they had PBI data for all four years of the study (2017 to 2020). The soil and bedrock information for these counties are given in Table 3-4. In the North, Counties Donegal and Sligo are characterised by metamorphic rocks and sandstones with peat soil. The South of County Galway consists of granite, and in the North, metamorphic rocks dominate. The South and South West of Ireland (County Cork) are dominated by sandstone and shales. There are nine major soil types in Ireland (Fay et al., 2007). Each soil type has a different drainage capacity. Alluvial soils are formed in river or lake deposits and are poorly drained. Brown Earths are a well-drained soil found in Co. Galway. Rendzinas are shallow soils found in Sligo. Donegal is dominated by peat, which contains high organic matter. The South of Ireland consists of Brown Podzolic soils rich in Iron, Aluminium, and Acid Brown Earths. The exact locations of the farms are not known because of General Data Protection Regulation (GDPR), which is a regulation on data protection and privacy in the European Union (EU) and the European Economic Area (EEA) and gives individuals control over their personal information.

Table 3-4 Soil type and bedrock for counties hosting the 179 farms used in Chapter 6

County	Soil	Bedrock
Donegal	Peat- high organic matter	Metamorphic rocks and sandstones
Sligo	Rendzinas - shallow soils	Metamorphic rocks and sandstones
Galway	Brown Earths - well-drained soil	Granite and metamorphic rocks
Cork	Brown Podzolic and Acid Brown Earths - rich in Iron, Aluminium	Sandstone and shales

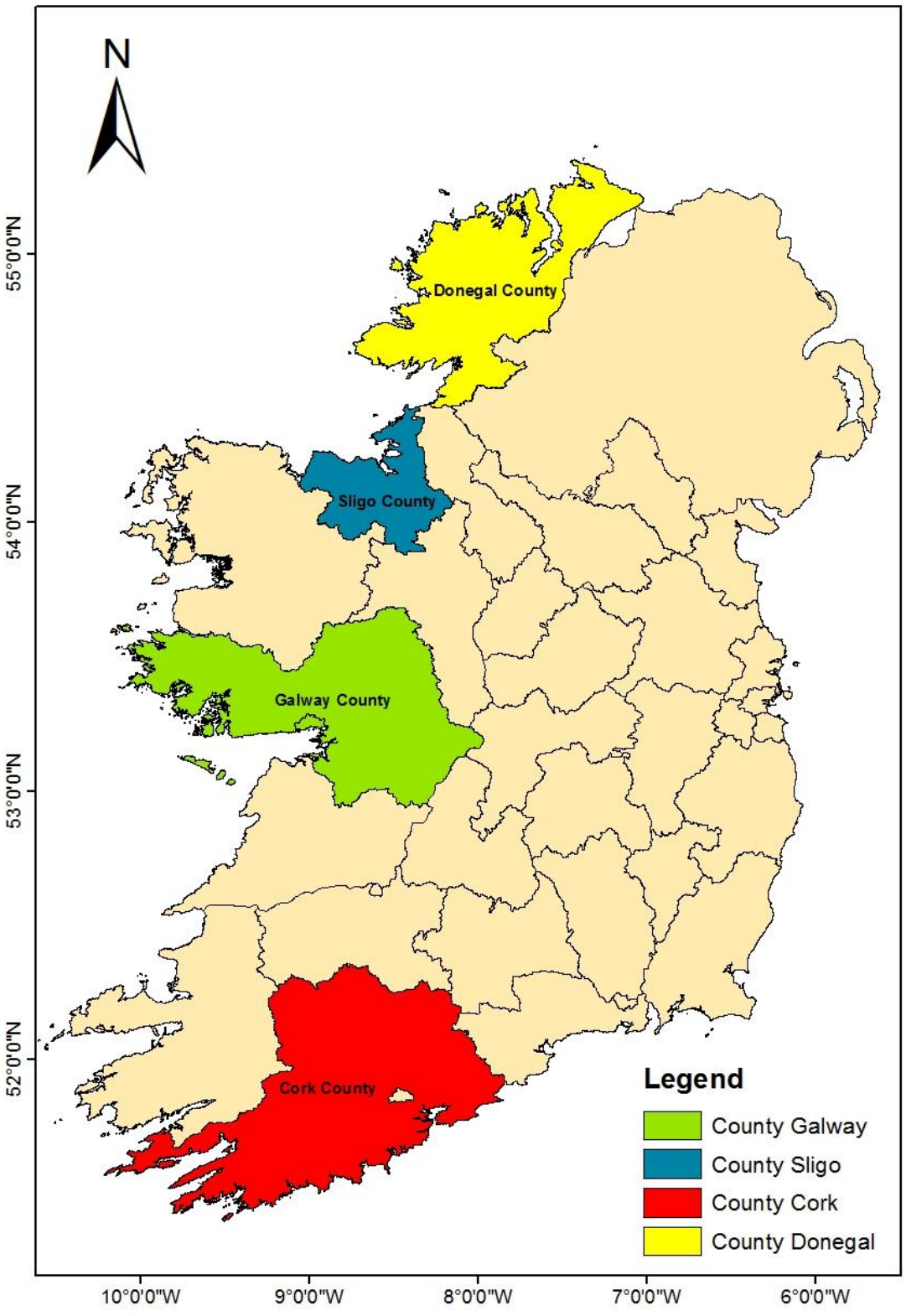


Figure 3-7 Location of farms used in Chapter 6.

The sites cover four counties- Donegal (Yellow), Sligo (Blue), Galway (Green) and Cork (Red).

3.4 Datasets

3.4.1 Landsat 8 data & processing

Landsat 8 Operational Land Imager (OLI) collection 2 multispectral imagery was downloaded from the USGS Earth Explorer website for the 8 Teagasc farms described in Section 3.3.2. The cloud-free images were downloaded as surface reflectance products, as processed by USGS with 30 m spatial resolution and 16 days temporal resolution. OLI has nine spectral bands with wavelengths ranging from 0.435 µm to 2.294 µm, as given in Table 3-5. Ireland is covered by paths 205-209 and rows 21-24. The surface reflectance data were converted into NDVI using equation 3-1, where B_{NIR} is the near-infrared band (band 5), and B_{Red} is the red band (band 4).

$$NDVI = \frac{(B_{NIR} - B_{Red})}{(B_{NIR} + B_{Red})} \qquad \text{Equation 3-1}$$

Table 3-5 Landsat 8 OLI bands and their names, spatial resolution and wavelengths

Band	Band name	Spatial resolution	Wavelength (µm)
Band 1	Coastal/aerosol	30 m	0.435-0.451
Band 2	Blue	30 m	0.452-0.512
Band 3	Green	30 m	0.533-0.590
Band 4	Red	30 m	0.636-0.673
Band 5	NIR	30 m	0.851-0.879
Band 6	SWIR-1	30 m	1.566-1.651
Band 7	SWIR-2	30 m	2.107-2.294
Band 8	Pan	15 m	0.503-0.676
Band 9	Cirrus	30 m	1.363-1.384

3.4.2 Sentinel 2 data & processing

Sentinel 2 is a constellation with two satellites. Sentinel 2A was launched in June 2015 and Sentinel 2B in March 2017. Sentinel 2 has a Multispectral Instrument (MSI) with 13 spectral bands (from 0.458 µm to 2.28 µm), as shown in Table 3-6. It has a swath width of 290 km and a 5-day temporal resolution. As the farm boundaries were not available for the PBI farms due to GDPR restrictions, the satellite data were processed by a Teagasc colleague who had access to the spatial location information and provided to this project as a vector of average NDVI values for each of the 179 farms.

The field level Sentinel 2 imagery was derived by cropping Sentinel 2 Level 2A tiles using polygons of the field outlines. The process was comprised of multiple steps. Level 1C Sentinel 2 imagery was downloaded from *the Plateforme d'Exploitation des Produits Sentinels* (PEPS) platform maintained by the French National Centre for Space Studies (CNES). A complete timeline covering the entire Island of Ireland ranging from 2017 to 2020 was acquired.

In an initial check, all polygons were tested for cloud cover for each acquisition. Cloud detection was carried out using the cloud mask created by the multi-temporal Multi-Mission Atmospheric Correction and Cloud Screening (MACCS), Atmospheric and Topographic Correction (ATCOR) Joint Algorithm (MAJA) v.3.3, developed by CNES (Hagolle et al., 2010). The MAJA algorithm produces a cloud mask raster, as well as a sensor footprint mask. Any part of a field polygon that overlapped with the cloud mask or fell outside the sensor footprint was not included for extraction of that specific polygon.

While the MAJA algorithm applies an atmospheric correction to produce Level 2A images, these images showed a large number of negative surface reflection rates, rendering many images unsuitable. Instead, Level 2A images were produced using the Sen2Cor atmospheric correction algorithm (v 2.8) (Louis et al., 2016). The Sen2Cor algorithm outputs Level-2 imagery as a set of single-band raster data.

A Python 3.6 processing chain was developed to crop the Sentinel 2 raster data, using the ‘rasterio’ package to read and crop the Sentinel 2 images and the ‘geopandas’ package to read and select the polygons for cropping. All bands were extracted separately for each valid polygon (i.e. no cloud cover and falling fully within the sensor footprint). To avoid issues with the different resolutions of the separate bands (10 m resolution: Bands 2, 3, 4 and 8; 20 m resolution: Bands 5, 6, 7, 8A, 11, 12; 60 m resolution bands were not included in the analysis), a 20 m buffer was applied to the polygon initially. The buffer ensures that for each 10 m pixel, all according to 20 m pixels were extracted as well. Then the 20 m bands were upscaled to 10 m using nearest neighbour resampling. Finally, all extracted bands were stacked into a single multi-band Geotiff. Band statistics for each image were calculated using the ‘numpy’ package in R. The sensor specifications of Landsat 8 and Sentinel 2 are shown in Table 3-7.

Table 3-6 Sentinel 2 bands and their names, spatial resolution and wavelengths

Band	Band name	Spatial resolution	Wavelength (µm)
Band 2	Blue	10 m	0.458-0.523
Band 3	Green	10 m	0.543-0.578
Band 4	Red	10 m	0.65-0.68
Band 5	Red-edge	20 m	0.698-0.713
Band 6	Red-edge	20 m	0.733-0.748
Band 7	Red-edge	20 m	0.765-0.785
Band 8	NIR	10 m	0.785-0.899
Band 8A	NIR narrow	20 m	0.855-0.875
Band 11	SWIR	20 m	1.565-1.655
Band 12	SWIR	20 m	2.1-2.28

Table 3-7 Specifications of Landsat 8 and Sentinel 2 satellites

	Landsat 8	Sentinel 2
Wavelength	0.435 – 1.384 µm (VIS – SWIR)	0.443 - 2.150 µm (VIS – SWIR)
Orbit	705 km (sun-synchronous)	786 km (sun-synchronous)
Interval	16 days	5 days
Mode	Multispectral Imager	Multispectral Imager
Swath width	185 km	290 km
Product	Level-2A Surface Reflectance	Level-2A Surface Reflectance
Spatial resolution	30 m	10 - 20 m
Operator	U.S. Geological Survey (USGS)	European Space Agency (ESA)

3.4.3 Meteorological data

Meteorological data were downloaded from Met Éireann, the Irish Meteorological Service[1], which operates a national network of instruments to record meteorological data. There are four different weather station types, which differ in the available instrumentation and the measurement interval. There are 4 manned and 20 automatic weather stations, 60 climatological stations, and 500 rainfall stations as shown in Figure 3-8. Manned and automatic synoptic stations provide hourly data for air temperature, rainfall, wind speed, relative humidity, global solar radiation and soil temperature. Climatological stations provide minute-by-minute and hourly data on air and grass temperature, humidity and rainfall. The extensive network of rainfall stations provides daily and monthly data.

[1] https://www.met.ie/climate/available-data/historical-data. Accessed 16th May 2018

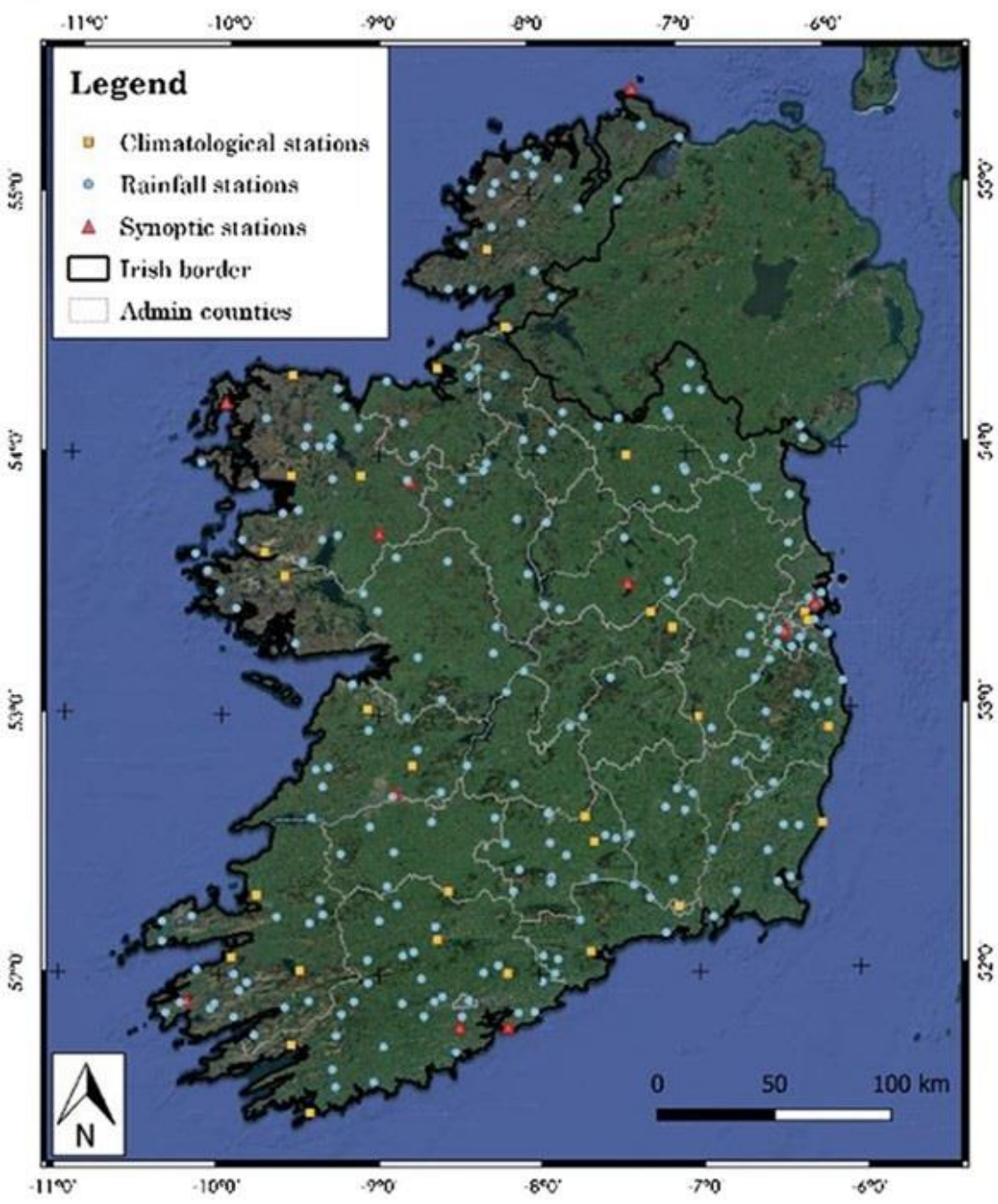


Figure 3-8 Irish meteorological stations.

Climatological stations (yellow squared boxes), rainfall stations(blue circles) and manned/automatic synoptic stations (red triangles) (Falzoi et al., 2019)

The mean, maximum and minimum air temperature, rainfall, soil moisture deficit, global radiation, potential evapotranspiration, and evaporation were used for this study. For some sites, such as the Teagasc farms, all the variables listed above could be obtained for 2017-2020. However, as new weather stations are installed and

others discontinued over time, some datasets are incomplete, so when the model was expanded to a national level, only widely available rainfall data could be used.

3.4.4 PastureBase Ireland data

Ground data is required to validate the grass growth rate models developed in this thesis. Pasturebase Ireland (PBI) is an online grass budgeting decision support system (DSS) for Irish farmers, whereby the farmers record grass growth rates in their fields. These data are uploaded to the national grassland database that is available via online and mobile applications and allows farmers to maintain up-to-date on-farm grass growth records. PBI subscribers include all Teagasc research farms, commercial farms, and some family-run, owner-operator farms. PBI has been described as an example of citizen science (Shalloo et al., 2018), where a distributed network of farmers collect data as they walk through their farm and then upload it into the web application, which can be used to gather a national picture of grass growth. Subscription to PBI also allows farmers to participate in broader discussion forums that inform farmers about cutting edge grassland research and innovation.

Grass records are added weekly for fields and paddocks using different methods (see Section 1.5.1) depending on the farmer's preference. These include a rising plate meter and visual inspection ("eyeballing") method. A paddock's status is also entered based on whether they are available for grazing, being grazed, or being kept for silage. Nutrient management and history can also be recorded. PBI uses this information to calculate a "grass wedge", an essential tool in rotational grazing systems that shows farmers whether paddocks are over-performing (have surplus grass) or under-performing (require intervention) (see Figure 3-9).

The grass wedge ranks paddocks from highest to lowest grass cover in kg of dry matter per hectare of grass (kg DM ha^{-1}) for a given week. Figures 3-9 to 3-11 show a grass wedge with a demand line (red line) to calculate pre-grazing yield and demand for the current stocking rate. It joins two points corresponding to the pre-and post-grazing growth rate. The highest point is the target pre-grazing (kg DM ha^{-1}), and the lowest point is the post-grazing grass growth rate. Target pre-grazing yield is

calculated as the product of the stocking rate (cows ha^{-1}) and demand (kg DM $cow^{-1}day^{-1}$) by the rotation length (in days) plus the target residual grass yield (kg DM ha^{-1}). The width of the columns in Figure 3-9 for the grass wedge depends on the paddock area.

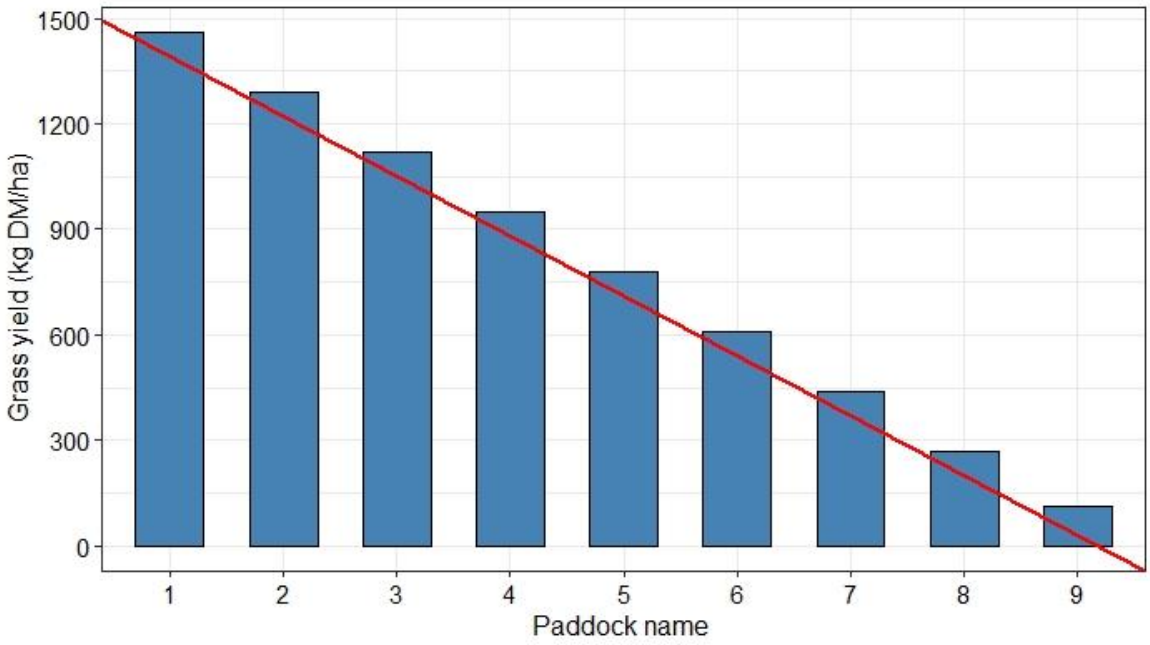


Figure 3-9 Grass wedge showing a scenario of the on-target wedge.
On target means a paddock's grass growth rate meets the farm's demand. Red line is a demand line, which connects pre- and post-grazing yield line and blue bars are the grass yield for each paddock.

According to Dilon and Kennedy (2009), there are three scenarios identified by the grass wedge: on-target, surplus or deficit.

i. On-target: All paddocks meet the demand line, as shown in Figure 3-9, an "ideal" case. There is usually a mixture of grass surplus and grass deficit on a typical farm throughout the season.

ii. Surplus wedge: Surplus wedge means that the paddocks are above the demand line, which means that those paddocks have a surplus of grass that needs to be removed. It is important to note that the pre-grazing yield should not be exceeded, and the post-grazing yield should be achieved. For example, paddocks 1, 2, 3, and 7 in Figure 3-10 have surplus grass, and 4, 5, 6, 8 and 9

are on-target. In this case, surplus paddocks exceed the pre-grazing yield and should be addressed by the farmer. The surplus grass can either be grazed or removed as silage for winter.

iii. Deficit wedge: A deficit wedge occurs when the paddocks are below the demand line. For example, in Figures 3-11, the farmer immediately addresses paddocks 1, 2, 3 and 4 in deficit. The paddocks with a lower grass growth rate than the demand can be supplemented using fertilizer, or those areas can be closed for grazing until they reach the demand level.

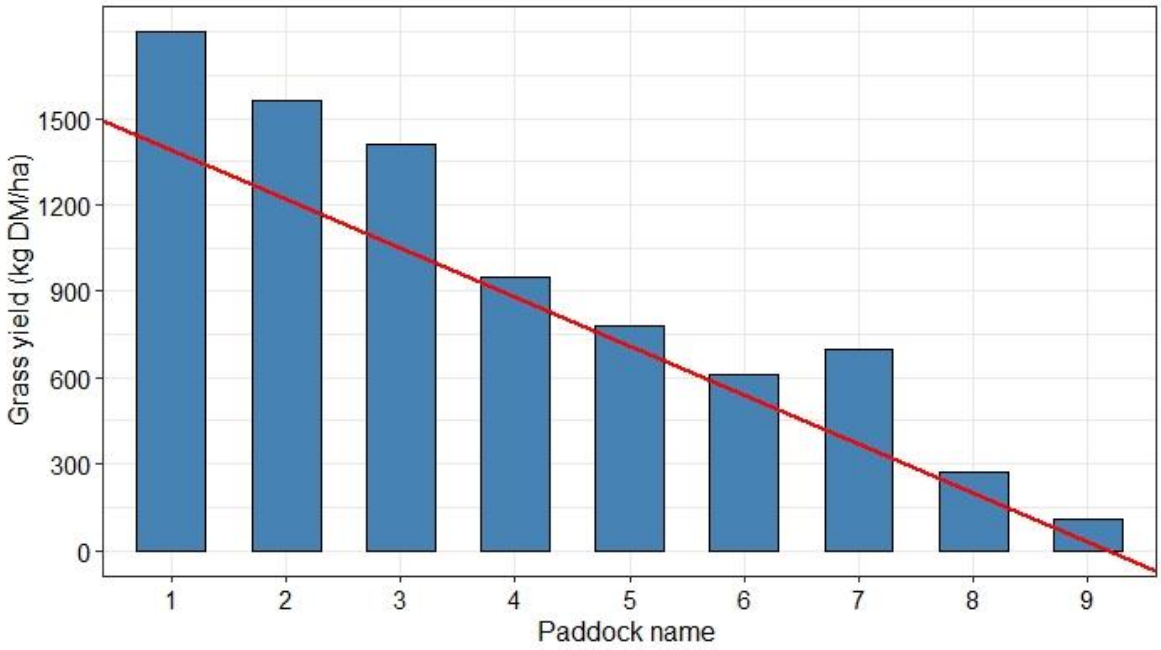


Figure 3-10 Grass wedge showing surplus condition on a farm.
The grass growth rate on paddocks 1, 2, 3 and 7 is more than the demand showing a surplus. Red line is a demand line, which connects pre- and post-grazing yield line and blue bars are the grass yield for each paddock.

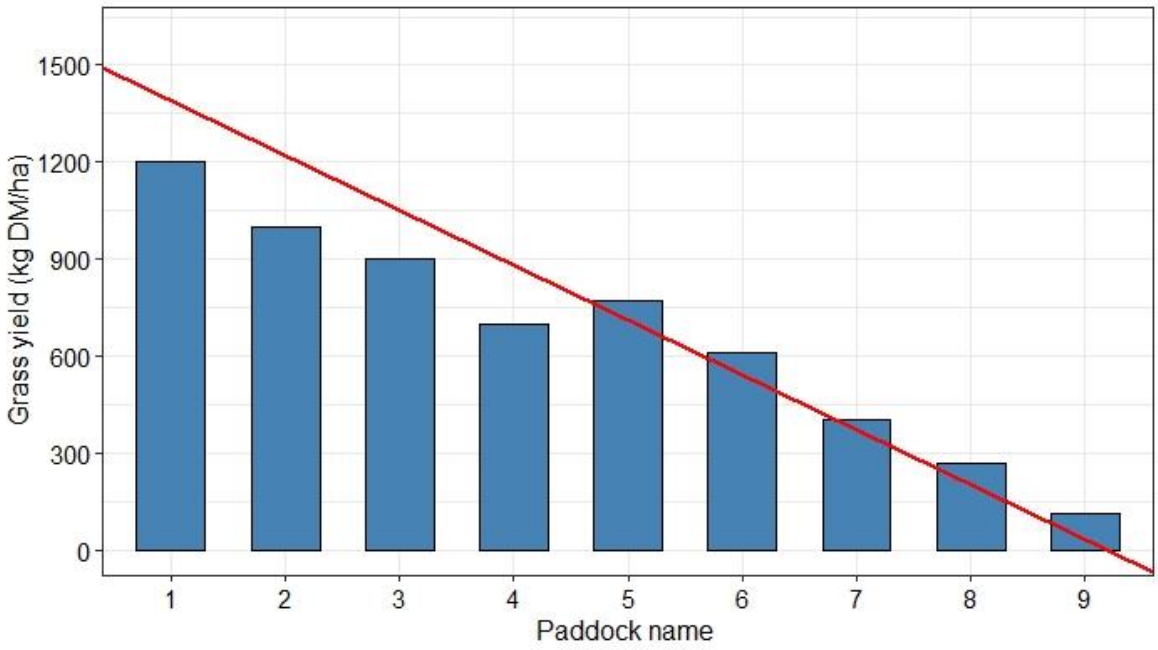


Figure 3-11 Grass wedge showing deficit condition on a farm.
The grass growth rate in paddocks 1, 2, 3 and 4 is less than the demand, which shows the farm's deficit condition. Red line is a demand line, which connects pre- and post-grazing yield line and blue bars are the grass yield for each paddock.

Figure 3-12 illustrates a real-life example of a grass wedge for Moorepark Farm, taken from PBI for 25th May 2020. The width of each column is proportional to the paddock area, for example, Paddocks C37B and C34 are 1.01 ha and 0.87 ha, whereas Paddock C39D is only 0.25 ha. Having a paddock area included in the wedge can help farmers make more informed decisions on matching the stocking rate to a particular paddock. Paddock C28A was being grazed, so it is shaded as brown. All the other paddocks are not grazed and are either surplus (C29B, C31A, C32C, C28A, C34 and C31B) or on target (C37B, C41C and C38B), except for Paddocks C39D and C40B, which are in deficit.

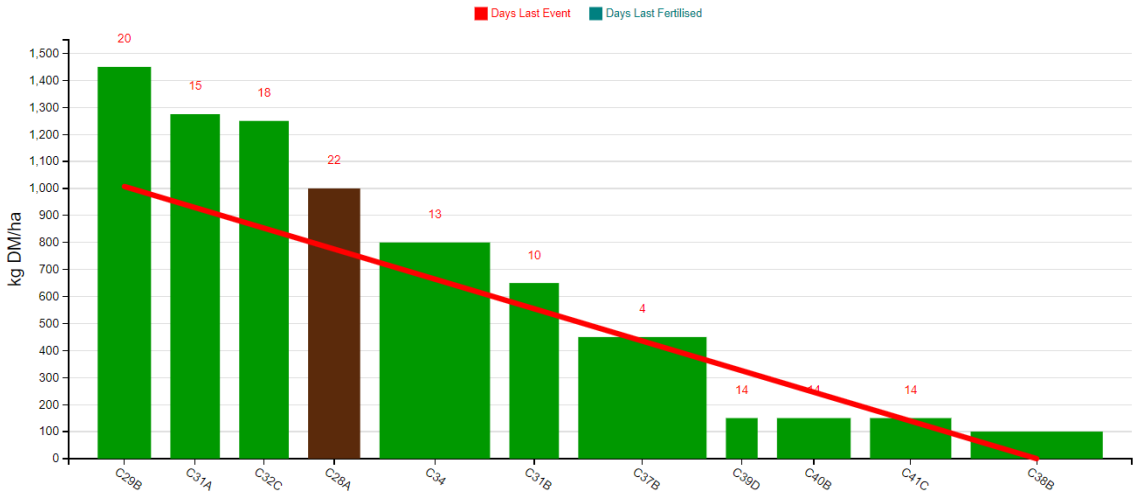


Figure 3-12 Grass wedge from PBI on 25th May 2020.
It ranks paddocks from highest to lowest grass cover (kg DM ha^{-1}). All the paddocks are shown as green meaning ungrazed, except for C28A, as it is being grazed at the moment. The width of each column represents the area of the paddock. The red line is called the demand line, which connects pre- and post-grazing yield.

Figure 3-13 is a screenshot of the PBI application showing the corresponding Table of data from which the wedge is drawn. The user must complete some parts, for example, paddock area, current grass cover (kg DM/ha), current status ("being grazed", "silage cutting", "grass available", "reseeding" or "other enterprise-area used for other than cattle grazing"), and daily growth (kg DM ha^{-1}day^{-1}). The number of feed days is the number of days a paddock will last based on current cover and demand. For example, in Figure 3-13, paddock C29B, C28A and C38B will last for two days with current grass cover. Days Last Fertilizer indicates how many days since the paddock was last fertilized, which is generated automatically, and it is column blank in the figure indicating that these paddocks have not been fertilised this year. The action column is for silage cutting, if the paddock is being cut for silage, it can be set to cut now. Feed days represents the number of days a paddock will last with the current amount of grass it has. The Last Event represented when the paddock was last grazed or cut. The grass growth rates (kg DM ha^{-1}day^{-1}) shown in the daily growth column are calculated for all paddocks with the status 'Grass' using Equation 3-2.

$$\text{Grass Growth rate} = \frac{\text{Current grass cover-Previous date grass cover}}{\text{Current date-Previous date}} \qquad \text{Equation 3-2}$$

PADDOCKS	AREA	COVER	STATUS	DAILY GROWTH	RESULTS ON 18/05/2020	COMMENT	FEED DAYS	GRAZING NO YTD	DAYS LAST FERTILISER	LAST EVENT	ACTION
C29B - C29B	0.42	1450	Grass	68	975 - Grass		2.0	2		Graze Date: 09/10/2020	Change to Silage: Cut Now Cut Later
C31A - C31A	0.39	1275	Grass	104	550 - Grass		1.6	3		Graze Date: 01/11/2020	Change to Silage: Cut Now Cut Later
C32C - C32C	0.39	1250	Grass	86	650 - Grass		1.6	3		Graze Date: 21/10/2020	Change to Silage: Cut Now Cut Later
C28A - C28A	0.41	1000	Being Grazed	(39)	1150 - Grass		1.3	2		Graze Date: 12/10/2020	
C34 - C34	0.87	800	Grass	86	200 - Grass		2.3	3		Graze Date: 22/10/2020	Change to Silage: Cut Now Cut Later
C31B - C31B	0.39	650	Grass	71	150 - Grass		0.8	3		Graze Date: 27/10/2020	Change to Silage: Cut Now Cut Later
C37B - C37B	1.01	450	Grass	(104)	1000 - Being Grazed		1.5	3		Graze Date: 12/11/2020	Change to Silage: Cut Now Cut Later
C39D - C39D	0.25	150	Grass	14	50 - Grass		0.1	0		Graze Date: 11/10/2020	Change to Silage: Cut Now Cut Later
C40B - C40B	0.58	150	Grass	14	50 - Grass		0.3	0		Graze Date: 04/10/2020	Change to Silage: Cut Now Cut Later
C41C - C41C	0.64	150	Grass	14	50 - Grass		0.3	0		Graze Date: 21/11/2020	Change to Silage: Cut Now Cut Later
C38B - C38B	1.04	100	Grass	(86)	900 - Grass		0.3	2		Graze Date: 18/11/2020	Change to Silage: Cut Now Cut Later

Figure 3-13 Screenshot of the information contained in PBI.
It shows the area of each paddock, grass cover or yield (kg DM ha^{-1}), status (grass, being grazed and silage), feed days (number of days a paddock will last with current grass cover), number of previous grazings, the day last fertilized, previous date of grazing or silage cutting, action (to change a paddock status to silage if there is an excess of grass) and daily growth rate (kg DMha^{-1}) for each paddock.

PBI also provides overall average farm values such as grass growth rate, demand, cover, grazing area, livestock unit and grass wedge option. The screenshot for overall farm data from PBI for Moorepark farm for December 2020 to February 2021 is shown in Figure 3-14. It records information including experiment, treatment, grass growth rate (kg DM ha^{-1}day^{-1}), demand ha^{-1} (kg DM ha^{-1}day^{-1}), cover (kg DM ha^{-1}), grazing area (ha), livestock numbers and livestock unit per hectare and grass wedge option.

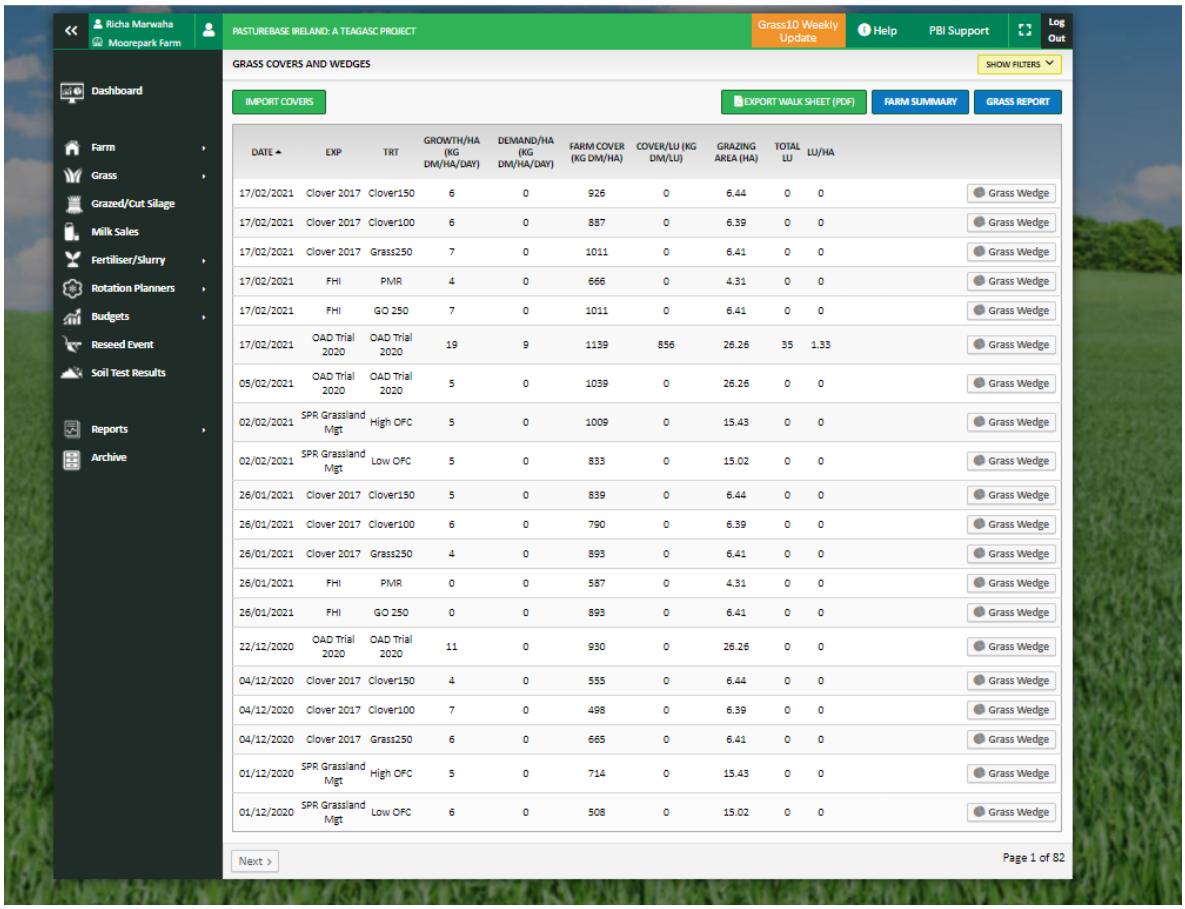


Figure 3-14 A screenshot from PBI for Moorepark Farm.
It shows the whole farm records such as experiment, treatment, grass growth rate, demand, cover, grazing area, livestock unit, and grass wedge option.

PBI provides a spring rotation planner (SRP) and an autumn rotation planner (ARP) to help farmers better utilise grass during critical times when supply may be short. During the spring period, the grass growth available is usually less than farm demand, due to which grass should be managed optimally. The SRP is a tool that helps farmers plan their first grazing of the year. The planner divides a farm into weekly portions known as paddocks and ensures that grass is provided throughout the spring, ideally until early April, depending upon the weather conditions when grass growth supply equals demand.

After spring, October is the most critical month where farmers plan for winter housing, as the grass growth rate slows down and the weather becomes wetter and colder. The ARP is intended to keep animals grazing grass for as long as possible and set up a final grazing rotation to ensure sufficient growth during winter to provide adequate grass in the following spring. The ARP subdivides a farm so that allocated paddocks match the available grass. PBI can help farmers decide when to close paddocks for grazing based on soil type, geographic location and grass availability. Using the "60:40 Rule", 60% of the total farm area should be closed for grazing by early November, while the remaining 40% should be closed between November and the beginning of the housing period (Dillon et al., 2018a). In this way, 60% of the farm has sufficient time to grow and supply grass for the following spring, and these would be the first paddocks to be grazed. The remaining 40% would be available for grazing later in the spring.

In summary, PBI provides three essential tools in one place: spring rotation planner, grass wedge during the primary season, and autumn rotation planner to assist farmers in Ireland. All these tools can be used if grass growth rate measurements have been made on the farm. PBI has an elaborate website structure, so retrieving the PBI data for analysis is not easy, and as the farms are recorded using only farm names, there is no information about the exact location or farm/field boundaries. In the following chapters, several grass growth models driven by satellite and meteorological data and validated using PBI data for the locations described in this chapter will be presented.

4

Chapter 4 Determining the Influence of Growing Degree-Days on Grass Growth Rate in Ireland.

4.1 Introduction

4.1.1 Need to monitor grass growth

In Ireland, livestock is primarily fed with grazed grass during spring, summer and autumn, and with silage (and to a lesser extent concentrates) provided during the winter housing period (O'Mara, 2012). The widespread availability of cheap and abundant fodder source, coupled with a favourable climate and long growing season, gives Ireland an economic advantage over competitors in other regions (O'Donovan et al., 2002). Grass utilisation is defined as the amount of grass harvested/grazed per hectare on a farm (McCarthy et al., 2011). While the opportunity to grow grass is high, farmers' ability to utilize the grass-grown varies according to grass growth rate, stocking rate (number of livestock units (LU) per hectare (LU/ha)), and usage of supplements. Shalloo (2009) analysed the data from representative Irish dairy farms from 2003 to 2008 and found that approximately 44% of the difference in net profits/ha between farms can be explained by the amount of grass utilised. Increasing the volume of grass utilisation on farms, for example, through improved management practices (such as rotational feeding, fertilising, reseeding and grass measurement), can improve farms' profitability (Shalloo et al., 2011, Ruelle et al., 2018a). The stocking rate of a farm is directly related to grass utilisation. For maximum utilisation of grass grown, it is essential to match the herd demand with the grass-grown. A high stocking rate means that there will be much more livestock per hectare of the farm, and there will not be enough grass to feed the cows. When the stocking rate is low, there will not be enough cows per hectare, and grass-grown will be wasted. The optimum stocking rate is when the amount of grass grown is sufficient to meet livestock demand (McCarthy et al., 2011).

4.1.2 Factors affecting the grass growth rate

As outlined in Section 1.3.3, several climatic or weather-related factors can influence grass growth rate, including air and soil temperature, rainfall, soil moisture content (excess or deficit) and growing degree-days (GDD) (Brereton and Keane, 1992). Weather influences grass in two ways. Firstly, it sets limits on the length of the growing season, and secondly, it controls grass growth rates during the growing season (Keane, 1986).

The growth rates and phenological development of grass are influenced by soil moisture deficit and precipitation. Soil moisture deficit (SMD) is the amount of water (mm) needed to bring the soil water content to the field capacity. Soil water content is the amount of water present in the soil. Field capacity (FC) is the amount of water soil can hold after the excess water has been drained away (Twarakavi et al., 2009). Water deficit is represented by positive soil moisture values, whereas negative values mean excess water in the soil. This surplus will drain away in time through surface runoff and percolation. Soil moisture depends on the soil type and the weather conditions (Schulte et al., 2005). Well-drained soil drains excess water immediately, moderately well-drained soil takes up to 24 hours, and poorly drained soil requires several days (Schulte et al., 2005). When rainfall is high during the winter months, SMD is zero for well-drained soil and less than zero for moderately and poorly drained soil.

These interrelated meteorological factors can have a significant impact on the start of the growing season. For example, traditionally, farmers look to put out cattle after winter housing by specific calendar data. In Ireland, it was traditional to let cattle out by mid-March (and usually around St. Patrick's Day on 17th March) (Green, 2019). However, grass development does not respect calendar dates, as growth can be delayed or expedited by meteorological conditions in late winter or early spring.

4.1.3 Growing degree days

Grass development in spring is affected by the amount of heat accumulation known as GDD. The concept of GDD was first introduced by Réaumur (1735), in which he stated that temperature affected the organism growth. A certain number of accumulated GDD are required to initiate growth, development and maturity depending on species. The GDD concept has been used for organisms such as insects (Hodgson et al., 2011), pests (Luedeling et al., 2011) and to predict crop's growth and maturity, for example, soybean (Kessler et al., 2020), wheat (Li et al., 2012), cotton (DeLaune et al., 2020) and grass (Calvache et al., 2020). The basic concept of GDD is that plant development will only occur when air temperatures exceed some minimum threshold. GDD is often strongly correlated with other factors, for example, soil type, rainfall and solar radiation (Hutchinson et al., 2000).

GDD is a temperature-based weather index to estimate crop growth. GDD is used as a scalar value that is input into a regression model. The essential assumption is that each plant has a base temperature below which growth stops and above which growth increases each day for every degree above this temperature until an upper limit is reached. Base temperature varies by crop type (for example, spring wheat and peas 4.5 ºC, oats 6.1 ºC and potatoes 7.2 ºC) (Burke, 1968). For temperate grasses such as perennial ryegrass, the air temperature at which growth starts is typically 5-6 ºC (Burke, 1968). Traditionally, GDD values can be used as a benchmark to differentiate various crop growth stages (Frank and Hofmann, 1989). Similarly, changes in growth resulting from extreme weather may also be evident. In this work, GDD is used as a direct response of temperature to include the impact of drought in 2018 in Ireland.

GDD is calculated using daily mean air temperature data from which a base value is subtracted. The base values are unique to each species and represent the lower temperature threshold for growth. Similarly, a maximum temperature threshold is present, above which the growth stops. The standard equation for calculating GDD is given in Equation 4-1(McMaster and Wilhelm, 1997).

$$GDD = \left[\frac{(T_{max} + T_{min})}{2}\right] - T_{Base}$$ Equation 4-1

Where, T_{Base} : Base temperature below which growth stops (5.5 °C for grass)

T_{max} : Daily maximum temperature

T_{min} : Daily minimum temperature

For Irish grasslands, the base temperature is 5-6 ºC, and for this work, the base temperature is defined as 5.5 ºC, which is in the middle of the range (Green, 2019). Different methods have been proposed to model GDD and its spatial variability in Ireland (Burke, 1968, Hargy, 1997). Such methods include temperature and geographic data such as elevation, latitude and longitude of the stations using regression algorithms. For example, Fealy and Fealy (2008) used stepwise linear regression to model the relationship between GDD and distance, distance east and north from an origin point, and elevation. Using this approach, they mapped geographical variation in GDD within Ireland for three threshold temperatures (0º C, 5º C and 10º C) using 30 years of meteorological data (1961-1990) as shown in Figure 4-1.

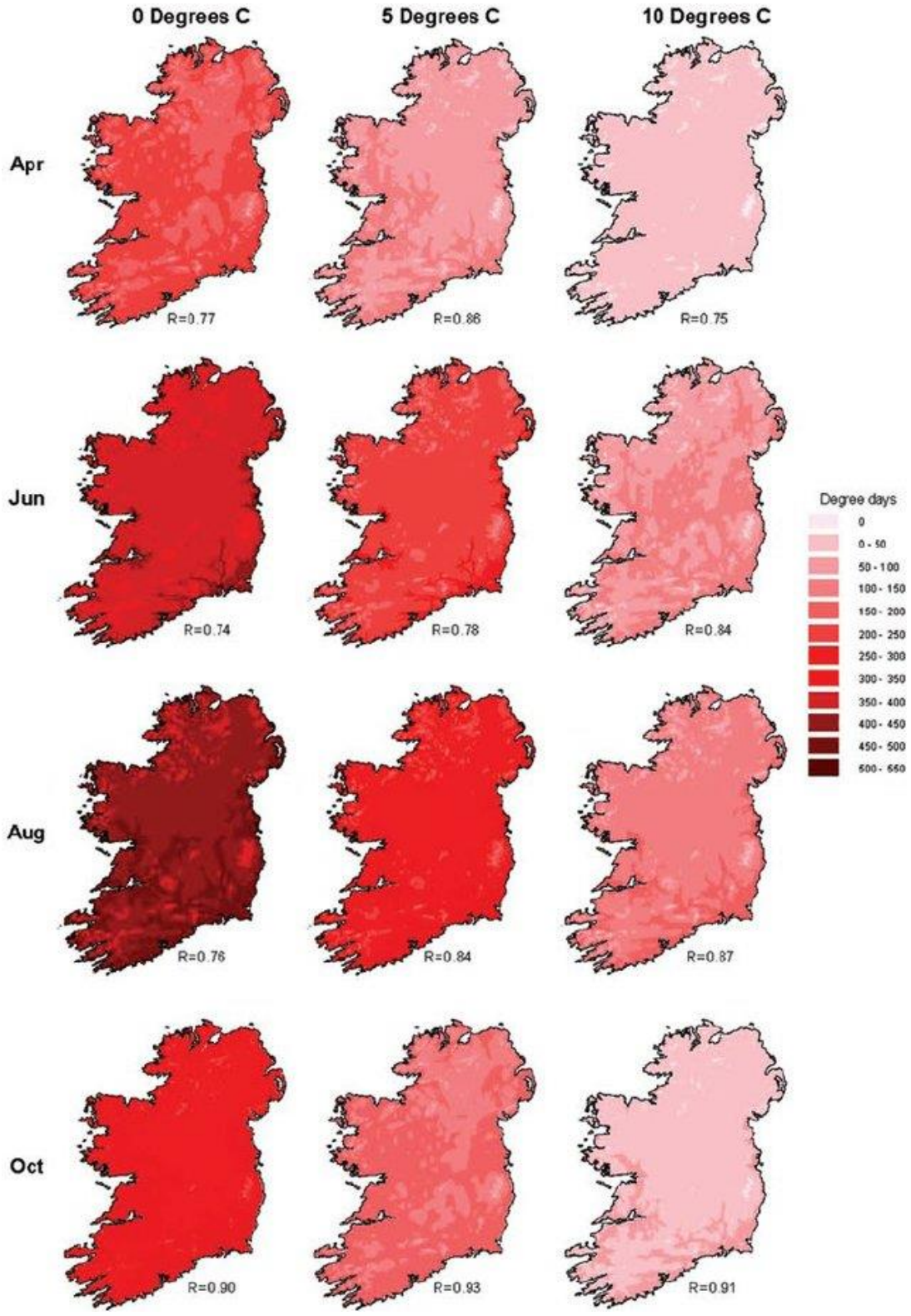


Figure 4-1 Average monthly accumulated GDD (1961-1990).
0ºC, 5ºC and 10ºC base temperature. The R-values are the correlation between observed values from the stations and predicted values from the regression model. (Source: Fealy and Fealy (2008))

McMaster and Wilhelm (1997) recommended that it is essential to clearly state the equations to calculate GDD as each method can lead to different results. They discussed two methods to calculate GDD using Equation 4-1 with the following conditions.

Method 1

If $$\left[\frac{(T_{max}+T_{min})}{2}\right] < T_{Base}$$ Equation 4-2

Then, $$\left[\frac{(T_{max}+T_{min})}{2}\right] = T_{Base}$$ Equation 4-3

Method 2

If $T_{max} < T_{Base}$ then $T_{max} = T_{Base}$ Equation 4-4

If $T_{min} < T_{Base}$ then $T_{min} = T_{Base}$ Equation 4-5

Both methods have been widely used to calculate GDD, and both generate approximated values giving different results. For example, the GDD for maize (base temperature = 10 ºC) and wheat (base temperature = 0 ºC) were calculated using both methods (McMaster and Wilhelm, 1997). However, the difference in the number of GDD using two methods was 9% for wheat and 28% for maize. Therefore, it is important to standardise the GDD equation and state the equations used.

In literature, GDD is associated with the phenology of crops and can be used to monitor various stages of development. Aslam et al. (2017) used GDD in combination with photoperiod to predict wheat developmental stages-emergence, tillering, stem elongation, flowering, and grain filling. The GDD was calculated using equations by Wang and Engel (1998). The crops were sown at a different time to investigate the impact of changing temperature and day length conditions. The GDD without photoperiod overestimated the days to flowering stage. Using the combination of the GDD with photoperiod, the days to flowering and maturity were closer to the observations. In a similar study, GDD was used to investigate wheat

phenology (McMaster and Smika, 1988), but the impact of soil water, cultivar, seeding rates, row spacing, rotation, and fertilizer on the phenology were also investigated. The GDD was related to soil water at the beginning of Jointing with the decreasing trend until maturity. GDD was sensitive to cultivar and row spacing as they affect light and nutrients, and water available to the wheat crop. GDD was not sensitive to fertilizer and planting date. GDD was used to predict the phenological stages of table grapes- budburst, flowering, and version, with the flowering stage best predicted with variability of only 4.4 days (Verdugo-Vásquez et al., 2017). GDD can help predict phenology (Aslam et al., 2017, Ahmad et al., 2017) and growing season length (Hastings et al., 2009) or as a heat stress indicator (Chen et al., 2018). GDD has also been used as a predictor variable to predict yields in maize, soybean, sorghum, spring wheat, winter wheat, and cotton (Kukal and Irmak, 2018).

This research aims to develop a national model based on GDD to estimate grass growth rate, using machine-learning with meteorological and satellite data. The GDD used in this model is a scalar quantity, which is a function of temperature rather than as a phenological factor. Meteorological variables influence the grass growth rate, whereas satellite-derived vegetation indices form a proxy for grass biomass and add spatial variability to the model. This work is for two years, i.e. 2017 and 2018, for Moorepark, County Cork. 2017 was a typical year with SMD values of 9.60 mm, compared to the 10-year average SMD of 8.20 mm, whereas 2018 was drought-affected with a high SMD of 17.49 mm during the summer, which is unusual for Ireland. The standard GDD equation (Equation 4-1) does not include the effect of extremely high temperature and SMD values. Therefore, the approach by Fealy and Fealy (2008) was modified to include conditions for high temperature and soil moisture deficit.

4.2 Study area & datasets

4.2.1 Study area

This chapter's study area was Moorepark, County Cork (50°7 N 8°16 W; (see Figure 3-1). Moorepark is a 220.77 ha dairy farm divided into 136 paddocks. Soil type within the farm ranges from sandy loamy to loamy and free-draining soils. In 2017, the grazing started on 18th January, and the last grazing was done on 9th December, i.e. a grazing season of 325 days. In 2018, the grazing started on 3rd February, and the last grazing was on 6th December, i.e. 306 grazing days. More details are provided in Section 3.3.1.

4.2.2 Meteorological data

Daily meteorological data from 2017-2018 were available from an on-site weather station operated by the Irish meteorological service, Met Éireann. Available meteorological data included rainfall (mm), evaporation (mm), potential evapotranspiration (mm), minimum, maximum, and mean air temperatures (°C), SMD (mm) and base temperature (°C).

The mean daily temperature and total daily rainfall for 2017 for Moorepark Farm are shown in Figure 4-2, The driest months are April until August, with the highest average monthly temperature in July (15.7 ºC). The winter months have an average of approximately 100 mm total rainfall. According to Met Éireann, the meteorological winter is from December (DOY 335 until 365) to February (DOY 1 until 59), spring from March to May (DOY 60 until 151), summer from June to August (DOY 152 until 243) and autumn from September to November (DOY 244 until 334).

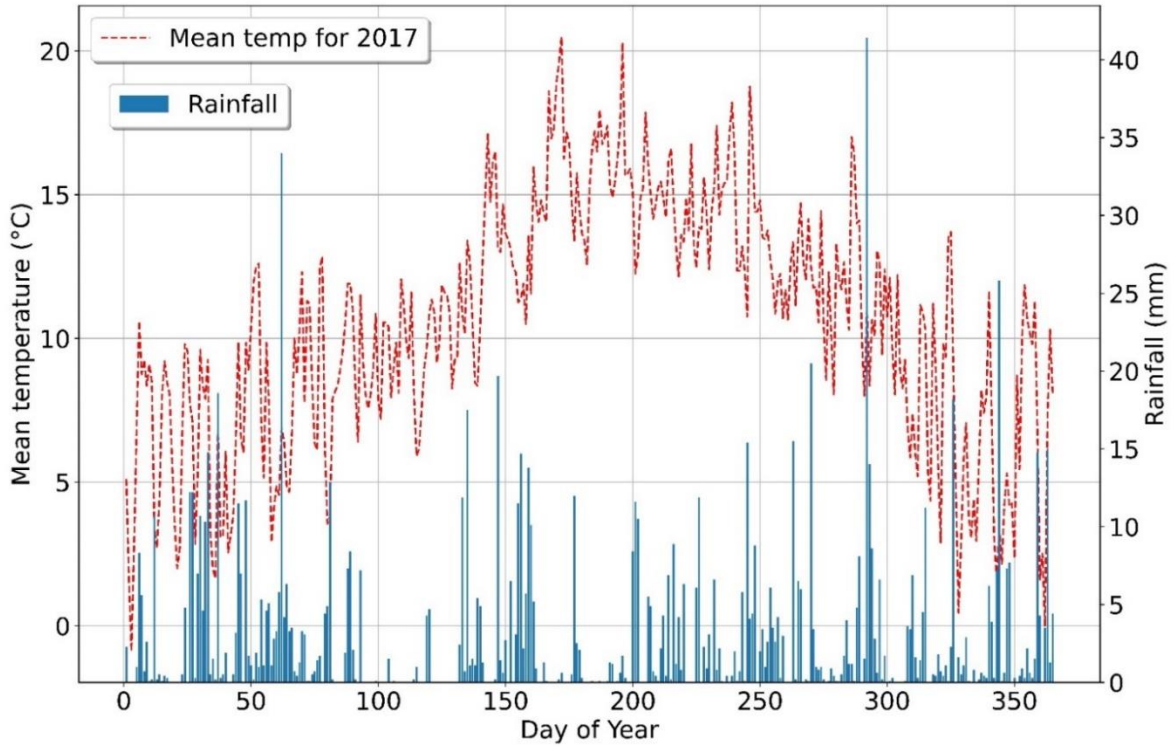


Figure 4-2 Mean temperature for Moorepark Farm in 2017.

The x-axis represents the day of the year, the y-axis on the left-hand side is mean temperature, and on the right side is rainfall. The mean temperature plot is shown in the red dotted curve, and rainfall is represented using the blue bar plot.

The mean temperature and rainfall for 2018 for Moorepark farm are shown in Figure 4-3. Between 28th February and 4th March, Ireland was affected by Storm Emma, a significant snowfall event bringing the low temperatures that delayed the start of spring. The lowest rainfall months in 2018 were May until October, with high mean summer temperatures indicating drought conditions. November, December and January were wet months. The highest maximum temperature in June was 26.9ºC in 2017 and 30.1ºC in 2018. The mean highest daily temperature was 20.5ºC in 2017 and 21.5 ºC in 2018. The highest total accumulation of rainfall for 2017 was 115.80 mm in March and was 174.80 mm in April 2018. The accumulated total rainfall was 219.70 mm for the summer of 2017 (June, July and August) and was 118.6 mm for 2018. The temperatures were higher during the summer of 2018 (16.36 ºC) than in 2017 (15.03 ºC).

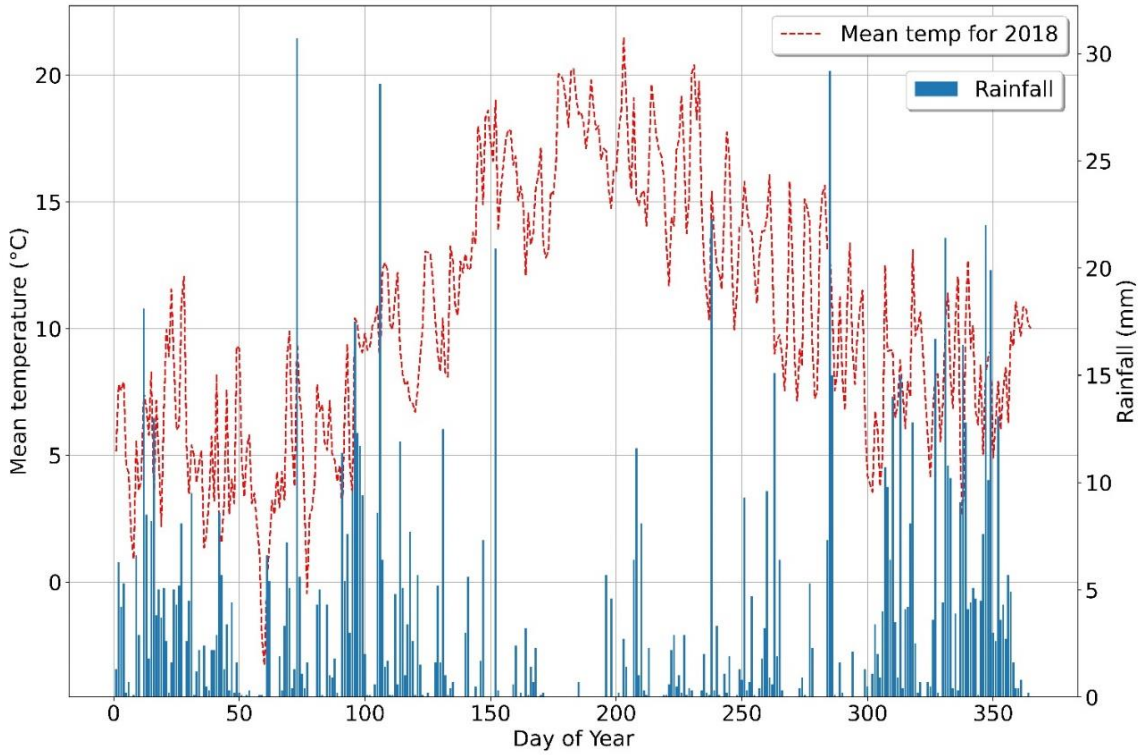


Figure 4-3 Mean temperature for Moorepark Farm in 2018.
Temperature is the red dotted line. Rainfall is the blue bar plot.
The x-axis represents the day of the year, the y-axis on the left-hand side is mean temperature, and on the right side is rainfall. The mean temperature plot is shown in the red dotted curve, and rainfall is represented using the blue bar plot.

4.2.2.1 Effect of SMD on GDD

Figure 4-4 presents the rainfall and SMD for 2017. The rainfall is represented in blue bar plots, and SMD is shown in the dotted red curve. The descriptive statistics for 2017 and 2018 were compared with the 10-year average from 2006-2016, which included mean, standard deviation, median, minimum and maximum values as shown in Table 4-1. There were 365 observations for 2017 and 2018. For ten years of data, there were 4013 observations except for potential evapotranspiration, evaporation and soil moisture deficit, which had data from 2011 until 2016. The relative percentage difference from the 10-year average for minimum and mean temperature was higher for 2017 (11.80% and 5.5 %) than 2018 (6.23% and 4.84%), showing that despite the long, warm summer of 2018, the previous year had a higher annual mean temperature. The SMD for 2018 was 113.30% higher than the 10-year average, whereas for 2017, SMD was 17.07% higher than average.

Table 4-1 Summary of the daily meteorological variables for Chapter 4.

Maximum, minimum and mean temperature, rainfall, solar radiation, air temperature, SMD, evaporation and potential evapotranspiration (PE) for 2017, 2018 and 10-year average (2006-2016)

Daily variables	N			Mean			Standard deviation			Median			Minimum			Maximum		
	2017	2018	10-yr	2017	2018	10-yr	2017	2018	10-yr	2017	2018	10-yr	2017	2018	10-yr	2017	2018	10-yr
Max temp (°C)	365		4013	14.29	14.48	13.89	4.49	5.66	4.72	14.20	13.70	13.90	3.60	-0.70	-4.70	26.90	30.10	28.90
Min temp (°C)				6.63	6.30	5.93	4.57	5.02	4.91	7	6.80	6.20	-5.70	-5.90	-12.3	15.90	16.90	17.80
Mean temp (°C)				10.46	10.39	9.91	4.26	5.04	4.54	10.70	9.95	10.15	-0.90	-3.30	-7.7	20.50	21.50	22.70
Rainfall (mm)				2.78	2.95	2.85	4.91	4.91	5.21	0.60	0.60	0.50	0.00	0	0	41.40	30.70	44.80
SMD (mm)			2166	9.60	17.49	8.20	16.12	27.12	14.35	1.70	2.80	2.70	-10	-10	-10	58.80	82.90	77.00
PE (mm)				1.48	1.58	1.39	1.05	1.25	0.99	1.30	1.20	1.20	0	0	0	4.60	5.40	5.10
Evaporation (mm)				2.02	2.15	1.92	1.44	1.66	1.37	1.70	1.60	1.60	0.10	0.10	0	6	7	6.60
Radiation (MJ/m2)			4012	930.35	994.87	919.83	686.66	776.69	677.96	756.00	780	753.50	44.00	39.00	20.00	2910.00	3065	3007

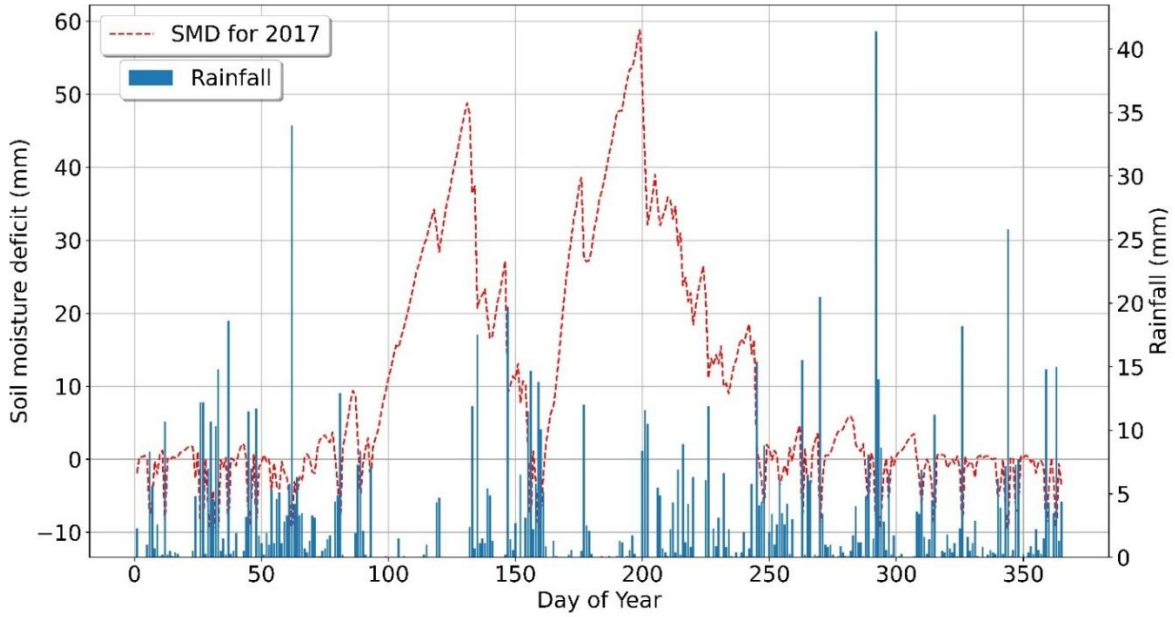


Figure 4-4 Soil Moisture Deficit for 2017.

SMD shown in red dotted line curve and rainfall shown in the blue bar plot for 2017. The x-axis is the day of the year, the y-axis on the right side is rainfall, and on the left side is soil moisture deficit.

Figure 4-5 presents the rainfall and SMD for 2018. The average values for maximum, minimum, and mean temperature, rainfall, potential evapotranspiration and evaporation were lower than the 10-year average. The SMD values are much higher for 2018 (17.49 mm) than the 10-year average (8.2 mm). The average daily global radiation value for 2018 was 994.87 MJ/m^2, which is 8.15% higher than the 10-year average values (919.83 MJ/m^2).

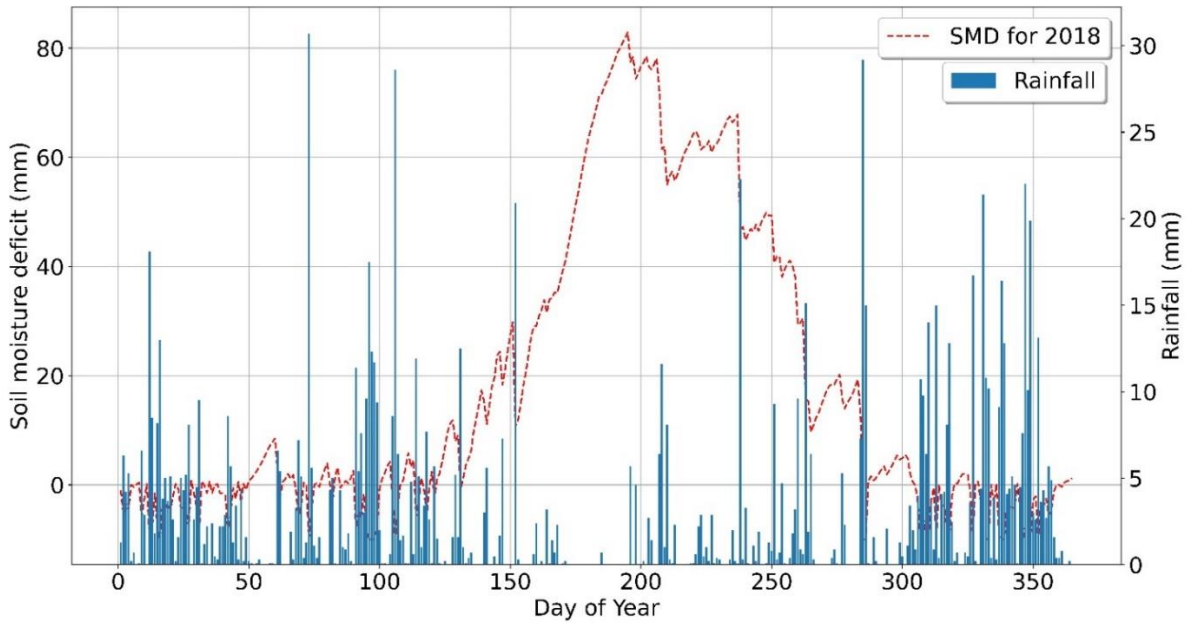


Figure 4-5 Soil Moisture Deficit for 2018.

SMD shown in red dotted line curve, and rainfall shown in the blue bar plot.

The rainfall is represented in blue bar plots, and SMD is shown in the dotted red curve. The x-axis is the day of the year, the y-axis on the right side is rainfall, and the one on the left side is soil moisture deficit.

According to Murphy (2020), agricultural drought refers to no crop growth due to increased soil moisture deficit. Daily SMD values between 50 to 75 mm indicate restricted growth, and SMD of more than 75 mm indicate drought conditions. The daily SMD values for 2017 and 2018 and the 2012 to 2016 average for Moorepark Farm are shown in Figure 4-6. The y-axis represents SMD in mm, and the x-axis is the day of the year. The SMD for 2017 is shown in a blue colour plot and 2018 in a cyan colour line plot, and the average as a red line plot. Compared to 2017, there are high SMD values in 2018 (>75mm), indicating drought conditions for 18 days from 7th July - 25th July 2018.

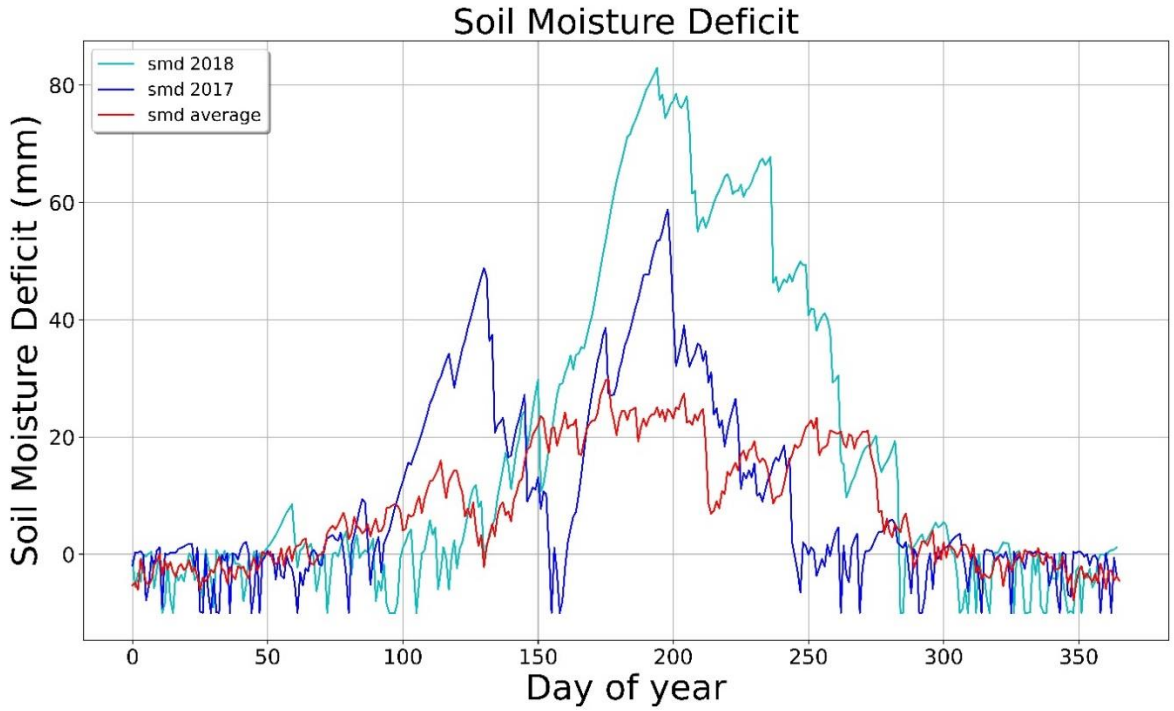


Figure 4-6 Inter-annual comparison of SMD (mm)

Comparison of 2017 (blue) and 2018 (cyan) versus average SMD from 2012-2016 (red)

4.2.3 Validation data

Grass growth rate (kg dry matter hectare^{-1} day^{-1}) values were used from PBI, discussed in more detail in Section 3.4.4. The data were collected weekly at Moorepark Farm during 2017 and 2018 using an RPM.

4.3 Methodology

4.3.1 Modified GDD equation

In this study, the GDD equation adopted from (Fealy and Fealy, 2008) was modified to include the upper and lower temperature thresholds as shown in Equation 4-6 and 4-7. For the onset of grass growth, the base temperature (T_{base}) used in this research was 5.5° C (Green, 2019). Above 15° C, grass growth starts to slow down, and temperatures above that threshold do not significantly contribute to grass growth development (Equation 4-7). Therefore, any maximum temperature above 15° C was set to 15° C. At 30° C, the grass-root starts to experience heat stress, which negatively contributes to the development. Ireland rarely experiences temperatures of 30° C and Moorepark had 30.1°C on 28th July 2018. At a temperature above 30° C,

the grass growth stops, and therefore GDD was set to zero (Equation 4-6). Moreover, if there is limited moisture in the soil, there is no growth of grass. Therefore, on days where SMD exceeds 70 mm, GDD was set to zero (Equation 4-6). Equations 4-8 to 4-11 are from (Fealy and Fealy, 2008)and the modification conditions added to the GDD calculations are in Equation 4-6 and 4-7.

Where T_{max}≥30 ºC or SMD≥70.0 mm

$$GDD = 0 \qquad \text{Equation 4-6}$$

Where T_{max}>15 ºC

$$T_{max}=15\ ºC \qquad \text{Equation 4-7}$$

Where $T_{min}>T_{base}$

$$GDD = \left(\frac{T_{min} + T_{max}}{2}\right) - T_{base} \qquad \text{Equation 4-8}$$

Where $T_{max}<T_{base}$

$$GDD = T_{base} - \left(\frac{T_{min} + T_{max}}{2}\right) \qquad \text{Equation 4-9}$$

Where, $T_{max} > T_{base}$, $T_{min} < T_{base}$ and $T_{mean} > T_{base}$

$$GDD = \left(\frac{T_{max} - T_b}{2}\right) - \left(\frac{T_b - T_{min}}{4}\right) \qquad \text{Equation 4-10}$$

Where, $T_{max} > T_{base}$, $T_{min} < T_{base}$ and $T_{mean} < T_{base}$

$$GDD = \left(\frac{T_b - T_{min}}{2}\right) - \left(\frac{T_{max} - T_b}{4}\right) \qquad \text{Equation 4-11}$$

4.3.2 Statistical analysis

Python 3.8 was used to calculate the correlations between all available meteorological variables and grass growth. The packages used were ‘Numpy’, ‘Matplotlib’, ‘Pandas’, ‘sklearn.metrics’ and ‘scipy.stats’. Correlation analysis was performed to identify the variables highly associated with each other and with grass growth. The Pearson’s coefficient of variation was calculated by dividing the standard deviation divided by the mean and expressing as a percentage. The ordinary least square (OLS) analysis model was implemented in Python using the ‘linear_model’ library from ‘sklearn’ package. In the models, grass growth was the dependent variable, and independent variables were air temperatures, rainfall, solar radiation, evaporation, and potential evapotranspiration. Separate regression models were developed for each year.

4.4 Results & discussion

4.4.1 Grass growth rate & GDD

The 10-day moving average of grass growth rate for 2017, 2018 and average from 2013-2016 are compared in Figure 4-7. The 2018 grass growth was well below 2017 values for Moorepark Farm, between DOY 20 and 110. For DOY 121-128, there was a dip in grass growth values because of silage cutting in 2018 and for average values. For DOY 130-150, the grass growth in 2018 was in excess of 2017 (mid to end May). The grass growth rate reduced from DOY 160 until 200 (June and July). The lowest grass growth rate during summer was 22 kg DM $ha^{-1}day^{-1}$ at DOY 202 (21st July 18) and 51.64 kg DM $ha^{-1}day^{-1}$ for 2017 at DOY 198 (17th July 17). At DOY 262, the grass growth rate values for 2018 reached close to the values of 2017, and the values remained very similar for the last two months of the year.

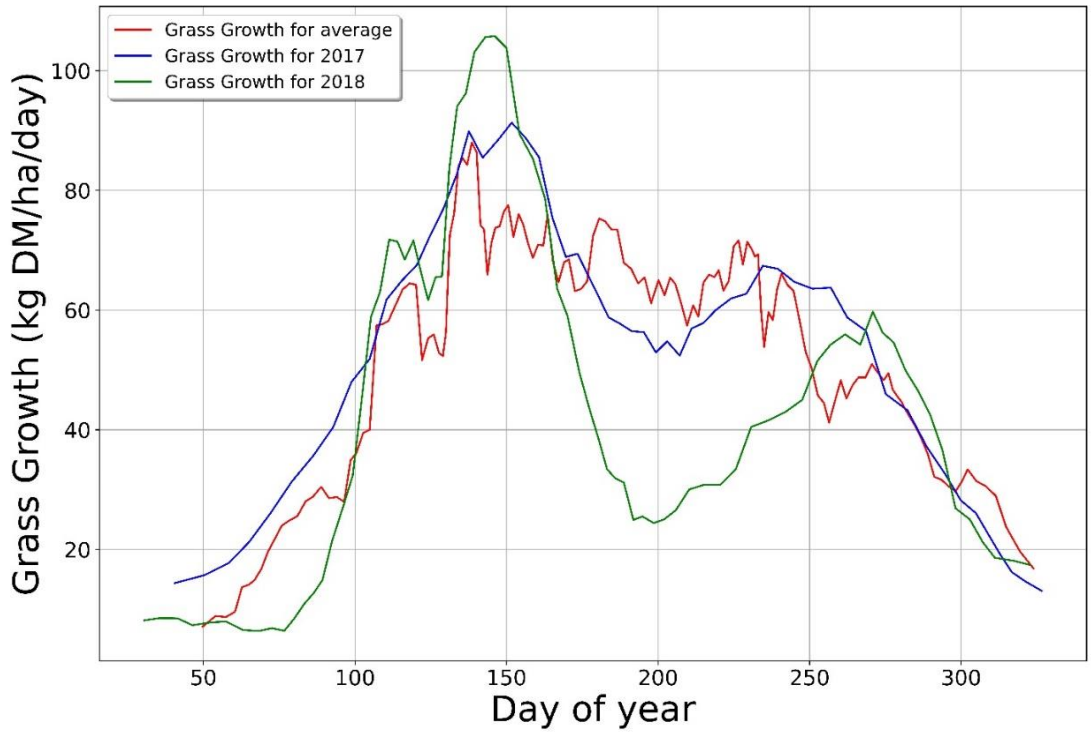


Figure 4-7 Grass growth rate for Moorepark farm (kg DM $ha^{-1}day^{-1}$). Curve for 2017 (blue), 2018 (green) versus 2013-2016 average (red). The x-axis is the day of the year, and the y-axis is the grass growth rate in kg DM ha^{-1} day^{-1}. The grass growth rate for 2017 is shown in a blue curve, for 2018 is shown in a green curve, and the average grass growth from 2013-2016 is in red.

Figure 4-8 shows the 7-day moving average of grass growth rate for 2017 and GDD as calculate using both the standard and modified equations. For 2017, a similar GDD trend is identified using both standard and modified formulae. GDD does not match growth in spring and summer, only in autumn because GDD is a function of only temperature and other factors such as rainfall, potential evapotranspiration and management effects are missing which are prominent during spring and summer. For spring (DOY 61- 152, i.e. 1st March- 31st May) average spring grass growth rate at Moorepark Farm for 2017 was 62.22 kg DM $ha^{-1}day^{-1}$. The modified GDD follows the same pattern as the standard GDD, but the values using the modified equations are lower by approximately 11.66% than using the standard method. The average GDD value from DOY 150-260 (30th May-17th September) for 2017 is 7.05 compared to 8.80 for 2018. From DOY 310 until 350, the GDD values were the same using both methods.

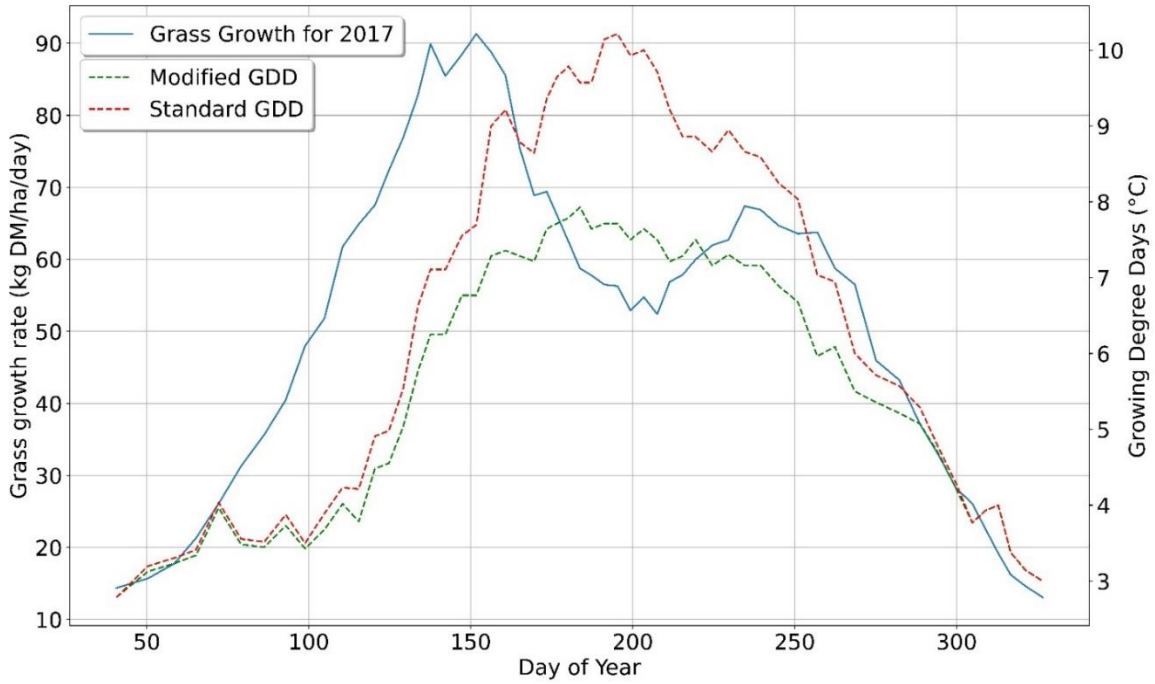


Figure 4-8 Comparison of the 7-day moving average of grass growth rate & GDD (2017). Grass growth (blue line). Standard GDD (red dotted) and modified GDD (green dotted). The modified GDD includes the impact of high SMD values. The x-axis is the day of the year, the left-hand y-axis is the grass growth rate, and the right-hand side y-axis is GDD.

Figure 4-9 shows a comparison of the 7-day moving average grass growth rate (blue plot) and GDD calculated using the standard (red dotted plot) and modified methods (green dotted plot) for 2018. The GDD values from DOY 50- 143 (19th February-22nd May) using both the methods had the same values. At the end of the grass growing season, i.e. from DOY 271 to 365, the GDD values using both the methods were the same. From DOY 184 to 207, the value calculated using the standard GDD method was 12.22, whereas GDD was 0 using the modified method because the SMD values were more than 75 mm and therefore defined as drought conditions. The GDD values increased from 0 to 10 for the modified method, and the values reduced from 12.22 to 10 using the standard method from DOY 208 to 247.

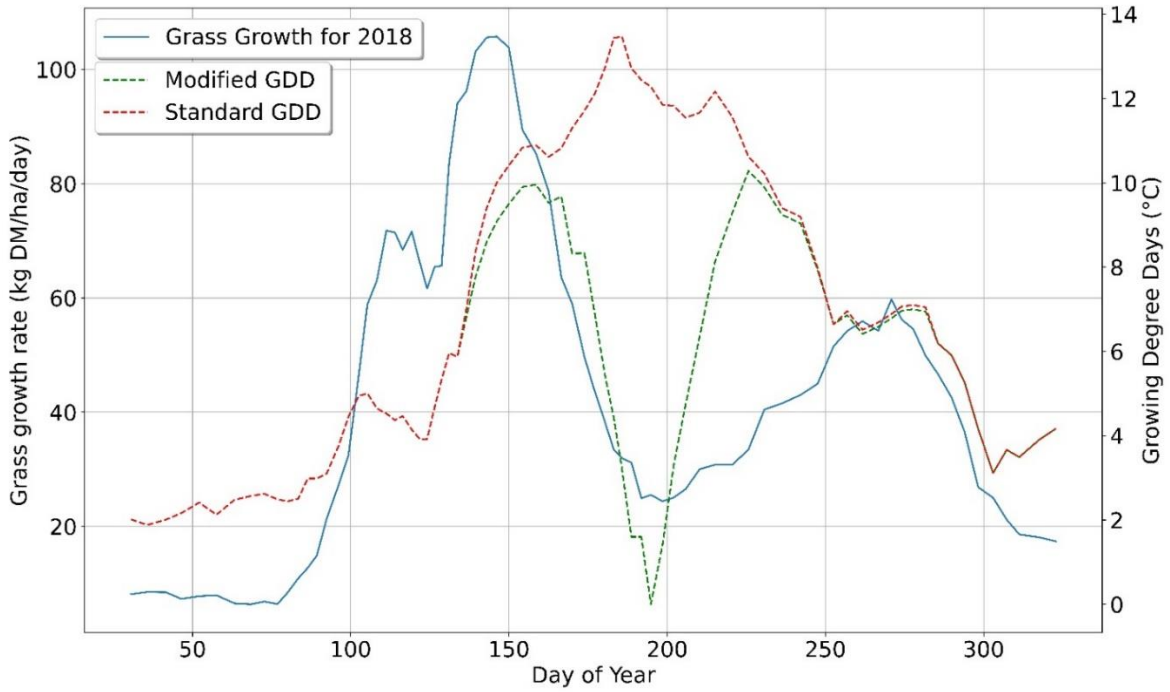


Figure 4-9 Comparison of the 7-day moving average of grass growth rate & GDD (2018). Grass growth (blue line). Standard GDD (red dotted) and modified GDD (green dotted). The modified GDD includes the impact of high SMD values. The x-axis is the day of the year, the left-hand y-axis is the grass growth rate, and the right-hand side y-axis is GDD.

4.4.2 Correlation between meteorological variables & grass growth rate

Pearson correlation coefficients were used to quantify the relationship between meteorological data and ground-based grass growth rate from PBI and identify the best performing variables. There were 63 values of weekly data for 2017. Seven out of the eight meteorological variables had statistically significant correlations with grass growth rate (Figure 4-10, Table 4-2), with rainfall being the only non-significant variable. The grass growth rate was positively correlated with all the independent variables except rainfall and cumulative GDD. Potential evapotranspiration and evaporation had the highest correlation with grass growth rate ($r = 0.65$, $p<0.01$) (Figure 4-10). The standard GDD ($r=0.59$, $p<0.01$), modified GDD ($r=0.58$, $p<0.01$), mean temperature ($r=0.60$, $p<0.01$), and solar radiation ($r=0.57$, $p<0.01$) had a similar correlation trend with grass growth rate. SMD ($r=0.50$, $p<0.01$) was statistically significant but with a lower correlation coefficient than the other variables. These results agree with Han et al. (2003), in which herbage growth was highly correlated with evaporation, temperature, and solar radiation.

There were 82 weekly values for 2018. Six out of eight meteorological variables had statistically significant ($p<0.01$) correlations with grass growth rate (Figure 4-11, Table 4-2), the exceptions being SMD and rainfall. The highest correlation of grass growth rate was with global solar radiation ($r = 0.43$, $p<0.01$), evaporation ($r=0.42$, $p<0.01$) and potential evapotranspiration ($r=0.42$, $p<0.01$) (Figure 4-11). Notably, these values are all lower than the 2017 values.

Table 4-2 Correlations between grass growth and meteorological data.
Pearson correlation coefficients (r) and p-values between meteorological data (n=63) and grass growth rate for 2017 and 2018.

Meteorological variables		r	
		2017	2018
1	Potential evapotranspiration (PE)	0.65, p≤ 0.01	0.42, p≤ 0.01
2	Evaporation (evap)	0.65, p≤ 0.01	0.42, p≤ 0.01
3	Modified Growing degree days (GDD)	0.59, p≤ 0.01	0.38, p≤ 0.01
4	Standard growing degree days (GDD)	0.59, p≤ 0.01	0.26, p≤ 0.01
5	Mean temperature	0.60, p≤ 0.01	0.33, p≤ 0.01
6	Solar radiation	0.57, p≤ 0.01	0.43, p≤ 0.01
7	Soil moisture deficit (SMD)	0.50, p≤ 0.01	0.05, p=0.64
8	Rainfall	-0.07, p= 0.60	-0.14, p=0.22

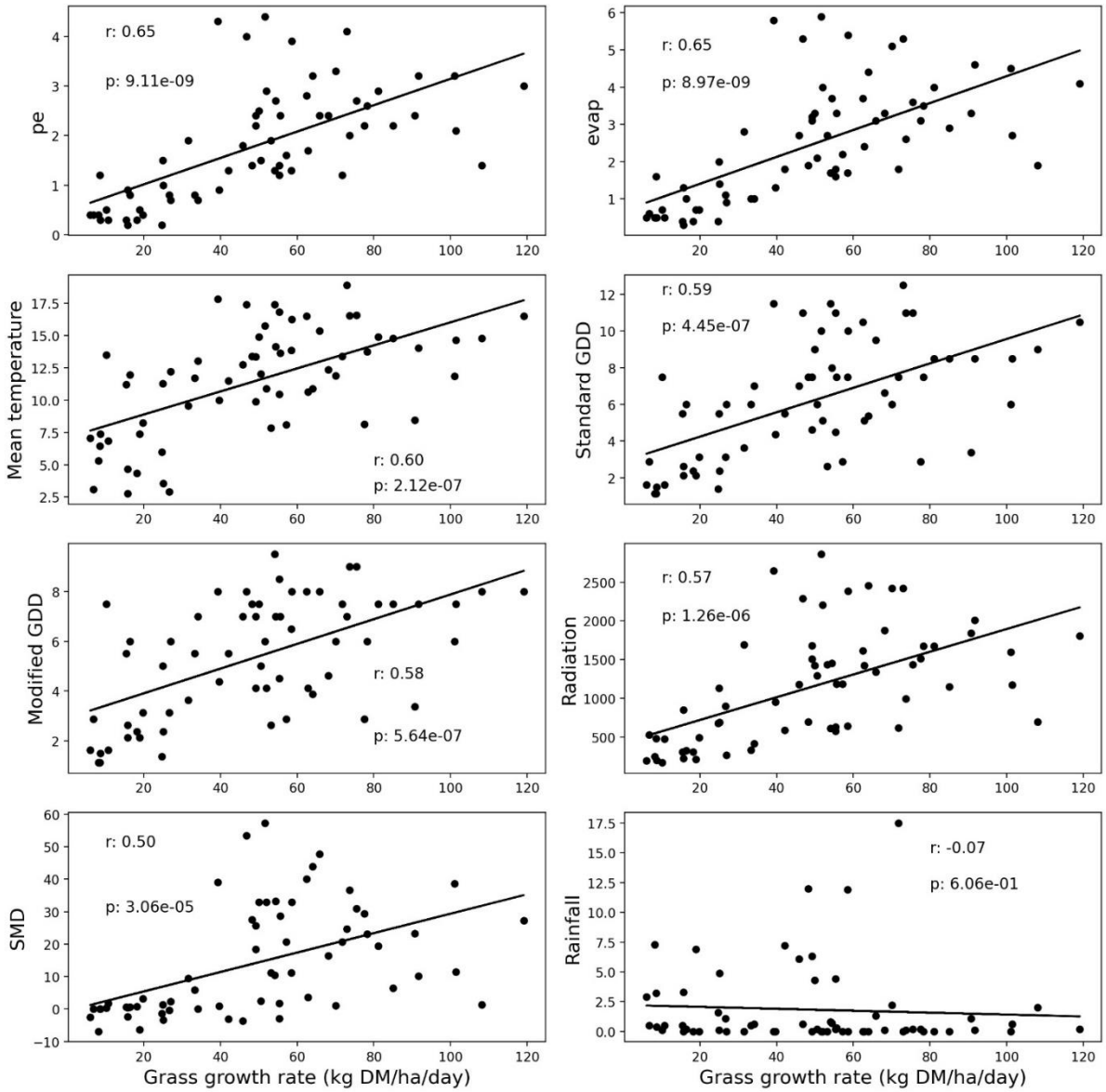


Figure 4-10 Grass growth rate vs. meteorological data (2017)

All values except rainfall are significant at p<0.01. GDD is growing degree-days, SMD is soil moisture deficit, 'pe' is potential evapotranspiration, and 'evap' is evaporation

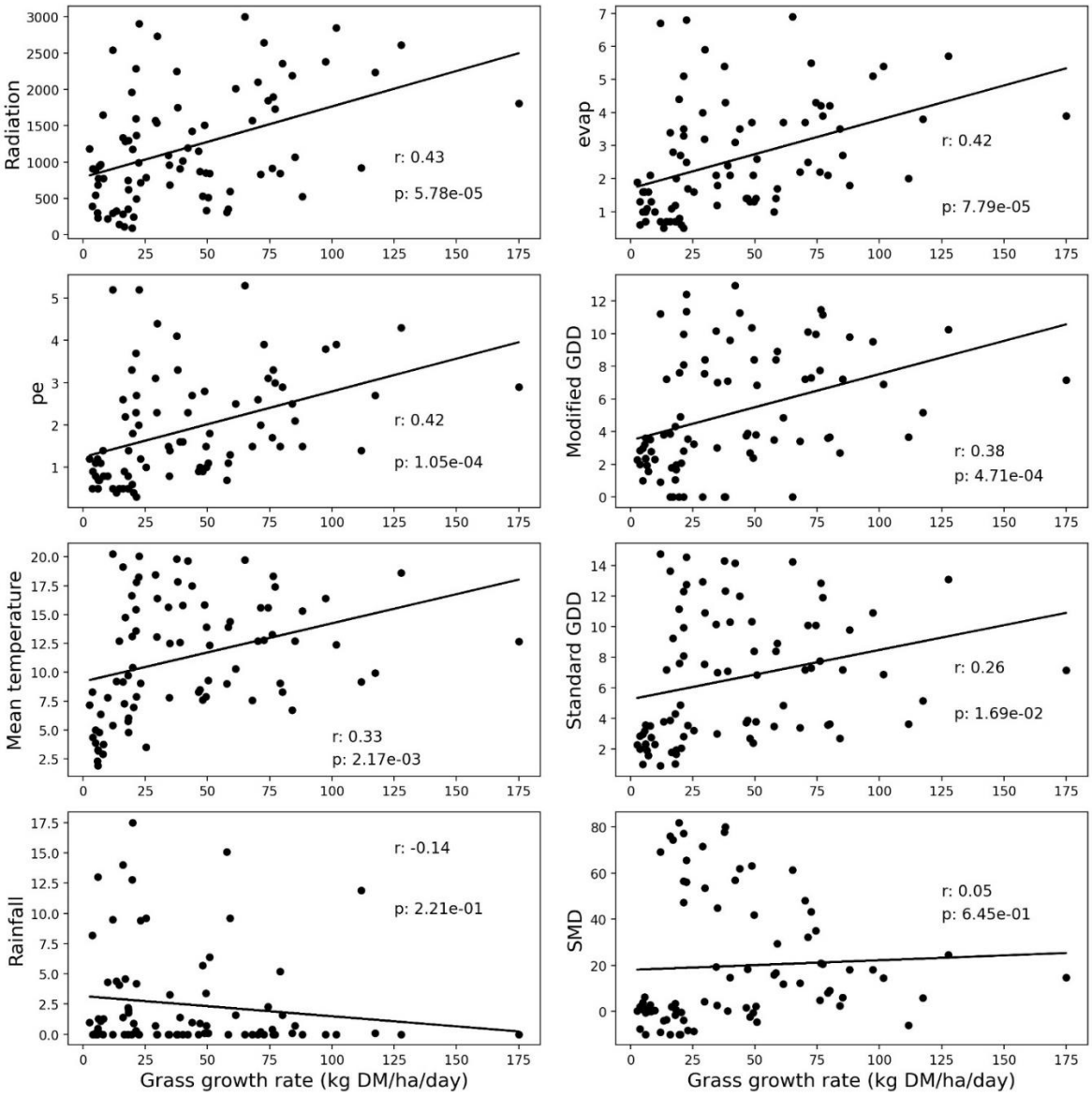


Figure 4-11 Grass growth rate vs. meteorological data (2018)

All values except rainfall, SMD and standard GDD are significant at p<0.01. GDD is growing degree-days, SMD is soil moisture deficit, 'pe' is potential evapotranspiration, and 'evap' is evaporation

4.4.3 Ordinary least square regression (OLS)

The 2017 OLS model was trained using the seven significant variables from Table 4-2. Figure 4-12 displays the scatterplots between measured and predicted grass growth rate values, and the points are coloured according to the season (winter is red, spring is orange, summer is green, and autumn is blue). The R^2 and RMSE based on a regression analysis of grass growth rate against meteorological data were 0.53 and 18.90 kg DM $ha^{-1}day^{-1}$. The predicted grass growth rate correlated strongly with the actual grass growth rate.

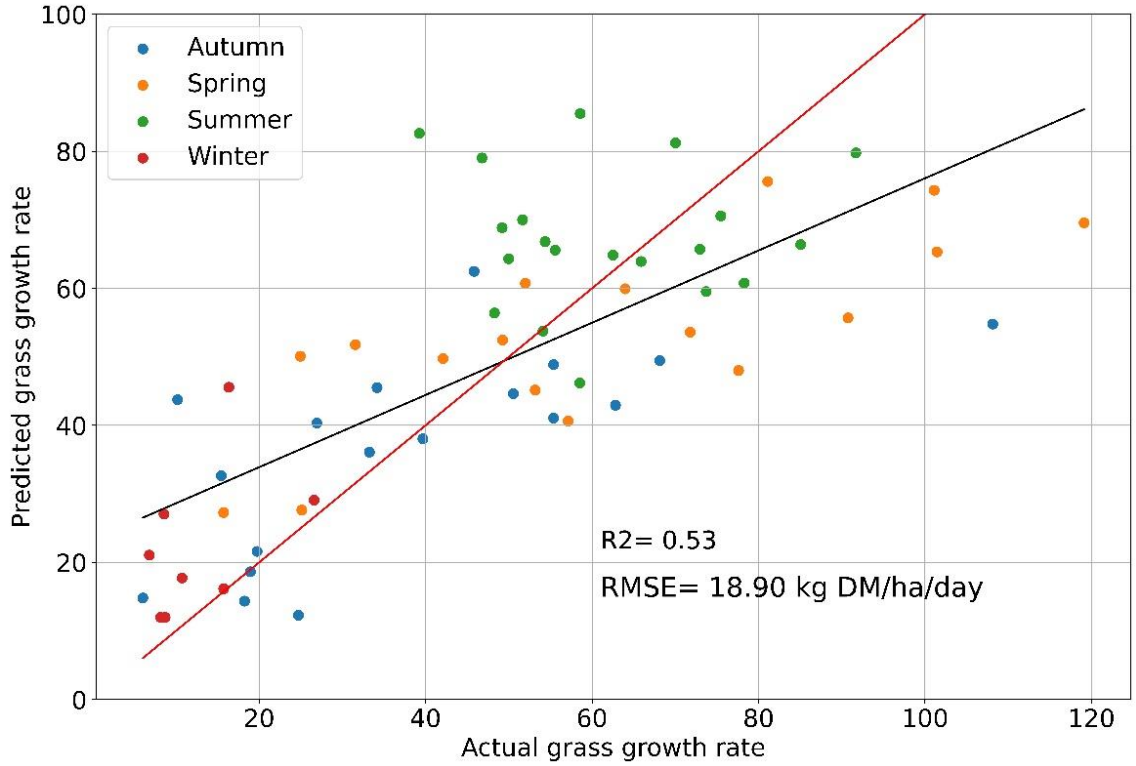


Figure 4-12 Actual vs. predicted grass growth rate at Moorepark (2017). Only those significant variables from OLS were used. The red line is the 1:1 line, while the black line is the modelled regression line. The points corresponding to winter are red, spring is orange, summer is green, and autumn is blue.

The model over-predicts between 0 and 40 kg DM $ha^{-1}day^{-1}$ and is more variable above 40 to 80 kg DM $ha^{-1}day^{-1}$ and underestimates above 80 kg DM $ha^{-1}day^{-1}$, which is evident from the dispersion of the samples around the 1:1 line. The grass growth values for winter were concentrated in the lower end of the plot and were over-predicted, and for summer, the values were also over-predicted. There was no apparent overestimation or underestimation for autumn. The values during spring were over-estimated for an actual grass growth rate of less than 55 kg DM $ha^{-1}day^{-1}$, whereas values above 55 kg DM $ha^{-1}day^{-1}$ were under-predicted. A scatterplot between actual and predicted grass growth rates for summer from June (DOY=152) until August (DOY=243) with seven significant variables is shown in Figures 4-13 with R^2 of 0.16 RMSE of 12.49 kg DM $ha^{-1}day^{-1}$.

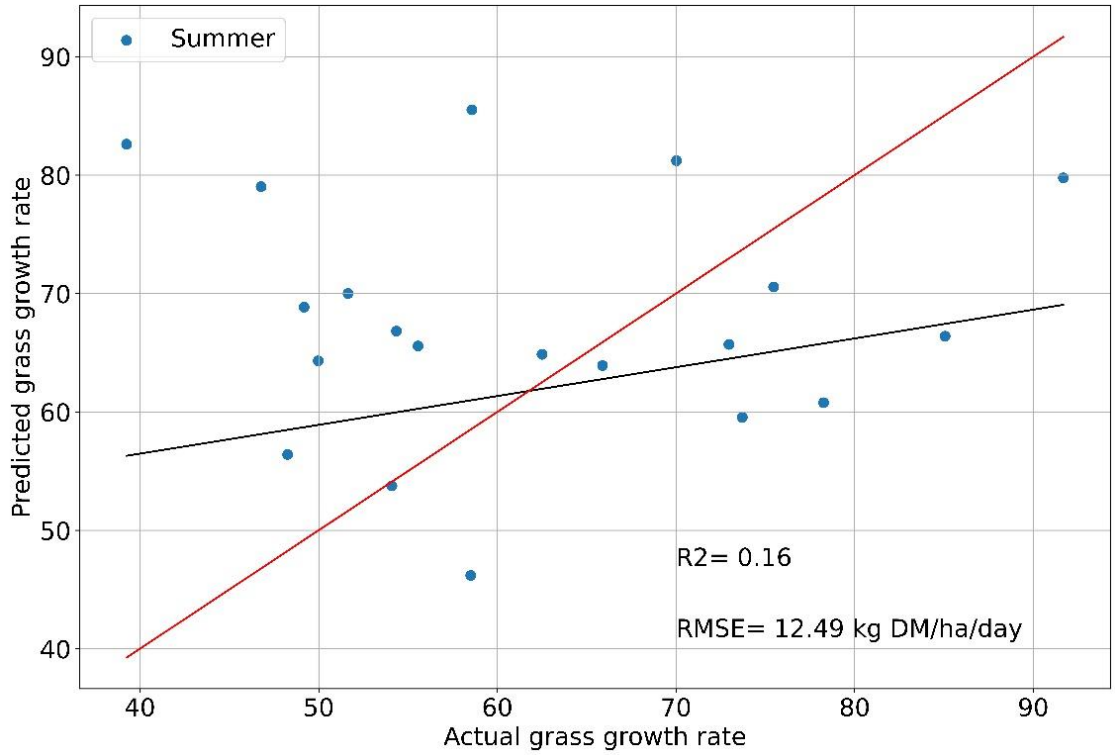


Figure 4-13 Actual vs. predicted grass growth rate (June-Aug 2017).
The red line is the 1:1 line, while the black line is the modelled regression line.

The same analysis was done for 2018 (Figure 4-14 and 4-15) with the six significant variables identified in Table 4-2. The R^2 and RMSE based on a regression analysis of grass growth rate against meteorological data were 0.36 and 27.02 kg DM $ha^{-1}day^{-1}$. The grass growth during winter (red) mainly was over-predicted. The data for spring covered the whole range of values with the scatter around the regression line. The autumn values were close to the 1:1 line with some under-predicted outliers. The summer values were scattered around 1:1 line with some over-predicted outliers. The values for summer were analysed separately in Figure 4-15. The plot shows that the grass growth for summer was over-predicted with R^2 of 0.57 and RMSE of 15.48 kg DM $ha^{-1}day^{-1}$.

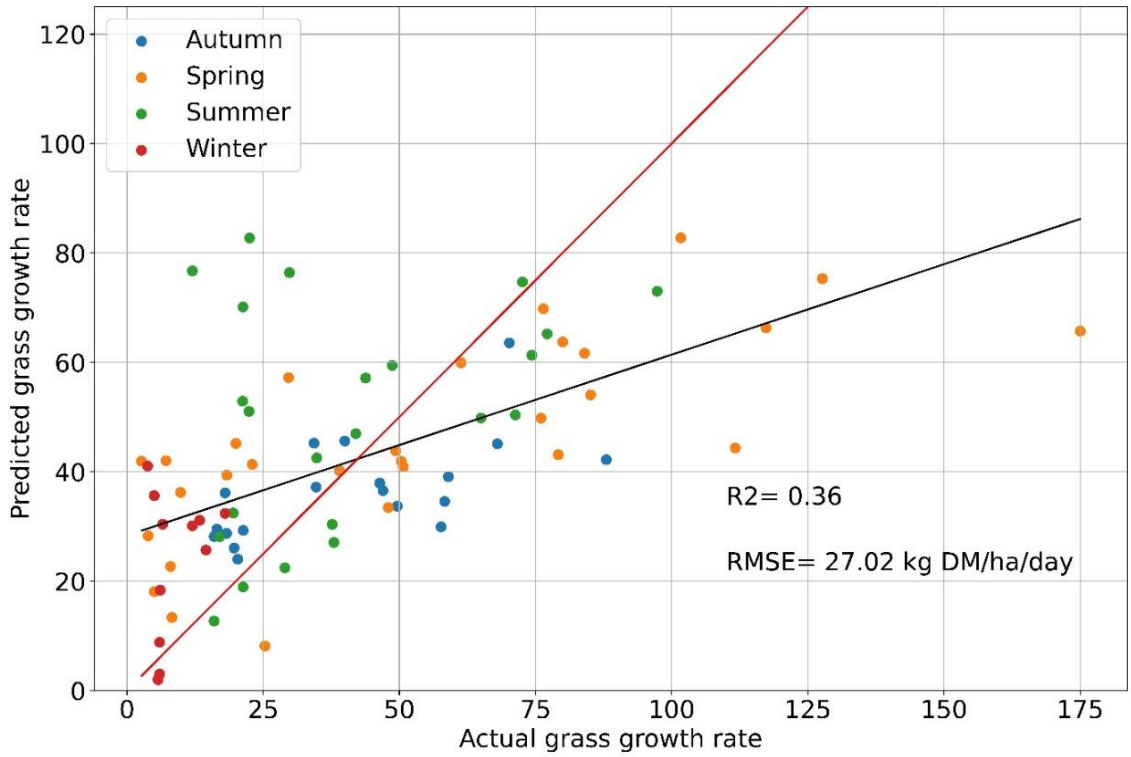


Figure 4-14 Actual vs. predicted grass growth rate at Moorepark (2017). The significant variables from OLS were used. The red line is the 1:1 line, while the black line is the modelled regression line. The points corresponding to winter are red, spring is orange, summer is green, and autumn is blue.

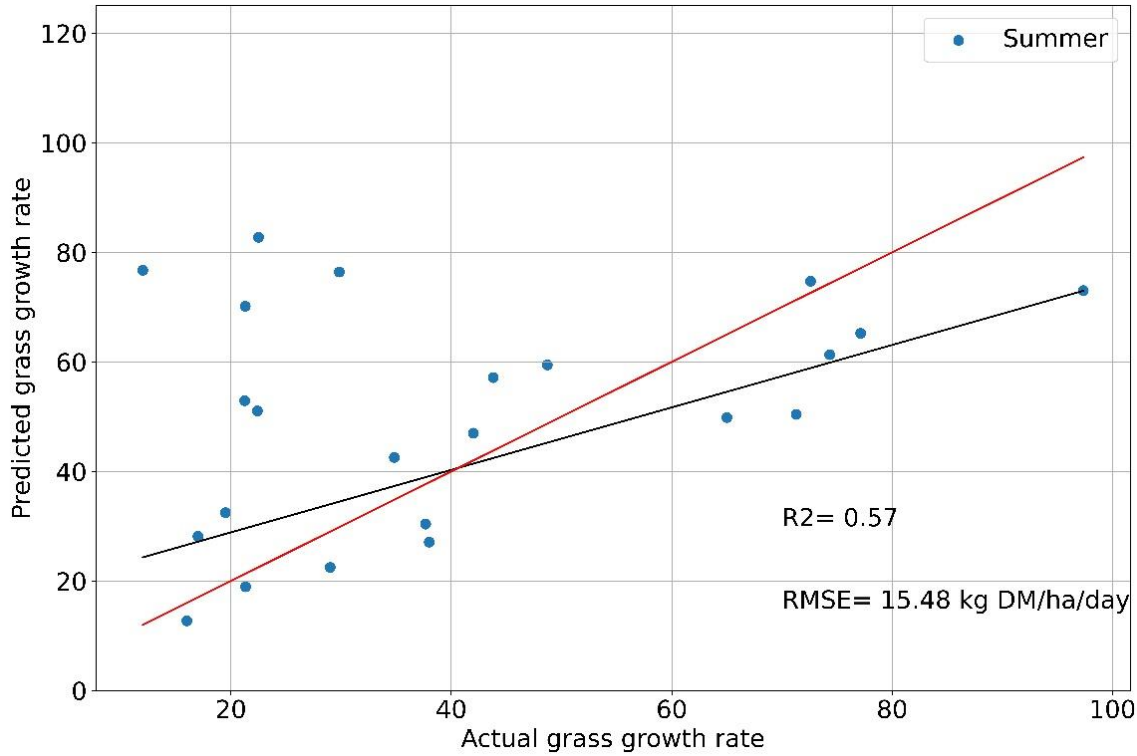


Figure 4-15 Actual vs. predicted grass growth rate (Jun-Aug 2017).
The red line is the 1:1 line, while the black line is the modelled regression line.

The results for 2017 and 2018 are summarised in Table 4-3. Two types of input combinations were presented – the whole year and the data from June until August. The summer DOY from 152 to 243 were explored in more detail with R^2 and RMSE. The R^2 and RMSE for June until August were 0.16 and 12.79 kg DM $ha^{-1}day^{-1}$ for 2017 and 0.57 and 15.48 kg DM $ha^{-1}day^{-1}$ for 2018.

Table 4-3 Results for 2017 and 2018 showing R^2 and RMSE (kg DM $ha^{-1}day^{-1}$)

	R^2		RMSE	
	2017	2018	2017	2018
Significant variables	**0.53**	0.36	**18.90**	27.02
Significant variables for June to August	0.16	0.57	12.79	15.48

4.5 Discussion

4.5.1 Effect of meteorological data on grass growth rate

Grass growth responds more strongly to temperature early than late in the season, resulting in a peak in spring and probably explaining why the GDD models cannot model this peak (Wingler and Hennessy, 2016). The mean temperature for 2017 and 2018 was compared with the 10-year average from 2006-2016, as shown in Figure 4-16, with months on the x-axis and mean temperature on the y-axis. The average from 2006-2016 is in red, 2017 in blue and 2018 in green. In February and March, the average temperature for 2018 was lower than average by 29.94% and 29.89%, whereas from April until August, the temperatures for 2018 were higher than average (by 2.48% for April, 10.26% for May, 14.70% for June, 10.51% for July and 4.51% for August).

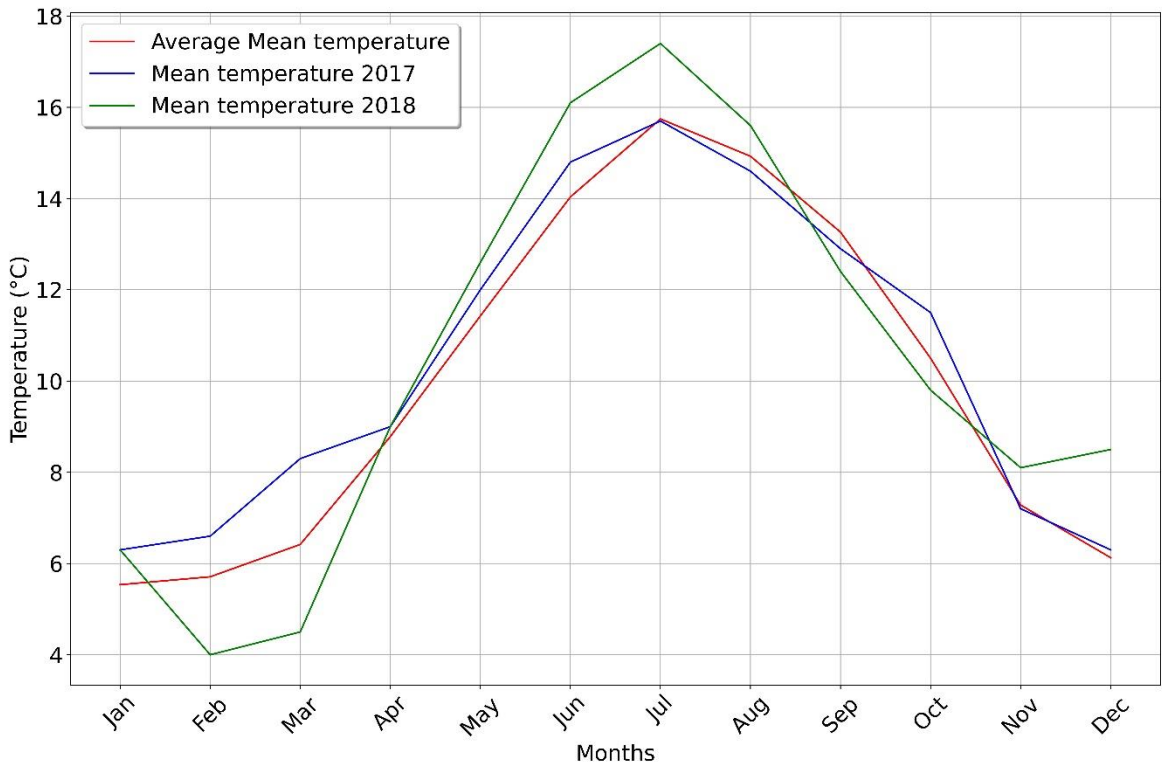


Figure 4-16 Mean monthly temperature at Moorepark Farm.
Data averaged for 2006 to 2016 (red plot), for 2017 (blue plot) and 2018 (green plot).

The total monthly precipitation for 2017 and 2018 was compared with the 10-year average from 2006-2016, as shown in Figure 4-17, with months on the x-axis and precipitation on the y-axis. The average from 2006-2016 is in blue, 2017 in orange and 2018 in a green bar plot. In April, the precipitation values are high by 103.31 mm than average for 2018, whereas for 2017, the rainfall was lower than average by 52.18 mm. From May until October, the rainfall for 2018 was below average (lower by 20.86% for May, 58.92% for June, 37.89% for July, 46.98% for August, 14.06% for September and 27.63% for October). Consequently, the high temperatures along with low accumulated rainfall led to poor grass growth in 2018.

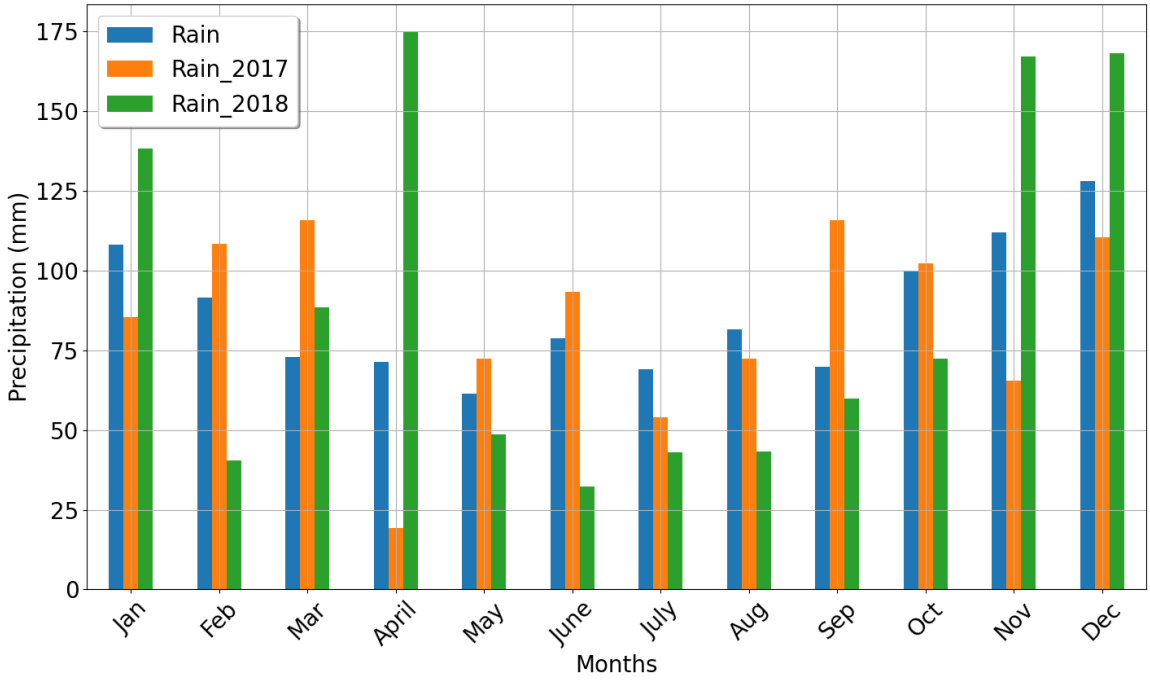


Figure 4-17 Total monthly precipitation at Moorepark Farm.
Data averaged for 2006 to 2016 (blue bar plot), for 2017 (orange bar plot) and 2018 (green bar plot).

The effect of poor weather conditions can be seen from the grass growth rate curve for 2017 and 2018 and its comparison with the average grass growth from 2013 until 2016 in Figure 4-18 (data before 2013 was not available on PBI). During 2017, the grass growth was lower than average in November and December, and there were poor ground conditions, which affected the grazing of cattle and led to early housing. As a result, the animals had to rely on fodder during that period, which affected the fodder availability for the beginning of 2018 and led to an increased need for animal feed (Dillon et al., 2018b).

During May, the grass growth rate for 2018 reached a peak and exceeded the average values due to high solar radiation and temperatures. The rainfall was not correlated with the grass growth rate (p-value of 0.60 for 2017 and 0.22 for 2018) in both years, suggesting that the combination of meteorological factors are responsible for grass growth variability rather than a single variable. Although SMD does not show a strong correlation with the grass growth rate, it is an essential factor affecting the grass growth during 2018, especially from DOY 152 (1st June) to DOY 243 (31st August), as shown previously in Figure 4-5.

The drought of 2018 affected the whole country but the local effects varied from farm to farm depending on the soil type (Falzoi et al., 2019). The soil at Moorepark Farm is well-drained and therefore does not hold water for long, leading to low grass growth rates.

When the grass growth curve was compared with the standard and modified GDD outputs, it was noticed that the modified method accumulated fewer days during summer than the standard method for 2017. For 2018, the grass growth rate is higher in May and June than in 2017 because of higher temperatures and higher solar radiation. The daily average temperature for July and August was 16.5 °C in 2018 and 15.16 °C in 2017, and the total monthly average rainfall was 63.2 mm and 43.1 mm, respectively. As a result, the SMD values were much higher for 2018 (66.50 mm) than in 2017 (30.30 mm) in July and August.

4.5.2 Regression models to predict grass growth rate

The OLS regression model was developed between the significant variables and grass growth. The model included both the standard GDD values and the modified ones, and the individual effect of each GDD was not included in the analysis as both were positively correlated (p value≤0.01) with the grass growth rate. The model for 2017 could explain 53% of the variance in the predicted grass growth rate. By contrast, the model for 2018 could explain only 36% of the variability in the grass growth rate.

The highly correlated variables for grass growth rate prediction for 2017 were potential evapotranspiration and evaporation with r of 0.65, whereas, for 2018, the most important variables were solar radiation and evaporation with an r of 0.43 and 0.42. Evapotranspiration is the combination of evaporation and transpiration accounting for the movement of water to air. The potential evapotranspiration is the amount of evaporation occurring if a sufficient amount of water is available to the grass. If there is not enough water in the soil when the temperature and radiation are high, the grass plants experience difficulty extracting water from the soil and reducing the transpiration (Keane and Collins, 1986). These factors are important during the summer, which can limit the water available for grass growth.

The results presented here are for a single experimental farm with constant management intervention such as grazing and cutting of grass. Developing a national model for Ireland means including more farms in the model that do not have such a high level of grassland management but inevitably also a reduction in data availability. Earth observation data could be incorporated into the model to overcome this limitation, as they provide additional information on the actual conditions on the ground, such as the effects of management (grazing and silage cutting) and drainage that are lacking in the meteorological data model (Green, 2019).

Another limitation of this method is the use of a linear regression model, which assumes that there is a linear relationship between the dependent and independent variables. Several variables were demonstrated to have a low correlation with grass growth in 2018 in particular, indicating a non-linear relationship between them. Machine-learning algorithms can better detect and accommodate the non-linear relationships between variables (Schwalbert et al., 2020). In the next chapter, the potential for using machine-learning models with Earth observation data for grass growth estimation will be explored.

4.6 Conclusions

There is a direct relationship between grass development and GDD (a scalar value proportional to temperature), which explains 53% of the variability in the model in 2017, but only 36% of variability for 2018, which had more extreme weather conditions. Moreover, the error in the 2017 model was 30% lower than in the 2018 model, with RMSE values of 18.90 kg DM $ha^{-1}day^{-1}$ and 27.02 kg DM $ha^{-1}day^{-1}$ respectively. This chapter developed a method that better captures real world grass growth conditions by including high temperature thresholds and SMD into the GDD model. Both the standard and modified GDD were equally significant factors in terms of their correlation to the grass growth rate in 2017, but both had a much reduced correlation in 2018 although the modified GDD performed slightly better than the standard GDD calculation. Since 2018 was an exceptional year, the summer (DOY 152 - 243) was analysed separately. The 2017 model performed poorly with R^2 of 0.16 and RMSE of 12.79 kg DM $ha^{-1}day^{-1}$ and R^2 of 0.57, and RMSE of 15.48 kg DM $ha^{-1}day^{-1}$ for 2018. The predicted grass growth rate for the summer of 2018 was 57% along the 1:1 line with some over-predicted values. The grass growth rate for 2017 was over-predicted compared to actual grass growth rates of less than 60 kg DM $ha^{-1}day^{-1}$ and under-estimated for values of more than 60 kg DM $ha^{-1}day^{-1}$. There were no management information available that relates to grazing and silage cutting which could help explain why the weather driven grass growth rate differed from the actual values. The model in this chapter demonstrated that meteorology alone cannot reliably predict how the grass will grow, but in conjunction with additional information is a useful input to better understanding grass management.

5

Chapter 5 Development of a Machine-learning Model for the Estimation of Grass Growth Rate in Ireland

5.1 Introduction

This chapter explores how machine-learning (ML) can improve grass growth prediction by combining agro-meteorological data with synoptic satellite data from two current EO satellites, Landsat 8 and Sentinel 2.

This chapter outlines a performance assessment of the empirical biophysical "Brereton model" and ML models for Moorepark farm (Section 3.3.1, Figure 3-1 in Chapter 3). The ML algorithms assessed are an adaptive-neuro fuzzy inference system (ANFIS) and random forest (RF), as described in Section 5.2.4. These models include satellite vegetation indices and agro-meteorological data for eight research farms in the south of Ireland (Munster province) (Section 3.3.2, Figure 3-4 of Chapter 3).

5.1.1 Decision support system

In Section 1.3, the high spatial and temporal variability of grass growth in Ireland was discussed. This inherent natural variability in growth can result in poor grass utilisation (where the grass is not grazed or harvested on time) and can be exacerbated if farms are not well managed (i.e. grass is not routinely measured and proper paddock allocation to the animals maintained). Section 3.4.4 explained how measuring grass growth could help farmers quantify the volume of grass available in individual fields or paddocks. By measuring grass, farmers can synchronise on-farm demand with fresh grass supply and address supply imbalances quickly. With

dedicated grass budgeting services, farmers can quickly see where surpluses or deficits exist and help them understand what intervention may be necessary (grazing, harvesting for winter fodder, fertilizing and reseeding) (Murphy et al., 2018). There are several grassland decision support systems (DSS) currently available that can help farmers with fodder budgeting, for example, PastureCoach in New Zealand (Dalley and Geddes, 2012), DSS-Ecopay in Germany (Sturm et al., 2018) and PastureBase Ireland (PBI) here in Ireland (Hanrahan et al., 2017). An essential part of all DSS is providing accurate and reliable predictions of grass growth over a period.

Section 1.5 outlined several methods of estimating the grass growth rate on Irish farms. In summary, these include *in-situ* methods such as the cut and dry method, eyeballing and rising plate meters. They also include biophysical models such as the Johnson & Thornley model (Johnson and Thornley, 1983), the Jouven model (Jouven et al., 2006), the Brereton model (Brereton et al., 1996b) and the MoSt model (Ruelle et al., 2018b). The Irish farmers that use the PBI DSS typically use *in-situ* methods, which are simple measurements of the actual growth rate. Ideally, data are recorded weekly for individual paddocks, but in reality, there are often insufficient or missing data, which impacts data quality (Gargiulo et al., 2020, Schirmel, 2021). Furthermore, weekly data entry for several paddocks may be time-consuming and burdensome for the farmer.

For predicting growth, biophysical simulations can model increasingly complex interactions between crop physiological processes, environmental conditions such as weather and soil, and management style (Feng et al., 2020). Such models are typically used by researchers and can be very accurate, but they are dependent on the accuracy of input data and, crucially, lack a spatial context (Kasampalis et al., 2018).

5.1.2 Biomass estimation using Earth observation

Increasingly, satellite images are being used to overcome the shortcomings in *in-situ* measurements and in biophysical simulation models to provide spatially continuous information on biomass over large geographical areas. Accurately estimating grass

growth rate using EO sensors requires a consistent, longitudinal archive of imagery to be able to capture seasonal phenology and temporal changes in biomass over time, for example, as a result of mowing or grazing. The United States' National Aeronautics and Space Administration (NASA) has led the way in providing long-term EO data, with the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor providing imagery since 1999. A low spatial resolution but high spectral and temporal resolution sensor as compared to Sentinel 2 (10-20 m), MODIS captures 36 visible and infrared bands every 1-2 days at 250 m to 1 km spatial resolution. MODIS 8-day composites at 250 m resolution are widely used to map land cover (Baeza and Paruelo, 2020) or biomass (Gao et al., 2020a) and mitigate against the impact of missing data due to cloud cover. Ali et al. (2017) used MODIS composites to estimate grass growth at two locations in Ireland (Moorepark, Co. Cork and Grange, Co. Meath) from an ANFIS regression model. The spatial scale of MODIS is ideal for large-scale biomass estimation but is less well suited to farm-, field- and paddock-scale estimates of growth where the low spatial resolution means multiple fields can be captured within one pixel, alongside non-grassland land cover, for example, hedgerows, buildings and surface water (Ali et al., 2017). Higher spatial resolution imagery is required, such as the USGS Landsat 8 mission at 30 m or the ESA Copernicus Sentinel 2 data (10 m to 20 m) to achieve more detailed estimates of grass growth.

Like MODIS, the Landsat mission is a well-known, widely accessed archive of global EO imagery with continuous data collection since 1972. The joint NASA and United States Geological Survey (USGS) program currently has two satellites in orbit, with a third planned for launch in September 2021 (Wulder et al., 2016, Markham et al., 2016). Landsat 7 was launched in 1999 but has operated with a permanent failure of the scan line corrector since 2003, resulting in banding within images (Markham et al., 2004). Landsat 8, launched in 2013, has two sensors- the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) that provide nine visible and infrared bands at 30 m spatial resolution and two thermal bands at 100 m resolution. The Landsat mission has a higher spatial resolution than MODIS but a return period of 16 days. The planned launch of Landsat 9 into a

complementary orbit will reduce the revisit time to 8 days, doubling the volume of data available and increasing the chances of cloud-free scenes. The ESA Copernicus Sentinel 2 mission is a two-satellite constellation launched in 2015 (Sentinel 2A) and 2017 (Sentinel 2B). The Sentinel satellites represent a massive shift in the availability of high-resolution EO data. The improved spatial, spectral and temporal resolution relative to Landsat 8 provides new opportunities for grassland mapping and biomass estimation globally (Kolecka et al., 2018, Wang et al., 2019).

As discussed in Section 1.5.3, multispectral bands in EO imagery may be combined into data products, called spectral or vegetation indices, to map vegetation fraction or plant health (Marti et al., 2007). The Normalised Difference Vegetation Index (NDVI), for example, relies on low reflectance in red wavelengths (due to chlorophyll absorption) and high reflectance in near-infrared (NIR) wavelengths (due to scattering of light within the leaf structure) (Xie et al., 2018). NDVI is a unitless measure commonly used in crop biomass studies (Lai et al., 2018, Guan et al., 2019). Todd et al. (2010) demonstrated how vegetation indices (VI) were essential predictors of rangeland biomass in the United States and how the choice of VI should be based on consideration of the specific reflectance characteristics of grass canopy and underlying soil. The rangelands were in a semi-arid area with a high proportion of bare soil, unlike managed temperate grasslands where there is no exposed soil, but nevertheless the large number of VI that have been derived remain testament to the better performance of some VI under particular conditions.

5.1.3 Machine-learning for biomass estimation

When estimating biomass using EO or ancillary data and regression modelling, the choice of algorithm can considerably impact the estimate's accuracy. Traditional regression methods, such as those introduced in Section 1.5.3, have been used widely in the past; however, machine-learning regression models are increasing in popularity. Traditional parametric methods are widely used and are straightforward to implement using various fitting functions, for example, linear (Zeng and Chen, 2018), nonlinear (Gourley et al., 2017), polynomial (Demanet et al., 2015), power (Ma et al., 2019) and exponential (Ge et al., 2019). However, these models have significant limitations rendering them unsuitable for complex, multidimensional

datasets common for different farm management scenarios (Chen et al., 2021). Machine-learning (ML) is a branch of artificial intelligence (AI) that can learn to recognise patterns in input data based on previous experience without being specifically programmed to do so (Attaran and Deb, 2018). ML algorithms can overcome limitations in parametric regression models, particularly for non-linear relationships (Solyali, 2020). Through open-source software such as Python and R, access to machine-learning algorithms has widened. Several algorithms have become popular in Earth observation, for example, random forest (RF) for crop yield (Jeong et al., 2016), support vector machine (SVM) for grassland biomass (Clevers et al., 2007), or adaptive neuro-fuzzy inference system (ANFIS) for grass growth (Ali et al., 2017b).

Many studies have demonstrated how effective ML regression models can be for grassland applications. Berger et al. (2020) used random forest regression to estimate grassland biomass in savannas with Landsat 8 and Sentinel 2 separately. Three models were developed with spectral bands and NDVI using Landsat 8 for 2016 and Sentinel 2 data for 2017 and 2018. Using spectral bands and NDVI, the models performed the poorest for Landsat 8 data with a relative RMSE of 38.91% and 38.24%, compared to Sentinel 2 with relative RMSE from 22.37% to 26.27%. The poor performance of the Landsat 8 model could be because of the lower resolution and drought effect in 2016 (Berger et al., 2020). Chen et al. (2021) used an artificial neural network model to estimate pasture biomass using Sentinel 2 and daily climate variables (rainfall, mean temperature and vapour pressure deficit) at a 5 km resolution with an R^2 of 0.6 (Chen et al., 2021). Grass biomass can be estimated indirectly using various biophysical parameters such as Leaf Area Index (LAI) (Punalekar et al., 2018) and grass height (Zhang et al., 2021b). Nickmilder et al. (2021) tested a broad range of ML algorithms to estimate grass height to calculate grass biomass using Sentinel 1 and Sentinel 2, as well as meteorological data (daily rainfall and degree-days) and reference data from a rising plate meter (RPM). The Sentinel 2 raw bands and vegetation indices (Enhanced Vegetation Index-EVI and Soil and Atmospherically Resistant Vegetation Index 2 –SAVI2) were the most significant variables, and Sentinel 1 was the least significant. The model with the combination of all three variables had the lowest RMSE with 19 mm of the grass

height. While several studies have demonstrated the potential of ML regression, only a few have compared ML estimates of biomass with biophysical models. Schwieder et al. (2020) compared random forest (RF) and a soil-leaf-canopy (SLC) radiative transfer model with Sentinel 2 bands for estimating aboveground biomass and LAI. Both performed equally well with average biomass estimates of 2.23 t/ha (RF) and 2.49 t/ha (SLC). Similar results were obtained in a study by Quan et al. (2017), where a biophysical model (R^2 of 0.64 and RMSE = 42.67 g/m^2) performed better than three empirical methods for grass biomass retrieval using vegetation indices from Landsat 8: exponential regression (R^2 of 0.48 and RMSE = 41.65 g/m^2), partial least squares regression (PLSR) (R^2 of 0.55 and 37.79 = g/m^2) and ANN (R^2 of 0.43 and RMSE 46.26 = g/m^2). The regression models with NDVI data suffered from saturation problems at high biomass values (resulting in a very high RMSE of 250g/m^2).

It has been established that ML algorithms can be used to estimate biomass, and these methods have already been applied in Ireland, albeit at coarse spatial resolutions and for a limited number of sites. This chapter expands previous research by Ali (2016) to explore how accurately ML algorithms can estimate grassland biomass on Irish farms using Landsat 8 and Sentinel 2 VI and ancillary meteorological data as predictor variables. Two ML algorithms (RF and ANFIS) were assessed. Furthermore, ML models were juxtaposed with a leading biophysical simulation model for grass biomass estimates, the Brereton model. Specifically, the research presented in this chapter seeks to:

- Explore the influence of EO and meteorological variables on grass biomass estimation;
- Compare Brereton model and machine-learning models for estimating biomass; and
- Compare the accuracy of the RF and ANFIS algorithms for estimating the grass growth rate.

5.2 Materials & Methods

5.2.1 Study area

The study area is described in detail in Chapter 3. To recap, this experiment used data from eight farms situated in different parts of Ireland (Figure 3-4 and Table 3-2). Between the eight farms, the annual average temperature ranges from 14.80–18.50 °C, with mean annual rainfall between 1200–1800 mm, and sunshine duration of 1238.7-1868.7 h (Met Éireann 2013). These values are each farm's annual mean and can be higher than the national average.

5.2.2 Datasets

This study used satellite VI imagery from Landsat 8 and Sentinel 2 and agro-meteorological data from Met Éireann synoptic weather stations from 2017-2018 (see Figure 3-8 in Chapter 3). The ground truth data for this experiment were *in-situ* grass growth rates inputted to PastureBase Ireland for each of the eight farm locations. The variables used for this study are listed in Table 5-1 below. All the input variables used in the Brereton model are presented in Table 5-2.

Table 5-1 Input variables used for the ML regression models

Variables	Description	Unit
T_{min}	Daily minimum temperature	°C
T_{max}	Daily maximum temperature	°C
T_{mean}	Daily mean temperature	°C
G	Global radiation	J/cm sq
P	Potential evapotranspiration	mm
smd	soil moisture deficit	mm
E	evaporation	mm
R	Rainfall	mm
GDD	Growing degree days	°C
NDVI	Normalised vegetation index	Unitless
NDRE	Normalized Difference Red Edge Index	Unitless
Growth	Grass growth rate (reference data)	Kg DM $ha^{-1}day^{-1}$

Table 5-2 Input variables used for the Brereton model

Variables	Description	Unit
T_{mean}	Daily mean temperature	°C
G	Global radiation	J/cm sq
P	Potential evapotranspiration	mm
smd	Soil moisture deficit	mm
A	Actual evapotranspiration	mm
R	Rainfall	mm

5.2.2.1 Vegetation indices

Cloud-free Landsat 8 surface reflectance level-2 (Collection 2) images (path 205 to 208 and row 22 to 24) were downloaded from the USGS Earth Explorer website[2] and converted into an NDVI. The Sentinel 2 data were downloaded from the Copernicus Open Access Hub platform from ESA website[3] and converted into surface reflectance in SNAP (version 7.0). The 20 m reflectance bands were resampled using bilinear interpolation to 10 m. NDVI (Tucker, 1979) and NDRE (Fernández-Manso et al., 2016) were produced from the resampled Sentinel 2 bands (Equations 5-1 and 5-2). A zonal statistics function in ArcGIS was used to generate mean VI values for each of the eight farms.

$$\text{NDVI} = \frac{(B_{NIR} - B_{Red})}{(B_{NIR} + B_{Red})} \qquad \text{Equation 5-1}$$

$$\text{NDRE} = \frac{(B_{NIR} - B_{Veg\ red\ edge})}{(B_{NIR} + B_{Veg\ red\ edge})} \qquad \text{Equation 5-2}$$

Where, B_{NIR} is the near-infrared band (band 5 for Landsat 8 and band 8 for Sentinel 2), B_{Red} is the red band (band 4 for both Landsat 8 and Sentinel 2). Veg red edge is vegetation red-edge band, which is band 5 for Sentinel 2.

[2] http://earthexplorer.usgs.gov. Accessed 12th October 2018

[3] https://scihub.copernicus.eu/dhus/#/home. Accessed 28th October 2018

5.2.2.2 Temporal matching of satellite data with ground data

Extensive cloud cover over Ireland often limits the amount of optical EO imagery available, with only a small number of images being available for each of the specific farm locations. The number of Landsat 8 and Sentinel 2 images for each farm and corresponding PBI data measurements are shown in Table 5-3. PBI grass growth rates were matched with available satellite images. As the satellite imagery was temporally sparse, a 7-day window before and after each PBI entry was considered for Landsat 8 and Sentinel 2 imagery. If an image fell within seven days of a satellite image, then PBI growth rate and VI values were matched. If a satellite image was greater than seven days from a PBI entry, it was rejected. To perform this matching, a rolling join function carried seven days backwards and forwards was implemented in R Statistical Software using the 'roll' function in the 'data.table' package (version 1.14.0). The bespoke script used to perform this join is included in the archive of scripts in Appendix 5.1.

Table 5-3 Number of cloud-free Landsat 8 and Sentinel 2 images (2017-2018) & available PBI data

Farm	Landsat 8	Sentinel 2	PBI measurements
Moorepark, County Cork	19	23	148
Curtins, County Cork	17	18	82
Kilworth, County Cork	16	16	135
JC, County Wexford	34	19	77
Kildavin, County Wexford	27	14	61
Tuohy, County Galway	15	14	78
INZAC, County Galway	16	14	75
Newford, County Galway	16	13	79

5.2.3 Brereton model

The Brereton model, also known as Johnstown Castle Grass Model, is a static, empirical model that uses meteorological conditions to predict grass growth (Equation 5-3). As listed in Table 5-2, the inputs to this model are rainfall, solar radiation, mean daily temperature, potential evapotranspiration, actual evapotranspiration, and soil moisture (Brereton et al., 1996a). The model estimates the grass growth rate based on solar radiation received at the surface. The growth rate depends on the efficiency of grass to convert incoming radiation into energy, which is a factor of mean temperature and rainfall at a given location. Both positive and negative soil moisture deficits (SMD) are also limiting factors as they can hamper the grass growth if soils are saturated or too dry. The output from the model is a prediction of the grass growth rate over a given period. For this study, the Brereton model was scripted using Python 3.7. This script is included in Appendix 5.2.

$$\frac{\mathrm{Y_a}}{\mathrm{Y_p}} = \mathrm{a} + \mathrm{b}\frac{\mathrm{E_a}}{\mathrm{E_p}} \qquad \text{Equation 5-3}$$

Where, a = 0.20 and b = 0.80

Y_a= Actual grass growth rate without irrigation

Y_p= Actual grass growth rate with irrigation

E_a= Actual evapotranspiration

E_p= Potential evapotranspiration

5.2.4 Machine-learning algorithms

The ML algorithm requires clean and normalised data. Cleaning of data involves removing the missing values and removing the duplicate rows. Cleaning was followed by feature selection. When training an ML model, not all the variables will be relevant, and therefore, the optimal variables relevant to the prediction are chosen. Scaling is used to standardise the data between zero and one as the range of input data varies widely. The input data are split into two parts, namely training data and testing data. Training data are used to train the ML model, and the model is evaluated using the testing data, which has not previously been used in the model.

5.2.4.1 Adaptive neuro-fuzzy inference system

ANFIS takes advantage of artificial neural network (ANN) and fuzzy logic (Jyh-Shing and Jang, 1993). ANFIS has been implemented for many remote sensing applications, for example, landslide susceptibility mapping (Aghdam et al., 2017), drought monitoring (Khosravi et al., 2017), flood susceptibility mapping (Tien Bui et al., 2018), and crop photosynthetic rate estimation (Valenzuela et al., 2017). For grasslands applications, ANFIS was previously used to estimate pasture biomass in Ireland (Ali et al., 2017b) and has also been used to map grazing intensity (Salski and Holsten, 2006). ANFIS is a type of fuzzy systems, which uses feed-forward network with five significant layers- input fuzzification, fuzzy rules, method, output aggregation, and output defuzzification.

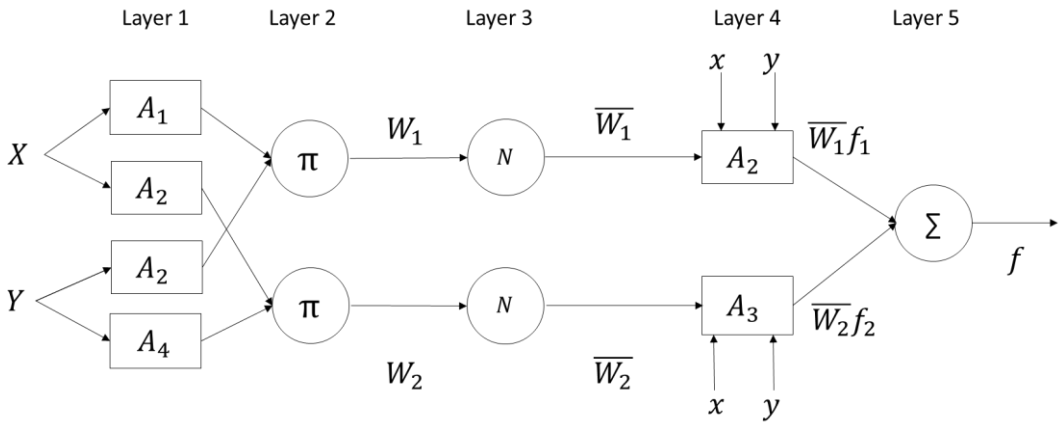


Figure 5-1 Architecture of Adaptive Neuro-Fuzzy Inference System (ANFIS).

A generalised ANFIS architecture is presented in Figure 5-1, where there are two inputs, X and Y, and one output, f.

Layer 1: This is the fuzzy layer that performs fuzzification of inputs X and Y.

Layer 2: This is a product layer with fixed nodes. It multiplies the node function to the input signals to get the output signal.

Layer 3: This is the normalisation layer that normalises the output from Layer 2.

Layer 4: This is an adaptive node.

Layer 5: The defuzzification layer, which has fixed nodes. It gives the overall output.

5.2.4.2 Random forest

RF is a non-parametric supervised classification and regression algorithm that has become very popular in EO studies (Breiman, 2001). RF is an ensemble decision tree algorithm that builds many trees independently from randomly sampled variables and uses a majority vote to determine the final label. As each tree is independent of the others, there is little correlation between trees. Therefore, RF is superior to a single decision tree as the final prediction is an aggregation of predictions from each tree. RF is relatively simple to parameterise, requiring only two parameters, the required number of trees (ntree) and the number of variables randomly sampled at each split (mtry). Both can be fine-tuned using a grid search function. The script was written in R statistical software using “caret” (v 6.0-88) packages for RF and tuned hyper-parameters random grid-search methods (see Appendix 5.4). The RF model was optimised by tuning hyper-parameters using random grid-search methods.

5.2.4.3 Feature selection

Feature selection is a method in machine-learning that reduces the number of input variables, leaving only those model variables most useful for prediction. Essentially, feature selection removes redundant predictors from a model, reducing training and processing time, and improving efficiency (Cai et al., 2018). Feature selection algorithms help to get the optimal number of variables required to train the model. RF has an internal method of calculating variable importance using out-of-the-bag

(OOB) error. While making the samples, data points were chosen randomly and with replacement, and the data points which fail to be a part of that particular sample are known as OOB points. In the package "caret", the VarImp function (for variable importance) was used to find the input variables with the most significant impact on mean squared error (MSE) by observing the effect on accuracy as variables are left out of models. ANFIS does not have an internal feature selection, therefore the Boruta package in R Statistical Software was used. Boruta is a wrapper model built around the RF algorithm. Comparative studies of different feature selection functions in R found that Boruta was computationally efficient with the lowest OOB error and low computational time (Speiser et al., 2019). Boruta duplicates an input dataset and randomly mixes values. These features are used to train an RF classifier, and Z scores are computed for each variable. The variables having a significantly higher Z score than the maximum Z score (MZ) among shadow features are "confirmed attributes". Attributes with a significantly lower Z score than MZ are "rejected attributes". Features (variable importance) that are neither confirmed nor rejected are called "tentative features". Confirmed attributes were used to train the regression models.

5.2.4.4 Error metrics

Five standard error metrics were calculated (detailed in Equations 5-4 to 5-8).

- Mean squared error (MSE) is the mean of the squared differences between predicted and expected target values in a dataset. The units of MSE are squared units, and often the root mean squared error is reported (Equation 5-4).

$$MSE= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i-\widehat{Y}_i\right)^2 \qquad \text{Equation 5-4}$$

- Mean absolute error (MAE) is the average of the absolute error values. Unlike MSE, the units of the error match the units of the predicted target value. While MSE (and RMSE) weight larger errors more than smaller errors (due to the square of the error value), MAE does not give more or less emphasis to different types of errors (Equation 5-5).

$$MAE = \frac{\sum_{i=1}^{n} \left|\hat{Y}_i - Y_i\right|}{N} \qquad \text{Equation 5-5}$$

- Root mean squared error (RMSE) is an extension of MSE where the square root of the error is calculated. The units of RMSE are the same as the original units of the value being predicted (Equation 5-6).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2}{N}} \qquad \text{Equation 5-6}$$

- Symmetric mean absolute percentage error (SMAPE) is an accuracy metric based on relative error (Equation 5-7).

$$SMAPE = \frac{1}{N} \sum_{i=1}^{N} \frac{\left|Y_i - \hat{Y}_i\right|}{\left(|Y_i| + \left|\hat{Y}_i\right|\right)/2} \qquad \text{Equation 5-7}$$

- The coefficient of determination (R^2) is a measure of how well the predictions fit the data (Equation 5-8).

$$R^2 = 1 - \sum_{i=1}^{N} \frac{\left(Y_i - \hat{Y}_i\right)^2}{(Y_i - \mu)^2} \qquad \text{Equation 5-8}$$

In each of the equations above, Y is the actual grass growth value, $\hat{Y}$ is the predicted grass growth value, N is the number of observations, i is the iterations, and μ is the mean value of the actual grass growth values.

5.2.5 Methodology

The methodology for the models described in this chapter is illustrated in Figure 5-2.

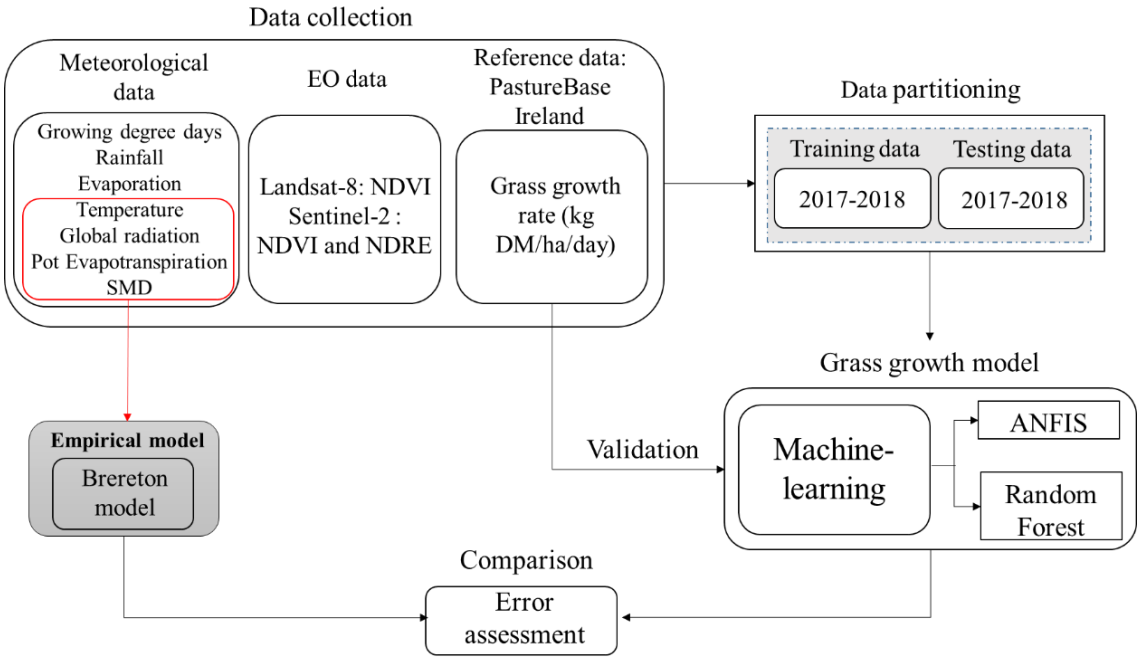


Figure 5-2 Methodology for Brereton, ANFIS and RF model

Individual databases were compiled for Landsat 8 and Sentinel 2 VI containing the weather data, modified GDD and the respective VI for both years (2017 and 2018). The Landsat 8 database comprised 131 rows (corresponding to image acquisition dates) and 11 individual columns. The Sentinel 2 database comprised 160 rows (dates) and 12 individual columns. The discrepancy between the number of rows between the Landsat 8 and Sentinel 2 databases is because of cloud cover and different revisit times (16 days for Landsat 8 and 5 days for Sentinel 2).

The data from 2017 and 2018 were randomly split into non-overlapping training (70%) and testing (30%) datasets. The same training data were used for both ANFIS and RF models. The error metrics listed in Section 5.2.4.4 were used to indicate the reliability of each model. A seed value is a defined starting point for random number generation and was set for all the models to ensure the reproducibility of results. Using a seed value ensures the model reproducibility.

The input data to ANFIS is in the form of a matrix (m × n), where m is the number of observations and n is the number of variables. Various combinations for the parameters were tested, and the parameters with the lowest RMSE were selected. Maximum iterations were set to 100, and the step size was chosen as 0.1, which are the default values.

A repeated k-fold cross-validation process was used to determine the optimal value of ‘mtry’. Repeated k-fold cross-validation is a resampling procedure to evaluate how an ML model is likely to perform on data not used to train the model. The general procedure for k-fold cross-validation is to split the dataset into k groups randomly. Data are further split into k groups for each unique group, using one group as a test set and the remaining groups as training data. A model is fit and evaluated multiple times, and a mean result from all folds is determined. The ‘mtry’ value with the lowest RMSE was selected. k values from 1 to 15 were tested, and the one with the lowest RMSE selected (k-values and RMSE are given in Appendix 5.5).

The ‘ntree’ parameter can be any value, and increasing the number of trees will result in better accuracy up to a point. If the number of trees is increased beyond this threshold, there will be no significant impact on model accuracy, but the computing time will be increased. Each tree is constructed using a subset of the original training data (bootstrapping), leaving some observations “out-of-bag” (OOB) (Breiman, 2001, Lee et al., 2020). These OOB observations are used to calculate the model's prediction error (Janitza and Hornung, 2018). OOB error is plotted against the number of trees to estimate the optimal number of trees, and the value with the lowest error was chosen.

5.3 Results

5.3.1 Brereton model

The Brereton model was implemented in Python 3.8 and was developed for three sites in Ireland: Moorepark, Co. Cork, Athenry in Co. Galway, and Johnstown Castle, in Co. Wexford. The model was run for each site for two years separately, i.e. 2017 and 2018, as the model can only be run for a single year with daily parameters. The grass growth rate from the Brereton model was compared with the data from the PBI.

5.3.1.1 Moorepark

The grass growth rate from PBI with the predictions from the Brereton model for 2017 and 2018 are shown in Figure 5-3 A and C, and the corresponding scatter plot is shown in 5-3 B and D, respectively.

In 2017, the Brereton model over-predicted for winter (January, February and December, i.e. DOY 1 to 59 and from 335 to 365) and for spring (March until May, i.e. DOY from 60 to 151) by approximately 6.58 kg DM $ha^{-1}day^{-1}$ and 35.51 kg DM $ha^{-1}day^{-1}$. The model under-predicted by 11.43 and 20.01 kg DM $ha^{-1}day^{-1}$ for summer (June, July and August, i.e. DOY from 152 to 243) and autumn (September, October and December, i.e. DOY from 244 to 334). The Brereton model captures the same overall trend as the PBI data. In 2018 also, the Brereton model over-predicted for winter, spring and summer by approximately 15.78, 42.0 and 6.43 kg DM $ha^{-1}day^{-1}$. On the other hand, the model under-predicted 15.73 kg DM $ha^{-1}day^{-1}$ for autumn.

The relationship between the Brereton model and PBI values in each of the years were explored using scatterplots. For 2017 and 2018, the R^2 was 0.26 ($p<0.05$) and 0.33 ($p<0.05$), respectively. For 2017, the model had variability in the predictions, whereas for 2018, there was a cluster of values between 0 and 50 kg DM $ha^{-1}day^{-1}$ with few outliers. The error metrics for all the farms for 2017 and 2018 are given in

Table 5-4. The RMSE value for Moorepark for 2017 (41.68 kg DM $ha^{-1}day^{-1}$) was lower than 2018 (50.74 kg DM $ha^{-1}day^{-1}$).

Table 5-4 Grass growth rate (kg DM $ha^{-1}day^{-1}$) output from the Brereton model (2017-2018).

*Five accuracy metrics are presented- R^2, MSE, RMSE, SMAPE and MAE. Key: Mp = Moorepark farm, Jc = Johnstown Castle farm and Ath = Athenry farm. Lowest RMSE and highest R^2 value in each year in **bold**.*

Metric	2017			2018		
	Mp	Jc	Ath	Mp	Jc	Ath
R^2	**0.26** p< 0.05	0.03 P = 0.3	0.06 P = 0.16	**0.33** p< 0.05	0.17 p< 0.05	0.16 p< 0.05
MSE	1737.87	6772.33	3643.57	2575.05	2580.02	3106.26
RMSE	**41.68**	82.29	60.36	**50.74**	50.79	55.73
SMAPE	74.14	77.36	77.84	82.14	75.33	83.13
MAE	31.04	52.65	37.26	32.42	29.92	34.79

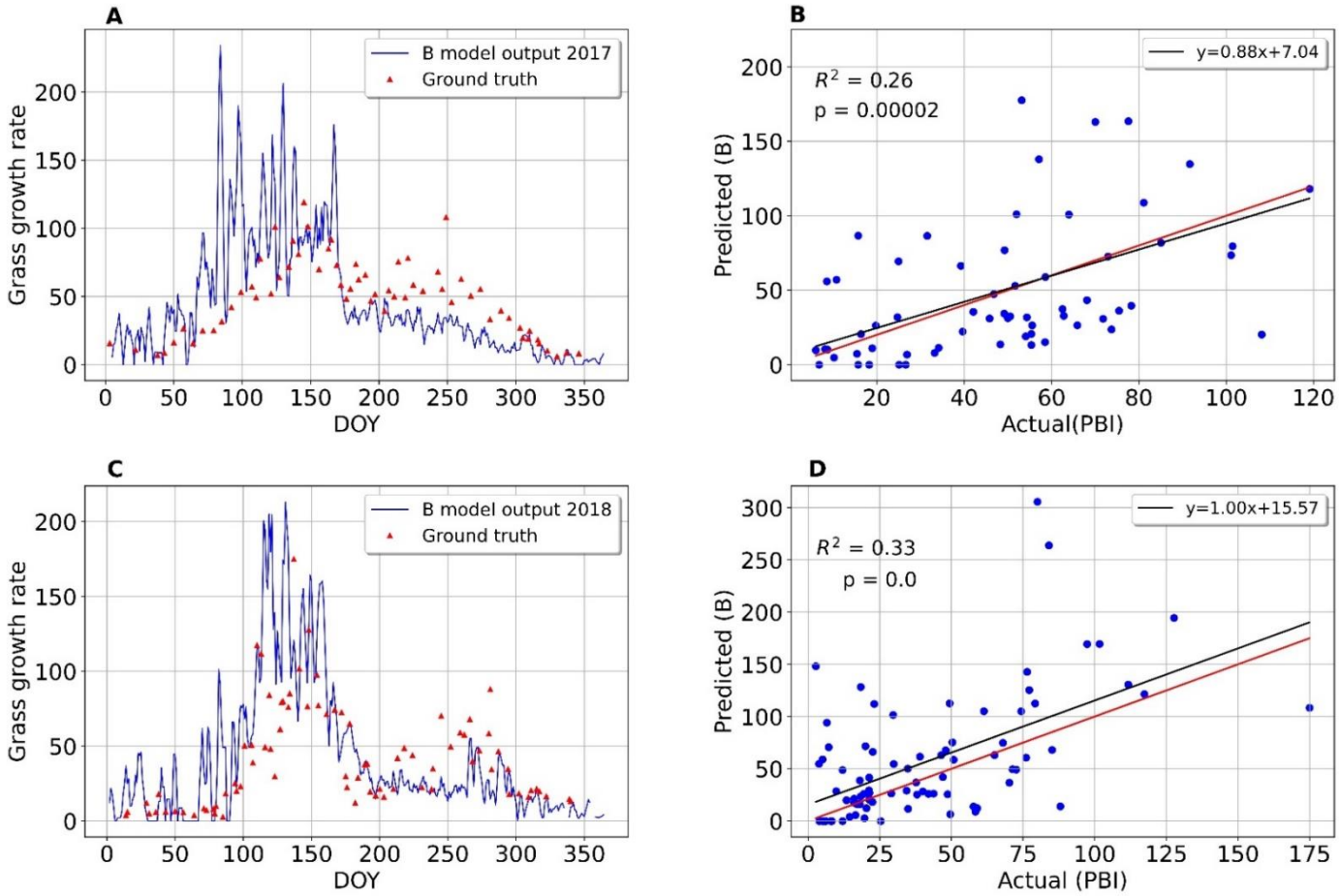


Figure 5-3 Predicted grass growth at Moorepark using the Brereton model (2017-2018).

A and B represent the results for 2017, and C and D are those for 2018. The corresponding scatter plot between predictions from the Brereton model -Predicted (B) and PBI data - Actual (PBI) is shown. The grass growth rate values from the Brereton model are shown as a blue line, and red triangles represent the reference data from PBI, where the x-axis is the day of the year (DOY) and the y-axis is the grass growth rate values in kg DM $ha^{-1}day^{-1}$. The 1:1 line is shown in the scatter plot in red.

5.3.1.2 Johnstown Castle

Analysis for Johnstown Castle is illustrated in Figure 5-4 A and C, and the corresponding scatter plot is shown in B and D, respectively. In 2017, the model over-predicted for winter and spring by 36.45 kg DM $ha^{-1}day^{-1}$ and 107.82 kg DM $ha^{-1}day^{-1}$ approximately. Conversely, the model under-predicted by 1.58 and 15.08 kg DM $ha^{-1}day^{-1}$ for summer and autumn. In 2018 also, the Brereton model over-predicted for winter, spring and summer by 11.45, 52.16 and 3.05 kg DM $ha^{-1}day^{-1}$ on an average. On the other hand, the model under-predicted by 13.47 kg DM $ha^{-1}day^{-1}$ for autumn. For 2017 and 2018, the R^2 value between the grass growth rate from PBI and Brereton model is 0.03 ($p = 0.3$) and 0.17 ($p = 0.01$), respectively. There is a positive relationship between the predictions and the reference data for 2017 and 2018. There is a linear relationship between the two variables except for a cluster of outliers values between 25 and 55 kg DM $ha^{-1}day^{-1}$ grass growth rate for 2017. There is a weak linear relationship between variables for 2017 as the slope is 0.57, whereas, for 2018, there is a strong relationship as the slope is closer to 1 (0.98), which is similar to Moorepark but a different offset. The 2017 predictions have p values of less than 0.05, meaning that the relationship is not statistically significant. 2018 has a statistically significant relationship with a p-value of less than 0.05. The RMSE value for 2017 (82.29 kg DM $ha^{-1}day^{-1}$) was higher than in 2018 (50.79 kg DM $ha^{-1}day^{-1}$).

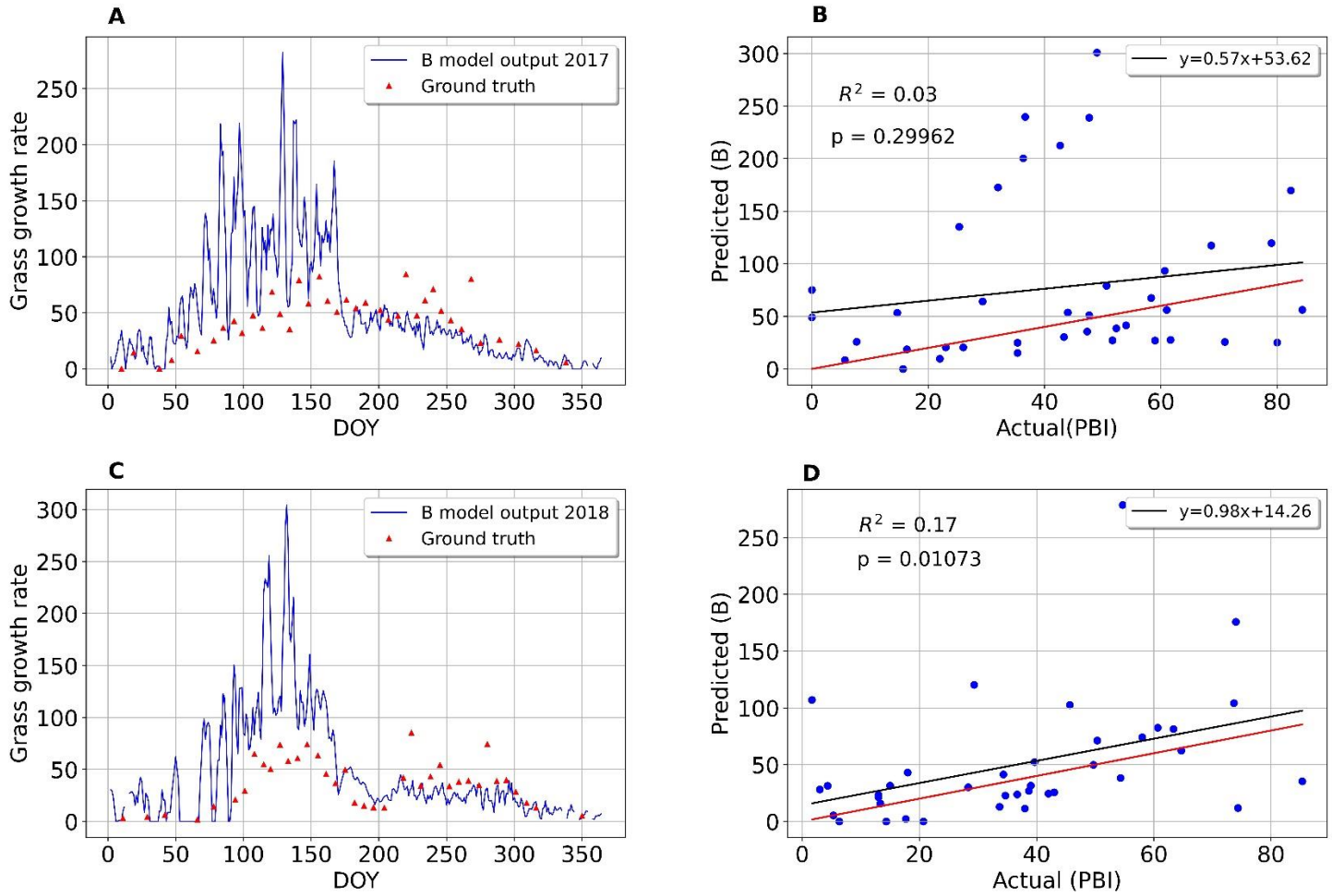


Figure 5-4 Predicted growth at Johnstown Castle from Brereton model (2017-2018)

A and B represent the results for 2017, and C and D are those for 2018. The corresponding scatter plot between predictions from the Brereton model -Predicted (B) and PBI data - Actual (PBI) is shown. Key: Brereton model (blue line) and PBI (red triangles), and DOY.

5.3.1.3 Athenry

The results for Athenry for 2017 and 2018 are shown in Figure 5-5. The grass growth curves are shown in Figure 5-5 A and C, and scatter plots are shown in B and D. In 2017, the model under-predicted for winter by 5 kg DM $ha^{-1}day^{-1}$. The model over-predicted by 11.43 kg DM $ha^{-1}day^{-1}$ for spring. For summer and autumn, the model under-predicted by 3.13 and 19.80 kg DM $ha^{-1}day^{-1}$. In 2018 also, the model under-predicted for winter and autumn by approximately 1 and 6.80 kg DM $ha^{-1}day^{-1}$. For spring and summer, the model over-predicted by 56.13 and 3.37 kg DM $ha^{-1}day^{-1}$.

The R^2 values for Athenry farm for 2017 and 2018 are 0.06 (p-value = 0.15) and 0.16 (p-value < 0.05), respectively. There is a very weak relationship between the predictions and the actual values as the correlation coefficient was 0.23, whereas, for 2018, the relationship was weak with a correlation coefficient of 0.4. The direction of the relationship for both 2017 and 2018 is positive. The shape of the scatter plot of grass growth rate values for both 2017 and 2018 are linear, with some outliers. The relationship between the predictions for 2017 is not statistically significant as the p-value is more than 0.05, whereas the predictions for 2018 are statistically significant as the p-value is less than 0.05. The RMSE (60.36 kg DM $ha^{-1}day^{-1}$) and MAE (37.26 kg DM $ha^{-1}day^{-1}$) values for Athenry farm for 2017 were higher than in 2018 (RMSE=55.73 kg DM $ha^{-1}day^{-1}$, and MAE=34.79 kg DM $ha^{-1}day^{-1}$).

Moorepark farm had the lowest overall RMSE of 41.68 kg DM $ha^{-1}day^{-1}$ in 2017 and 50.74 kg DM $ha^{-1}day^{-1}$ than Johnstown Castle (82.29 for 2017and 50.79 kg DM $ha^{-1}day^{-1}$ for 2018) and Athenry farm (60.36 for 2017 and 55.73 kg DM $ha^{-1}day^{-1}$ for 2018).

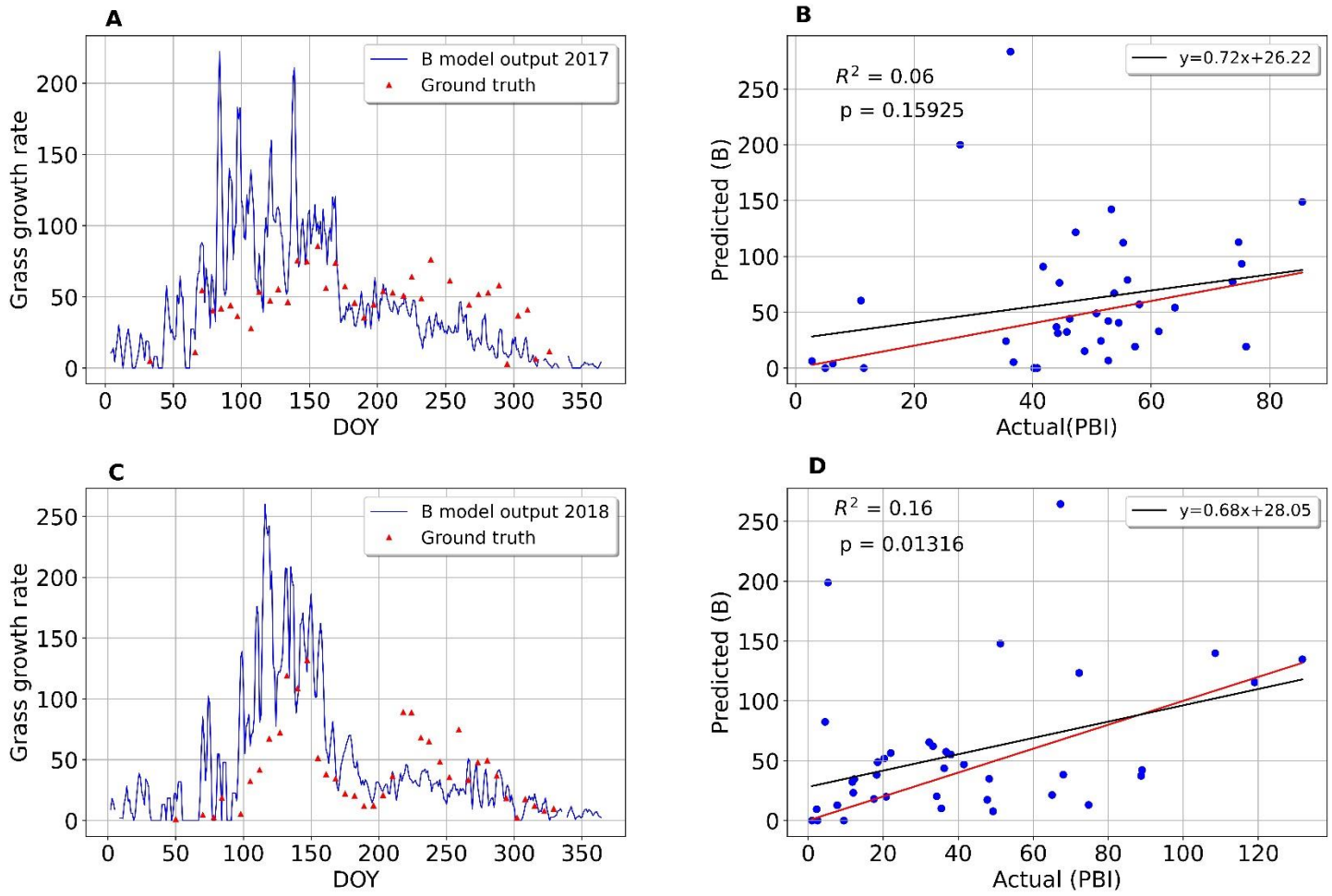


Figure 5-5 Predicted grass growth at Athenry using Brereton model (2017-2018).

A and B represent the results for 2017, and C and D are those for 2018. The corresponding scatter plot between predictions from the Brereton model -Predicted (B) and PBI data - Actual (PBI) is shown. Key: Brereton model (blue line) and PBI (red triangles) and DOY.

5.3.2 Machine-learning models

5.3.2.1 Feature selection

The Boruta and VarImp functions were used for feature selection for Landsat 8 (Figure 5-6 and 5-7) and Sentinel 2 databases (Figure 5-8 and 5-9). The optimal variables for the Landsat 8 model were global radiation, potential evapotranspiration (pe), rainfall, soil moisture deficit, evaporation, GDD, maximum (max), minimum (min), and mean (mean) temperature (green box plots). Both the Boruta and VarImp functions rejected NDVI as an essential predictor variable (red box plot). The pe was the most important variable using both Boruta plot and VarImp function. The blue box plots randomly duplicate original variables called "shadow" variables, which acts as a threshold to select the attributes. Boruta calculates the Z score for each input variable and their shadow variables. If the Z-score of the variables is more than the shadow attributes, the variables are labelled as necessary, whereas if the Z-score is lower than the shadow variables, those variables are marked unimportant and are rejected.

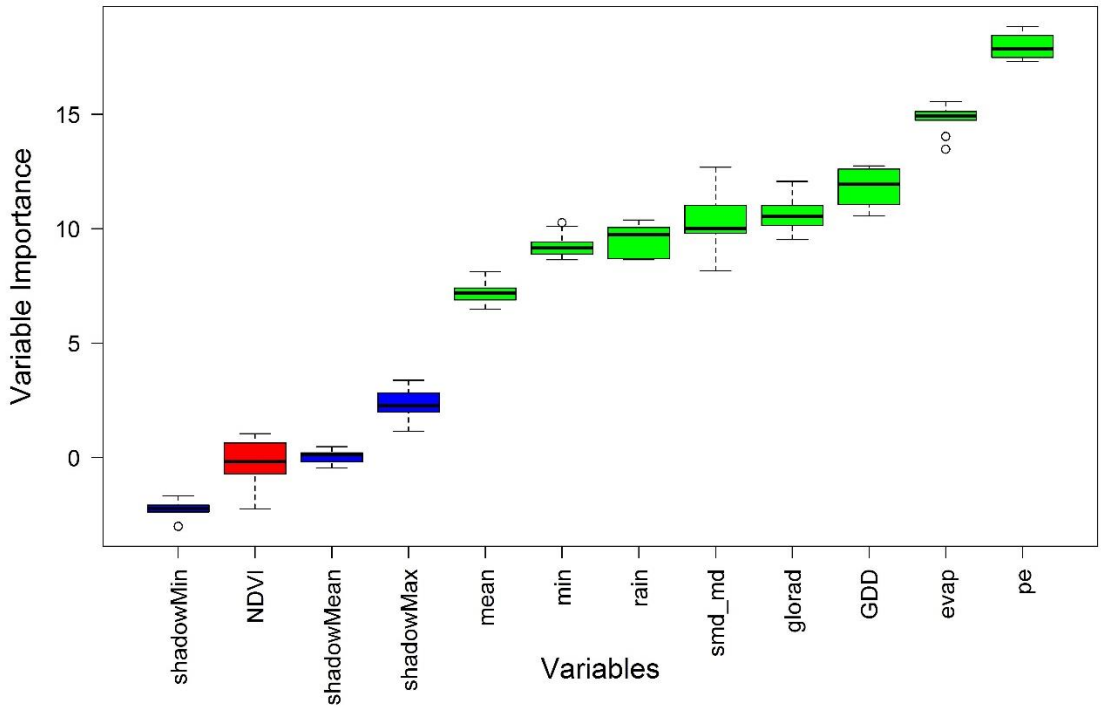


Figure 5-6 Boruta plot for the Landsat 8 database.

NDVI is a rejected variable (red boxplot), whereas the meteorological variables are all important (green boxplots), the blue boxplots indicate the threshold of importance. Mean and min refers to mean and minimum temperature, pe refers to potential evapotranspiration, evap refers to evaporation, rain is rainfall, and GDD is growing degree days.

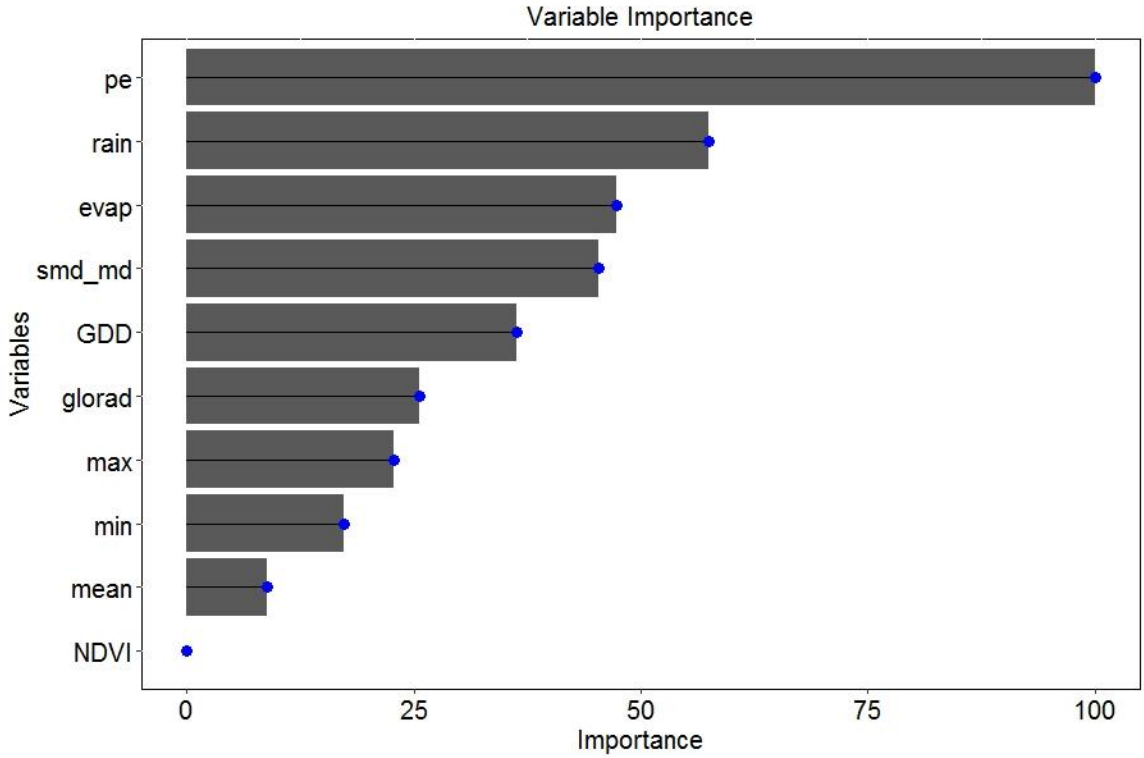


Figure 5-7 Landsat 8 VarImp output from the 'Caret' package
NDVI has zero importance value, which means that it is rejected. Mean and min refers to mean and minimum temperature

For the Sentinel 2 database, the plots in Figure 5-8 and 5-9 show that potential evapotranspiration, evaporation, GDD, SMD, NDVI, NDRE, mean (mean), maximum (max) and minimum (min) temperature were the most critical factors. Rainfall was designated as tentative, as shown in the yellow box-plot, which means that Boruta cannot confirm its significance with the desired confidence within the default number of random forest runs. For VarImp, rainfall was the rejected variable. The 'pe' is very similar to 'evap' in Sentinel 2 database, unlike for Landsat 8.

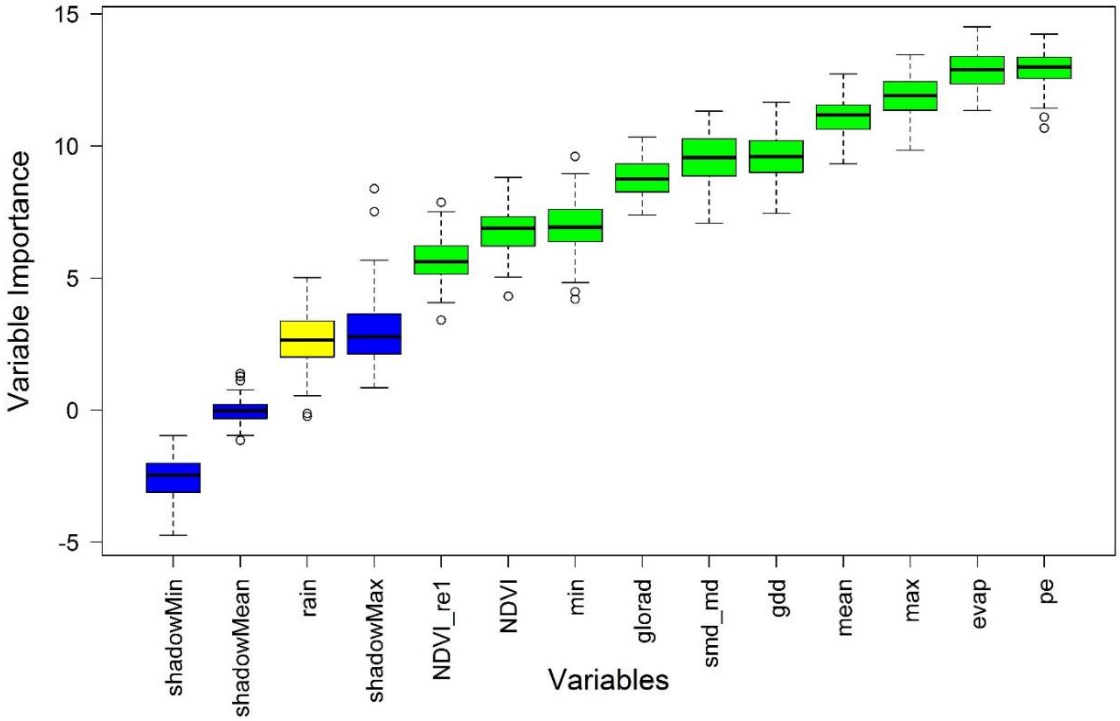


Figure 5-8 Boruta plot for Sentinel 2 database.
Mean and min refers to mean and minimum temperature.

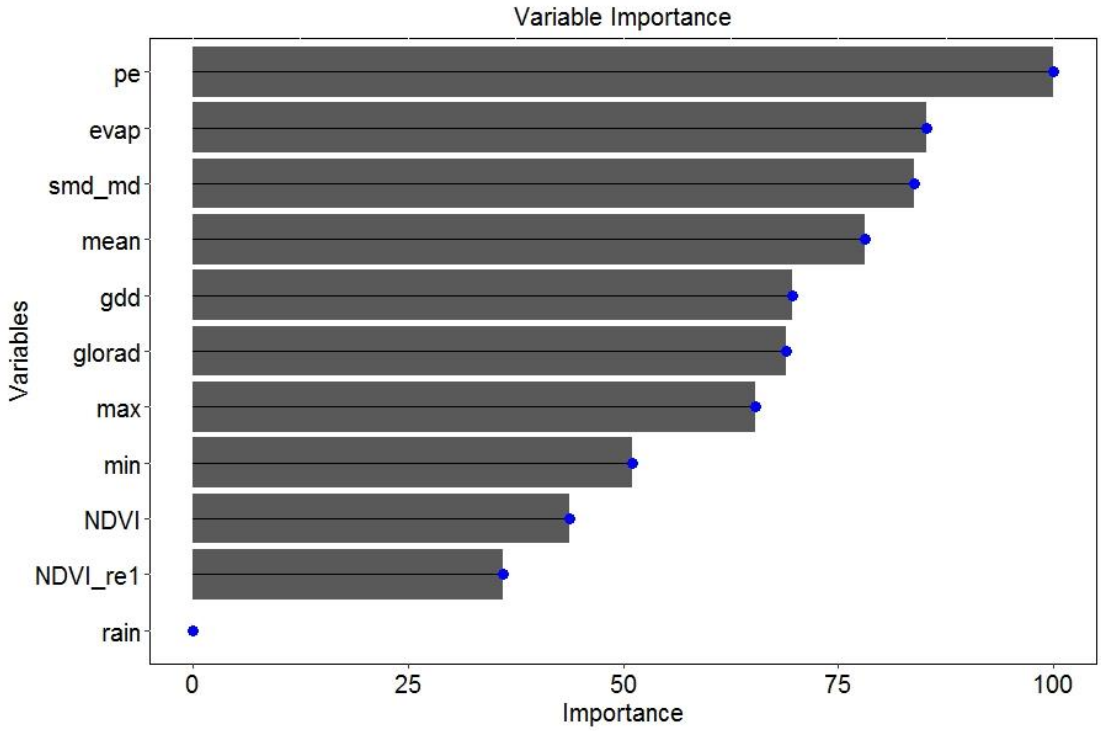


Figure 5-9 Sentinel 2 'Varimp' output plot showing the importance of all the variables. Rainfall has zero importance value, which means that it is rejected. Mean and min refers to mean and minimum temperature

Based on the outputs from the feature selection algorithms, six sets of parameters were defined, and each used with the ANFIS and RF models (Table 5-5). For the Landsat 8 database, two different combinations were evaluated – one with all variables and another with all the variables excluding NDVI. For the Sentinel 2 database, four combinations were formed- one using all the variables, a second model excluded rainfall, a third model excluded NDRE and rainfall, and a fourth model excluded NDVI and rainfall.

Table 5-5 Twelve models based on optimal variables from feature selection

Model	Sub-models
ANFIS- Landsat 8	Model 1- All variables
	Model 2- All variables except NDVI
ANFIS- Sentinel 2	Model 3- All variables
	Model 4- All variables except rainfall
	Model 5- All variables except NDVI and rainfall
	Model 6- All variables except NDRE and rainfall
RF- Landsat 8	Model 7- All variables
	Model 8- All variables except NDVI
RF- Sentinel 2	Model 9- All variables
	Model 10- All variables except rainfall
	Model 11- All variables except NDVI and rainfall
	Model 12- All variables except NDRE and rainfall

5.3.2.2 Models 1 & 2 (ANFIS Landsat 8)

The tuned ANFIS model was used to predict the grass growth values from testing data. The scatter plots for training (A and C) and testing data (B and D) are shown in Figure 5-10 with actual grass growth values from PBI on the x-axis and the predicted values from ANFIS on the Y-axis. R^2 values for training and testing data were 0.28 and 0.55 for model 1. Model 2, using only meteorological data, had a lower R^2 of 0.32 and 0.49 for training and testing stages than Model 1. The scatter plots for testing data for both years are similar in terms of the spread of the points along the 1:1 line. The values between 55-85 kg DM $ha^{-1}day^{-1}$ are over-predicted for testing data in 2017 and 2018. Model 1 (RMSE of 22.14 kg DM $ha^{-1}day^{-1}$) performed slightly better than model 2 (RMSE of 24.76 kg DM $ha^{-1}day^{-1}$). A similar outlier for testing data for both the model at an actual value of 130 and a predicted value of 90 kg DM $ha^{-1}day^{-1}$. The error metrics for all the models are given in Table 5-7.

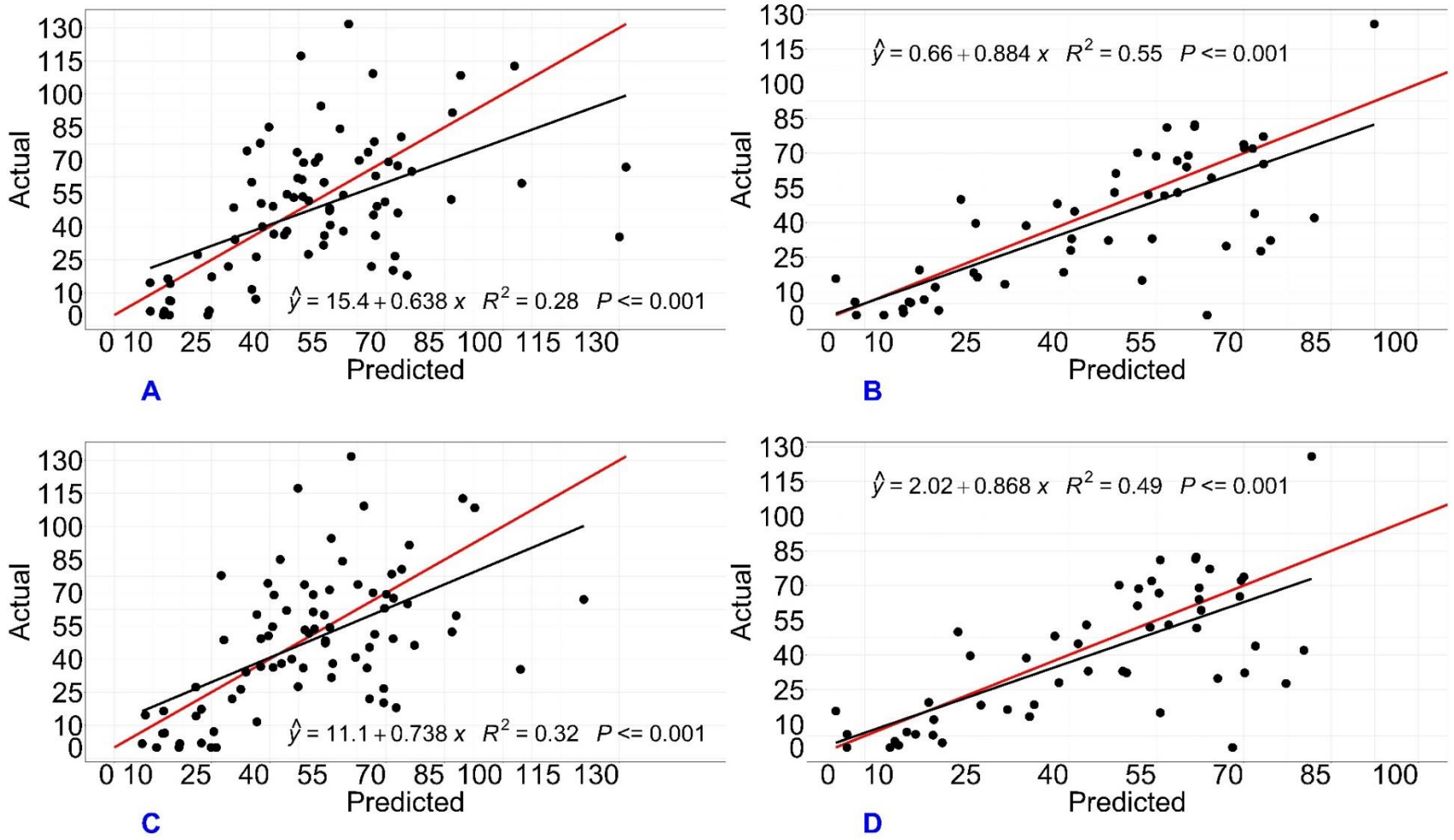


Figure 5-10 Actual vs. predicted grass growth for models 1-2.

Model 1: Training is Panel A. Testing is Panel B. Model 2: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$.

5.3.2.3 Models 3-6 (ANFIS Sentinel 2)

The scatter plots for training and testing data for all four models are shown in Figures 5-11 and 5-12, with actual grass growth values from PBI on the x-axis and the predicted values from ANFIS on the Y-axis. The scatter plot for training data are in Figure 5-11 A and C and testing data in B and D. The R^2 for the training data (0.41-0.44) for all the models (4-6) is lower than the testing data (0.47-0.52) except model 3 in which the R^2 for training data (0.33) is higher than the testing data (0.32).

The R^2 was lowest (0.32 for test data) when all the variables were used as input to the model. Models 5 and 6 had similar performance with R^2 of 0.51 for model 5 and 0.52 for model 6, and both the models are without rainfall. Model 5 with meteorological with NDVI (15.32 kg DM $ha^{-1}day^{-1}$) performed slightly better than model 6 with meteorological data and NDRE (15.04 kg DM $ha^{-1}day^{-1}$). The model with all the variables performed the worst among all models with an RMSE of 18.11 kg DM $ha^{-1}day^{-1}$.

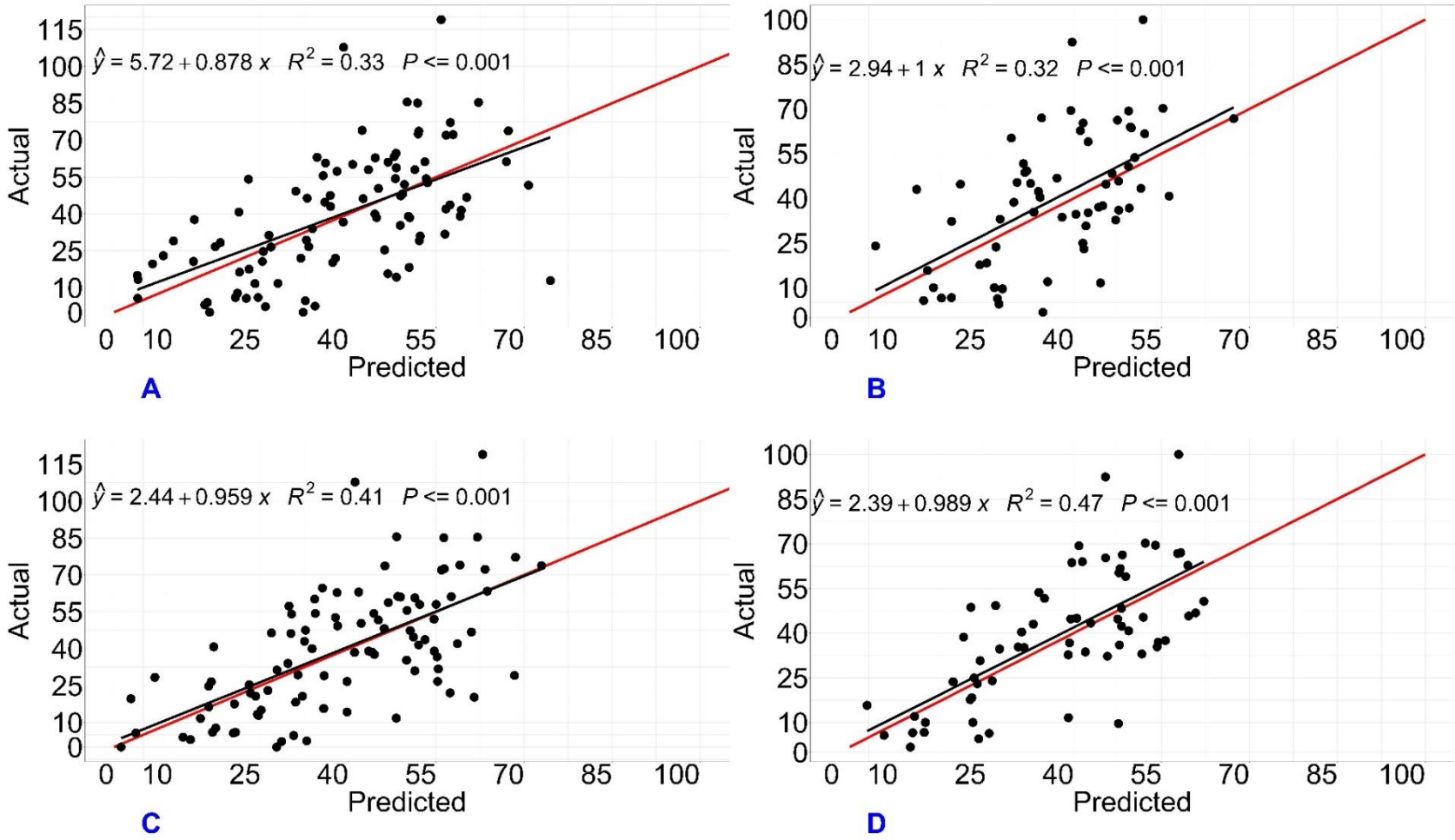


Figure 5-11 Actual vs. predicted grass growth for models 3-4.

Model 3: Training is Panel A. Testing is Panel B. Model 4: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$.

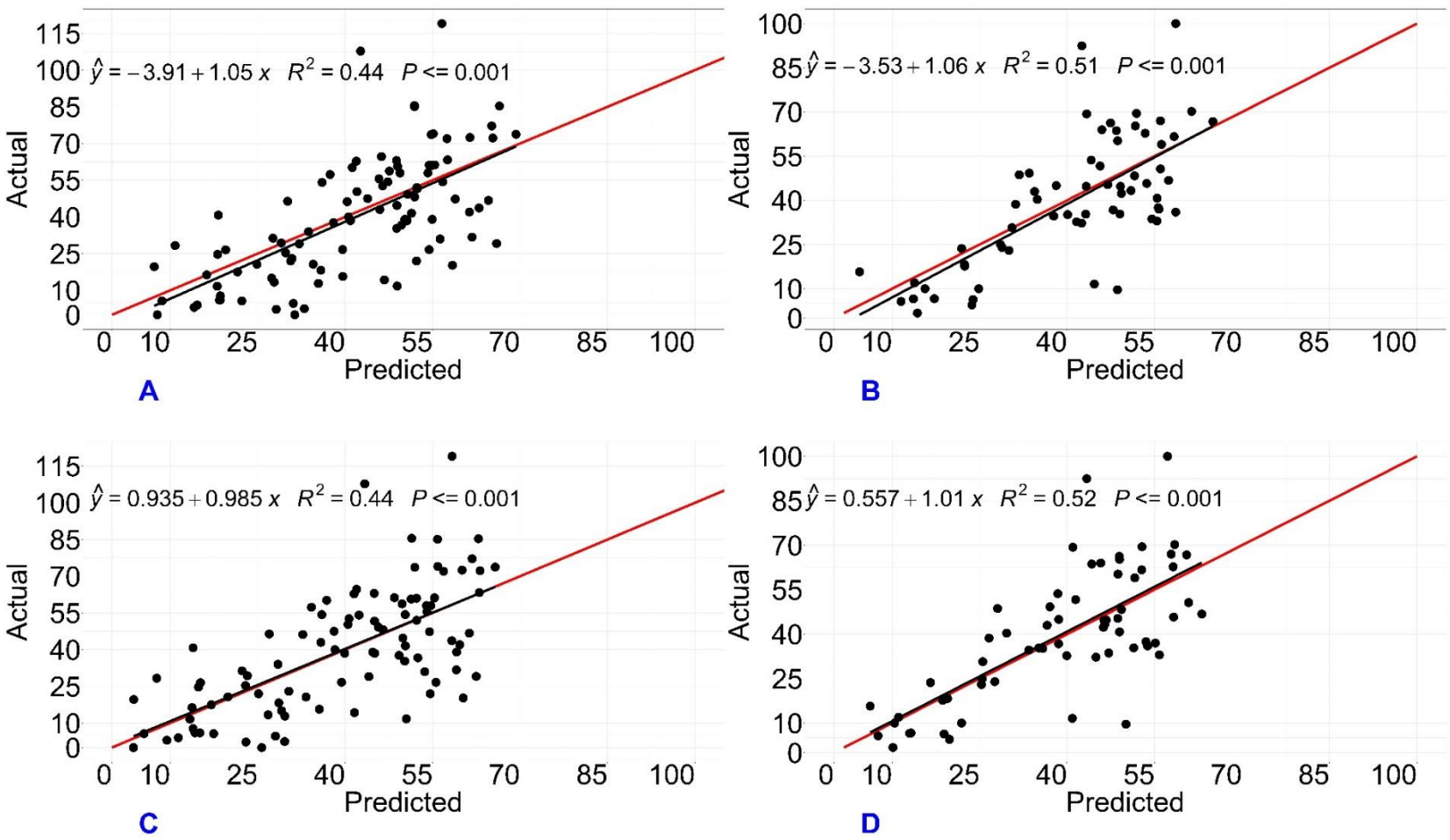


Figure 5-12 Actual vs. predicted grass growth for models 5-6.

Model 5: Training is Panel A. Testing is Panel B. Model 6: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$.

5.3.2.4 Models 7 & 8 (RF Landsat 8)

For tuning of RF, the essential parameters are- mtry and ntrees. To estimate the optimal number of trees, we plotted OOB error against several trees and the trees with the lowest error was chosen. The plot of OOB error with many trees with the mtry tables is shown in Appendix 5.6. The optimal number of trees with the lowest error is 219 trees with an average grass growth error of 22.28 kg/DM/ha/day. RF model uses cross-validation and chooses an mtry value with the lowest RMSE value. The mtry value of 10 was chosen in the final model, which has an RMSE value of 20.60 kg DM $ha^{-1}day^{-1}$. The difference between RMSE values for various mtry values was small. The comparison of Model 7 (all the variables) and 8 (without NDVI) optimal trees, grass growth rate value at an optimal tree and 'mtry' are shown in Table 5-6.

Table 5-6 Comparison of RF models 7 & 8.
R^2 for training and testing, optimal trees, grass growth rate value at optimal tree and mtry are shown for Model 7 (all the variables) and Model 8 (without NDVI).

Model	ntree	Grass growth rate value at optimal tree	mtry
7	219	22.28	10
8	130	21.72	9

The scatter plot for training, and testing data is shown in Figure 5-13, respectively, with actual grass growth values from PBI on the x-axis and the predicted values from RF on Y-axis. Figure 5-13 A and C are training data scatter plot for Model 7 and 8, and the testing data scatter plot are shown in B and D. R^2 was 0.92 for training data and 0.73 for testing data using model 7. For test data, the low values were predicted accurately. For model 8 the optimal number of trees with the lowest error was 130 trees with an average grass growth error of 21.72 kg/DM/ha/day, and an mtry value of 9 was chosen for the final model. R^2 was 0.90 for training data and 0.72 for testing data for model 8. It can be seen that the values of accuracy metrics are similar for both models. However, the RMSE for Model 7 (14.70 kg DM $ha^{-1}day^{-1}$) is lower than for model 8 (15.33 kg DM $ha^{-1}day^{-1}$).

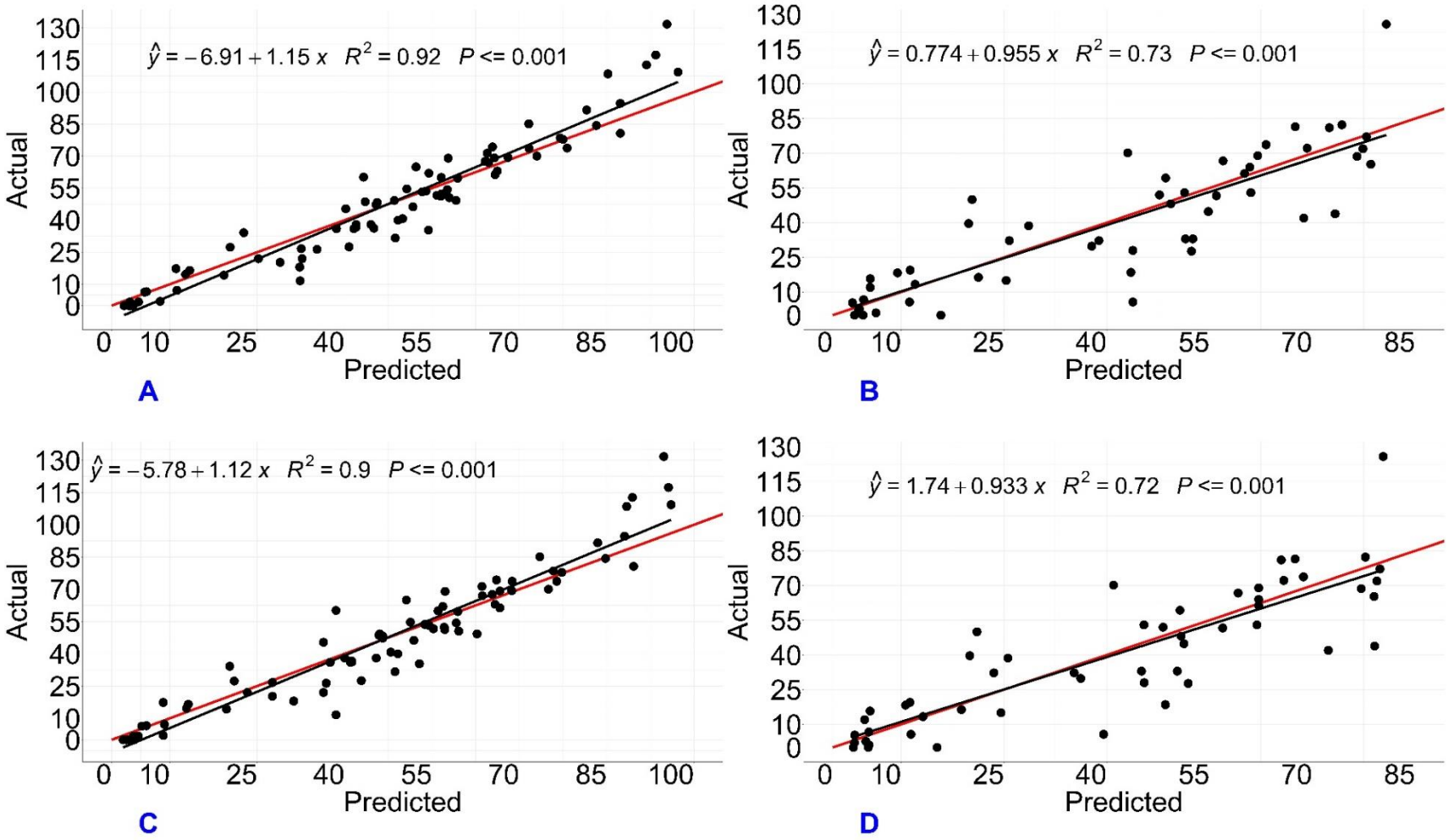


Figure 5-13 Actual vs. predicted grass growth for models 7-8.

Model 7: Training is Panel A. Testing is Panel B. Model 8: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$.

5.3.2.5 Models 9-12 (RF Sentinel 2)

For model 9-12, the optimal number of trees with the lowest error are 69, 103, 209 and 96 trees with an average grass growth error of 16.94, 17.28, 17.19 and 17.31 kg/DM/ha/day for model 9-12. For model 9, 10, 11 and 12 mtry value of 2 was chosen in the final model with an RMSE of 16.46, 16.47, 16.51 and 16.54 kg DM $ha^{-1}day^{-1}$.

The scatter plot for training and testing data for all the models are shown in Figures 5-14 and Figure 5.15. R^2 was 0.90 and 0.55, respectively, for model 9 and 10, R^2 was 0.91 and 0.56 for model 11, and R^2 was 0.91 and 0.55 for model 12, respectively. Although it was observed that R^2 values are significantly high for the testing stage using Landsat 8 data (0.72-0.73) than using Sentinel 2 (0.55-0.56) and the RMSE of the Sentinel 2 model (14.65-14.94 kg DM $ha^{-1}day^{-1}$) is lower than Landsat 8 (14.70-15.33 kg DM $ha^{-1}day^{-1}$). The R^2 for training data was similar for all the models (0.90-0.91). For all the four models for testing data, the values below 40 kg DM $ha^{-1}day^{-1}$ were close to the actual data with the values along the 1:1 line. The values above 40 kg DM $ha^{-1}day^{-1}$ had more variability in the data with under-prediction.

The error metrics for all the models for testing data is shown in Table 5-7. The lowest R^2 of 0.47 for testing data was for model 6, and the highest value of 0.73 is for model 7. For training data, the lowest R^2 of 0.28 was for model 1, whereas models 7 until 12 had the highest R^2 from 0.90-0.92. Model 2 had the highest (24.76 kg DM $ha^{-1}day^{-1}$), and model 10 had the lowest RMSE (14.65 kg DM $ha^{-1}day^{-1}$) of all the models.

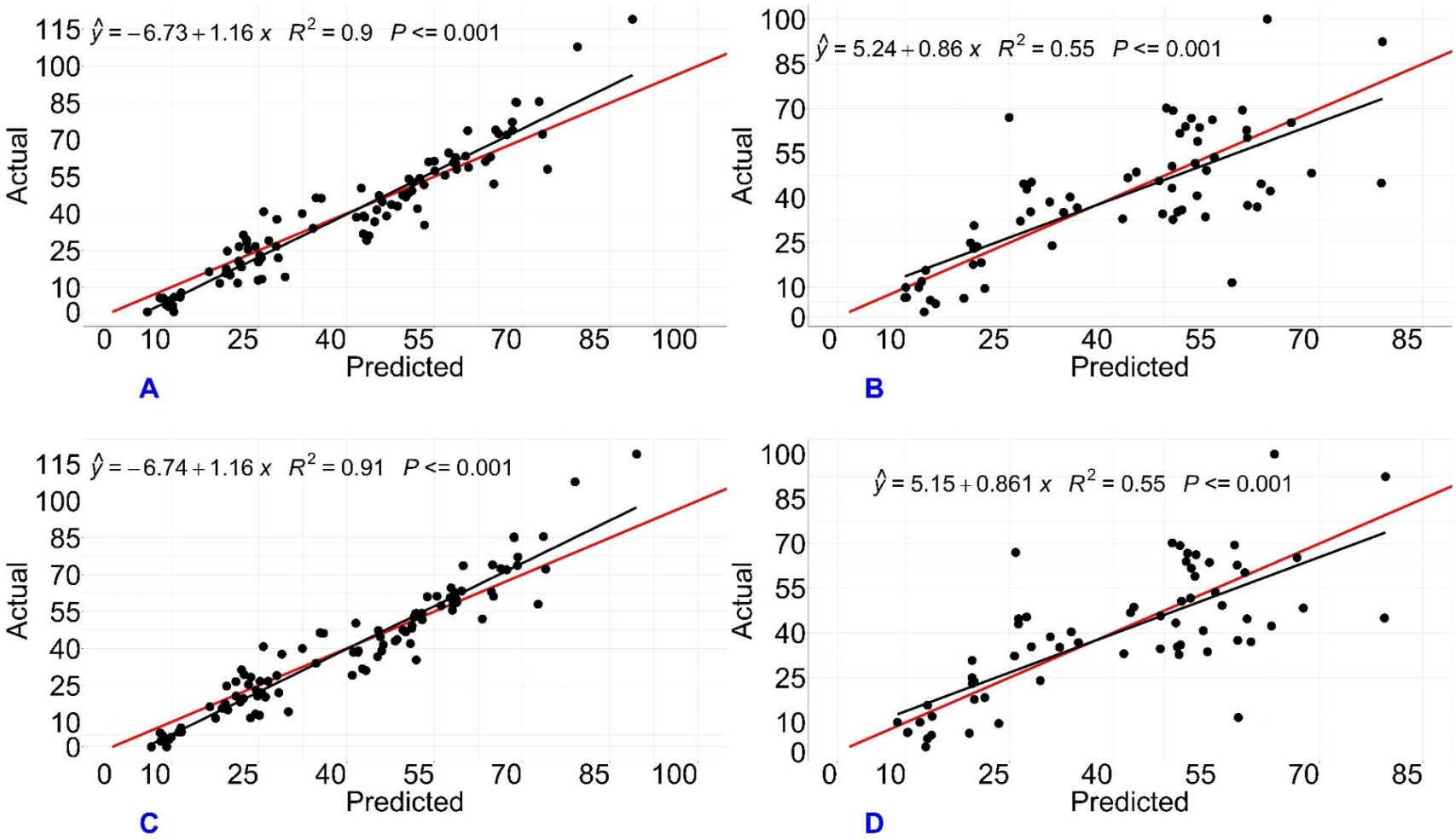


Figure 5-14 Actual vs. predicted grass growth for models 9-10.

Model 9: Training is Panel A. Testing is Panel B. Model 10: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$.

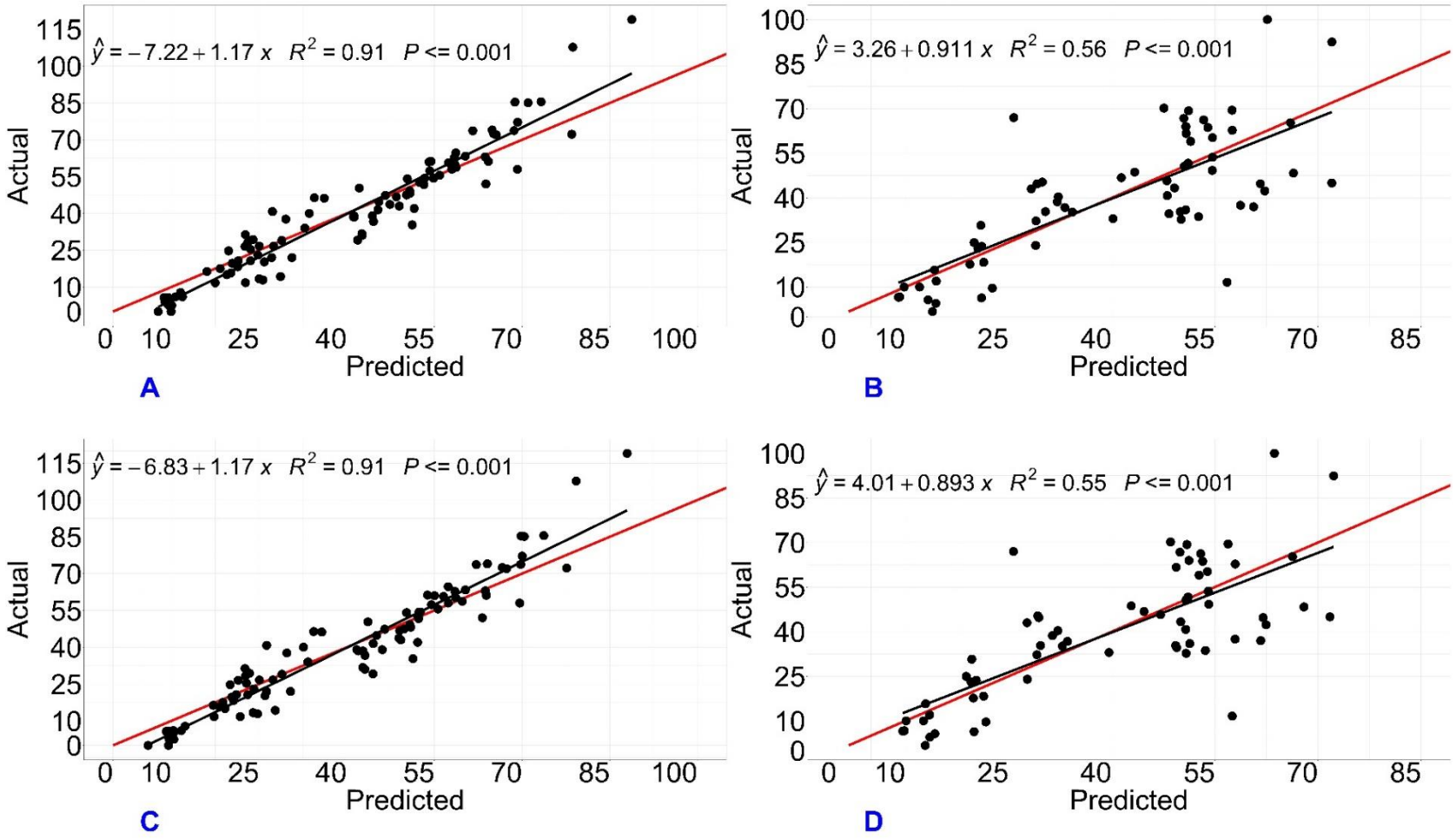


Figure 5-15 Actual vs. predicted grass growth for models 11-12.

Model 11: Training is Panel A. Testing is Panel B. Model 12: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$.

Table 5-7 Error metrics for machine-learning models

Showing R^2, MSE, MAE, RMSE, SMAPE (all in kg DM $ha^{-1}day^{-1}$) for all the models 1-12 arranged in ascending order by RMSE.

Model	R^2 Training data	R^2 Testing data	MSE	MAE	RMSE	SMAPE
10	0.91	0.55	214.79	11.23	**14.65**	8.28
9	0.90	0.55	215.68	11.01	14.68	7.95
7	**0.92**	0.73	216.35	10.93	14.7	11.6
12	0.91	0.55	219.96	11.08	14.83	8.11
11	0.91	0.56	223.31	11.26	14.94	8.35
6	0.44	0.52	226.47	11.19	15.04	8.32
5	0.44	0.51	234.79	11.82	15.32	9.16
8	0.9	0.72	235.16	11.31	15.33	11.75
4	0.41	0.47	254.4	12.55	15.95	9.5
3	0.33	0.32	327.99	14.53	18.11	11.38
1	0.28	0.55	490.34	16.49	22.14	Nan
2	0.32	0.49	473.58	16.49	24.76	Nan

5.4 Discussion

5.4.1 Limitations of ANFIS model

For ANFIS model, the testing data had higher R^2 and better error statistics than the training data, which is unusual. The possible reason for this could be data leakage which means unintentional use of testing data into the training data (Just et al., 2020). There could be two reasons for the data leakage specifically for this work. The first one is that the dataset from 2017-2018 was divided randomly into training and testing data. This creates a problem in time-series data because the data from 2018 are used to predict for 2017 creating a bias. A solution for this is cross-validation by dividing the data into k-folds and using k-1 data for training and 1 for testing the model. Cross-validation is not available for ANFIS in R. Secondly, the data were normalized before splitting into training and testing data. While normalizing, the data are divided by the mean value of the whole dataset including the test data, which can influence the training data. One possible solution for this is to normalize the training and testing data separately. Another solution to avoid leakage is to keep a validation data to test the trained model for a completely unseen dataset, which was followed in Chapter 6. In Chapter 6, the data from 2107-2019 were divided into training and testing data and the data for 2020 were used to validate the model which were completely unseen by the trained random forest.

To test this data leakage effect, the data were divided into 2 parts - 2017 for training and 2018 for testing the ANFIS model. It was seen that training data had better error metrics than the testing data. The results shown in Section 5.3.2.2 and 5.3.2.3 were a case of data leakage and therefore the testing data had better R^2 than the training data. In this example, training data had better R^2.

The RMSE for model 1 was 29.17 kg DM ha^{-1} day^{-1} and for model 2 RMSE was 34.36 kg DM ha^{-1} day^{-1}. The scatter plots for training (A and C) and testing data (B and D) are shown in Figure 5-16 with actual grass growth values from PBI on the x-axis and the predicted values from ANFIS on the Y-axis. R^2 values for training and testing data were 0.28 and 0.55 for model 1. Model 2, using only meteorological data, had a lower R^2 of 0.32 and 0.49 for training and testing stages than Model 1.

The RMSE for model 3 was 20.71 kg DM ha^{-1} day^{-1} and for model 4 RMSE was 19.44 kg DM ha^{-1} day^{-1}. The scatter plots for training (A and C) and testing data (B and D) are shown in Figure 5-17. R^2 values for training and testing data were 0.28 and 0.55 for model 1. Model 2, using only meteorological data, had a lower R^2 of 0.32 and 0.49 for training and testing stages than Model 1.

The RMSE for model 5 was 19.75 kg DM ha^{-1} day^{-1} and for model 6 RMSE was 17.76 kg DM ha^{-1} day^{-1}. The scatter plots for training (A and C) and testing data (B and D) are shown in Figure 5-18. R^2 values for training and testing data were 0.28 and 0.55 for model 1. Model 2, using only meteorological data, had a lower R^2 of 0.32 and 0.49 for training and testing stages than Model 1. The scatter plots for testing data for both years are similar in terms of the spread of the points along the 1:1 line.

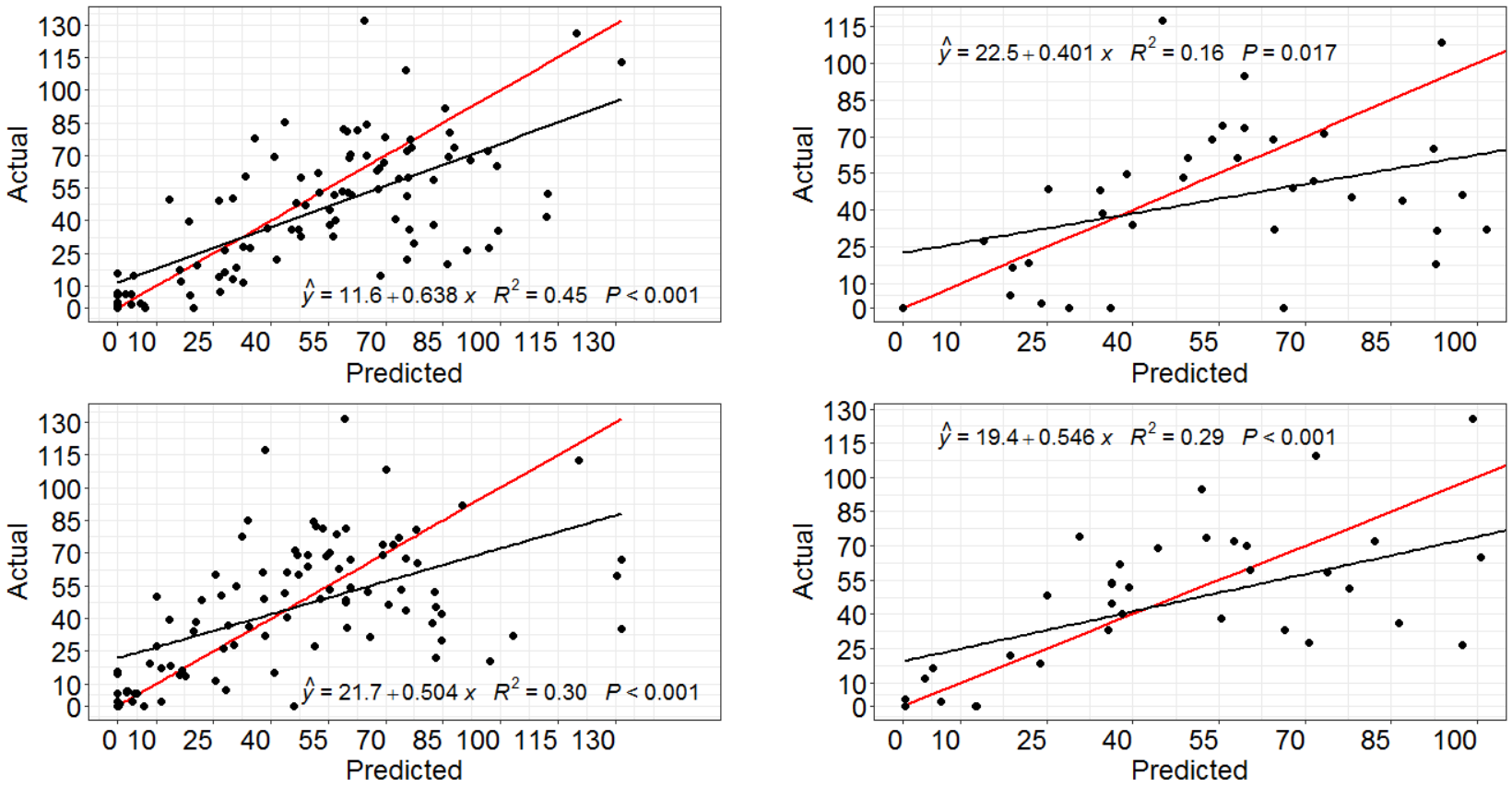


Figure 5-16 Actual vs. predicted grass growth for models 1-2.

Model 1: Training is Panel A. Testing is Panel B. Model 2: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$

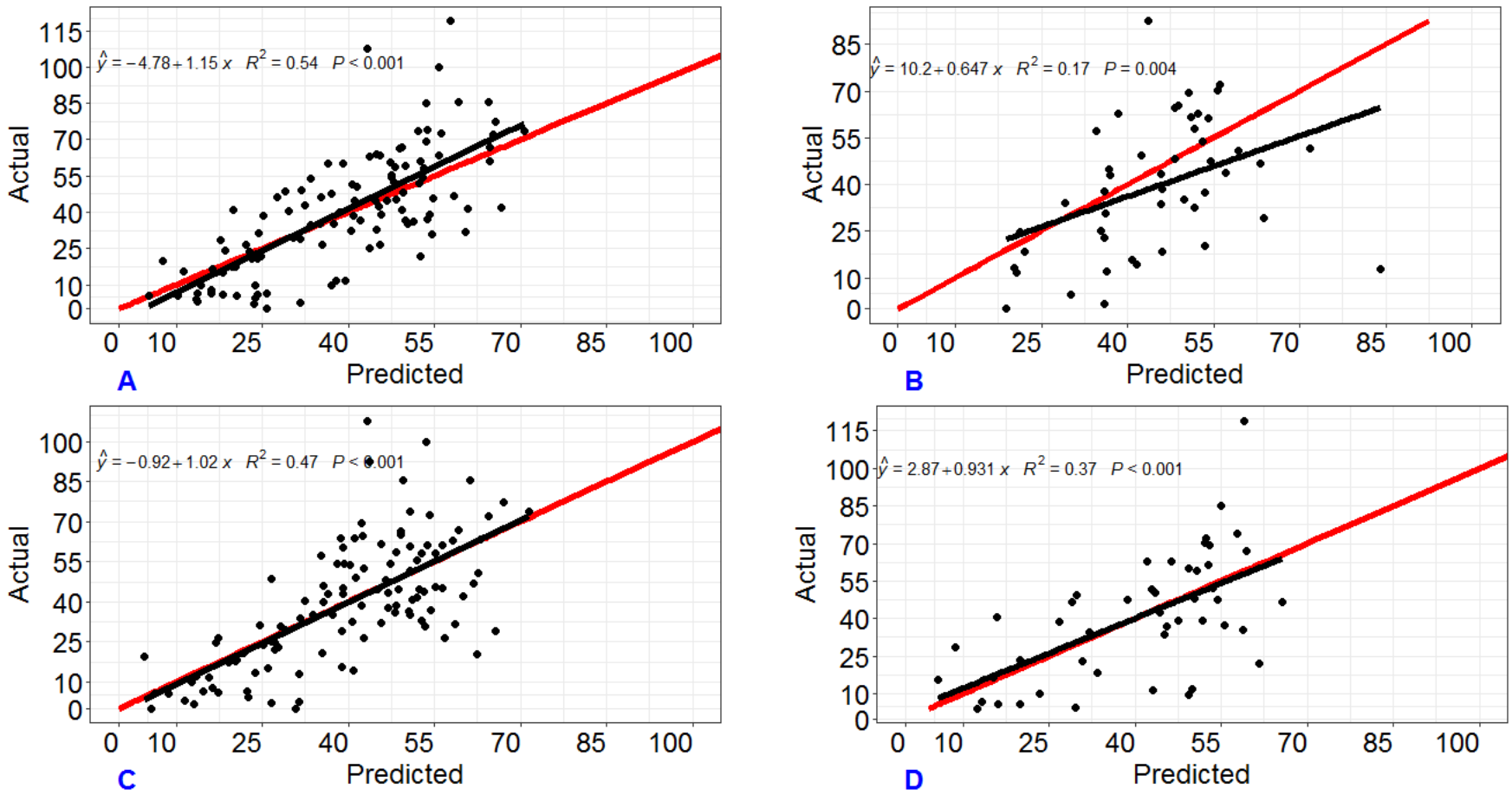


Figure 5-17 Actual vs. predicted grass growth for models 3-4.

Model 1: Training is Panel A. Testing is Panel B. Model 2: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$

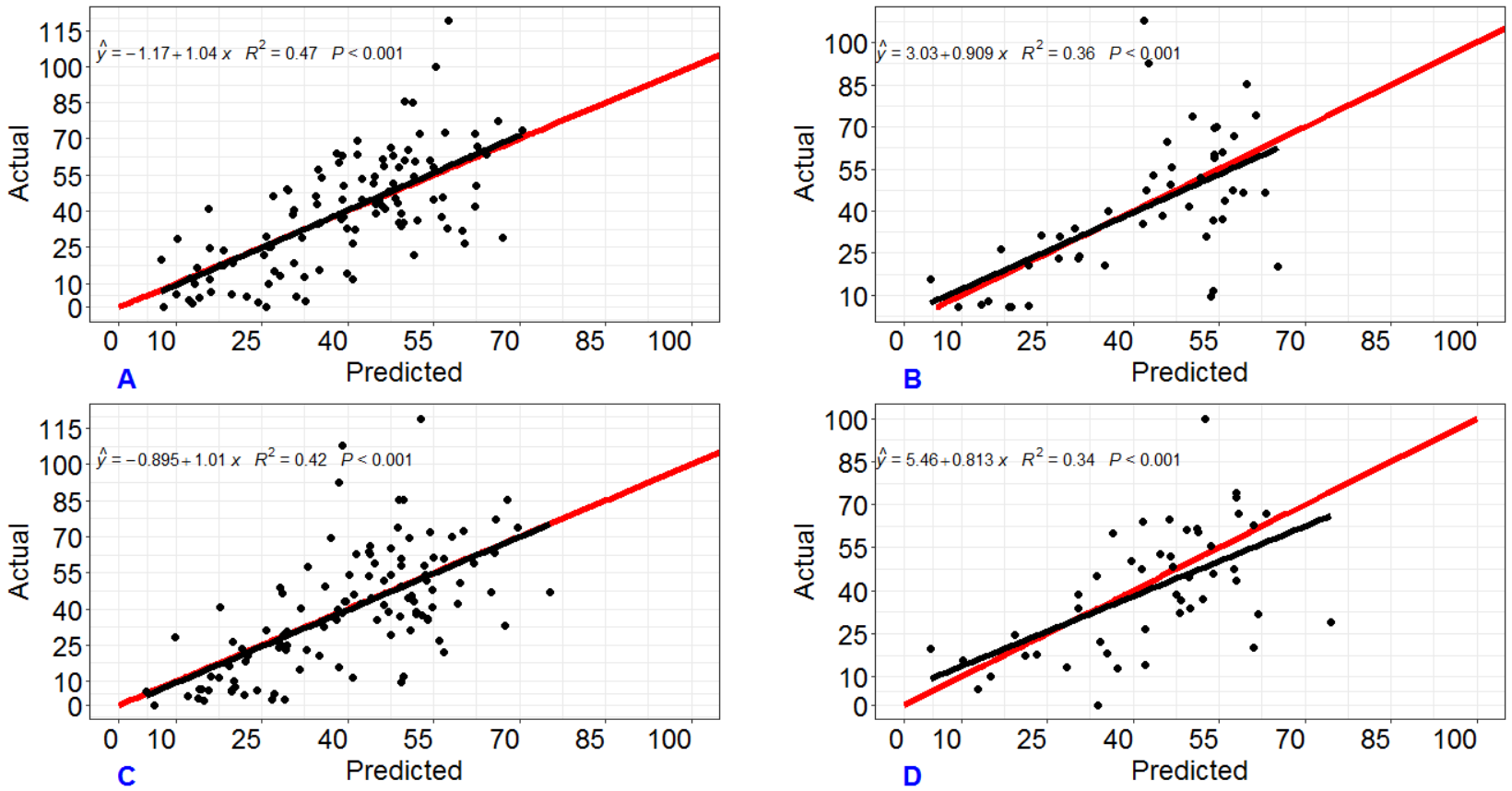


Figure 5-18 Actual vs. predicted grass growth for models 5-6.

Model 1: Training is Panel A. Testing is Panel B. Model 2: Training is Panel C. Testing is Panel D.

Grass growth in kg DM $ha^{-1}day^{-1}$

5.4.2 Model performance

The Brereton model was developed at the individual farm level (Moorepark, Johnstown Castle and Athenry), and its performance was compared for 2017 and 2018. The Brereton model worked better on Moorepark farm than on the other two for both the years. The R^2 was higher for Moorepark was 0.26 in 2017 and 0.33 in 2018 than the other two farms (0.03-0.17). During summer (June-August), the model under-predicted for 2017 and over-predicted, the grass growth rate for 2018 for all the farms (Table 5-8) as the model works by using the solar radiation values and converting it into dry matter. The difference between the actual and predicted values for Moorepark was the highest among all the farms with 11.43 kg DM $ha^{-1}day^{-1}$ of under-prediction in 2017 and 6.43 kg DM $ha^{-1}day^{-1}$ of over-prediction in 2018. Moorepark is a highly managed farm in which the grass growth can vary in paddocks but in this work, the average values of the whole farm were used to compare the predicted values.

Table 5-8 Difference in actual & predicted growth for three farms (2017-2018). Figures in kg DM $ha^{-1}day^{-1}$. In 2017 the growth was under-predicted, while in 2018, it was over-predicted (negative values).

Farm	2017	2018
Moorepark	11.43	6.43
Johnstown Castle	1.58	3.05
Athenry	3.13	3.37

The Brereton model is an empirical, and it tends to perform poorly when applied to new data. This would be a significant drawback in developing the Brereton model as a national model. The model also produces the same growth rate for farms sharing a meteorological station, which is also true for the ML models that exclude the EO data. The grass growth is variable depending on factors such as meteorological and management such as grazing and silage for winter. There were some concerning errors within the model, for example, for Moorepark it over-estimated the peak season growth rate by 18.24 kg DM $ha^{-1}day^{-1}$ for 2017 and by 45.15 kg DM $ha^{-1}day^{-1}$ in 2018 in May (DOY = 121 to 151). Hurtado-Uria et al. (2012) reported similar accuracy and found that the model over-predicted the summer grass growth values.

Barrett et al. (2004) used the Brereton model over two sites with contrasting grass growing conditions and observed that the model performed differently in both sites. It performed better at the site with drought conditions than without drought. However, in this work, the model performed slightly better for 2018 with drought conditions for Johnstown Castle (RMSE=50.79 for 2018 and 82.29 kg DM $ha^{-1}day^{-1}$ for 2017) and Athenry than in 2017 (RMSE=55.73 for 2018 and 60.36 kg DM $ha^{-1}day^{-1}$ for 2017). There were two machine-learning models (model 2 and 8) which used only meteorological data without EO data, and both of them performed better than the Brereton model. These models had higher R^2 (0.49 and 0.72) and lower RMSE (24.76 and 15.33 kg DM $ha^{-1}day^{-1}$) than the Brereton model (R^2 from 0.02-0.33 and RMSE from 41.68-82.29 kg DM $ha^{-1}day^{-1}$).

The issues with Brereton can be mitigated by introducing the EO imagery. The SMD is included in the Brereton model, but it is not a grass growth-limiting factor at more than 70 mm SMD. In contrast, the effect of extreme conditions like drought as experienced in 2018 was included in the ML models. The result was introduced in the growing degree-day variable with no growth for over 30 ºC maximum temperature and more than 70 mm of soil moisture deficit.

The ANFIS model performed poorly for all the farms. Model 2 (ANFIS), with all the variables except NDVI from Landsat 8, had the highest RMSE of 24.76 kg DM $ha^{-1}day^{-1}$. The lowest RMSE (15.04 kg DM $ha^{-1}day^{-1}$) was for model 6 (ANFIS) with all the variables from Sentinel 2 database except NDRE and rainfall. For the RF model, model 8 (with all the variables except Landsat 8 NDVI) had the highest RMSE of 15.33 kg DM $ha^{-1}day^{-1}$ and model 10 (all the variables from Sentinel 2 database except rainfall) had the lowest RMSE of 14.65 kg DM $ha^{-1}day^{-1}$.

For the machine-learning models using EO data, an error might have been reduced if consistent satellite imagery occurred between years. Ireland experiences extensive cloud cover all year, limiting the number of multispectral data available. The RMSE of all the models is plotted in Figure 5-16. A comparison of error metrics showed how RF outperformed ANFIS using both Landsat 8 and Sentinel 2 data for all the

farms. The RMSE for RF (14.65-15.33 kg DM $ha^{-1}day^{-1}$) was lower than ANFIS (15.04-24.76 kg DM $ha^{-1}day^{-1}$).

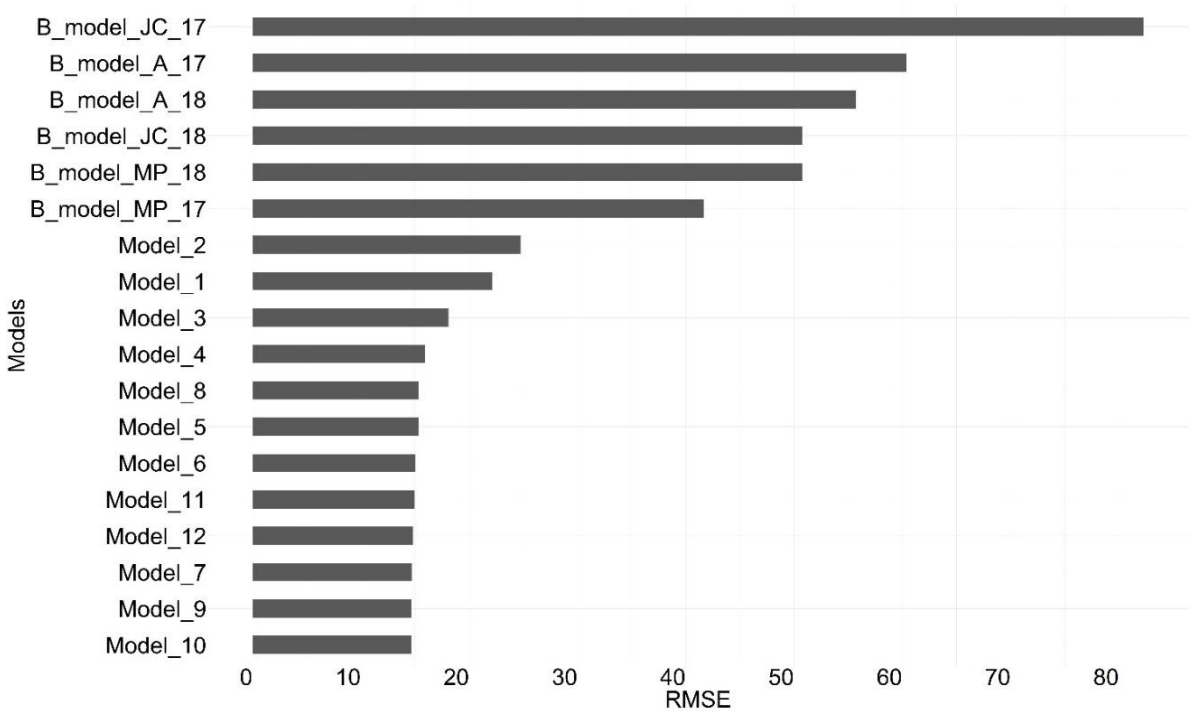


Figure 5-19 RMSE for testing data of all the Brereton, ANFIS and RF models. Model 1 to 12 are ML models, and B_model are Brereton models

The plot also shows that the Brereton model performed poorly with the highest RMSE values for all the farms. The experiments described in the previous section have shown how ML models provide greater accuracy in predicting grass growth rate than the Brereton Model. The ML models consistently had higher R^2 and lower RMSE (R^2=0.32-0.73 and RMSE=14.65-24.76 kg DM $ha^{-1}day^{-1}$) and better results than the Brereton model (R^2=0.03-0.33 and RMSE=41.68-82.29 kg DM $ha^{-1}day^{-1}$).

5.4.3 Variable importance

When using Landsat 8 database, NDVI was the rejected variable by both ANFIS and RF. The reason is the revisit time of Landsat 8, which is 16 days. If there is a cloudy day on the day of acquisition, there is a gap of over a month in the subsequent image acquisition. The data gaps due to the cloudy images can affect the model's ability to capture the management events. Therefore, the models with NDVI from Sentinel 2 performed better than the NDVI from Landsat 8. Although Landsat 8 NDVI was a rejected variable, the model performance using all the variables and the models with NDVI performed with similar accuracy.

For Sentinel 2 database, NDVI and NDRE were significantly important, and the rainfall was the least significant variable. The rainfall was a tentative variable, which means that the Boruta algorithm could not decide its importance. The rainfall decreases as the season changes from spring to summer. The variation in rainfall affects the grass growth rate. For example, low rainfall with high temperature and high solar radiations can lead to drought-like conditions, such as in 2018. The impact of rainfall is reduced when there are no extreme conditions such as high temperatures, radiation and evapotranspiration. The rainfall can affect a farmer's decision to turn out the animals for grazing (Green, 2019).

NDVI was more important in the variable importance than NDRE (a red edge vegetation index). However, the model performance with NDVI (RMSE 14.83 kg DM $ha^{-1}day^{-1}$) and NDRE (14.94 kg DM $ha^{-1}day^{-1}$) was similar. Sibanda et al. (2015) found that the vegetation indices using red-edge bands were more important than the raw bands of Sentinel 2 using variable importance scores in estimating grassland. In their study, there was a strong correlation between grass biomass and red edge bands from Sentinel 2, such as band 5 (0.705 µm), 6 (0.740 µm) and 7 (0.783 µm). We used 0.705 µm wavelength as a red-edge band for the calculation of NDRE. The red-edge spectrum's reflectance provides a better and higher value than the low reflectance values in the red region.

5.4.4 Sensor performance

Landsat 8 and Sentinel 2 satellites provide an opportunity for grass biomass modelling. Landsat 8 has been operational since 2013 with Operational Land Imager (OLI) sensor providing 30 m multispectral images and 16 days revisit time. The Sentinel 2A was launched in 2015 and Sentinel 2B in 2017, providing 10 m and 20 m multispectral images with 5 days revisit time. The different spatial resolution of both the satellites limits their combined use. Although they have similar spectral characteristics, the difference in radiometry can make it difficult to use them together. The variation in radiometry is due to the different acquisition time affecting the illumination conditions (Mandanici and Bitelli, 2016). In addition, the cloud cover and shadow effects can reduce the number of images acquired in a month because of the different revisit time of satellites.

Sentinel 2 has red-edge bands – band 5, 6, 7 and 8 A, which offers a great potential for grass biomass estimation, and is missing in Landsat 8. Red-edge bands are correlated to vegetation chlorophyll, which can be an indication of healthy vegetation. In a study by Ramoelo et al. (2015), the essential variables from the random forest were red-edge at 0.705 µm followed by short-wave infrared at 1.610 µm and 2.190 µm. In the vegetation spectral response, the red-edge band at 0.705 µm has high reflectivity.

5.4.5 PBI as a source of ground truth data

PBI is a web and mobile application based database providing farm data such as weekly grass growth rate, grass cover, dates for grazing and silage cutting at farm and paddock scale (Hanrahan et al., 2017). More details are given in Section 3.4.4. The farmers collect the data, which can lead to the operator error affecting the final measurements. In addition, visual assessment can have a bias, which can skew the final grass growth entered into PBI. Another method to estimate grass growth rate is RPM, which measures the grass height and can be entered into the PBI database. PBI predicts grass growth rate from height measurements using a linear relationship between biomass and height. Since the grass growth is estimated indirectly, a model prediction error is related to RPM apart from operator error.

In a study by Hanrahan et al. (2017), the visually estimated data from PBI for Moorepark farm was evaluated against the data from the cut and dry method in 2014. The RMSE for experiments with silage cuts between PBI and ground data was 2270 kg DM/ha and R^2 of 0.58, whereas the RMSE was 2266 kg DM/ha and R^2 was 0.88 for the data without silage. Murphy et al. (2021) compared linear regression and machine-learning (random forest) to estimate grass growth using height from RPM. There were 17 inputs to the models, such as management data- grass height, fertilizer application, grazing rotation and meteorological data. The RF performed the best with and without meteorological variables with an RMSE of 262 and 243 kg DM/ha. The addition of meteorological data to both the models reduced error by 0.9% for linear regression and 1.5% for RF.

5.4.6 Grass growth curve

An example of the grass growth curve from Model 10 (one of the best models) using RF with all the inputs except rainfall is shown in Figure 5-17. The farm shown is Moorpark farm. The actual grass growth from PBI is shown using black points, and RF output is shown in red points. As it is clear from Figure 5-17 that there are many missing data points. For May, at the peak season, the RF model over-predicted the values. There were not enough satellite images during the winter. The start and end of grass growth measurements vary for the same farm every year. There is a variability of grass growth rate between 2017 and 2018, which can be due to the meteorological conditions, management and soil type. For 2018, there were no satellite data available, due to which the peak values are missing from the curve. The model had data points mostly in autumn. The predictions are affected by the lack of satellite data. The rest of the farm's grass growth curves are shown in Appendix 5.7.

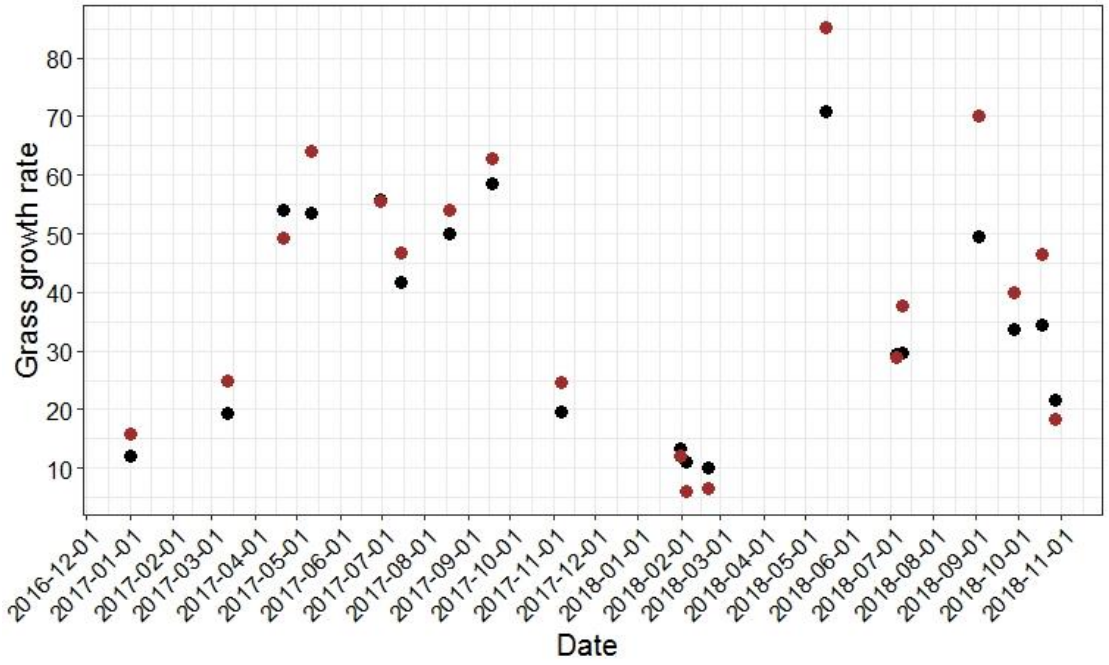


Figure 5-20 Grass growth curve for Moorepark (2017-2018) using RF Model 10. The actual grass growth from PBI is shown using black points, and RF output is shown in red points

Another example of a grass growth curve from Model 10 for Johnstown Castle farm is shown in Figure 5-18. The actual grass growth from PBI is shown using black points, and RF output is shown in red points. There are two curves one for 2017 and the other for 2018. For 2017, in May, the first primary peak is predicted accurately by the model. However, from September until December, the grass growth values were under-estimated. In 2018, the model over-estimated the values in May and autumn; the values were under-estimated. RF's performance increases with the increasing sample size. Therefore, the model with Johnstown Castle farm data performed better than the model with Moorpark farm data, as more data points were available for Johnstown Castle.

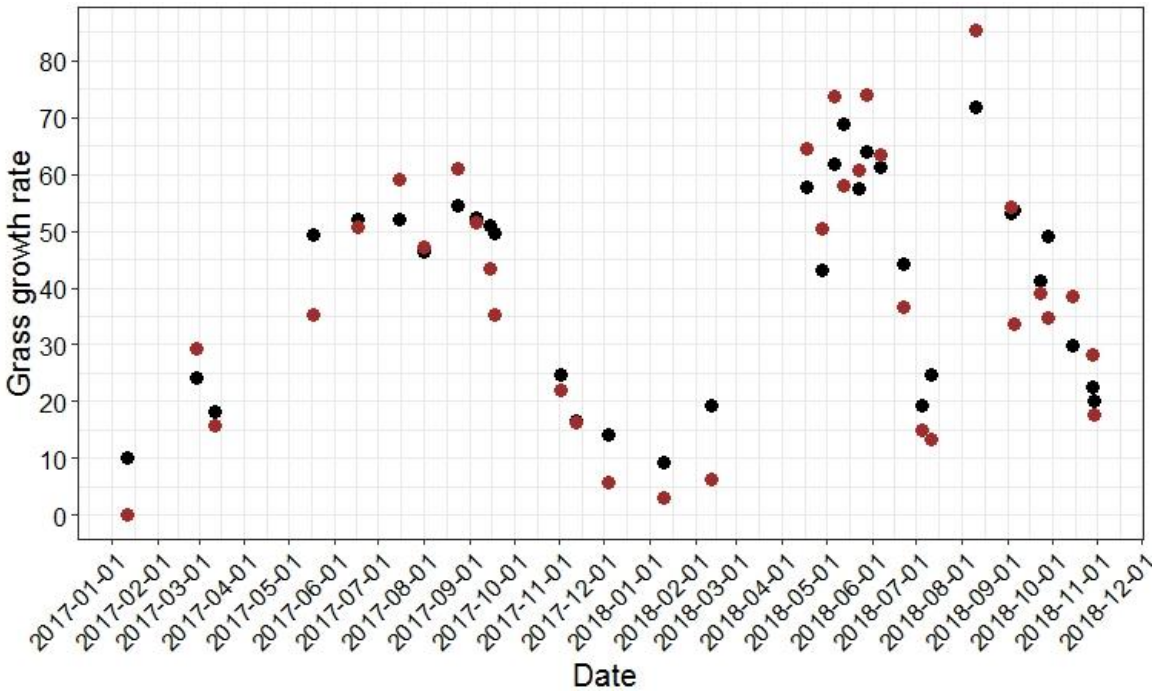


Figure 5-21 Grass growth curve for Johnstown Castle (2017-2018) using RF Model 10. The actual grass growth from PBI is shown using black points, and RF output is shown in red points

The grass growth curve from Model 10 for Athenry farm is shown in Figure 5-19. The actual grass growth from PBI is shown using black points, and RF output is shown in red points. There are two curves, one for 2017 and the other for 2018. There were no satellite images before May 2017 and before April 2018. The model over-predicted for May-July, whereas it under-predicted from August-October in 2017. For 2018, for April and May, the values were under-predicted, whereas the values were over-predicted for the rest of the year.

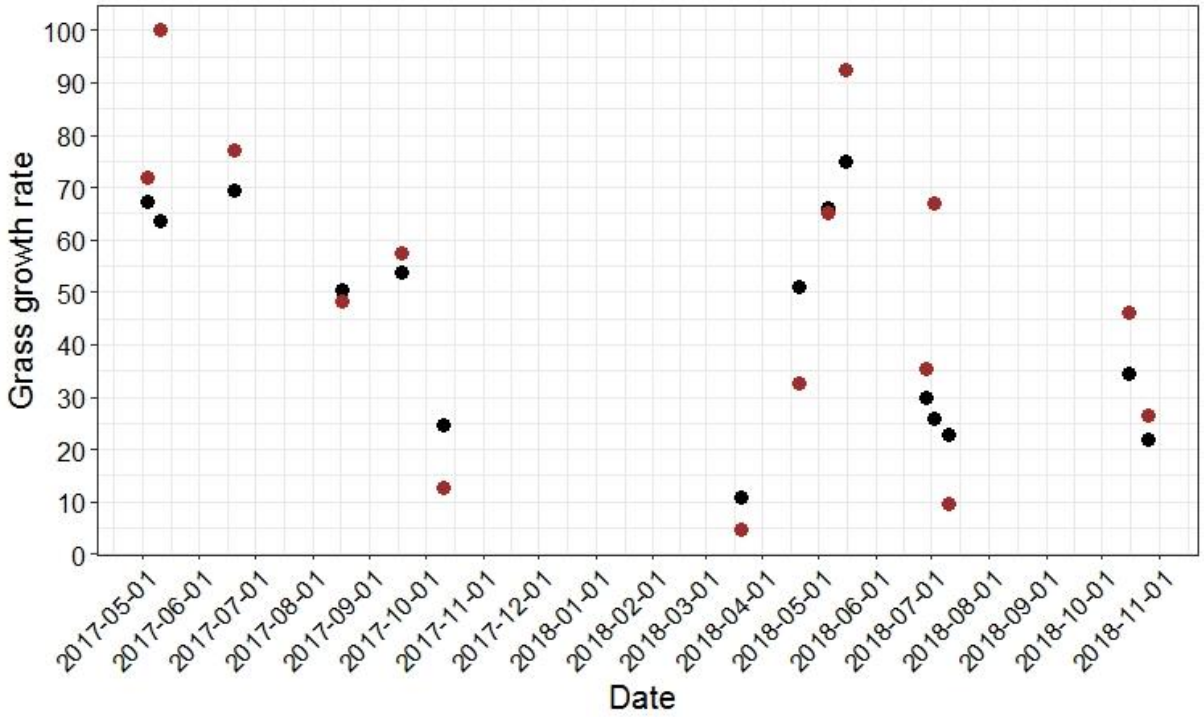


Figure 5-22 Grass growth curve for Athenry farm (2017-2018) using RF Model 10. The actual grass growth from PBI is shown using black points, and RF output is shown in red points

5.4.7 Uncertainties & sources of error

Several error sources and uncertainties in the model can arise because of locations and number of farms, satellite, meteorological, and field data. For optical satellite data, clouds and shadows are a major constraint limiting the number of data points in a year. There is missing data during the primary grass-growing season, and the satellite data capture dates are uncoordinated with the field data collection. The farmers collect the field data stored in the PBI database, which is based on citizen science. The ground data can have error while collecting and entering PBI data. The data collection instrument also plays an important role, as some farmers opt for a visual estimation, which is not as accurate as RPM measurements. There is a systematic error related to RPM and the cut and dry method. All these errors are inevitable. In this study, Sentinel 2 was used, 10 m, to match the farm scale as possible. The temporal matching of satellite data and ground data is performed to overcome the mismatch, as the grass growth does not change within 7 days. The advantages and disadvantages of using different modelling strategies are not fully demonstrated because the sample size is small. ANFIS model requires repeated

training to obtain an optimal neural network, which requires more computer time. The RF model can be applied to large dataset because of its ability to resist overfitting and deal with high-dimensional data.

5.4.8 Applications & future developments

The grass-growth prediction model has many applications, such as grazing management and avoids fodder shortage due to extreme climate events such as floods and droughts (discussed in Section 1.4.1). Accurate and timely measurement of grass biomass has a potentially significant role in helping farmers achieve effective grazing management practice. One of the different issues with various grass growth models is their actual application, hindering their real-time farm usage. The models need to be simple and easy to use. For DSS, the fundamental requirement is moderate accuracy for efficient farm management.

Currently, there are eight farms in the model. This work will be expanded to include more farms from PBI, which needs further work. The uncertainty in the model can be reduced by including other satellite images to fill the gap due to cloud cover, such as microsatellites. The microsatellite can provide on-demand data, but it is not free. The Vegetation and Environment monitoring on a New Micro-Satellite (VENμS) is a microsatellite launched in 2017 with a 2-day revisit time and 5m spatial resolution. Liao et al. (2019) used the VENμS data to fill the gaps in Sentinel 2 data to estimate corn biomass. With the launch of Landsat 9, the volume of data used for training the model will increase, which will improve prediction accuracy, as ML models are data-driven, which means that the more the training data, the more will be the model's predictive power. The UAV data provides high spatial and temporal resolution, which can be used for biomass modelling at paddock and farm scale. UAV can help to understand the variability of grass growth between paddocks. De Rosa et al. (2021) used NDVI from UAV with random forest to estimate grass biomass with R^2 of 0.68 and RMSE of 17.40%. However, the processing of UAV data is complex, which might need high computing power.

5.5 Conclusion

Accurate measurement of grass growth rate is an integral part of optimising the efficiency of a farm. This chapter compared an empirical biophysical simulation model, the Brereton model, and two ML algorithms, ANFIS and RF, for predicting grass growth on eight Irish farms. This analysis demonstrated that ML models with RMSE in the range 14.65-24.76 kg DM $ha^{-1}day^{-1}$ outperformed the biophysical simulation model with RMSE in the range 41.68-82.29 kg MD/ha/day for predicting grass growth on these selected farms. The ML models were developed for three different regions- West, South and Southeast of Ireland, indicating that the model can be applied across different farm locations in different geographical locations. The Brereton model is site-specific and needs to be developed for each site with their respective input parameters.

The RF model yielded the best results with low RMSE (around 14.65 kg DM $ha^{-1}day^{-1}$) and a variance of 0.55% than ANFIS with RMSE of 15.04 kg DM $ha^{-1}day^{-1}$, explaining 52% of the variance in grass growth rate. Vegetation indices played an essential role in estimating the grass growth rate at the farm scale. Random Forest was the most accurate machine-learning technique in grass growth rate estimation, with R^2 values reaching 0.55 for test data. The model establishes a good foundation for developing this method to predict grass yield across a more extensive area. In the next chapter, a methodology to expand this work to a national level will be discussed.

6

Chapter 6 A National Model for Biomass Estimation

6.1 Introduction

In previous chapters, the importance of grass as a fodder source has been outlined. While grass growth is abundant in Ireland, utilisation of grown grass by farmers is low. Routine grass measurements allow farmers to make informed and timely decisions about their grass resource to match their feed available and their livestock demand, for example, whether it should be grazed or cut for silage, whether fertiliser is required or whether the grazing interval could be increased/decreased (see Section 1.3.2) (French et al., 2006, O'Brien et al., 2015). Grass measurement as a grassland management tool can help farmers improve their grass production and utilisation and build some resilience into their farming system to mitigate the impact of extreme or unseasonal weather such as flooding or drought, which can lower the grass production and lead to lower utilisation rates.

In 2018, a summer heatwave caused a prolonged drought that resulted in reduced grass yields across Ireland (Falzoi et al., 2019). As a direct result of the drought, there was a 52% increase in expenditure on fodder and supplemental feeds, which increased the price of milk by 11% per litre (Dillon et al., 2018b). Similarly, in 2012, below-average temperatures and above-average rainfall in summer, followed by extended winter conditions in 2013, resulted in more significant expenditure on fodder (Green, 2019). With expected climate change scenarios, mean annual temperatures are expected to increase by 1.3-1.6 ºC and lower rainfall (0-11%) for summer months, grass and silage production will be affected, requiring farmers and stakeholders to plan accordingly (Nolan and Flanagan, 2020).

Grass growth is highly variable, spatially and temporally, both within and between farms but Ireland currently does not have a system to estimate pasture growth using satellite images on a national scale. Tools such as PBI help farmers record the weekly grass growth and feed in an online database system, but for a selection of farms (albeit a large and increasing selection).

Sentinel 2A and Sentinel 2B were launched in 2015 and 2017, respectively, each having 10-20 m spatial resolution with a five-day repeat period. The improved spatial resolution compared to the MODIS data means Irish fields can be mapped in greater detail. One 10 m pixel of Sentinel 2 covers an area of 0.01 ha, and 625 (10m) pixels of Sentinel 2 can fit into one 250 m pixel of MODIS. Chen et al. (2021) used Sentinel 2 imagery and climate variables in a multilayer perceptron (MLP) neural network to estimate grass biomass for five dairy farms. The study used eight bands in the visible and near-infrared region, two short-wave infrared (SWIR) wavelengths, and the red-edge NDVI. The satellite data and ground data measurements were paired if they were less than 3-days apart. The meteorological data used were precipitation, maximum and minimum temperature, solar radiation and vapour pressure deficit. Two models were developed using satellite data and satellite data with climate data. When the climatic variables were added to the model, the RMSE decreased from 406 to 356 kg DM ha^{-1}, and R^2 increased from 0.51 to 0.62. Some factors were not included in these models, such as management on farm (grazing, tilling) and field-based error, which is a limitation of their work.

The objective of the present study was to develop and evaluate the accuracy of national farm-scale grass growth predictions using EO imagery and weather variables, with a focus on identifying the strengths and weaknesses of Sentinel 2 imagery for modelling grass growth under Irish conditions. A random forest model was used with satellite data from Sentinel 2, meteorological data from Met Éireann, and grass growth data as a reference from PastureBase Ireland (PBI).

6.2 Materials & study area

6.2.1 Study area

Before developing a national model, several models were tested for a group of 10 commercial farms in the County Limerick, Tipperary and Cork, as discussed in Section 3.3.3. The dataset will hereafter be referred to as the 'Commercial dataset', and the model will be called the 'Commercial model'.

For the national model, calibration data were collected for farms distributed across four Irish counties, Co. Cork, Co. Donegal, Co. Galway and Co. Sligo (see Figure 3-7 in Section 3.3.4). The farms were extracted from PBI, and their location anonymized by providing each farm with a random number without spatial location. The exact locations and boundaries of these farms are not known because of the EU's GDPR law of data privacy.

6.2.2 Data acquisition and pre-processing

6.2.2.1 Sentinel 2

Two vegetation indices were used for the model, a normalised difference vegetation index (NDVI) and a normalised difference red-edge index (NDRE). These indices were derived from 10 m spatial resolution, atmospherically corrected Level-2A Bottom of Atmosphere (BOA) reflectance Sentinel 2 images (as discussed in Section 3.4.2). Each farm consisted of several paddocks, and NDVI and NDRE values for each paddock were provided as a vector with randomly assigned farm numbers. Due to the nature of grass production, as outlined in Section 1.3.2, some paddocks performed better than the others depending on the grazing pattern, fertilizer application, drainage and soil type. A hypothesis was developed in which the vegetation indices could be divided into three parts based on quartile values to reflect growth patterns. The low performing paddocks were those in the lower quartile (25th percentile), medium-performing paddocks fell in the second and third quartiles (between the 25th and 75th percentiles), with high performing paddocks in the fourth quartile (75th percentile). According to this hypothesis, lower vegetation index values would relate to lower grass growth rate, with higher vegetation index values

corresponding to higher grass growth rates. This hypothesis was tested for the 10 commercial farms and then applied to the 179 farms of the national dataset.

Hereafter the lowest-performing paddocks are referred to in the text as NDVI1 and NDRE1 depending on the vegetation index. Medium performing paddocks are referred to as NDVI2 and NDRE2, with high performing paddocks referred to as NDVI3 and NDRE3. The mean and standard deviation of all categories of paddocks were calculated. Mean, standard deviation, median, minimum and maximum values of NDVI and NDRE and Sentinel 2 bands 2, 3, 4, 5, 6, 7, 8, 8A, 11, 12 were calculated and are listed in Table 6-1.

Table 6-1 Input variables from satellite data for the national data. NDVI 1/NDRE 1, NDVI 2/NDRE 2 and NDVI 3/NDRE 3 are values for low, medium and high values paddocks, respectively.

Variable
NDVI 1: mean, std. dev.
NDVI 2: mean, std. dev.
NDVI 3: mean, std. dev.
NDRE 1: mean, std. dev.
NDRE 2: mean, std. dev.
NDRE 3: mean, std. dev.
NDVI: Mean, median, std. dev, min, max
NDRE: Mean, median, std. dev, min, max
Band 2, 3, 4, 5, 6, 7, 8, 8A, 11, 12: Mean, median, std. dev, min, max.

6.2.2.2 Meteorological data

Rainfall data (in mm) were obtained at daily time-steps using the nearest synoptic station to each farm. The rainfall values were amalgamated into three variables according to the number of days of rain accumulation; Rain 3, Rain 5 and Rain 7 refer to rain accumulation over three, five and seven days, respectively.

6.2.2.3 PastureBase data

179 farms in PBI with four full years of data from 2017 to 2020 were chosen for this work and cleaned into an analysis-ready format:

1. Missing values were removed.
2. Records, where cover and grass growth rate equalled zero, were removed.
3. Records less than or equal to 100 kg DM $ha^{-1}day^{-1}$ were selected because some records had unrealistically high grass growth values due to data entry error.
4. Records with all the years 2017-2020 were extracted.

After cleaning, the data were stored in a CSV file used in the random forest regression model in R statistical software.

6.3 Random forest model development

The methodology for the proposed model is illustrated in Figure 6-1. First, three CSV files of the satellite, rainfall and PBI data were created- one containing all the data for the national model called Agmodel, the second CSV file had the data from January to June for Agmodel 1, and the third file had the data from July until December for Agmodel 2. As explained in Section 5.2.2.2, a technique was developed to select the nearest PBI data within seven days of the date of the satellite data. The difference of days between the PBI and satellite data acquisitions was also calculated.

The input dataset was scaled to bring all the values between 0 and 1. The data from 2017 to 2019 were split (70:30) into training data and testing data using 5-fold cross-

validation (CV). CV shuffles the data randomly and the training data set are split into 5 folds with k-1 (4) folds used for training and the fifth subset for testing. In the first iteration, the first subset is retained for testing, and the rest of the folds are used for training the model. The splitting continues until all the folds are used as training and testing data. The 5-fold CV was performed to ensure that data are split evenly with no spatial or temporal bias. The data from 2020 were used as a validation of the trained model. The accuracy of the model was described using the coefficient of determination R^2, MAE, MSE, and RMSE as described in Section 5.2.4.4.

There are two peaks in grass growth in Ireland. In the primary peak in May, growth can reach ~100 kg DM $ha^{-1}day^{-1}$. In the second peak in August, maximum grass growth rates values can be ~60 kg DM $ha^{-1}day^{-1}$. There is a gradual decrease in growth thereafter and over winter. An example grass growth curve from Moorepark is illustrated in Figure 6-2 for 2017. In building a national model, the data were split into two halves- January to June and July to December to exploit these two peaks.

A model was developed for each half of the year. Agmodel 1 was used from January until June, and Agmodel 2 was used from July until December. For the first half of the year, the grass growth rate values were high, and as the weather is better than the second half, there were many cloud-free satellite images. During the second half of the year, the grass growth values started decreasing, and the number of cloud-free satellite images were typically lower.

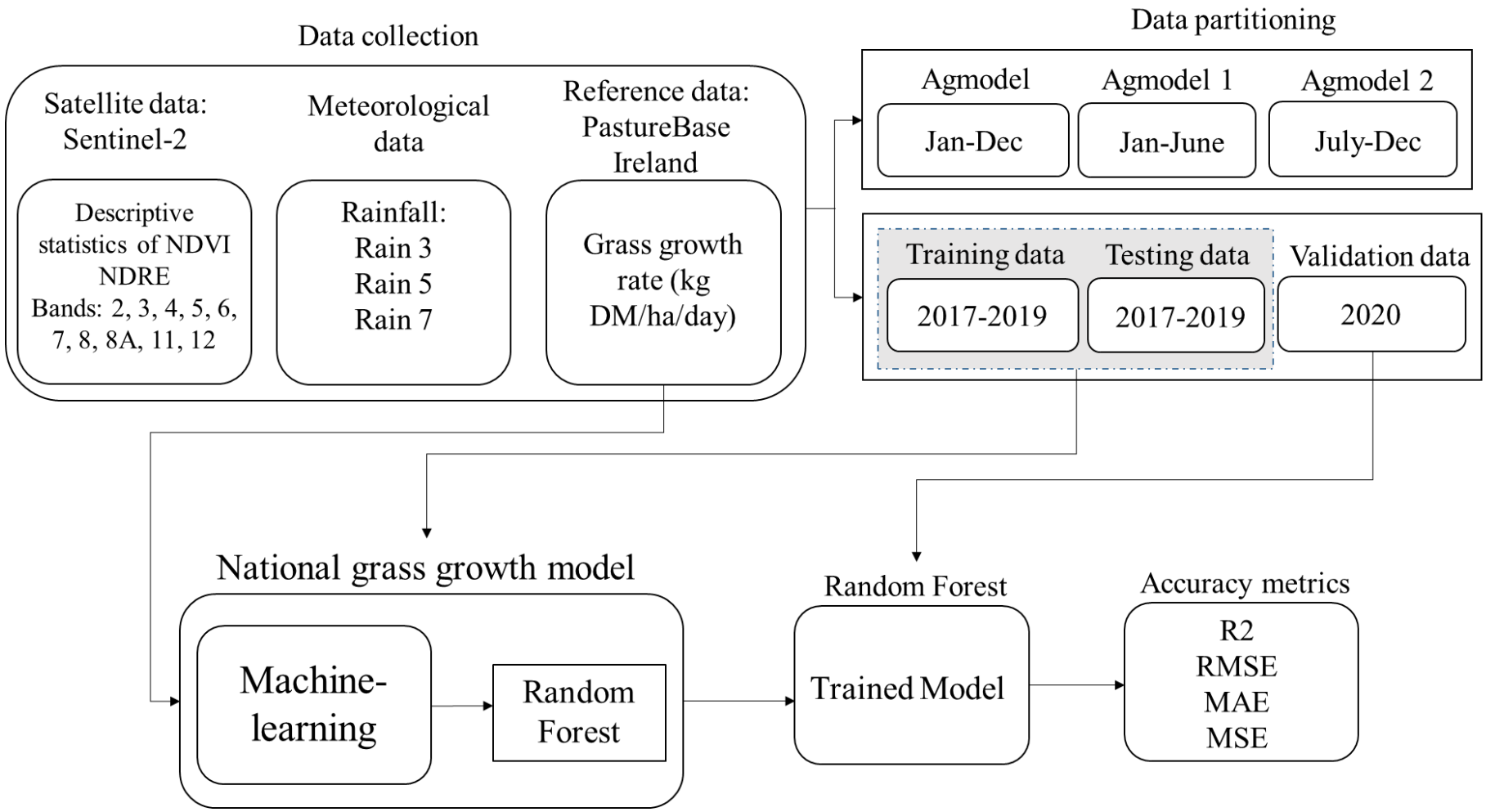


Figure 6-1 Methodology flowchart

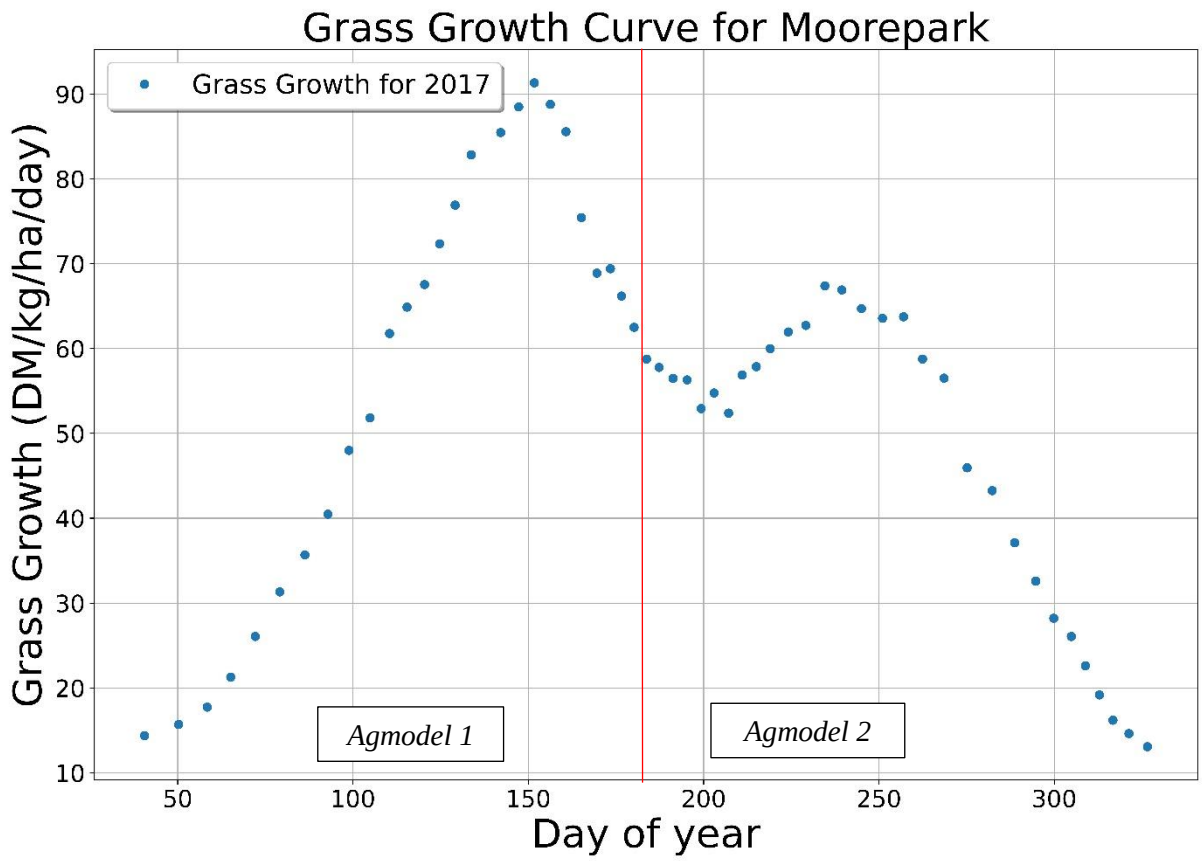


Figure 6-2 Grass growth curve from Moorepark, Co. Cork (2017).
The solid red vertical line is a demarcation between the two models.
Agmodel 1: DOY 1 until 182 (i.e. January to June).
Agmodel 2: DOY 183 to 365 (i.e. July to December).

6.3.1 Feature selection

As discussed in Section 5.2.4.3, feature selection is an essential pre-processing step to identify the features that contribute the most to the predictions from the model. The package used to get the importance values for each variable in the national model is 'ranger', a fast implementation of the random forest using the 'variable.importance' function in R, which also calculates the accuracy metrics for each feature. The 77 satellite and rainfall variables were input into the random forest, and the most important for each of the three models identified.

6.3.2 Hyper-parameter tuning & data splitting

The hyper-parameters required for training the random forest, ‘ntree’ and ‘mtry’ described in Section 5.2.4.2, were selected using 5-fold cross-validation. At the end of the five iterations, the results were averaged to select the best parameters for the model, which were then used with the complete training data.

6.3.3 Model evaluation

To test the power of the model to predict the grass growth for years outside the calibration period, the model was developed using data from 2017-2019 and then used to predict the 2020 growth. The model was evaluated using the metrics previously discussed in Section 5.2.4.4.

6.4 Results

The random forest model developed for the10 Commercial is described first, followed by the national model.

6.4.1 Commercial dataset

The Commercial farms had 19 variables, as shown in Table 6-1. The feature importance of all the variables is shown in Figure 6-3. DOY had the highest importance. The red-edge vegetation index (referred to as NDRE_re in the plot) had a higher importance value than NDVI. The lowest importance variables were the daily rainfall values.

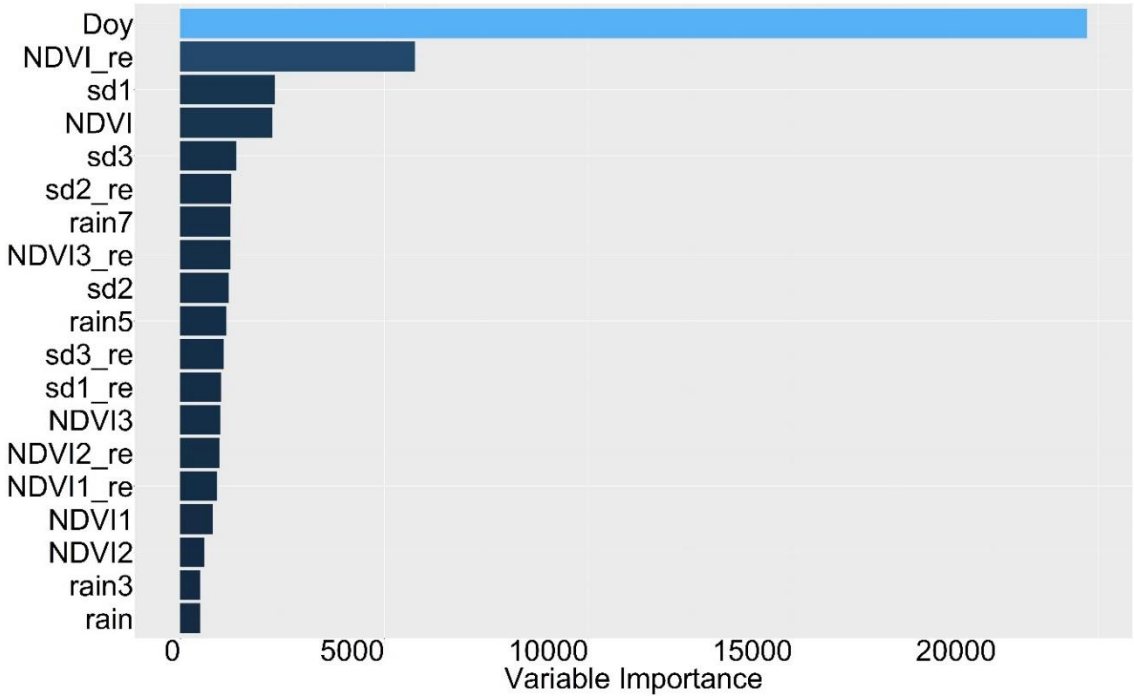


Figure 6-3 Feature importance for Commercial dataset

1/2/3 refers to low/medium/high performing paddocks

The scatter plot for predicted grass growth rate from training data is shown in Figure 6-4 A and testing data in Figure 6-4 B. The R^2 for training data was 0.94 and for testing data it was 0.61 with $p<0.01$.

Values of less than 25 kg DM $ha^{-1}day^{-1}$ formed one cluster, with values greater than 40 kg DM $ha^{-1}day^{-1}$ formed another cluster for test data. The values with more than 40 kg DM $ha^{-1}day^{-1}$ had more spread and variation than the values less than 25 kg DM $ha^{-1}day^{-1}$. The residuals were calculated between the actual and predicted grass growth rate for the Commercial farms are shown in Appendix 6-1. The MAE for testing data was 13.40 kg DM $ha^{-1}day^{-1}$, MSE was 281.43 kg DM $ha^{-1}day^{-1}$, and the RMSE was 16.77 kg DM $ha^{-1}day^{-1}$.

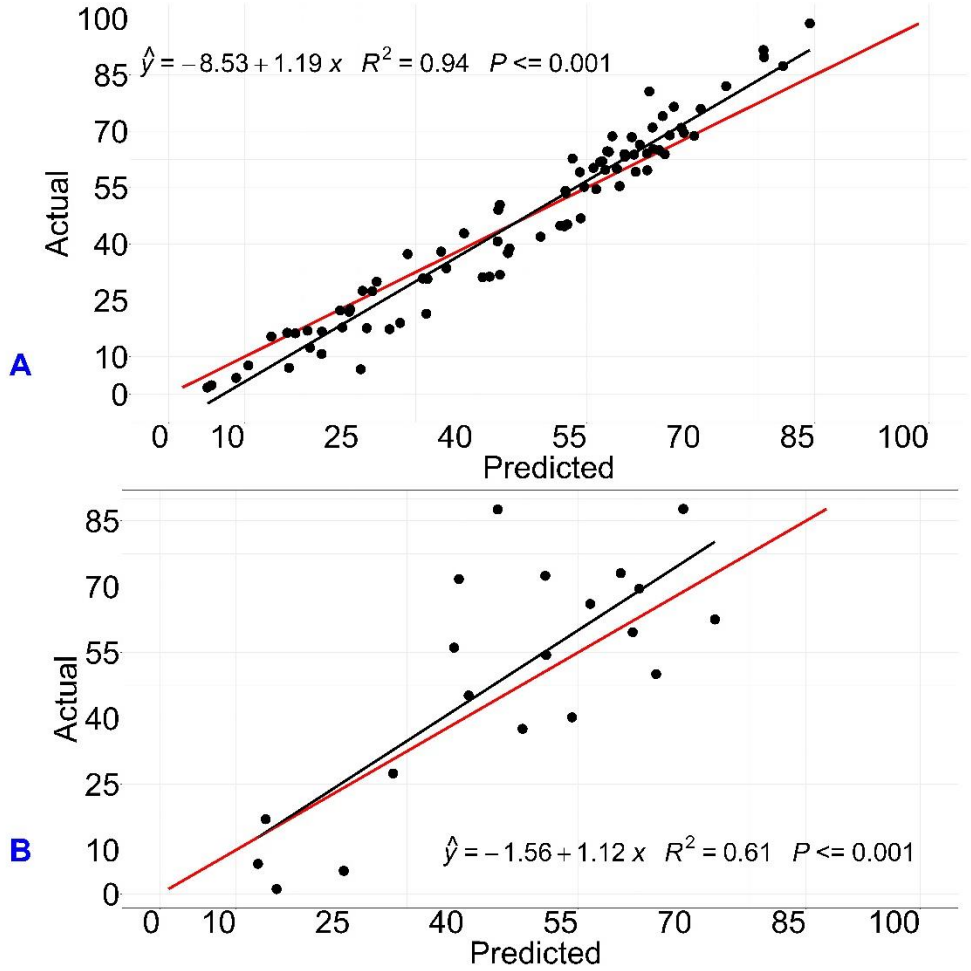


Figure 6-4 Actual vs. predicted growth on commercial farms. RF with all variables. Graph shows training (A) and testing datasets (B). The red line is a 1:1 line.

6.4.2 Agmodel- National dataset

Variable importance for the 77 variables in the national dataset was calculated. The DOY, which represents the phenology of the grass through the year, had the highest importance, therefore, the feature importance shown in Figure 6-5 is calculated after excluding DOY. The importance is shown in decreasing order, with the most important variables at the top being Mean_08, which is the mean value of NIR band 8, and Median_11, the median value of SWIR band 11.

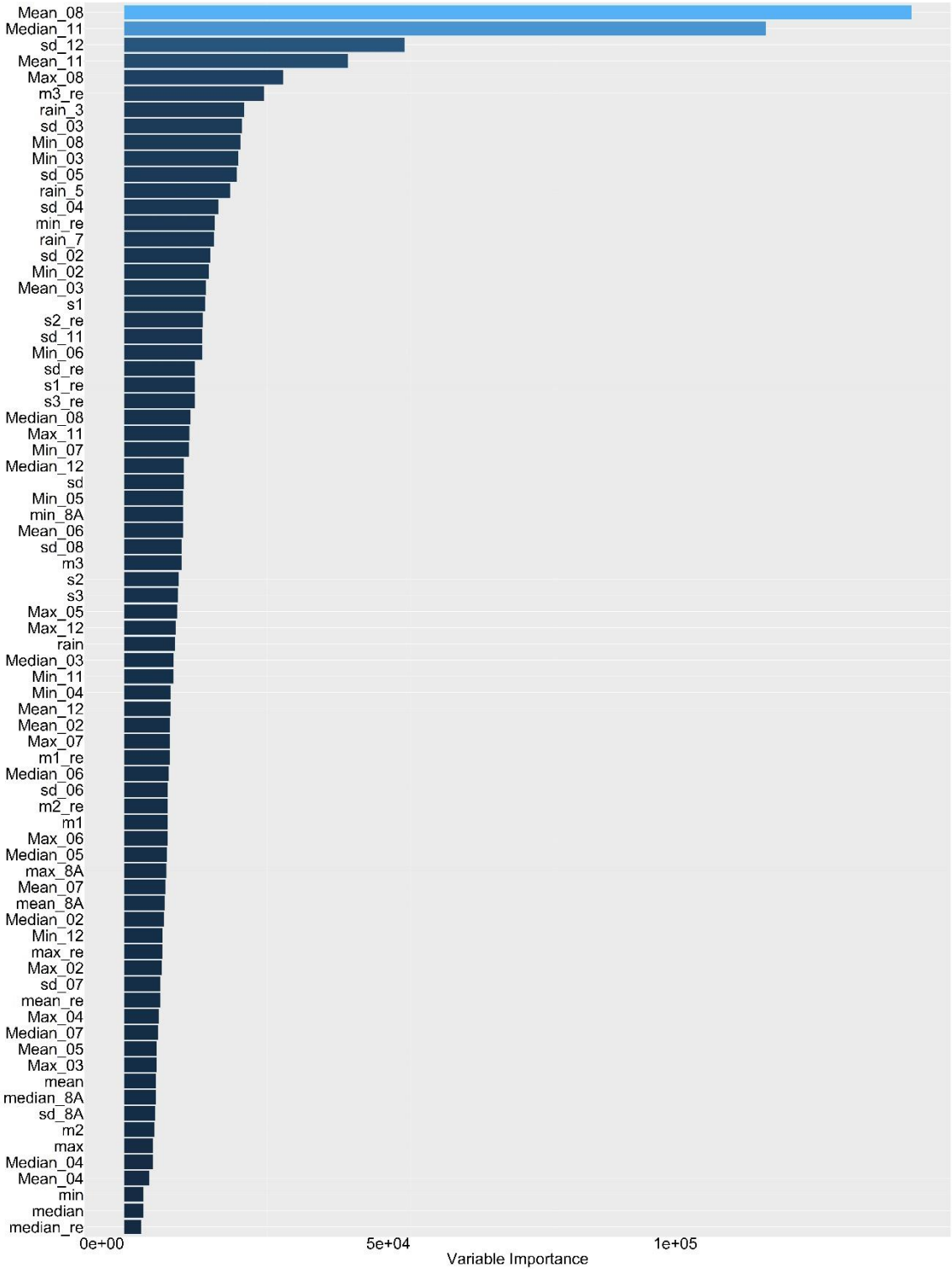


Figure 6-5 Variable importance in national model.

Feature importance using the 'ranger' library

Random forest models were developed for the top two to twenty most important variables, including DOY, with accuracy metrics calculated at each iteration (Table 6-2 with the lowest RMSE and highest R^2 in bold). The model with the lowest RMSE was chosen as the final model. The model with twenty variables had the lowest RMSE of 15.005 kg DM $ha^{-1}day^{-1}$ and MSE of 225.16 kg DM $ha^{-1}day^{-1}$, and only a slightly higher MAE compared to the lowest value, of 11.34 kg DM $ha^{-1}day^{-1}$. The overall RMSE value of all the variables was similar.

Table 6-2 Accuracy metrics for the national test dataset

R^2, MAE, MSE, RMSE. The metrics are shown from 20 to 2 significant variables with lowest RMSE and highest R^2 in bold.

No. variables	R^2	MAE	MSE	RMSE
20	**0.63**	**11.34**	**225.16**	**15.00**
19	0.63	11.35	225.52	15.017
18	0.63	11.32	225.51	15.01
17	0.62	11.43	228.67	15.12
16	0.63	11.34	226.49	15.04
15	0.62	11.46	228.87	15.12
14	0.62	11.51	231.60	15.21
13	0.62	11.43	228.85	15.12
12	0.62	11.50	229.87	15.16
11	0.61	11.61	235.02	15.33
10	0.62	11.50	231.83	15.22
9	0.61	11.65	236.57	15.38
8	0.61	11.78	237.52	15.41
7	0.60	11.96	241.52	15.54
6	0.60	11.82	240.75	15.51
5	0.60	11.82	242.65	15.57
4	0.60	11.85	241.49	15.54
3	0.59	12.17	245.73	15.67
2	0.56	12.30	268.23	16.37

The scatterplot for the training data with top 20 variables is shown in Figure 6-6 A and for testing data in B. The x-axis represents the predicted grass growth rate from Agmodel using the national data, and the actual grass growth values from PBI are shown on the y-axis. The 1:1 is shown as a red line, and the black line is the regression line. The training data R^2 was 0.96 with low variability around the 1:1 line. The points are coloured according to the year 2017 (orange), 2018 (green) and 2019 (blue). The values for all the years are distributed equally throughout.

The R^2 for test data was 0.63. The model could predict the lower range grass growth values with higher accuracy than the values of more than 50 kg DM $ha^{-1}day^{-1}$ for which the variability in the prediction increases. The 2017 values were mostly under 25 kg DM $ha^{-1}day^{-1}$ and more than 55 kg DM $ha^{-1}day^{-1}$. The residual plots are provided in Appendix 6-1.

To compare with Agmodel 1 and Agmodel 2, Agmodel results were divided into two parts- data from January until June and from July until December. The R^2 for Agmodel from January until June was 0.91 and RMSE was 8.36 kg DM ha^{-1} day^{-1}. For second half of Agmodel 2, the R^2 was 0.86 and RMSE was 8.22 kg DM $ha^{-1}day^{-1}$.

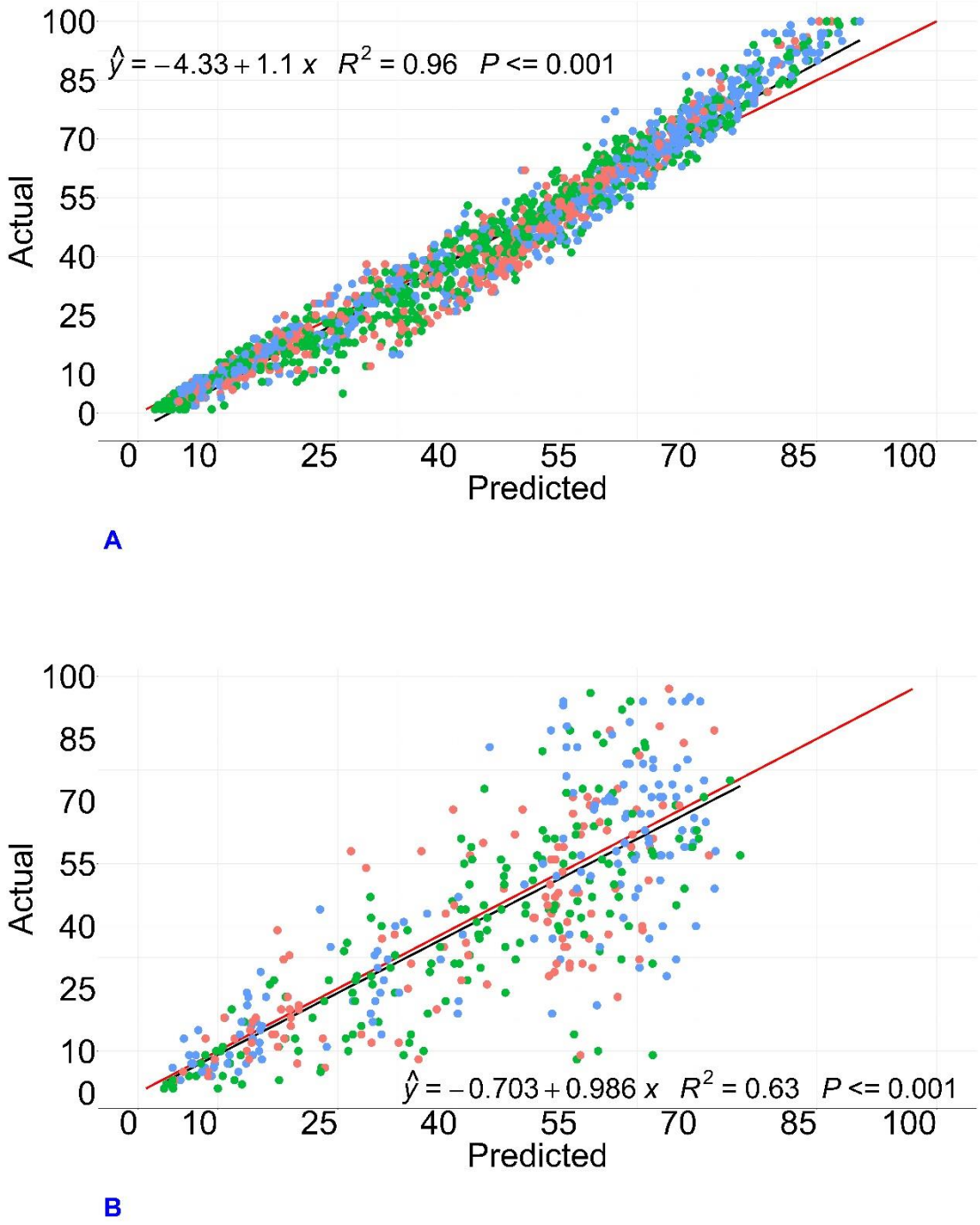


Figure 6-6 Actual vs. predicted grass growth in the national model.

Training data with 20 significant variables. 2017 points (orange), 2018-(green) and 2019 (blue). Panel A represents the model with training data and panel B represents the model with testing data.

6.4.3 Agmodel 1-First part (January- June)

The features and their importance in terms of MSE are plotted in Figure 6-7. The highest importance variable was the maximum value of band 11, and the lowest importance variable was the median value of NDVI. The model was trained from 20 variables to two variables one by one to see how much the RMSE changes. The error metrics for input variables from twenty until two are given for test data in Table 6-3. The accuracy metric values for all the input combinations were similar. The final model was with the top 18 variables and had the highest R^2 (0.74), MAE (11.30 kg DM $ha^{-1}day^{-1}$), MSE (241.15 kg DM $ha^{-1}day^{-1}$) and RMSE (15.52 kg DM $ha^{-1}day^{-1}$) for the test data.

The scatterplot for training data (Figure 6-8 A) and test data (Figure 6-8 B) between grass growth rate from PBI and Agmodel 1 output is shown in Figure 6-8. The actual grass growth rate are on the y-axis, and the modelled output are on the x-axis. The R^2 was 0.97 between actual and predicted values for training data and 0.74 for the test data. There were three clusters for the year 2017 for test data- one under 20 kg DM $ha^{-1}day^{-1}$, the second one at around 35 kg DM $ha^{-1}day^{-1}$ and the third cluster above 50 kg DM $ha^{-1}day^{-1}$. There were no values for 2017 between 30-50 kg DM $ha^{-1}day^{-1}$ for 2018 and 2019. The values for 2018 below 25 kg DM $ha^{-1}day^{-1}$ were over-predicted, whereas the values above 50 kg DM $ha^{-1}day^{-1}$ had large variability. The 2019 values below 30 kg DM $ha^{-1}day^{-1}$ were on the 1:1 line, and some values were under-predicted, whereas the values above 50 kg DM $ha^{-1}day^{-1}$ were over- and under-predicted equally. This model had a similar behaviour as the Agmodel with over-prediction of more than 50 kg DM $ha^{-1}day^{-1}$. The values of less than 40 kg DM $ha^{-1}day^{-1}$ were closely following the regression line.

As explained in Section 5.2.2.2, PBI data were paired with satellite data if it falls within seven days. The difference of days between PBI and satellite data acquisitions was calculated, and a histogram was plotted as shown in Figure 6-9. It was observed that the maximum amount of data had a difference of zero which means that satellite and PBI data collection was on the same day.

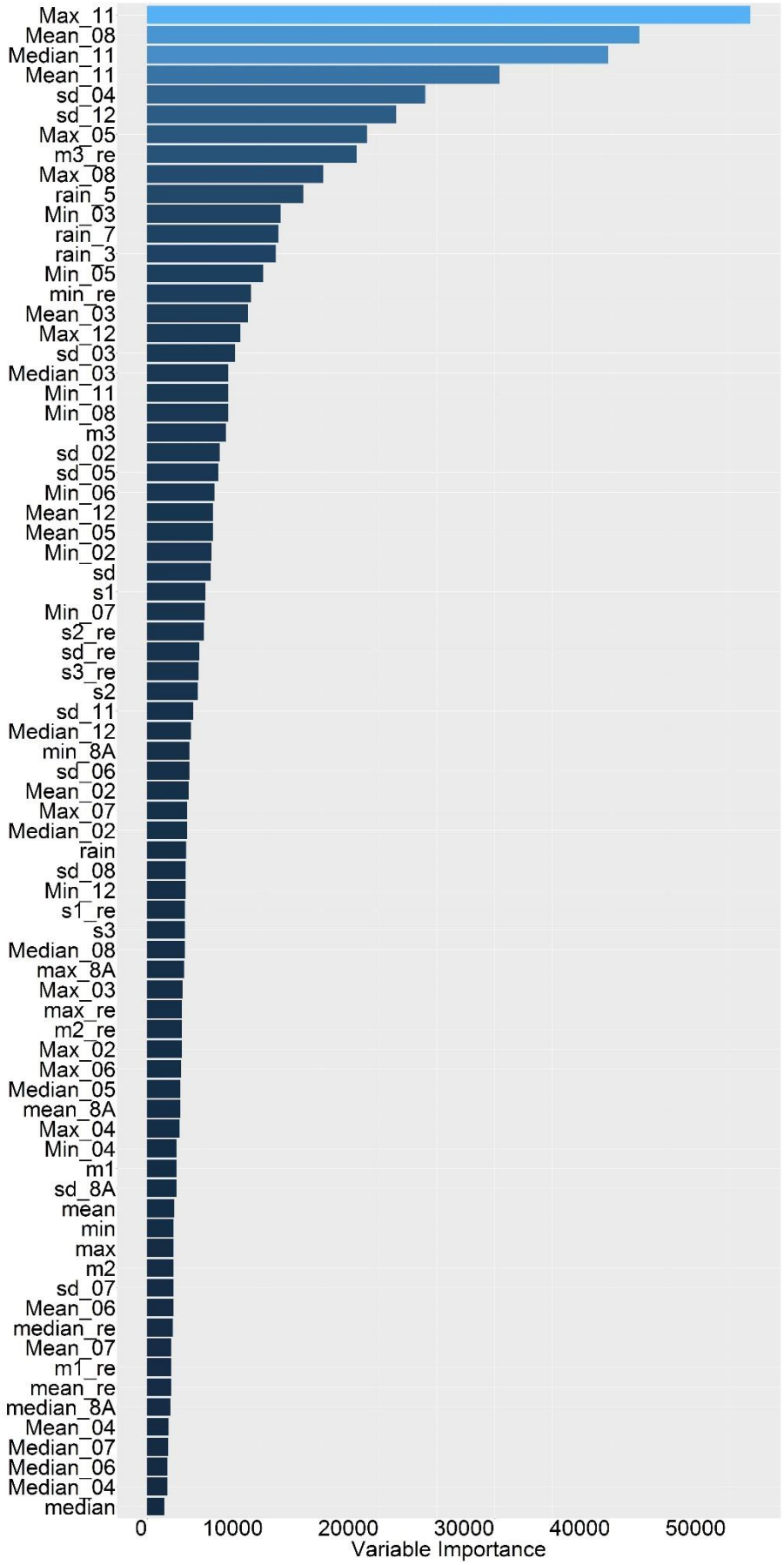


Figure 6-7 Feature importance for Agmodel 1

Table 6-3 Accuracy metrics for test data for Agmodel 1

R^2, MAE, MSE, RMSE. The metrics are shown from top 2 to 20 significant variables. The lowest RMSE (kg DM $ha^{-1}day^{-1}$) was for the top 18 significant variables. The final model with lowest RMSE and highest R^2 is highlighted in bold.

No. of Variables	R^2	MAE	MSE	RMSE
20	0.73	11.33	243.80	15.61
19	0.73	11.30	243.41	15.60
18	**0.74**	**11.30**	**241.15**	**15.52**
17	0.73	11.32	243.26	15.59
16	0.73	11.34	246.95	15.71
15	0.74	11.42	241.53	15.54
14	0.72	11.46	245.81	15.67
13	0.73	11.49	243.80	15.61
12	0.72	11.69	250.47	15.82
11	0.72	11.65	250.51	15.82
10	0.71	11.65	254.30	15.94
9	0.72	11.59	252.53	15.89
8	0.70	12.28	268.06	16.37
7	0.71	11.84	257.07	16.03
6	0.70	12.04	262.86	16.21
5	0.71	11.88	256.74	16.02
4	0.72	11.80	251.31	15.85
3	0.68	12.54	286.38	16.92
2	0.69	12.07	279.21	16.70

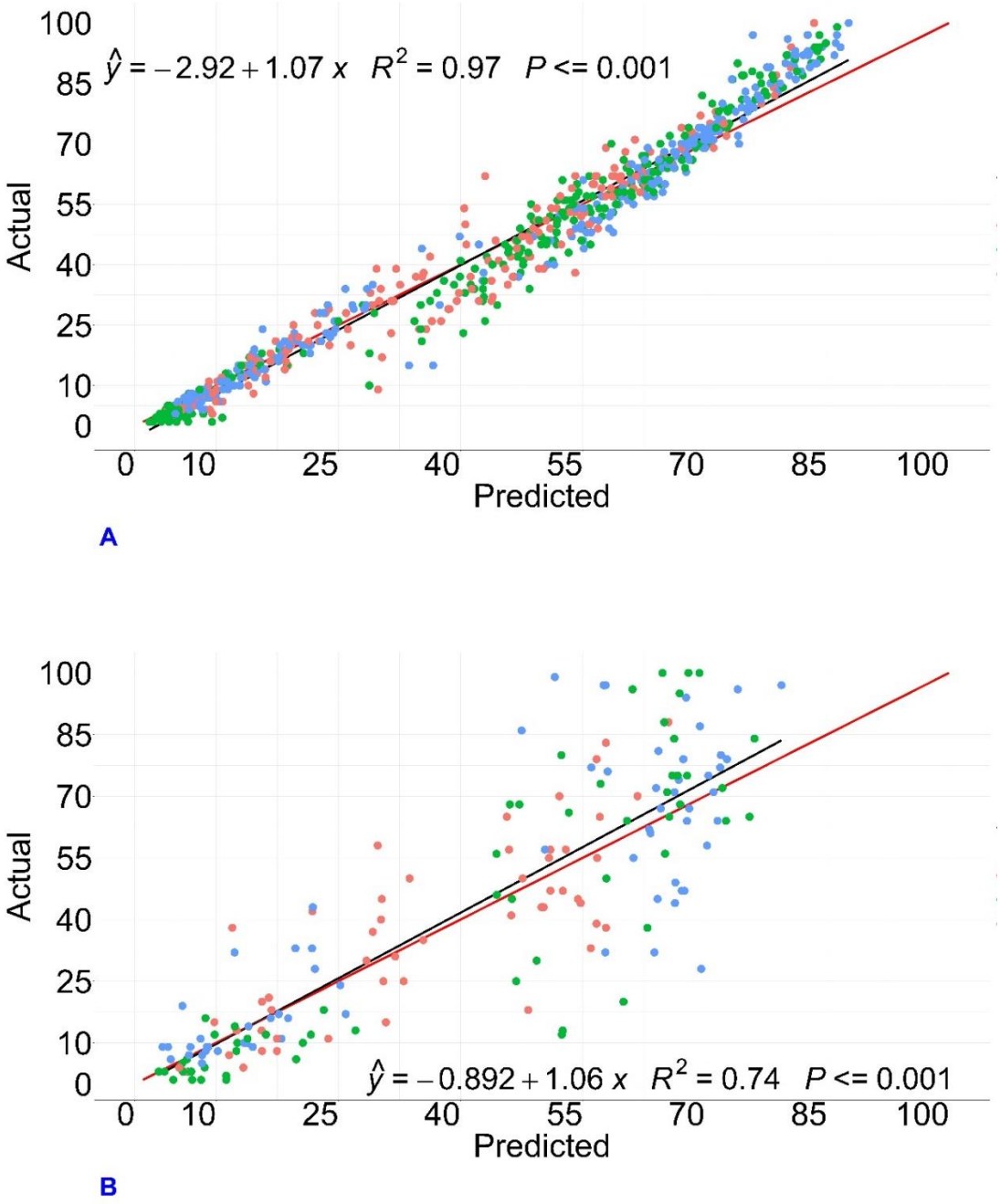


Figure 6-8 Actual vs. predicted grass growth for Agmodel 1 national model. Training data with 18 significant variables. 2017 points (orange), 2018-(green) and 2019 (blue). Panel A represents the model with training data and panel B represents the model with testing data.

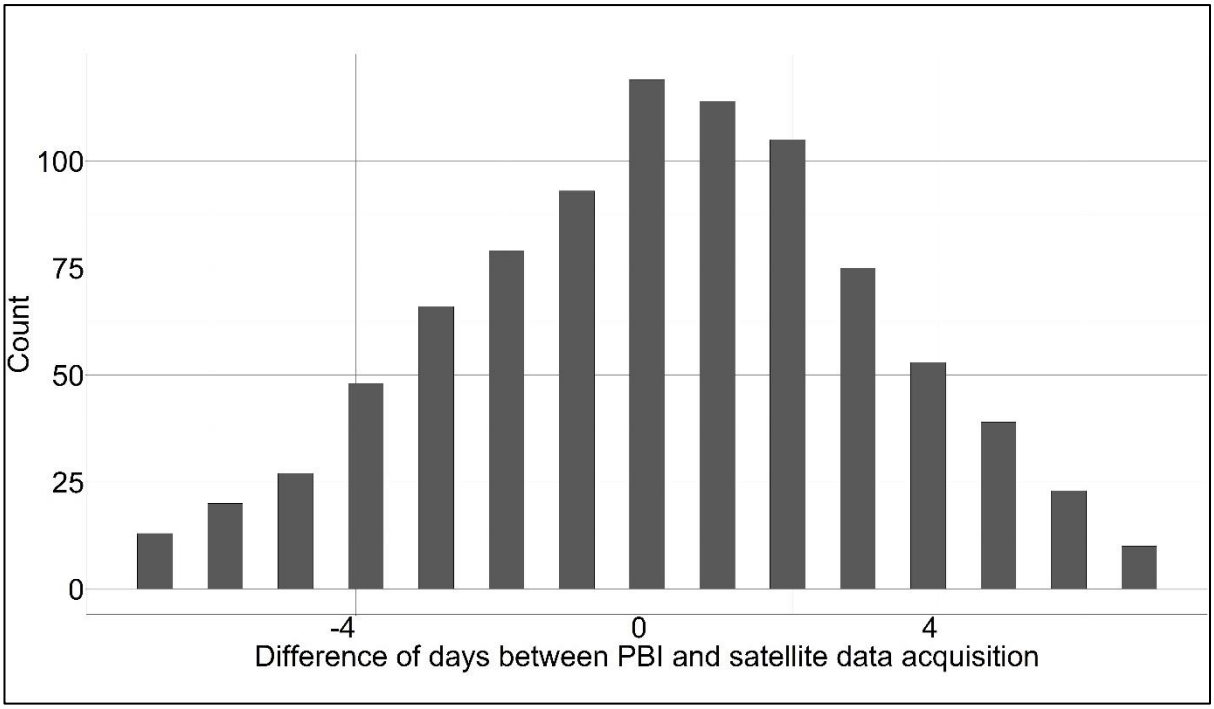


Figure 6-9 Histogram of difference of days between Sentinel 2 and PBI.

6.4.4 Agmodel 2-Second part (July-Dec)

Agmodel 2 used the data by dividing the national data from July until December, the second half of the grass-growing season. The feature importance of all the variables is calculated as shown in Figure 6-10. The median value of band 8 had the highest importance, whereas the minimum value of NDVI had the lowest importance. The model was run for all the variables and leaving the least important variable each time. The RMSE for the variables from 77 to 22 was similar and not shown in Table 6-4. The RMSE for all the variables was similar, with the lowest RMSE of 13.74 kg DM $ha^{-1}day^{-1}$ for the top 17 variables. For the top 17 variables for test data, the R^2 was 0.58, MAE was 10.81 kg DM $ha^{-1}day^{-1}$, MSE was 188.80 kg DM $ha^{-1}day^{-1}$, and RMSE was 13.74 kg DM $ha^{-1}day^{-1}$, which was the final model.

The scatterplot for training data is shown in Figure 6-11 A and test data in Figure 6-11 B. The training and testing data had an R^2 of 0.95 and 0.58 between the actual and predicted grass growth rate. The grass growth values for less than 40 kg DM $ha^{-1}day^{-1}$ had less variability and the values of more than 40 kg DM $ha^{-1}day^{-1}$ had higher variability. Unlike Agmodel 1, there was no clear clustering of the data for 2018 and 2019. The year 2017 had few values at around 10 kg DM $ha^{-1}day^{-1}$ and highly variable values for more than 40 kg DM $ha^{-1}day^{-1}$. There was just a single value for 2017 between 25 and 40 kg DM $ha^{-1}day^{-1}$. The 2018 values were scattered throughout without any clear pattern. The 2019 values were around 1:1 line for less than 40 kg DM $ha^{-1}day^{-1}$ and were under-predicted for more than 40 kg DM $ha^{-1}day^{-1}$.

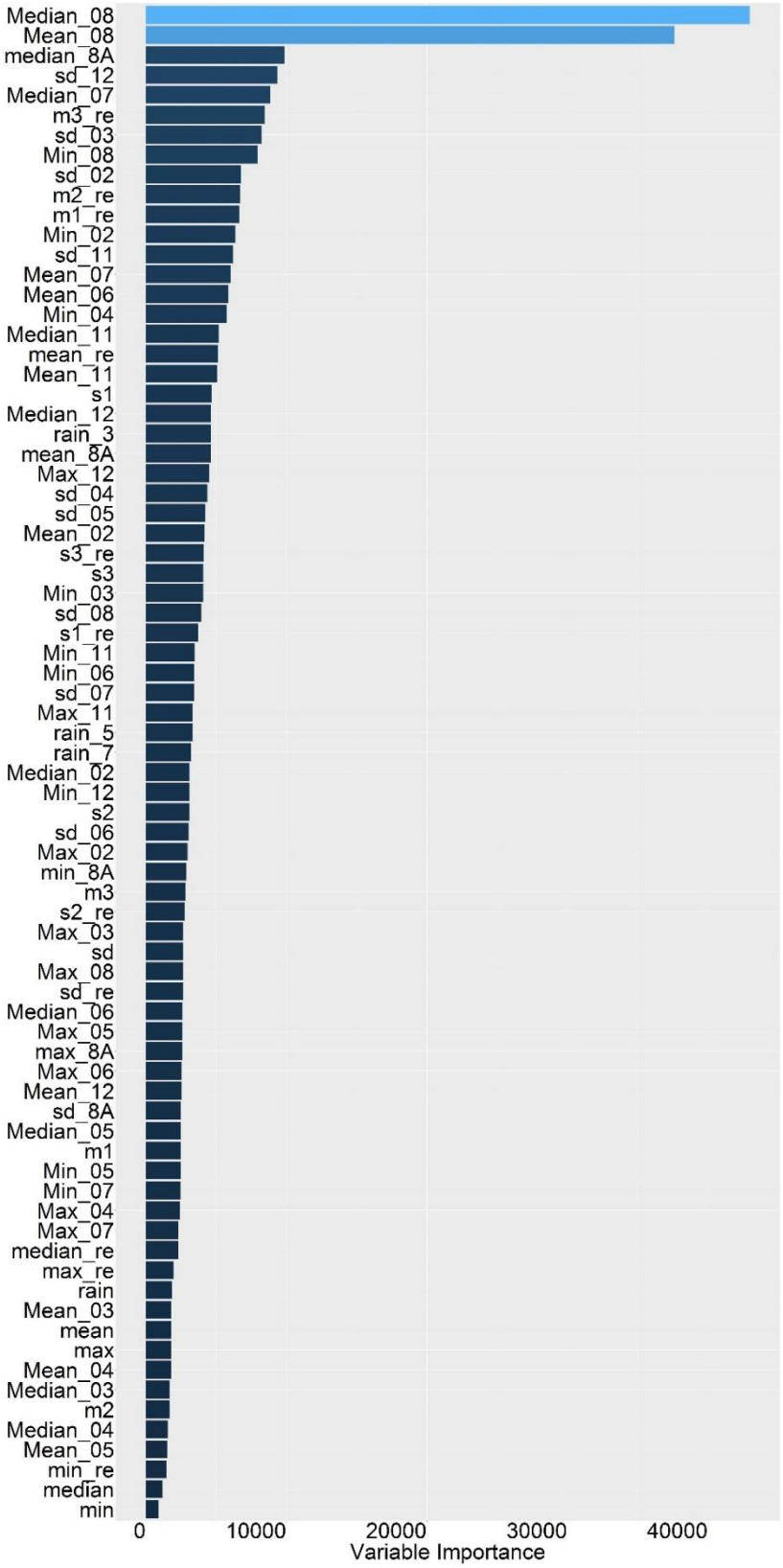


Figure 6-10 Variable importance for Agmodel 2

Table 6-4 Accuracy metrics- R^2 for test data for Agmodel 2

R^2, MAE, MSE, RMSE. The metrics are shown from the top 2 to 20 significant variables. The lowest RMSE (kg DM $ha^{-1}day^{-1}$) was for the top 17 significant variables and is highlighted in bold.

No. variables	R^2	MAE	MSE	RMSE
20	0.56	10.97	192.75	13.88
19	0.57	10.84	189.37	13.76
18	0.57	10.83	189.65	13.77
17	**0.58**	**10.81**	**188.80**	**13.74**
16	0.56	10.97	197.60	14.05
15	0.56	10.95	196.35	14.01
14	0.55	11.01	198.31	14.08
13	0.55	11.01	198.57	14.09
12	0.55	10.86	198.93	14.10
11	0.56	10.86	195.01	13.96
10	0.56	10.87	196.42	14.01
9	0.57	10.80	191.90	13.85
8	0.55	11.03	198.97	14.10
7	0.54	11.11	206.22	14.36
6	0.53	11.05	210.52	14.50
5	0.52	11.18	213.70	14.61
4	0.51	11.40	217.83	14.75
3	0.52	11.32	215.03	14.66
2	0.50	11.55	222.08	14.90

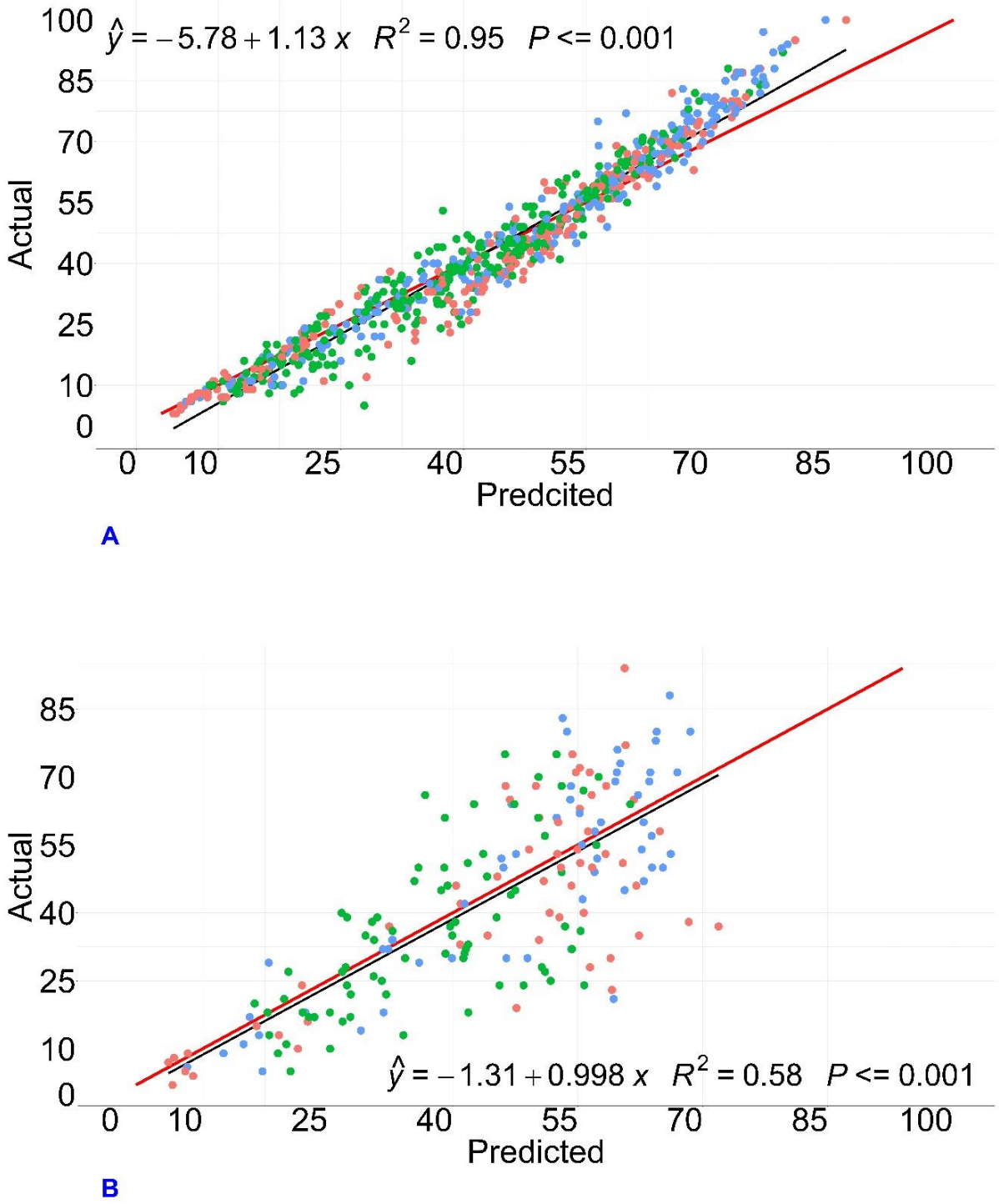


Figure 6-11 Actual vs. predicted grass growth for Agmodel 1 national model. Training data with 17 significant variables. 2017 points (orange), 2018-(green) and 2019 (blue). Panel A represents the model with training data and panel B represents the model with testing data.

6.5 Validation of final model 2020 data

The three national models were developed in Section 6.4.2 to 6.4.4 were run for the data from 2020 which was kept aside as validation data to validate the models.

6.5.1 Agmodel- National model

The trained Agmodel from Section 6.4.2 was used to predict the grass growth rate for 2020 data. The scatterplot for the data is shown in Figure 6-12. As the grass growth rate values cease, the variability in prediction increases. There were three clusters of data points in the plot, which could be because of the hypothesis of dividing the NDVI data into 3 quartiles. The values less than 20 kg DM $ha^{-1}day^{-1}$ were clustered together and are closer to the regression line; the grass growth rate values between 25 to 45 kg DM $ha^{-1}day^{-1}$ formed another cluster with higher variability than the values lower than 25 kg DM $ha^{-1}day^{-1}$. The grass growth rate cluster above 55 kg DM $ha^{-1}day^{-1}$ 1 had the highest variability. The R^2 for the testing data was 0.63, MAE was 12.55 kg DM $ha^{-1}day^{-1}$, MSE was 283.28 kg DM $ha^{-1}day^{-1}$, and RMSE was 16.83 kg DM $ha^{-1}day^{-1}$.

The difference of days of ground and satellite data acquisition is plotted as a histogram in Figure 6-13. The majority of the values are between -4 and 4, which means the grass growth rate values were collected 4 days before or after the Sentinel 2 images. The difference of days was plotted against the mean residual values from the Agmodel in Figure 6-14. The standard deviation values are plotted as error bars. The mean residual values are the highest for the higher difference in days such as -7 and +7 days. That means the residuals are affected by the difference in data acquisition dates. The difference of days was also plotted against the absolute values of residuals in Figure 6-15. A massive cluster of values is between -4 and +4 days with few outliers.

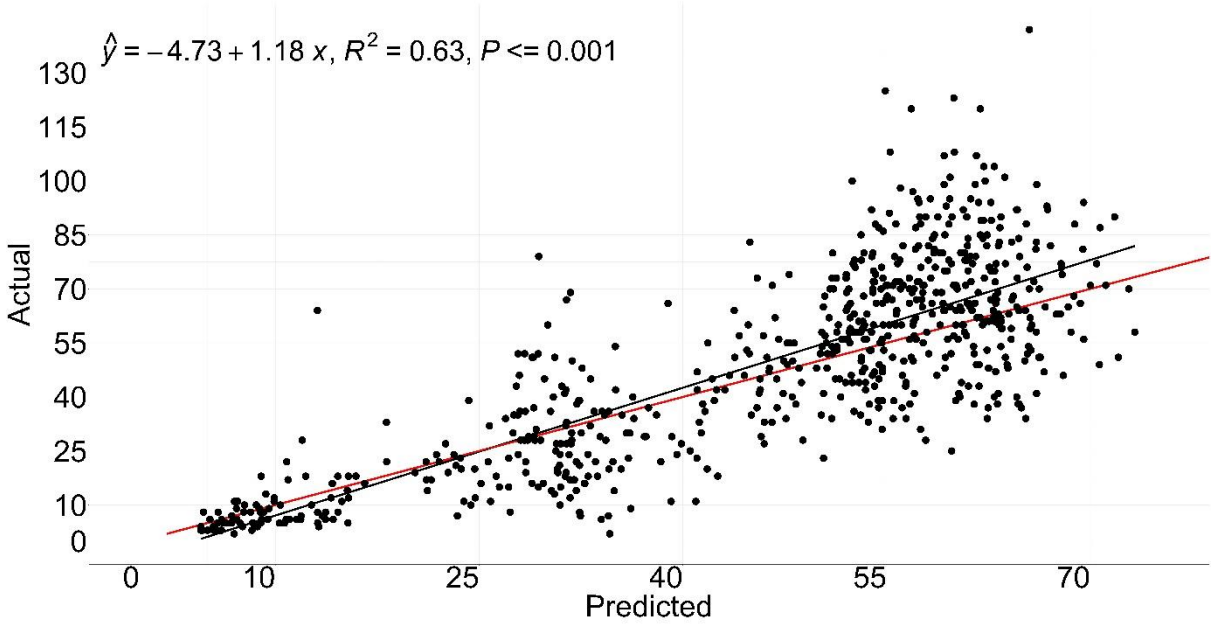


Figure 6-12 Actual vs. predicted grass growth rate for 2020 data using Agmodel

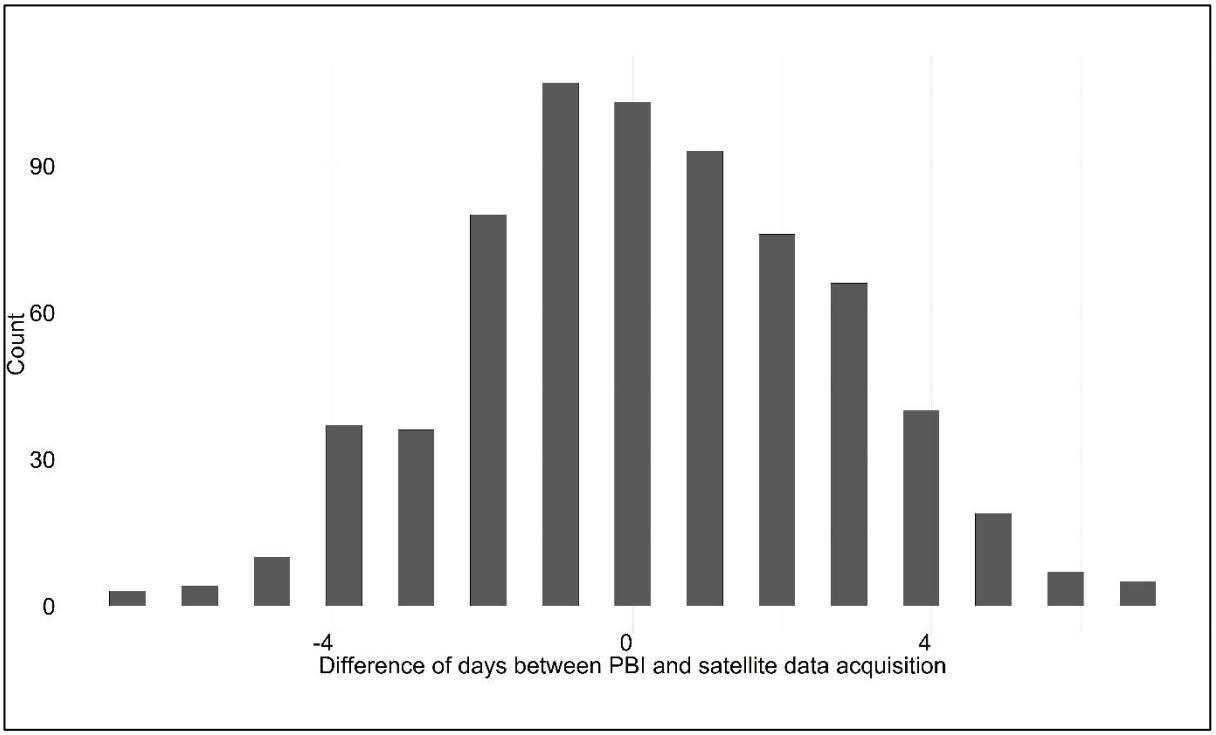


Figure 6-13 Difference of acquisition days between Sentinel 2 & PBI (Agmodel)

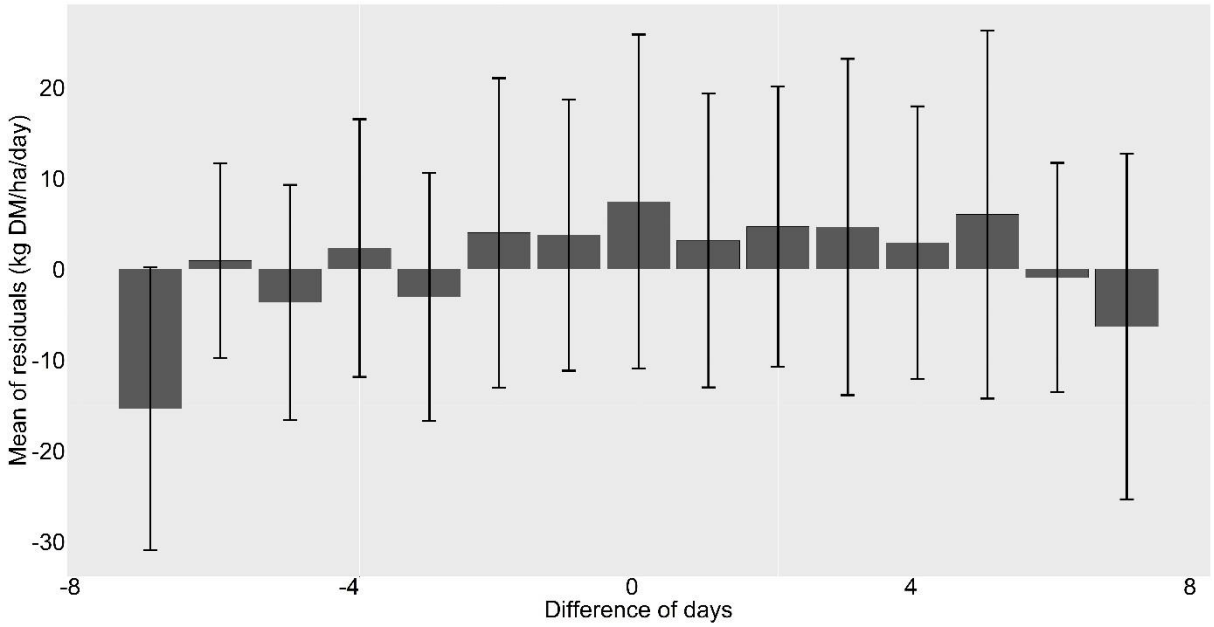


Figure 6-14 Residuals vs. difference of data acquisition days (Agmodel)

The standard deviation values are plotted as error bars.

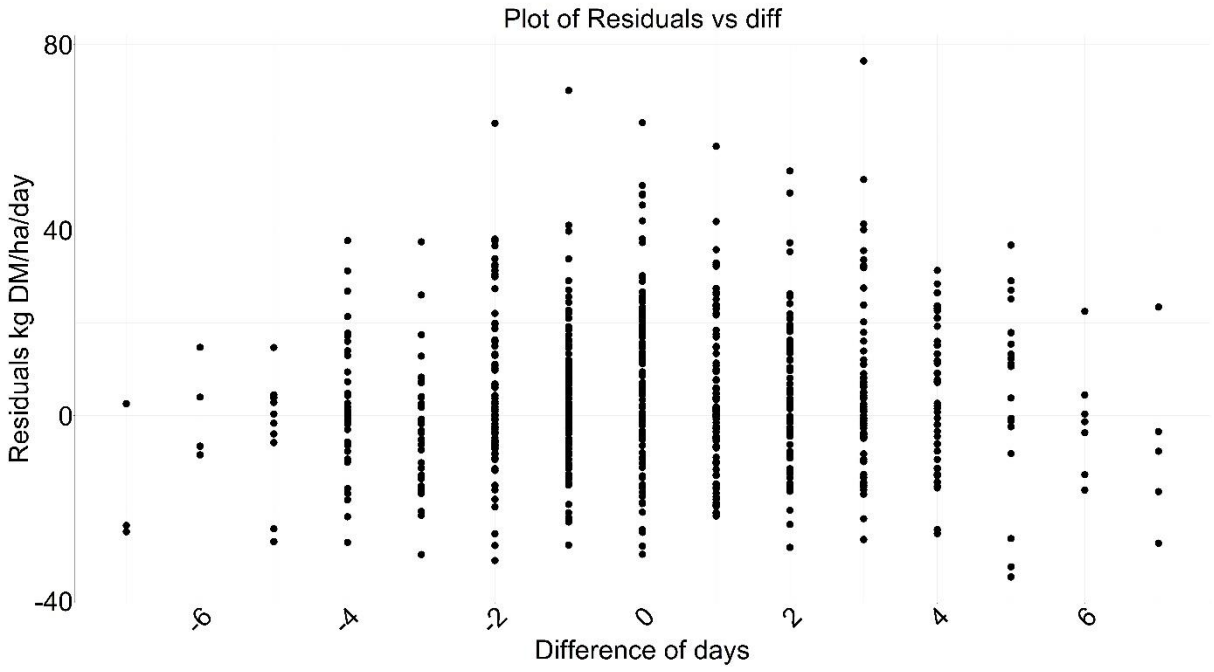


Figure 6-15 Difference of acquisition days between Sentinel 2 & PBI vs. residuals

6.5.2 Agmodel 1- First part (January-June)

The Agmodel 1 trained in Section 6.4.3 is used to predict the values for 2020 validation data. The scatter plot for the predicted and ground values is shown in Figure 6-16. Similar to the Agmodel results, this plot has three clusters. The lower grass growth rate values were predicted with higher accuracy than the higher grass growth values of more than 40 kg DM $ha^{-1}day^{-1}$. The model under-predicted values more than 40 kg DM $ha^{-1}day^{-1}$. The R^2 of the testing data was 0.7, MAE was 12.49 kg DM $ha^{-1}day^{-1}$, MSE was 318.01 kg DM $ha^{-1}day^{-1}$, and RMSE was 17.83 kg DM $ha^{-1}day^{-1}$.

When the mean values of residuals were plotted against the difference of days in Figure 6-17, the overall model under-predicted grass growth to see the influence of the difference of days on the prediction values rate. The mean residual values were high for the highest difference of days, i.e. ±7 days. The absolute values of residuals were plotted against the difference of days in Figure 6-18. The residuals are mostly centred on ±2 days, but there are some outliers with high residuals for ±7 days.

The histogram of the difference of days is plotted in Figure 6-19. As the first half of the grass-growing season has more images due to better weather conditions in spring and summer, most images were found within ±3 days of the ground data.

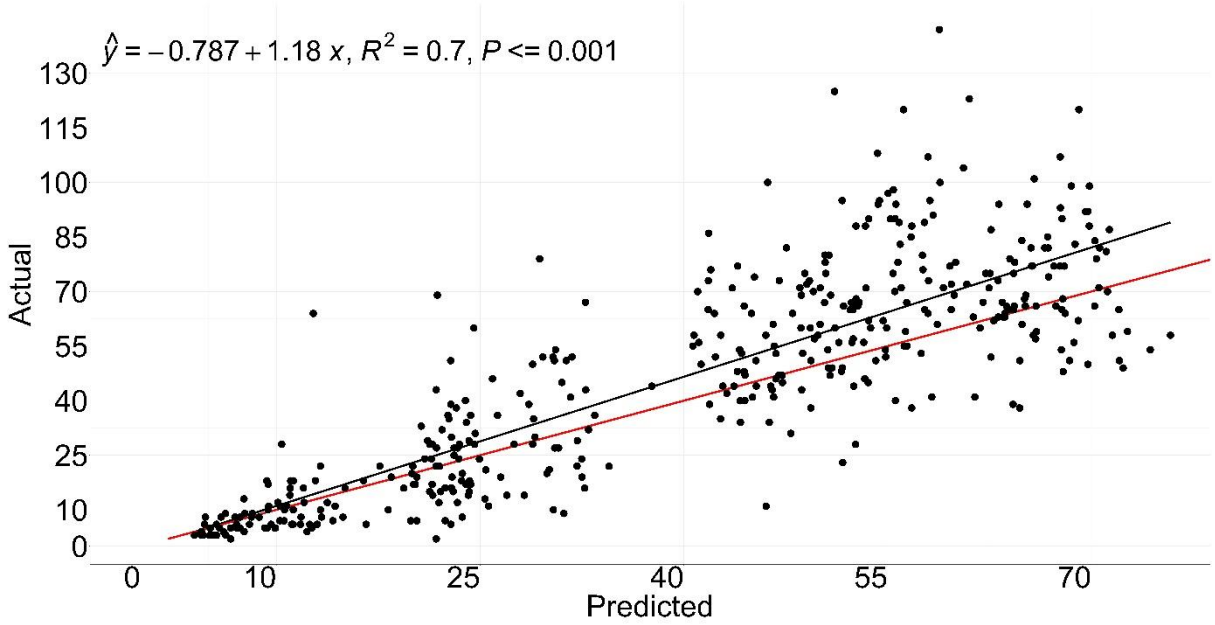


Figure 6-16 Actual vs. predicted grass growth (2020) using Agmodel 1.
Redline is the 1:1 line, and the black line is the regression line.

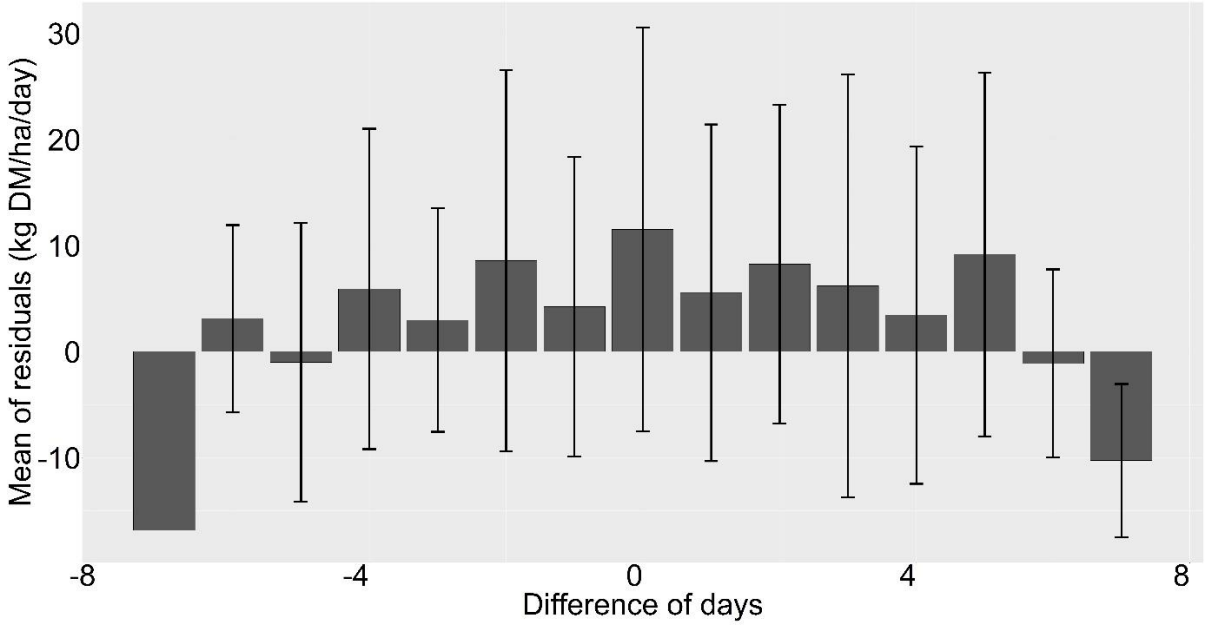


Figure 6-17 Residuals for Agmodel1 vs. difference of days between Sentinel 2 & PBI.
The standard deviation is shown as error bars.

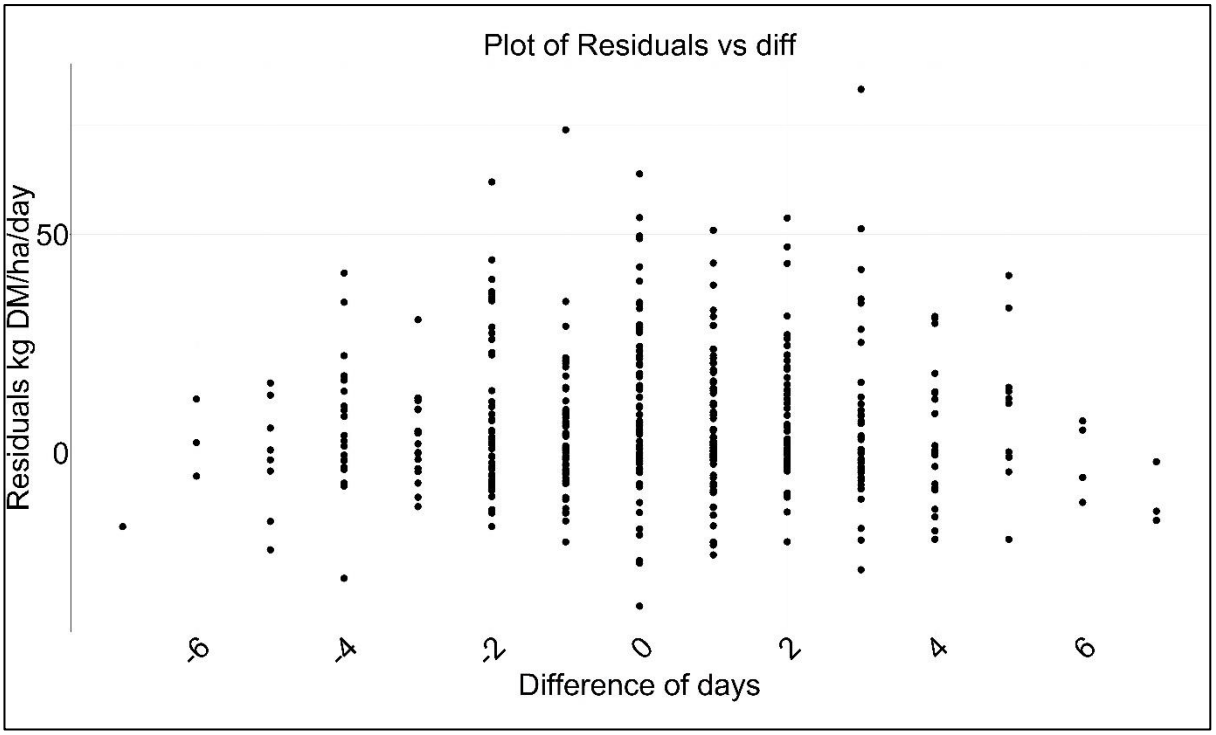


Figure 6-18 Residual values (kg DM $ha^{-1}day^{-1}$) for Agmodel 1 vs. difference of days between Sentinel 2 & PBI.

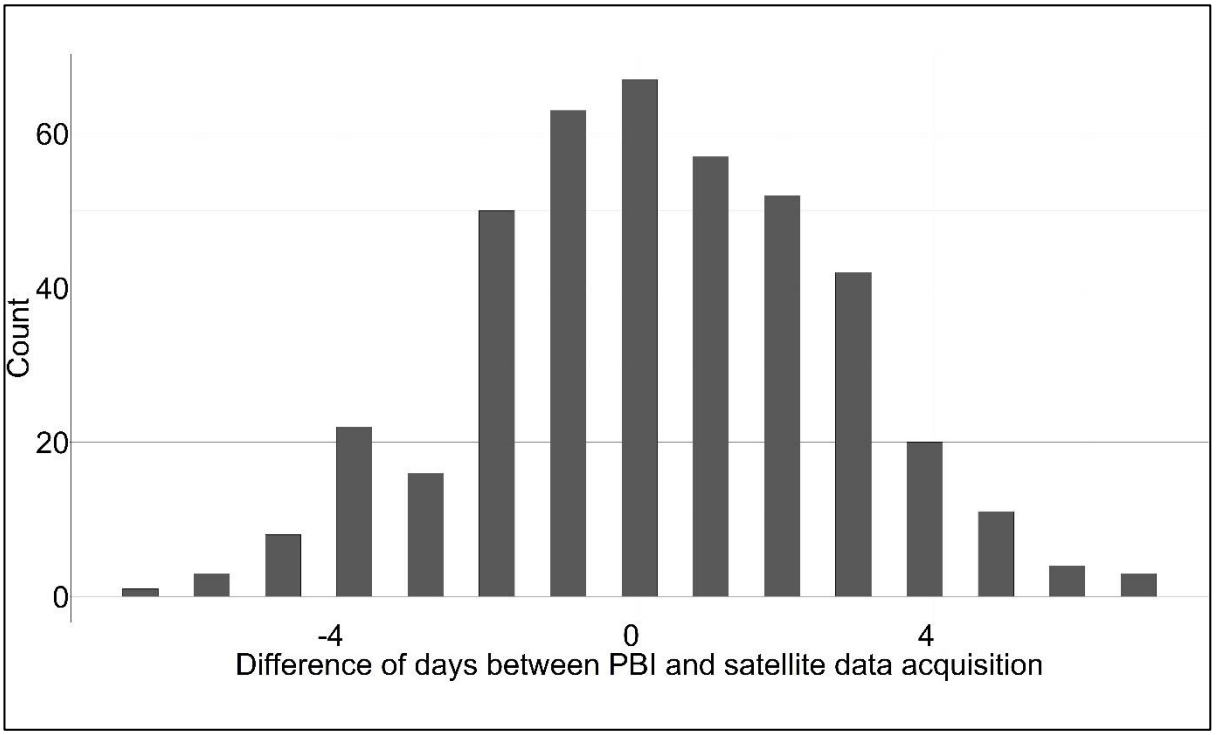


Figure 6-19 Histogram of difference of acquisition days between Sentinel & PBI (Agmodel 1)

6.5.3 Agmodel 2- Second part (July-December)

The scatterplot for the validation data is shown in Figure 6-20. The R^2 was 0.36, MAE was 13.14 kg DM $ha^{-1}day^{-1}$, MSE was 273.50 kg DM $ha^{-1}day^{-1}$, and RMSE was 16.53 kg DM $ha^{-1}day^{-1}$. The predictions were equally distributed under and above the regression line without any clear clustering of data. The mean residual values were plotted against the difference of days, as shown in Figure 6-21. The difference of days between ±4 days had low average residual values, whereas the high mean residual values were observed for 5 and 7 days of difference.

The absolute residual values were plotted against the difference of day, as shown in Figure 6-22. The majority of the residual values lies between ±4 days; as the difference of days increases, the residual values increases. The histogram of the difference of days was plotted as in Figure 6-23. The majority of the data had grass growth from PBI collected a day after satellite data acquisition. Very few data points had more than ±4 days of difference.

The overall results for validation of the three models are shown in Table 6-5. The validation data were from the 2020 year until the end of October.

- RMSE for Agmodel was 16.83 kg DM $ha^{-1}day^{-1}$.
- RMSE for Agmodel (first half) was 17.27 kg DM $ha^{-1}day^{-1}$.
- RMSE for Agmodel (second half) was 16.08 kg DM $ha^{-1}day^{-1}$.
- RMSE for Agmodel 2 was 16.53 kg DM $ha^{-1}day^{-1}$.
- RMSE for Agmodel 1 higher at 17.83 kg DM $ha^{-1}day^{-1}$.

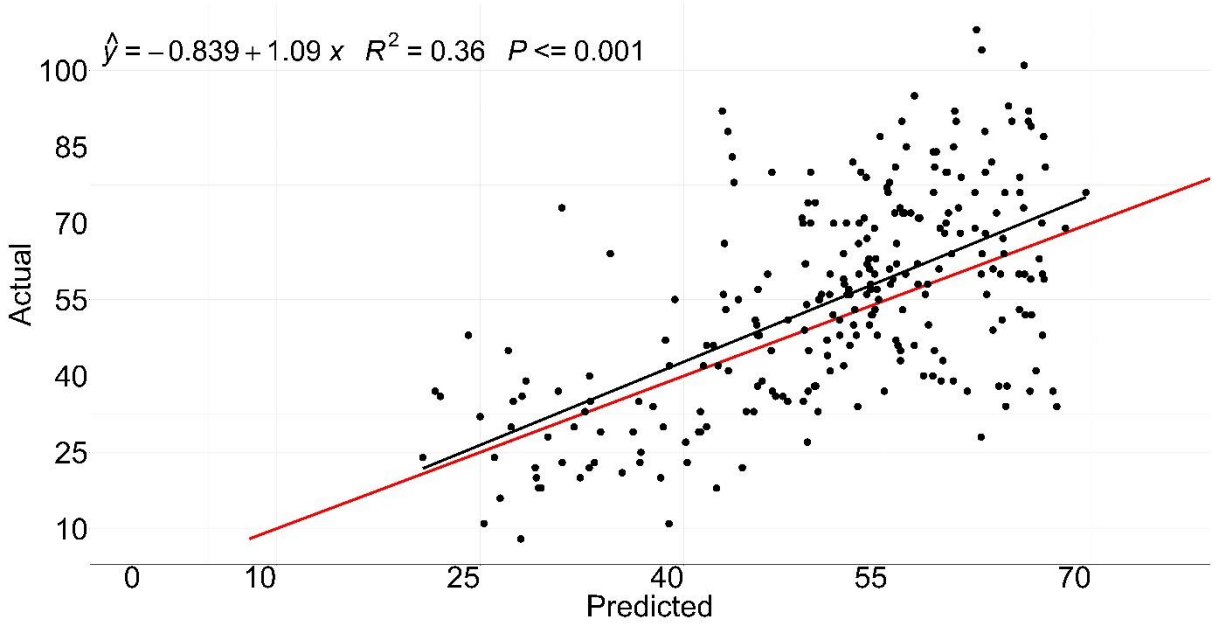


Figure 6-20 Actual vs. predicted grass growth rate (2020) using Agmodel 2. Redline is the 1:1 line, and the black line is the regression line.

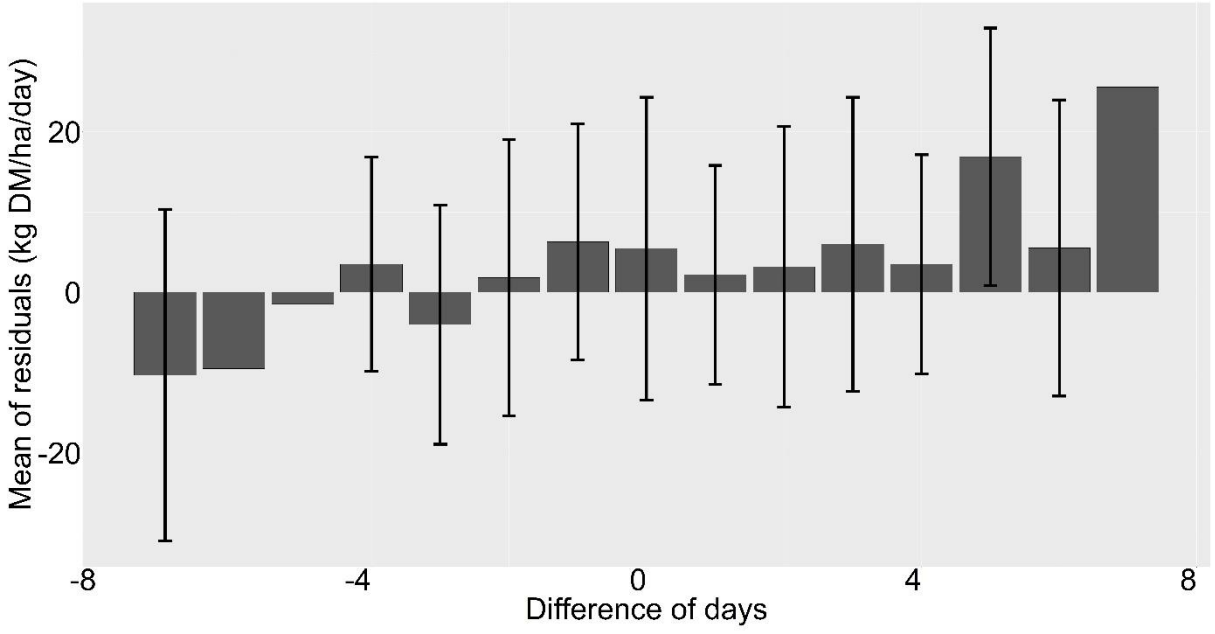


Figure 6-21 Mean residuals for Agmodel1 vs. difference of days between Sentinel & PBI. The standard deviation is shown as error bars.

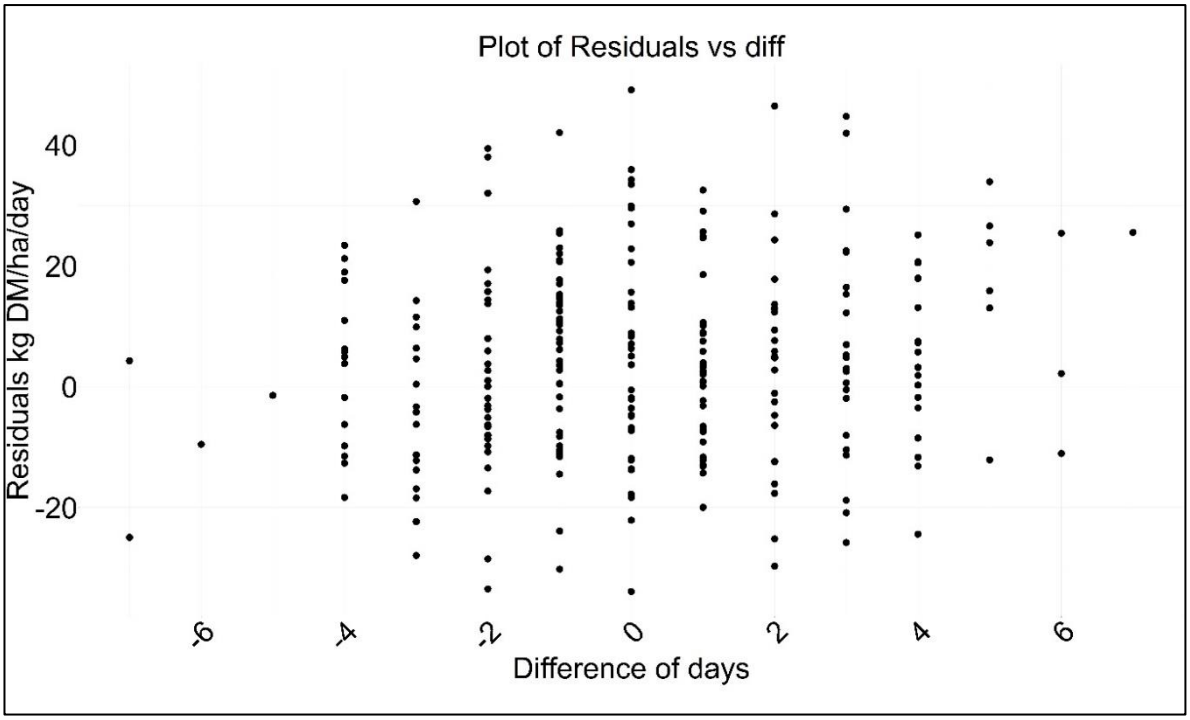


Figure 6-22 Residual values from Agmodel 2 vs. difference of days between Sentinel 2 & PBI.

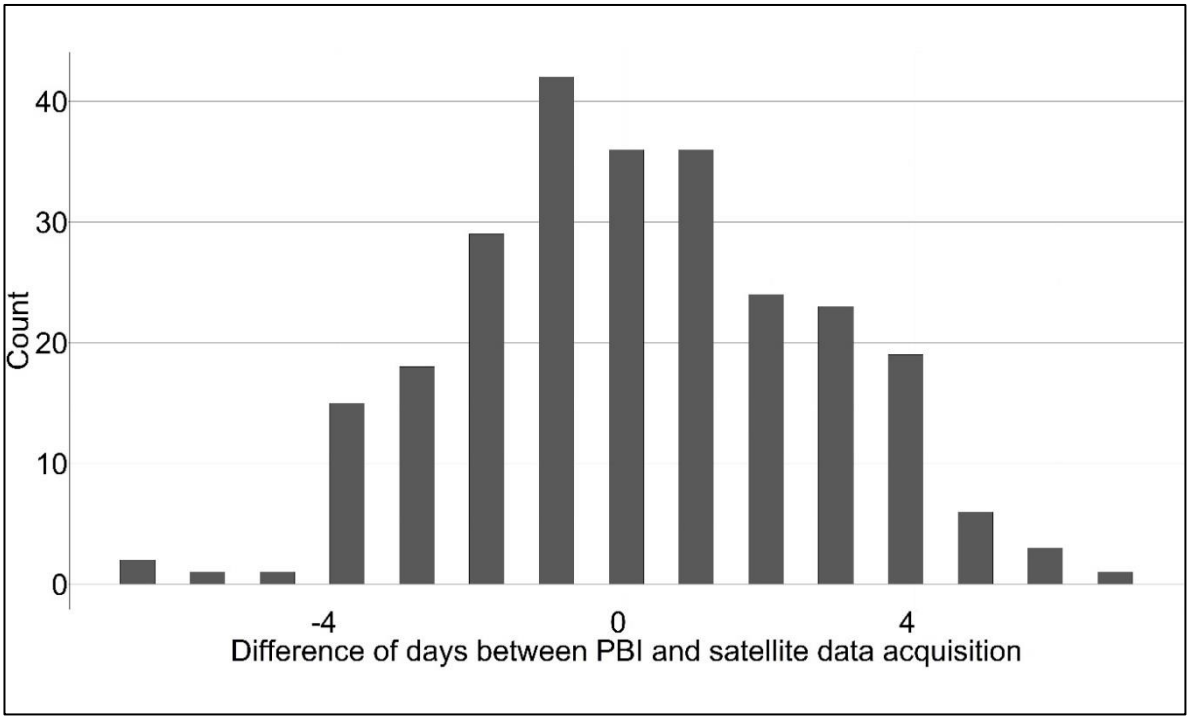


Figure 6-23 Histogram of difference of acquisition days between Sentinel 2 & PBI (Agmodel 2)

Table 6-5 Comparison of accuracy metrics for the national model R^2, MAE, MSE and RMSE (kg DM $ha^{-1}day^{-1}$) for 2020 validation data using ($p<0.01$ in each case)

	R^2	MAE	MSE	RMSE
Agmodel	0.63	12.55	283.28	16.83
Agmodel (First half)	0.68	12.36	298.58	17.27
Agmodel (Second half)	0.41	12.84	258.88	**16.08**
Agmodel 1	**0.70**	12.49	318.01	17.83
Agmodel 2	0.36	13.14	273.50	16.53

6.6 Discussion

A machine-learning model was developed to estimate the grass growth rate for 179 farm sites distributed across Ireland. The model was using the only Sentinel 2 satellite imagery and daily rainfall data. A shortage of satellite data because of cloud cover prevented expanding this model for all the farms registered with the PBI.

The approach adopted here builds on the machine-learning modelling approach of Ali et al. (2017b), where satellite data from two farms were used to estimate grass growth rate. In this study, there was access to an extensive network of farms in County Donegal, Galway, Sligo and Cork in Ireland. The 179 farms covered geographically north to south of Ireland, covering various enterprises, soil type, drainage and different management on farms. The model was developed using data from 179 farms registered with PBI from 2017 to 2020.

The current model is a significant step towards developing national grass modelling capability here in Ireland. However, when complete data becomes available, this model's accuracy should improve, as the machine-learning models are data-driven. Unlike conventional crop biomass models, such as the Brereton (Brereton and Keane, 1992), the Johnson & Thornley (Johnson and Thornley, 1983), the Jouven

(Jouven et al., 2006) and Moorepark St Gilles model (MoSt GG model) (Ruelle et al., 2018b), the EO-based model outlined here does not require site-specific information on soil type, initial conditions, topography, or farm management data. All these factors vary spatially in a heterogeneous environment. The main disadvantage of such models is that they lack the spatial element and need to be calibrated for each location with varying soil type, weather and management conditions. It is challenging to use biophysical models at all points as they are difficult to parameterise (Donohue et al., 2018).

The machine-learning models presented here have an advantage over the conventional models in that once trained on an extensive training set, they can predict growth for unseen data in subsequent years. Another advantage of Agmodel is that it is scalable, from paddock/farm-scale to countrywide scale. The models used in the present study were not calibrated for a specific farm. All the farms were divided randomly for training and testing the model. The national model here is extensive and can be used for any farm in the country with good performance (RMSE between 13.74 and 15.52 kg DM $ha^{-1}day^{-1}$).

6.6.1 National model limitations

The first limitation for the national model is the use of optical data. The models using vegetation indices tend to saturate at high grass growth values and the uncertainties in the predictions will be high when grass growth rates are high. The satellite data also suffers from cloud contamination leading to few useful images per year. For this work, Sentinel-2 was used and when the satellite data was not available, the corresponding ground observation was not used in the model. Such unavailability of data can lead to uncertainties and errors in the model. The time when Sentinel-2 will be unavailable due to clouds, the gap can be filled using UAV datasets, which can be flown on demand and can be incorporated in the future work. The SAR data can be used to capture the management events with the Sentinel-2 data and can help to fill the gaps when cloudy images occur.

The model presented in this work is not a final product and cannot be used by the farmers directly. This model is an attempt towards development of a national model to estimate grass growth, which can be challenging. This worked can be improved in the future by incorporation of the points as discussed in Section 8.2.

6.6.2 Model performance

The Commercial model performed satisfactorily with an R^2 of 0.61 for the test data. The model had DOY and NDRE as the top two variables. The approach followed for this model formed as a basis for the national model.

According to the feature importance for the national dataset, the variables were used to assess the effect of each variable on model output. The accuracy for all the models for test data does not change drastically while moving from the top 77 significant variables to two. The models with the top 20 variables were the best models with the lowest RMSE- Agmodel (15.00 kg DM $ha^{-1}day^{-1}$), top 18 variables for Agmodel 1 (15.52 kg DM $ha^{-1}day^{-1}$) and top 17 variables for Agmodel 2 (13.74 kg DM $ha^{-1}day^{-1}$) for testing data. Agmodel 2 had the lowest RMSE and lowest R^2 showing that data has high variability and is noisy. The trend indicates that the predictor variables still provides information about the grass growth rate even though data points fall further from the regression line.

The three trained models were applied to the data from 2020 for validation. Agmodel for 2020 data had an RMSE of 16.83 kg DM $ha^{-1}day^{-1}$. Agmodel for first half of 2020 data had an RMSE of 17.27 kg DM $ha^{-1}day^{-1}$. Agmodel for second half of 2020 data had an RMSE of 16.08 kg DM $ha^{-1}day^{-1}$. Agmodel 1 had the highest RMSE of 17.83 kg DM $ha^{-1}day^{-1}$, whereas Agmodel 2 had an RMSE of 16.53 kg DM $ha^{-1}day^{-1}$. The difference in RMSE between the models was not significant. The training accuracy was consistently higher for all the models because training data were seen by the random forest while training the models. The RMSE for the validation model with the year 2020 data was between approximately 16.08-17.83 kg DM $ha^{-1}day^{-1}$.

6.6.3 Variable importance

The most important variable for all the models was DOY. The second most important variable for the Commercial model was NDRE, a red edge vegetation index and sensitive to the chlorophyll content in grass. The essential variables for Agmodel were the near-infrared (NIR) band (Sentinel 2, band 8). The least essential variables were the median value of NDRE. For Agmodel 1, the best predictor was the maximum SWIR band (Sentinel 2, band 11 while the least important was the median value of NDVI. The median value of the NIR band was the most significant in estimating grass growth rate using Agmodel 2, whereas; median of NDVI was the least important variable. In all the models, NIR band was the most crucial variable. The reason for NIR and SWIR to be the most important predictors in grass growth estimation is because NIR and SWIR are sensitive to grass water content, which is correlated to chlorophyll content (Shoko et al., 2018). Random forest creates 'n' number of trees using out-of-bag samples with available input features as discussed in Section 5.2.4.2. It randomly shuffles the data and calculates feature importance for permuted out-of-bag samples and therefore for each model, variable importance can differ which is a case here.

Rainfall was the rejected variable for all three models as the grass growth represents the previous accumulation rather than the same-day rainfall which is the result was in Chapter 5. Rainfall accumulation for 3-, 5- and 7- days was in the top twenty variables for Agmodel and Agmodel 1. For Agmodel 2, rainfall accumulation for 3 days was in the top twenty, whereas; accumulation of rainfall for 5- and 7-days were rejected. The SWIR band is correlated to crop biomass and is not affected by saturation due to increasing biomass which can form a significant predictor variable in grass growth modelling (Jenal et al., 2021). The Sentinel 2 bands- NIR and SWIR were the most important predictors for estimating grass biophysical parameters such as canopy chlorophyll content and a fraction of absorbed photosynthetically active radiation (FAPAR), which can help in grass growth monitoring (Sakowska et al., 2016).

6.6.4 PBI as a source of ground truth data

The reference data used in the machine-learning model was taken from PBI which is based on citizen science. PBI is a web- /mobile- based DSS where farmers sign up and upload their farm data such as grass growth rate, grass available on a farm, grazing and silage cutting dates. The methods to collect farm information is either by using an RPM or by visual assessment. Hanrahan et al. (2017) conducted a study in which PBI estimates were compared with the actual grass biomass values calculated using a cut and dry method used as a benchmark. The study was conducted for two years, 2014-2015, in Moorepark Farm. The analysis showed an RMSE of 409 kg DM/ha and an R^2 of 0.84. There could be potential sources of error while estimating grass biomass visually or using a plate meter on a farm. The errors from these methods can be propagated into the machine-learning model, and its impact should be considered while assessing the accuracy of the national model.

6.6.5 Grass growth curve visually

For analysis of the grass growth curve, a farm is chosen for representation. The grass growth curve using Agmodel for Farm 9 is shown in Figure 6-24. The predicted values match the actual grass growth rate very well from February to April and for October. The values from July until August were under-predicted by the model. The grass growth values corresponding to a lower difference in days were more accurate than the high difference in days. The more the difference in days of acquisition of data from satellite and PBI, the less is the accuracy of prediction.

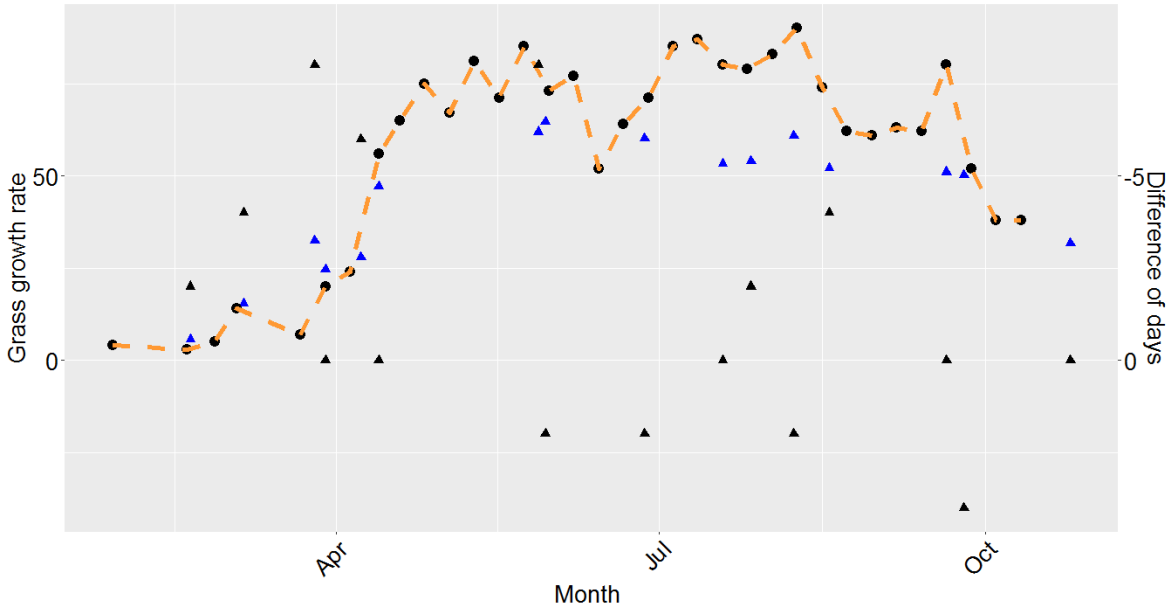


Figure 6-24 Grass growth curve for 2020 using Agmodel.
The grass growth rate from PBI is shown in a dotted orange curve. The predicted grass growth rate is shown as blue triangles, and the difference of days of acquisition between PBI and satellite data is represented in black triangles. The difference of days is shown on right hand axis.

The plot for predictions from Agmodel 1 and actual data from PBI for 2020 are shown in Figure 6-25. The model predicts the start of grass growing season well with predicted values closer to the actual values. There was no satellite data available in May for this farm, and it affected the overall prediction accuracy of the model. The accuracy of predictions is proportional to the difference in days between PBI and Sentinel 2.

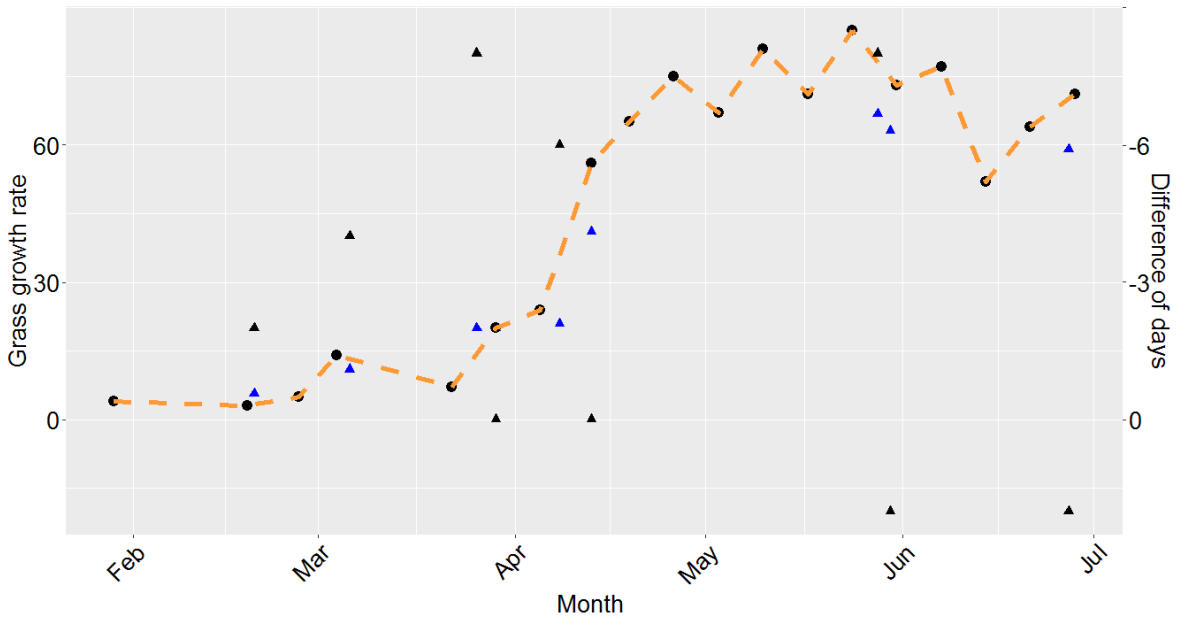


Figure 6-25 Grass growth curve for 2020 using Agmodel 1.
The grass growth rate from PBI is shown as an orange dotted curve. The predicted grass growth rates are shown as blue triangles. The black triangles represent the difference in days of acquisition between satellite and PBI data. The difference of days shown on right-hand axis.

The 2020 data from Agmodel 2 is shown in Figure 6-26. The actual grass growth rate values from PBI is shown using an orange dotted curve, and the modelled values are represented using blue triangles. The difference of days between PBI and Sentinel 2 is plotted on right axis. From July until August, the Agmodel 2 under-predicted the grass growth rate values. For October, the predictions are much accurate.

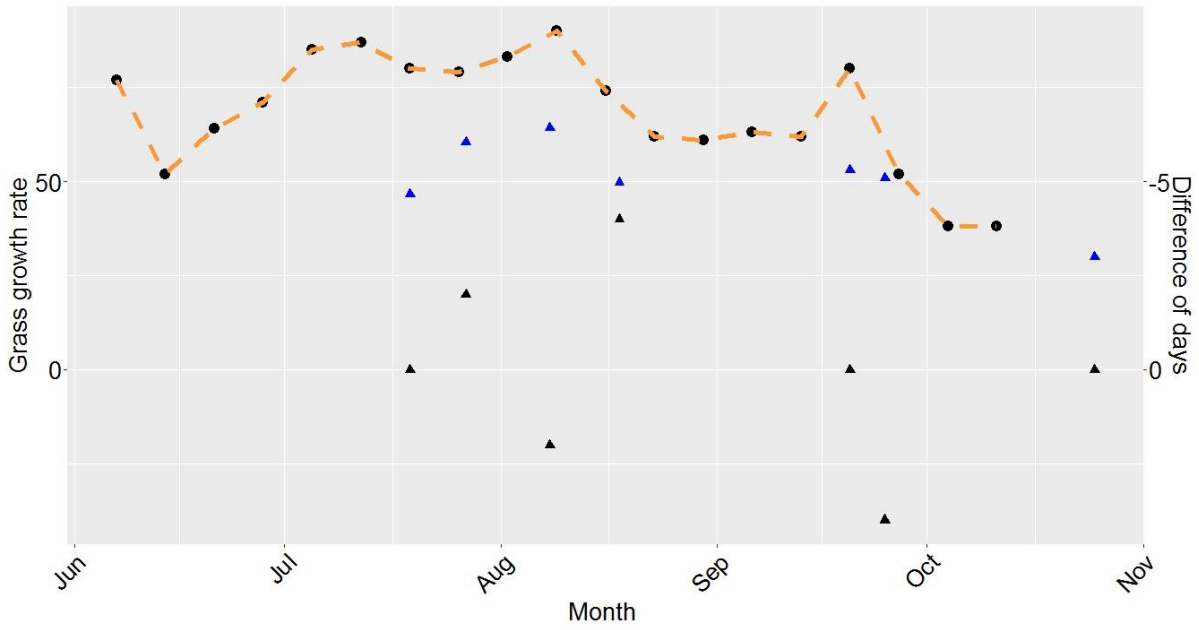


Figure 6-26 Grass growth curve for 2020 for Agmodel 2.
An orange dotted curve shows the PBI data. The difference of days is represented by black triangles and shown on right axis. The predicted grass growth rate values from Agmodel 2 are shown in blue triangles.

6.6.6 Variability in grass growth rate between individual paddocks

To analyse inter-paddock variability of grass growth rate, a subset of 20 farms were selected randomly from PBI. For all the farms, the grass growth rate from March 2019, which is the start of the grass growing season, was selected. The mean, standard deviation and median of grass growth rate values were calculated for a month. The mean values with standard deviations for each farm for March 2019 are shown in Figure 6-27. The figure not only shows the inter-paddock variability but also shows inter-farm variations for the same month.

Some farms had high standard deviations, such as farm number 495, 600, 6184, 6231, 6282, 6279, 6298, 6327, 6368, 70 and 919. There were four farms- 6235, 72, 74 and 973 with only one observation in March 2019. In March, the lowest mean grass growth rate was for farm number 132 (10 kg DM $ha^{-1}day^{-1}$), and the highest mean grass growth rate values were for farm number 6327 (40 kg DM $ha^{-1}day^{-1}$). There was a 30 kg DM $ha^{-1}day^{-1}$ difference between the lowest and highest grass growth rate values. These differences could be because of location, soil, weather and management.

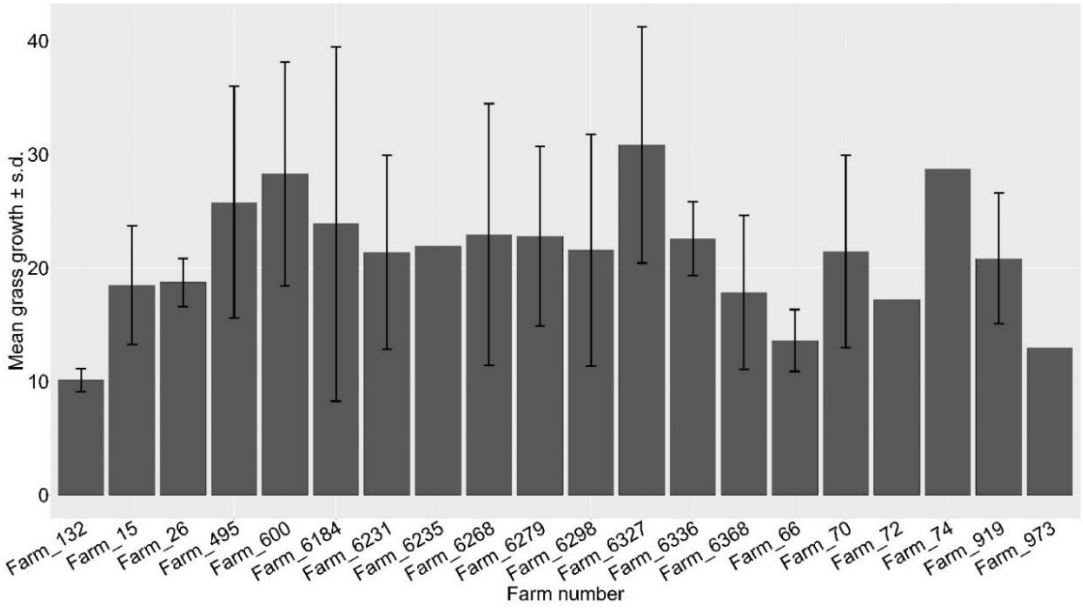


Figure 6-27 Grass growth variability in 20 farms selected from PBI.
The grass growth rate was for May in 2019. Error bars represent standard deviation.

The farm summary is shown in Table 6-6 with the area, average growth (kg DM $ha^{-1}day^{-1}$), stocking rate, start and end date of grazing, grazing season length and the peak growth date and value. There is a considerable variability in the area of the farms, the highest area is for farm number 6268 with 121.47 ha, and the lowest area is farm 600 with 21.94 ha. The average grass growth rate also varies on the farm, with the highest value of 78 kg DM $ha^{-1}day^{-1}$ (Farm 6235), and the lowest was 19 kg DM $ha^{-1}day^{-1}$ (Farm 973). Stocking rate is an essential factor on the farm, which can help to balance the feed supply and demand. Stocking rate means number of livestock units per hectare (LU ha^{-1}). Each farm had a different stocking rate. Farm 973 had the lowest stocking rate of 1.65 LU ha^{-1}, and the highest rate was 3.81 LU ha^{-1} for farm 6268. The grazing season was more than 300 for 5 farms- number 26, 6184, 919, 495 and 72, and for the rest of the farms, grazing season length was more than 229. The inter-farm differences in grass growth are due to differences in location, soil type, weather, and the start date of grazing, which can be observed in Table 6-6. It is important to note that Farm 6235 has a peak grass growth of 978.19 kg DM $ha^{-1}day^{-1}$ which is an example of human error while entering the data in PBI.

Table 6-6 Farm summary for the 20 randomly selected farms from PBI for 2019

Growth rate and peak growth in kg DM $ha^{-1}day^{-1}$. Stocking rate in LU/ha

Farm ID	Area (ha)	Average Growth	Stocking Rate	Start Grazing	Finish Grazing	Length (days)	Peak growth (date)	Peak growth
Farm_26	43.79	41	3.64	10-Feb	10-Dec	303	20-Jul	106
Farm_15	75.46	49	2.92	05-Feb	20-Nov	288	24-May	104
Farm_66	61.2	46	2.3	02-Feb	10-Nov	281	01-May	123
Farm_70	100.24	39	2.77	09-Feb	21-Nov	285	02-Jul	85
Farm_72	28.82	35	2.6	08-Jan	23-Dec	349	12-Aug	78
Farm_74	83.39	49	3.12	08-Feb	14-Nov	279	01-May	102
Farm_132	37.2	39	2.5	28-Feb	15-Nov	260	31-May	93
Farm_495	111.70	47	2.83	15-Jan	09-Dec	328	26-Jun	98
Farm_600	21.94	54	3.08	02-Feb	12-Nov	283	26-Jun	102
Farm_919	57.05	47	3.53	15-Jan	19-Nov	308	31-May	102
Farm_973	35.09	19	1.95	24-Jan	05-Nov	285	10-Mar	116
Farm_6184	75.73	44	3.03	08-Feb	09-Dec	304	22-May	97
Farm_6231	44.99	44	3.36	21-Jan	27-Oct	279	21-May	91
Farm_6235	54.25	78	3.41	01-Mar	14-Nov	258	01-May	**978.19**
Farm_6268	121.47	55	3.81	19-Mar	03-Nov	229	30-Jun	111
Farm_6279	58.94	48	2.96	06-Feb	04-Nov	271	27-Jun	169
Farm_6298	52.87	51	2.86	10-Feb	02-Nov	265	31-May	131
Farm_6327	46.12	28	2.99	04-Feb	05-Nov	274	20-May	94
Farm_6336	28.81	63	3.37	13-Feb	29-Oct	258	15-May	137
Farm_6368	98.99	44	2.76	10-Feb	12-Nov	275	17-Jul	104

6.6.7 Uncertainties & sources of model error

There can be several sources of uncertainties in the model, which are unavoidable. For this model, the only meteorological variables available were the rainfall data. There are over 500 rainfall stations in the country, whereas there are four manned and 20 automatic weather stations. To develop a national grass growth model for Ireland, it is optimal that weather stations are located in close proximity to the farms. In Chapter 5, a machine-learning model was developed using meteorological variables such as temperature, rainfall, global solar radiation, evaporation, soil moisture deficit, potential evapotranspiration and GDD. Temperature is more important in Ireland than rainfall. For national model, temperature data were not available and only rainfall data were available which could have led to uncertainties in the model.

This missing data can act as a source of error in the model. Ireland is cloudy for most of the days in a year, and Sentinel 2 used in this work can be affected by clouds and shadows. There was missing Sentinel 2 data because of clouds and shadows. The date of acquisition of satellite and ground grass growth data is not the same, which can be a source of error and ambiguity.

A significant limitation is the unavailability of the farm locations, the number of paddocks per farm and the farm boundaries. The satellite data for the farms were available as a vector. A hypothesis was made to get the paddocks with low, medium and high grass growth rate by dividing the NDVI vector into 3 quartiles- lower, middle and upper quartile. The NDVI values lower than the lower quartile were paddocks with low grass growth rate, and values higher than the third quartile were the highest grass growth rate paddocks. The medium grass growth rate values were between the first and third quartile.

The unavailability of farm information can produce uncertainty in the model. Knowing the farms' locations can help identify similar soil type, weather and management conditions, drainage, and paddock information. In Ireland, the Spring Rotation Planner (SRP) is a management tool discussed in Section 3.4.4, in which a

farm is divided into paddocks using a strip wire (Roche et al., 2017). Each paddock is grazed per week to ensure that sufficient grass is grazed according to the demand by livestock. The information about paddocks, such as area and grass growth rate, can help get a detailed model with paddock-level predictions. The paddock-level model can help allocate sufficient grass for grazing and allow other paddocks for grass regrowth.

6.6.8 Applications & future developments

An EO-based grass growth model could be used in grassland management DSS which can be used by the famers for feed budgeting strategies. The model developed covers a large geographical area in Ireland, covering four counties- Donegal, Galway, Sligo and Cork. These farms represent different enterprises in the country, a mix of soil types, systems, landscapes and topography. As more data is available from PBI and Sentinel 2, new farms can be added to the random forest training dataset by optimising hyperparameters of random forest. If the paddock grass growth rate is available, it can be incorporated into the model to give paddock-scale predictions. The MoSt GG model developed by Ruelle et al. (2018b) gives paddock-scale predictions, and to get the whole farm predictions, the paddock estimations have to be averaged. The model in this work can easily be scaled up to get paddock-scale predictions.

Machine-learning models are data-driven, relying on historical data for predictions, and their accuracy improves with a high number of input training data (Renault, 2019). The weather data availability such as temperature, rainfall, global solar radiation, evaporation, SMD and PE can help to improve the accuracy of the model. Internet of things (IoT) is a technology that can help collect weather data using proximal sensors on farms. IoT sensors can send the data wirelessly to phones or laptop, which can be used directly for analysis (Salam, 2020).

The synergistic use of Sentinel 1 and Sentinel 2 has been successfully used to discriminate pasture species in an active grazing farm using machine-learning models (Crabbe et al., 2020). The collaborative study helps to take advantage of each

sensor. Sentinel 1 can help capture textural information such as before and after grazing, which can affect the polarimetric scattering and Sentinel 2 capture the spectral information of the grass. Sentinel 1 can be included for future work, which could help identify defoliation events such as grazing and silage cutting.

6.6.9 Applications & future developments

An EO-based grass growth model could be used in grassland management DSS which can be used by the famers for feed budgeting strategies. The model developed covers a large geographical area in Ireland, covering four counties- Donegal, Galway, Sligo and Cork. These farms represent different enterprises in the country, a mix of soil types, systems, landscapes and topography. As more data is available from PBI and Sentinel 2, new farms can be added to the random forest training dataset by optimising hyperparameters of random forest. If the paddock grass growth rate is available, it can be incorporated into the model to give paddock-scale predictions. The MoSt GG model developed by Ruelle et al. (2018b) gives paddock-scale predictions, and to get the whole farm predictions, the paddock estimations have to be averaged. The model in this work can easily be scaled up to get paddock-scale predictions.

6.7 Conclusions

The national grass growth model can be used as a DSS to help farmers take on-farm decisions. Such models help to understand grass growth from a national perspective and manage grass in feed deficit. For example, extreme climatic conditions such as floods or drought can affect the grass growth rate leading to a feed deficit for livestock. For Irish grasslands, a machine-learning model to estimate grass growth exists, which was developed for two independent farms, which cannot be used as a national model.

This chapter developed a farm-scale national model using Random Forest to estimate the grass growth rate known as Agmodel. The model was developed using minimal inputs, i.e. descriptive statistics of rainfall and satellite data. To analyse the effect of spring and summer on the grass growth rate, the grass growth curve was divided into two parts to get two models. Agmodel 1 was developed from the rainfall and satellite data from January to June, and Agmodel 2 was used from July to December. The three models were validated using data from 2020. For farms across Ireland, Agmodel for first half of 2020 data had an RMSE of 17.27 kg DM $ha^{-1}day^{-1}$. Agmodel for second half of 2020 data had an RMSE of 16.08 kg DM $ha^{-1}day^{-1}$. Agmodel 1 had an RMSE of 17.83 kg DM $ha^{-1}day^{-1}$, whereas; Agmodel 2 had an RMSE of 16.53 kg DM $ha^{-1}day^{-1}$. Agmodel 1 performed better with R^2 of 0.70 than Agmodel 2 with R^2 of 0.36 because, during the first half of the grass-growing season, the grass growth rate is high and, as the weather is better than autumn and winter, more satellite images are available.

The strength of this model is that it is straightforward and more farms can easily be added to the existing dataset. The machine-leaning algorithms are data-dependent; therefore, as more data will be available in the future, such as from PBI and future satellite mission - Landsat 9 can easily be incorporated into this model (Li and Chen, 2020).

7

Chapter 7 Discussion

As the world population grows, there is increasing pressure on food growers to provide a greater volume of food. At the same time, there is a heightened awareness that industrialised agricultural production is economically and environmentally unsustainable. The wide availability of cheap and high-quality grass gives Ireland an advantage in sustainable food production economically. A key component in efficient utilisation of available grass is to increase the proportion of grass in the animals' diet. This involves first growing more grass and better utilising the grazed grass in a timely fashion or harvested as silage. To understand where there are surpluses or deficits in grass growth, farmers are encouraged to measure sward height to estimate available biomass regularly.

There are currently several methods of measuring grass, with various shortcomings. This thesis commenced with an in-depth review of pasture biomass estimation from *in-situ* on-farm methods commonly used by farmers to numerical models and EO-based remote sensing methods. It was evident from existing published literature that a combination of *in-situ* methods and EO imagery is now widely used for estimating grassland biomass at different scales (Wang et al., 2019, Zumo et al., 2021). EO-based methods using satellite imagery are commonly used, largely due to the strong relationship between vegetation indices and biomass and data availability. In the past, research has primarily had to use coarse spatial resolution imagery (typically MODIS with a ground sampling distance of 250 m). More studies that are recent have had access to higher spatial resolution data, including NASA's Landsat 8, but particularly ESA's Sentinel 2. The availability of higher spatial resolution of Sentinel 2 and its return period of 2-3 days at mid-latitudes has made detecting

grassland management interventions and pasture biomass estimation at paddock scale (<1 ha area) a possibility. At the same time, there have been ongoing developments in machine-learning algorithms that are suited to high dimensional data, such as time series EO and meteorological data.

This study examined the role of this new generation EO imagery for predicting grass growth nationally, comparing machine-learning regression models against conventional biophysical modelling methods. Previous research by Ali *et al*. (2017) showed an EO approach suited to Irish conditions. The current project built on this earlier research, experimenting with variables and algorithms over increasing scale, from a farm-scale model for Teagasc Research and commercial farms nationally to a broader model using a national network of farms providing data to the pasture management tool known as PBI.

In the wake of the extensive drought in 2018, an important element of the study was to understand the impact of extreme weather events on grass growth estimation. The summer of 2018 was characterised by lower than average rainfall and higher than average mean temperature, resulting in a high SMD. This resulted in a lower than average grass growth rate over the period, which is well illustrated in grass growth curves from the period. The analysis outlined in Chapter 4 indicated that in 2017 grass growth rate for Moorepark Farm, Co. Cork was significantly correlated with several meteorological variables but not rainfall. In 2018, growth was not significantly correlated with rainfall and SMD in correlation analyses. It is the combination of low rainfall and high temperature that led to low grass growth in 2018. Previous research has shown that rainfall in Ireland does not limit the grass growth rate (Green, 2019). However, in low rainfall/ high evapotranspiration periods, soils can become dry, creating water stress in plants. The soil at Moorepark is generally well-drained, which means soil water above the field capacity will drain quickly. Persistent periods of water stress will reduce grass growth rates, as will persistent excess soil moisture (Schulte et al., 2012).

The model developed in Chapter 4 found that in 2017 the GDD model explained 53% of the variability in grass growth, but in 2018 it only explained 36%. The model accounted for all meteorological variables likely to have an impact, and these data were both site-specific and timely as there is a meteorological station on-site. These results are very promising but suggest variables are missing from the model that could explain the remaining variability. This might include better characterisation of soil type between paddocks. There may be an error within the EO data, where non-grass reflectance contributes to vegetation index values. For example, road infrastructure between paddocks will lower NDVI values. The way pasture measurements were collected on the ground using an RPM may contribute to the unexplained variability. By only using a small sample of grass measurements aggregated by the paddock, there is no account for the spatial variation in growth that occurs within paddocks. It may also be that the linear regression model used does not adequately fit the data to the best model.

The ML algorithms are well suited to non-parametric, multidimensional data, and ensemble methods like RF are suitable because of their low bias, uncorrelated trees. GDD could be a useful variable in modelling growth nationally. However, detailed meteorological data for the majority of most Irish farms are limited. Currently, there are only 20 climatological stations in the Republic of Ireland. The problem of low number of climatological stations can be improved by building more stations or educating and enabling farmers to install on-site weather stations or sensors on farms nationally to collect detailed weather data for different soil types under different management systems.

A leading grass growth model, developed under Irish conditions and used widely in grassland research, is the Brereton model. The model predicts the grass growth using temperature, SMD, rainfall, and actual and potential evapotranspiration. In Chapter 5 it was observed that the Brereton model over-estimated growth during the phenological peak season (spring generally, but May in particular). The over-estimation could be because the Brereton model is based on converting solar radiation into dry matter without considering the leaf area index or leaf formation,

which contribute to the dry matter (Brereton and Keane, 1992). During the spring and summer, the amount of radiation is more than in autumn and winter and therefore, the Brereton model maybe be over-predicting. A previous comparative study by Hurtado-Uria et al. (2012) also reported that the Brereton model over-predicted grass growth. However, they found that all of the three models assessed (Brereton, Jouven and Johnson & Thornley) underperformed in Irish conditions. The issue is again the spatial resolution and distribution of climatological data in Ireland. Although there are vastly different management regimes, soil type, drainage and enterprise within a region, farms using the same climatological data from a local station will predict the same outcomes for all farms. The Brereton model also fails to account for farm management, how recently they might have occurred, or underlying environmental factors such as soil type. Data on soil type or the frequency or extent of management interventions were not available for the study to explain why the predicted rate differed from the actual rate. If these data were widely available, it could be incorporated into machine-learning models easily as encoded categorical data.

Where the Brereton model fails to represent continuous conditions on the ground, EO data can capture the effect (VI reflectance) of actual environmental, meteorological and management conditions on grass growth. As noted above, Sentinel 2 data has sufficient spatial resolution to detect variation in grass growth within and between paddocks or fields. Building on the previous research of Ali et al. (2017), ANFIS and RF machine-learning models were developed Sentinel 2 imagery and available meteorological data. It was found that RF using a range of meteorological data except rainfall was the best predictor of grass growth. This supported the findings of Hurtado-Uria et al. (2014), who also reported no strong relationship between rainfall and grass growth, but did find growth was correlated with temperature, solar radiation and evapotranspiration from January to March and from September to December. The research described in Chapter 5 explored the ability of Sentinel 2 and Landsat 8 imagery to estimate grass growth. With 30 m spatial resolution, the Landsat mission was the benchmark in moderate resolution, publicly-available EO data before 10 m resolution imagery became available when Sentinel 2A was launched in 2015. The machine-learning models developed using

Sentinel 2 imagery consistently outperformed those developed with Landsat 8 data, both for ANFIS and RF. The spatial resolution of Landsat 8 is a limiting factor, for example, where different land cover types are covered by a single pixel, therefore introducing error into the models. This was likely an issue where pixels cross the paddock boundaries. Another source of error is the lower temporal resolution. With a revisit time of 16 days, and the high likelihood cloud cover in Ireland on any given acquisition date, there is a greater chance that Landsat 8 does not accurately record grass growth, due to large temporal gaps arising from cloudy conditions on acquisition days.

The improved sensor specifications of Sentinel 2 improved the models. The additional red-edge bands in Sentinel 2 are reported to better characterise vegetation and biomass (Ramoelo et al., 2015). The 10 m spatial resolution of Sentinel 2 is much better at mapping paddocks than Landsat 8 as for every Landsat pixel, Sentinel 2 has 9 corresponding pixels. This is critical for mapping grassland in Ireland, where fields are generally small (average field size is ~2.5 ha) and irregularly shaped. The higher resolution makes it easier to demarcate and mitigate mixed pixels at field boundaries. The variation in grass growth rates due to the changes in management regimes between paddocks is challenging to monitor using Landsat 8, for example, where a pixel captures both high grass cover in one paddock and low cover in an adjacent paddock. However, Sentinel 2 is missing a thermal band that makes cloud detection more challenging and potentially introduces some error into Sentinel data where clouds are not completely removed (Tarrio et al., 2020).

The lowest RMSE attained by RF in initial tests on single farm data and incrementally increased with additional farms made it the algorithm of choice for developing a national model. By exclusively using remotely sensed or data, the number of training sites was increased to 179 representative farms using PBI data. This was the maximum number of sites available for the study. At the time of the experiment increasing the number of farms could significantly improve model accuracy, particularly in years that don’t have extreme weather. However, as more farm data were added to the model, there was a concomitant reduction in the

availability of meteorological data. As noted previously, there are only 20 climatological stations distributed around Ireland, so the national model could not be built with the same range of meteorological variables. In the national model, both DOY and season impact the grass growth (Edirisinghe et al., 2012). The relationship between the DOY and biomass can affect the surface reflectance and, in turn, vegetation indices. The high temperatures and solar radiation result in greater grass growth and the corresponding NDVI values will also be high. In winter, soil temperatures are low and likely wet, leading to low grass growth and the corresponding NDVI values will be low. The amount of biomass determines management strategies, influencing grazing rotation or silage cutting schedules. Despite the differences in R^2 and the variability in outcomes using different subsets of data, RMSE was less than 18 kg DM $ha^{-1}day^{-1}$ for all models.

A national model of grass growth rate is very desirable for Irish farmers. Variable meteorological conditions and soil conditions, farm enterprise, and management conditions change rapidly from farm to farm. The heterogeneity in all these factors and the ability to measure them accurately is likely to be a considerable source of error in the models developed within this study. The sensors installed in the farms to measure meteorological data can give accurate measurements and help to improve the model. If there are more climatological stations, they could provide more meteorological data and incorporate it into the model. On certain soils, the effect of rainfall will be significant depending on the soil drainage. Soils with impeded drainage will respond differently to well and moderate drained soils during heavy or persistent rainfall. Fitzgerald et al. (2008) discussed how water deficit and excess water could limit grass growth. In excess water conditions, well and moderately drained soils performed better than the poorly drained soils, which can experience persistent waterlogging (Schulte et al., 2012). Water stress situations due to high temperatures and low rainfall can limit the grass growth for well-drained soils as their water holding capacity is lower than the poorly drained soil. A range of farms with varying soil type in Irelands and the effect of rainfall can affect the grass growth differently.

Another obstacle to modelling grass growth was data unavailability, both for meteorological data and EO imagery. The first limitation was the unavailability of

meteorological data in Chapter 6 for the national model. Without the correct data on, for example, soil temperature and evapotranspiration, it is not possible to estimate growing degree-days accurately for farms nationally. Consequently, only rainfall was available for modelling. For example, this research and previous research, for example Hurtado-Uria et al. (2014), found rainfall is not an important meteorological variable for grass growth rate as temperature and evapotranspiration are. These variables were present in the models discussed in Chapter 5 where they were the most significantly model inputs. For the national model, these factors are missing, which is very likely to have lowered model accuracy. However, another limitation that cannot be avoided is a systematic error introduced by farmers while collecting grass height data using RPM. These errors can be propagated into the model leading to low accuracy. This was compounded by the fact that national data were anonymised, making it difficult to assess potential local issues impacting accuracy.

PBI is a useful resource but could be better used. All models consistently had a variability that was not explained by the available variables. This may be because of interventions such as grazing, silage cutting, or environmental factors such as soil type and soil drainage. For example, a certain level of management data, for example, grazing and silage cutting, is available in PBI. If this level of data could be better integrated into machine-learning models, it might mitigate the shortage of meteorological data. However, the necessity to anonymise the data is a further complication that also needs to be addressed,

The results in this thesis highlighted the importance of the factors governing the grass growth in Ireland-climate and EO data. When this PhD was started, there was not much data available in PBI. Furthermore, the meteorological data were available for only a few farm sites. This research highlights the issues caused by data sparsity with regards to accurately modelling grass growth. However, it also shows the potential of newly available data sources, such as Sentinel 2 satellite imagery through the Copernicus platform (which provides free images every with a revisit time as short as five days). These data have an immense potential improve national models. Currently, there are 3500 farms registered with PBI, which can be included

into the model. The modelling process can be automated by automatic downloading of satellite images, pre-processing and integrating them into the model to estimate grass growth and detect the management events. Such approach can help farmers by transforming the raw data into information, which is valuable to the farmers.

8

Chapter 8 Conclusion & Future Research

This research aimed to investigate whether machine-learning regression of optical EO satellite imagery in conjunction with meteorological data could accurately estimate grass growth rate across the whole of the Republic of Ireland. Previous research by Ali (2016) suggested that machine-learning models with vegetation indices from MODIS data can estimate the grass growth rate for Irish grasslands. The models were developed for two independent sites - the Teagasc research farms in Moorepark in Co. Cork and Grange in Co. Meath. It was suggested that the satellite data with machine-learning is a promising tool for grass biomass modelling. Available EO data has been growing since then with the launch of Sentinel 2A in 2015 and Sentinel 2B in 2017

The work presented in this thesis was aimed at developing a national model with Sentinel 2 and meteorological data to estimate grass growth rate with three specific objectives:

(i) Evaluate the role of growing degree-days (GDD) and climate data in simple grass growth rate models at a farm scale for one farm-

(ii) Compare the performance of the conventional grass growth model estimating grass growth rate at 8 farm locations against RF and ANFIS regression models and

(iii) Transfer the machine-learning model developed in (ii) above to a broader national scale using 179 representative farms which contribute grass measurements to PBI and explore the importance of spatial variation across Ireland

Each of these objectives was carried out at an increasing scale and used different combinations of data. The following chapter examines the outcomes of these

individual experiments and how collectively they have furthered our understanding of the strengths, opportunities and weakness of EO-based pasture biomass estimation in Ireland.

The work started with exploring the influence of meteorological datasets on grass growth rate. A Pearson correlation analysis was performed between grass growth data and meteorological including standard and modified GDD for Moorepark, County Cork for 2017 and 2018. The statistically significant variables from correlation analysis were used for ordinary least square regression for 2017 and 2018, respectively. The results were relatively poor, showing an R^2 of 0.53 in 2017 but did confirm that a range of meteorological data were significant even in simple linear models of Irish grass growth

The GDD terms were generally more important in the model output than individual meteorological data such as evaporation and PE. However, the models only explained half of the variability in growth rate over the two years, highlighting the importance of other factors such as management (recalling that Moorepark is a research farm and, as such, is subject to experimental changes). This model attempted to include the effect of extreme weather conditions such as high temperature and SMD in 2018, which led to fodder crisis due to low grass growth rates. The major limitation of this model is that it is based just on meteorological data is not enough to model the grass growth rate on intensively managed grasslands with variability between paddocks. The other factors such as spatial variation, soil variability, management on farms and actual environmental conditions are missing from the model, which might help to explain the rest of the variability in grass growth rate.

Another empirical model, which uses only meteorological data for grass growth estimation is the Brereton model. Brereton model was developed for a number of farms in Ireland. The relatively poor performance can be ascribed to the recording of meteorological data at sites distant from the grass measurements and the absence of soil and management terms in the Brereton model.

The Brereton model was compared to an initial machine-learning approach on the same data but now including EO data and meteorological data. The EO observation is essentially an empirical term that captures the local growing conditions (soil type and management). The inputs to ANFIS and RF were vegetation indices (Landsat 8 NDVI, Sentinel 2 NDVI and Sentinel 2 NDRE), modified GDD and meteorological data. The ANFIS model performed poorly for all the farms but better than the Brereton Model. The RF model using Sentinel 2 data performed the best. Better results were obtained using Sentinel 2 vegetation indices than Landsat 8. This is because of the low revisit time of Landsat 8 (16 days) compared to Sentinel 2 (5 days), which leads to the lower number of images acquisitions and a longer gap between acquisition and grass measurement. Another reason that Landsat 8 had a bit lower spatial (30 m) resolution than Sentinel 2 (10-20 m). During the study, farms with a higher number of satellite images estimated growth better than those farms with fewer points. For example, growth estimation at Johnstown Castle (34 Sentinel 2 images for 2017 and 2018) was better than that at Moorepark (19 Sentinel 2 images for 2017 and 2018) visually, as discussed in Section 5.4.5. Overall, the machine-learning algorithms performed better than the Brereton model. The random forest model from this study was further used to develop a national model for Ireland.

The national model was developed using RF with Sentinel 2 NDVI and NDRE along with rainfall. However, GDPR regulations required anonymising the farm data, so satellite data were provided without x, y coordinates. The absence of location data for these farms affected the final analysis as the data such as location, soil type, farm type, number of paddocks and management information could not be included but could be assumed to have improved the outcome. The random forest inputs were daily rainfall, 3-, 5- and 7- day rainfall accumulations, descriptive statistics of individual Sentinel 2 bands, NDVI and NDRE. When moving to national model, there was a major limitation of data unavailability. Only rainfall data were available for the farms in national model.

On comparing Agmodel 1 and Agmodel for the data from January until June, it was found that both the models performed equally well with RMSE of 17.83 and 17.27 kg DM ha^{-1} day^{-1} respectively. Similarly, Agmodel 2 and Agmodel with second half of the year data performed equally well with RMSE of 16.53 and 16.08 kg DM ha^{-1} day^{-1}.

Overall, Agmodel 1 and Agmodel for first half performed better than the Agmodel 2 and Agmodel for the second half as the number of data points are more than the autumn and winter period because the grass growth rate is high during the summer. A likely reason for poor predictive ability in winter and autumn is that grass growth and the sward density is lower than in spring/summer. Therefore, the model estimated poorly during these months. The study demonstrates the importance of location in grass growth. To understand grass growth, the observation data for that particular place and time is needed. It also highlights the importance of high-quality ground truth in machine-learning EO analysis as often in the literature. This was a relatively data-rich study compared with much of the literature with weekly data from 179 farms over 3 years, and yet the model could explain only 70% of the variance, and this is partly due because the 2017-19 ground truth did not (could not) capture all the potential variation in grass productivity in 2020. This issue will increase if it is intended to increase the model output to a wall-to-wall national weekly grass growth estimate. In the following section, possible research avenues to potentially expand and improve these estimates are discussed.

8.2 Future work

This research has demonstrated the potential in using EO and meteorological data for estimating grass growth in Ireland. While there is currently widespread availability of suitable EO data through the Sentinel 2 mission, some developments are still required before grass growth can be routinely estimated for operational application on Irish farms.

The key to accurate prediction of growth in the development of these machine-learning models is large volumes of cloud-free data and from the right time of the year. The cloudy data can cause gaps in the satellite data acquisition and lead to uncertainties in the model as discussed in Section 5.4.6. With the launch of new satellites in the coming years, the potential to model grass biomass will continue to increase with high accuracy. This can be supplemented with imagery from other sources, UAV or proximal sensors attached to vehicles or fixed within the field. Precision agriculture (PA) technologies continue to develop and are integrated into decision support systems to help farmers optimize their systems and maximize their resources (Higgins et al., 2019). The PA technologies involve integrating multiple sensors for monitoring grass and livestock grazing, which can help farmers manage the farm efficiently.

8.2.1 Upcoming satellites

The machine-learning models are data-driven, which means that their accuracy increases with the increase in data availability (Reichstein et al., 2019). With the launch of new satellites, more EO data will be available in the future, which can meet the growing demand for data. The major drawback for optical data in grassland management is unreliability due to cloud contamination. A method to solve this issue is the increased temporal frequency of the available satellites.

Landsat 9 will be launched as a joint programme of NASA and USGS with two sensors similar to Landsat 8 - Operational Land Imager 2 (OLI-2) imaging in visible, near, and shortwave infrared region of the electromagnetic spectrum, and the

Thermal Imaging Sensor 2 (TIRS-2) for thermal imaging providing 8-day global revisit cycle (Masek et al., 2020). The temporal resolution plays an important role, especially in countries with cloud cover issues, by providing more images per month. With the launch of Landsat 9, the revisit time will be reduced to 2.3 days.

Hyperspectral sensors can provide optimal spatial-spectral and temporal resolutions, which can be helpful for grassland applications as they provide narrow continuous spectral bands. A hyperspectral satellite known as Hyperspectral Precursor and Application Mission (PRISMA) was launched in March 2019 by an Italian space agency. It has a spatial resolution of 30 and 12nm spectral resolution from 400-2500 nm. There are some satellite data not available yet, but their simulated bands are used for vegetation applications. A satellite known as EnMAP (environmental mapping and analysis program) will be launched in 2021/2022 by the German space industry. It will have 230 spectral bands from 420-2450 nm of wavelength and a spatial resolution of 30m. It will have a repeat cycle of 4 days, which will be helpful in frequent grassland modelling. Another similar sensor is NASA's Hyperspectral Infrared Imager (HyspIRI) instrument covering spectral wavelength from 380-2510nm providing visible to infrared data at 30m resolution and thermal data at 60m resolution. There has been some work using simulated EnMAP and HyspIRI data. Sibanda et al. (2019) simulated the spectral response of EnMAP and HyspIRI sensors using a ground-based hyperspectral sensor to understand the water content in grasslands under different fertilizer regimes using regression models. The hyperspectral data can be used to derive narrow-band vegetation indices, which can be helpful in applications such as biomass estimation (Kong et al., 2019) and plant diversity mapping (Peng et al., 2019).

Another way is to use commercially available high temporal and spatial resolution microsatellite, which provides images daily. Recently there is a growing trend of commercially available satellite constellations consisting of micro, nano and cube satellites (Farrag et al., 2019). Micro, Nano and Cube satellite are small satellite with less than 500 kg weight. The main advantages are that they consist of three or more satellites imaging the same target from different angles simultaneously and are of

low launching cost and high resolution. Radkowski et al. (2021) presented the preliminary results with the Plantelab's Dove satellite images available at spatial resolution from 3-5 to map changes in managed grazed areas.

8.2.2 Unmanned aerial vehicle (UAV)

In this study, Sentinel 2 was used which has 10 m spatial resolution, whereas the ground data are very high-resolution point data. The use of UAV at very high resolution can match better with the ground data than satellite data. The increasing availability of unmanned aerial vehicles (UAV) systems provides an opportunity to monitor grassland farms frequently and with precise data. The UAV's can captured data from visible to infrared region to estimate grass height. Michez et al. (2019) used UAV to get 3D information- height and spectral information to monitor grazing on a farm. UAV can also monitor the pasture quality, in theory, which helps in targeted fertilization for paddocks with low grass growth.

The farmers can be trained to use the UAV on their farms and can monitor their farms regularly. The data can be transferred to cloud storage for analysis. However, there are few limitations to the use of UAV in Ireland. The regulation of maximum flying altitude from the Irish Aviation Authority (IAA) has hindered the use of UAVs on farms. The UAVs are limited by the windy and rainy conditions and by battery power.

Another type of UAV technology is the swarm UAV, a state of the art method involving a set of drones that work together and synchronise with other UAVs in the swarm (Carbone et al., 2018). The swarm UAVs can help to monitor large areas with precision. The concept of the swarm is inspired by nature, such as the flying pattern of birds and ant colonies. It can reduce the time to collect and reduce labour, which can play an important role in developing the national model. The swarm UAV is a promising technology but can be limited by cost and the time to deploy the drones. The number of drones for a swarm will depend on the cost, maintenance factor, and application type.

8.2.3 Proximal sensors

As discussed in Section 5.4.4, there can be some uncertainties in the ground data from PBI such as error in RPM, operator error, error while entering the data in PBI. One solution for this could be the use of proximal sensors on farms. The on-farm sensors can also reduce the uncertainties in the model due to the unavailability of meteorological data as discussed in Section 6.6.6. Meteorological data can be collected on farm such as temperature, solar radiation, potential evapotranspiration and evaporation. The proximal sensors are installed in the farms close to the object being sensed and can be fixed or on a moving vehicle as part of the so-called "Internet of Things" (IoT) (Sanjeevi et al., 2020). The automation of the farm can help in improved data collection, storage and analysis. The IoT can help farmers in managing their resources efficiently and improve productivity. The IoT system involves a wireless sensor network to monitor temperature, precipitation, humidity, and soil.

There can be various types of sensors such as optical, electrical and sonic. Such sensors can provide paddock-scale information and can help to identify inter-paddock grass variations. An example of a commercial internet of things approach is provided by the Irish Company Anuland (anuland.ie). They have many sensors collecting the data for soil and using a video camera to measure grass growth, transmitting the data to the cloud for processing. The data can be viewed on an app to help the farmers.

Another common sensor in development is a smartphone for grassland monitoring. Skovsen et al. (2019) developed a dataset using three cameras mounted on different platforms- two static platforms for plot trials and a moving platform for the field level image acquisition. A deep neural network segmentation algorithm was trained using synthetic images to classify into five labels- grass, clover (white and red), weed and soil. These models were further used on real images collected in the field to predict biomass.

Multiple proximal sensors on a farm can be connected and automated, known as the 'Internet of Things' (IoT) (Sanjeevi et al., 2020). The automation of the farm can help in improved data collection, storage and analysis. The IoT can help farmers in managing their resources efficiently and improve productivity. The IoT system involves a wireless sensor network to monitor temperature, precipitation, humidity, and soil.

The methods discussed are more efficient than the destructive method- clipping and weighing. The cameras can be deployed on smartphones and UAVs to capture demand-basis images to take field measurements. The images can be added to the machine-learning or deep learning models to relate the biophysical parameters and the biomass.

8.2.4 Synthetic Aperture Radar

Cloud-cover is a serious problem while using optical satellite data in Ireland as discussed in Section 5.4.6 and 6.6.6. The synergistic approach using both synthetic aperture radar (SAR) and optical data should be explored. SAR satellites (such as sentinel 1) can penetrate through the cloud but get affected by the background soil information and interfere with the grass growth data. With the increasing altitude, the SAR data is less accurate because of layover and shadow effects. The combination of both the sensors' data can overcome each of their limitations. SAR and optical satellites provide complementary information which can be used for grass biomass estimation (Wang et al., 2019). The combination of both the satellite data will allow frequent grass growth estimations (Ali et al., 2017a).

A project called GrassQ[4] by Teagasc and Maynooth University used a multi-sensor approach combining satellite imagery, UAV, ground and weather data (Murphy et al., 2019). GrassQ is a decision support system that provides real-time management information to farmers and researchers. The ground data were recorded using a rising plate meter (RPM) and near-infrared spectroscopy (NIRS). The multispectral images were captured using UAV and from Sentinel 2.

[4] *www.grassq.com.* Accessed 25th June 2021

8.2.5 Combination of biophysical models with machine-learning

Grass biophysical simulation models use physiological processes such as conversion of solar radiation into dry matter and capture the interaction between weather, grass and soil, and an example is the MoST model developed by Teagasc. The biophysical models are unable to model extreme weather conditions such as heat stress. These models also oversimplify the grass growth estimations over large areas with similar conditions such as soil type and weather, affecting the accuracy of estimation.

A hybrid approach can overcome the limitations of both crop model and remote sensing methods and improve accuracy. Recently, the use of a hybrid approach combining biophysical simulation models and machine-learning models has increased. For example, (Feng et al., 2020) used a hybrid model by integrating Agricultural Production System sIMulator (APSIM), climate data, NDVI from MODIS to estimate wheat yield. The satellite observations can help the actual conditions on the ground, which can act as a correction factor to the biophysical simulation models.

For grass growth estimation, the scale and frequency of measurement play an important role in deciding which sensor platform to use - satellite, aerial, UAV, proximal sensors, or the multi-sensor approach. The methods discussed above vary in spectral and spatial resolution. If the grass growth model is needed at paddock-scale, then proximal sensors and UAV with the high spatial and temporal resolution is required. The farm-level grass modelling needs frequent but medium-resolution imagery. In connecting Irish grasslands, a multi-sensor approach could help and improve the modelling accuracy because of the cloud cover issues. The upcoming new spaceborne can ensure the data availability, and the model could be trained for more datasets from and additional year's data could be included in the model. The on-farm sensors such as meteorological data and NDVI Sensors could also be used to get a high temporal resolution dataset to train the model.

The data assimilation techniques to include the EO data into the biophysical simulation models allow to improve the model and include the spatial information into the model. When the EO data is affected by clouds, the simulation model will give the estimations, and when the images will be available, they will be incorporated into the model. Apart from the optical data, the Sentinel 1 data can be incorporated into the model in the form of backscatter ratios and radar index, which can be useful for grass biomass and cutting detection (Holtgrave et al., 2020).

The PBI data could be modified to include the farm and paddock boundaries. The data retrieval could be improved as now a farm can be searched using the farmer's name and county. The overall farm data such as grass growth rate and grass cover cannot be exported, as there is no option available. The interface could be improved to include the location and number of paddocks and their information. 3500 farms on PBI are measuring grass every week. Many farmers are not measuring the grass growth on their farm and rely on their memory for decision-making. The farmers daily need to measure grass, calculate the demand according to the livestock. They need to utilise the amount of grass grown on their farms and plan forward for any unanticipated extreme climatic conditions. Grass measurement daily helps to identify surplus and deficit on paddocks, and timely action can be taken in such events. However, walking the farm weekly and uploading the data into a decision support system software can be daunting.

The management strategies need to be changed according to the changing farm conditions. There is a gap in understanding of the grass growth using the models with meteorological data. The EO data has the potential to fill that gap by providing valuable spatial and temporal information about grass growth and actual conditions on the ground, such as grazing, cutting and fertilizer application. The satellite data provides synoptic data covering a large geographical area. The temporal resolution of satellite data allows us to model the seasonal variation of grass growth due to climate and management factors.

This study's final RF regression model used EO and meteorological data for 179 farms nationally to estimate grass growth rate. The RMSE in grass growth estimation for the validation data using the year 2020 was 16-17 kg DM $ha^{-1}day^{-1}$ with peak growth rates of up to 110 kg DM $ha^{-1}day^{-1}$.

More satellite data and more farms in the model will lead to an improved accuracy model. The addition of paddock information can also give paddock-scale estimates available in PBI but restricted access. The national model developed can act as a decision support system for farmers. A grass wedge using satellite estimates can help give the best and worst performing paddocks or farms that can help farmers make decisions and manage their livestock sustainably. Moreover, due to changing climate conditions (extreme weather conditions), grass productivity is affected, which can be monitored using grass growth models. The model presents an opportunity to provide grassland spatial and temporal information, which can be valuable for national policy development.

References

ABBERTON, M., CONANT, R. & BATELLO, C. 2009. Grassland and carbon sequestration: management, policy and economics.

AGHDAM, I. N., PRADHAN, B. & PANAHI, M. 2017. Landslide susceptibility assessment using a novel hybrid model of statistical bivariate methods (FR and WOE) and adaptive neuro-fuzzy inference system (ANFIS) at southern Zagros Mountains in Iran. *Environmental Earth Sciences,* 76.

AHMAD, S., ABBAS, Q., ABBAS, G., FATIMA, Z., ATIQUE UR, R., NAZ, S., YOUNIS, H., KHAN, R. J., NASIM, W., HABIB UR REHMAN, M., AHMAD, A., RASUL, G., KHAN, M. A. & HASANUZZAMAN, M. 2017. Quantification of Climate Warming and Crop Management Impacts on Cotton Phenology. *Plants (Basel),* 6.

ALI, I. 2016. *Retrieval of grassland biophysical parameters using multi-temporal optical and radar satellite data* Doctor of Philosophy University College Cork.

ALI, I., BARRETT, B., CAWKWELL, F., GREEN, S., DWYER, E. & NEUMANN, M. 2017a. 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**,** 3225-3231.

ALI, I., CAWKWELL, F., DWYER, E. & GREEN, S. 2017b. 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**,** 3254-3264.

AMIES, A. C., DYMOND, J. R., SHEPHERD, J. D., PAIRMAN, D., HOOGENDOORN, C., SABETIZADE, M. & BELLISS, S. E. 2021. National Mapping of New Zealand Pasture Productivity Using Temporal Sentinel-2 Data. *Remote Sensing,* 13.

ANDERSSON, K., TROTTER, M., ROBSON, A., SCHNEIDER, D., FRIZELL, L., SAINT, A., LAMB, D. & BLORE, C. 2017. Estimating pasture biomass with active optical sensors. *Advances in Animal Biosciences,* 8**,** 754-757.

ARNTZEN, J. W., ABRAHAMS, C., MEILINK, W. R. M., IOSIF, R. & ZUIDERWIJK, A. 2017. Amphibian decline, pond loss and reduced population connectivity under agricultural intensification over a 38 year period. *Biodiversity and Conservation,* 26**,** 1411-1430.

ASLAM, M. A., AHMED, M., STÖCKLE, C. O., HIGGINS, S. S., HASSAN, F. U. & HAYAT, R. 2017. Can Growing Degree Days and Photoperiod Predict Spring Wheat Phenology? *Frontiers in Environmental Science,* 5.

ATTARAN, M. & DEB, P. 2018. Machine Learning: The New 'Big Thing' for Competitive Advantage. *International Journal of Knowledge Engineering and Data Mining,* 5.

BAEZA, S. & PARUELO, J. M. 2020. Land use/land cover change (2000–2014) in the Rio de la Plata grasslands: an analysis based on MODIS NDVI time series. *Remote Sensing,* 12**,** 381.

BALDI, A. & GOTTARDO, D. 2017. Livestock production to feed the planet. Animal Protein: A Forecast of Global Demand Over the Next Years. *Relations,* 5.

BARNETSON, J., PHINN, S. & SCARTH, P. 2020. Estimating Plant Pasture Biomass and Quality from UAV Imaging across Queensland's Rangelands. *AgriEngineering,* 2**,** 523-543.

BARRETT, B., NITZE, I., GREEN, S. & CAWKWELL, F. 2014. Assessment of multi-temporal, multi-sensor radar and ancillary spatial data for grasslands monitoring in Ireland using machine learning approaches. *Remote Sensing of Environment,* 152**,** 109-124.

BARRETT, P., LAIDLAW, A. & MAYNE, C. 2004. An evaluation of selected perennial ryegrass growth models for development and integration into a pasture management decision support system. *The Journal of Agricultural Science,* 142**,** 327.

BARRETT, P. D., LAIDLAW, A. S. & 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**,** 37-56.

BELLOCCHI, G., RIVINGTON, M., DONATELLI, M. & MATTHEWS, K. 2010. Validation of biophysical models: issues and methodologies. A review. *Agronomy for Sustainable Development,* 30**,** 109-130.

BENGTSSON, J., BULLOCK, J. M., EGOH, B., EVERSON, C., E., T. O'CONNOR, T., P.J., O. F., H.G., S. & LINDBORG, R. 2018. Grasslands—more important for ecosystem services than you might think. *Ecosphere,* 10.

BERGER, C., LUX, H., URBAN, M., SCHMULLIUS, C., BAADE, J., THIEL, C., WIGLEY-COETSEE, C. & SMIT, I. 2020. Annual Grass Biomass Mapping with Landsat-8 and Sentinel-2 Data Over Kruger National Park, South Africa. *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium.*

BOSSIO, D. A., COOK-PATTON, S. C., ELLIS, P. W., FARGIONE, J., SANDERMAN, J., P.SMITH, WOOD, S., ZOMER, R. J., UNGER, M. V., EMMER, I. M. & GRISCOM, B. W. 2020. The role of soil carbon in natural climate solutions. *Nature Sustainability,* 3.

BREIMAN, L. 2001. Random forests. *Machine learning,* 45**,** 5-32.

BRERETON, A., DANIELOV, S. & SCOTT, D. 1996a. Agrometeorology of grass and grasslands for middle latitudes.

BRERETON, A. & KEANE, T. 1992. The impact of weather on grassland farming. *Irish Farming, Weather and Environment***,** 125-135.

BRERETON, A. J., DANIELOV, S. A. & SCOTT, D. 1996b. *Agrometeorology of grass and grasslands for middle latitudes*, Geneva: World Meteorological Organisation.

BURKE, W. 1968. Growing Degree-Days in Ireland. *Irish Journal of Agricultural Research,* 7**,** 61-71.

CAI, J., LUO, J., WANG, S. & YANG, S. 2018. Feature selection in machine learning: A new perspective. *Neurocomputing,* 300**,** 70-79.

CALVACHE, I., BALOCCHI, O., ALONSO, M., KEIM, J. P. & 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.

CARBONE, C., GARIBALDI, O. & KURT, Z. 2018. Swarm robotics as a solution to crops inspection for precision agriculture.

CHARLTON, K. E. 2016. Food security, food systems and food sovereignty in the 21st century: A new paradigm required to meet Sustainable Development Goals. *Nutrition & Dietetics,* 73**,** 3-12.

CHEN, Y., GUERSCHMAN, J., SHENDRYK, Y., HENRY, D. & HARRISON, M. T. 2021. Estimating Pasture Biomass Using Sentinel-2 Imagery and Machine Learning. *Remote Sensing,* 13.

CHEN, Y., ZHANG, Z., TAO, F., PALOSUO, T. & RÖTTER, R. P. 2018. Impacts of heat stress on leaf area index and growth duration of winter wheat in the North China Plain. *Field Crops Research,* 222**,** 230-237.

CLEMENTINI, C., POMENTE, A., LATINI, D., KANAMARU, H., VUOLO, M. R., HEUREUX, A., FUJISAWA, M., SCHIAVON, G. & DEL FRATE, F. 2020. Long-Term Grass Biomass Estimation of Pastures from Satellite Data. *Remote Sensing,* 12.

CLEVERS, J., VAN DER HEIJDEN, G., VERZAKOV, S. & SCHAEPMAN, M. E. 2007. Estimating grassland biomass using SVM band shaving of hyperspectral data. *Photogrammetric Engineering & Remote Sensing,* 73**,** 1141-1148.

COMMISSION, E. 2020. Farm to fork strategy: for a fair, healthy and environmentally-friendly food system. *DG SANTE/Unit 'Food information and composition, food waste''*.

CONANT, R. T. 2010. Challenges and opportunities for carbon sequestration in grassland systems. *Integrated Crop Management.* Food and Agriculture Organization of the United Nations.

CRABBE, R. A., LAMB, D. & EDWARDS, C. 2020. Discrimination of species composition types of a grazed pasture landscape using Sentinel-1 and Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation,* 84.

CRABBE, R. A., LAMB, D. W., EDWARDS, C., ANDERSSON, K. & SCHNEIDER, D. 2019. A Preliminary Investigation of the Potential of Sentinel-1 Radar to Estimate Pasture Biomass in a Grazed, Native Pasture Landscape. *Remote Sensing,* 11.

CSO 2012. Census of agriculture 2010-final results.

CULLEN, B. R., ECKARD, R. J., CALLOW, M. N., JOHNSON, I. R., CHAPMAN, D. F., RAWNSLEY, R. P., GARCIA, S. C., WHITE, T. & SNOW, V. O. 2008. Simulating pasture growth rates in Australian and New Zealand grazing systems. *Australian Journal of Agricultural Research,* 59.

DAFM 2020. Annual Review and Outlook for Agriculture, Food and the Marine 2020. Economics and Planning Division.

DALLEY, D. E. & GEDDES, T. Pasture growth and quality on Southland and Otago dairy farms Proceedings of the New Zealand Grassland Association, 2012. 237-242.

DASSELAAR, A. V. D. P.-V., VLIEGHER, A. D., HENNESSY, D. & ISSELSTEIN, J. 2017. Grazing in a high-tech world. *Proceedings 5th Meeting EGF Working Group "Grazing" in Trondheim.* Wageningen: Wageningen University.

DE ROSA, D., BASSO, B., FASIOLO, M., FRIEDL, J., FULKERSON, B., GRACE, P. R. & 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.

DE VROEY, M., RADOUX, J. & DEFOURNY, P. 2021. Grassland Mowing Detection Using Sentinel-1 Time Series: Potential and Limitations. *Remote Sensing,* 13.

DELAUNE, P. B., MUBVUMBA, P., FAN, Y. & BEVERS, S. 2020. Cover crop impact on irrigated cotton yield and net return in the southern Great Plains. *Agronomy Journal,* 112**,** 1049-1056.

DEMANET, R., MORA, M. L., HERRERA, M. A., MIRANDA, H. & BAREA, J. M. 2015. Seasonal variation of the productivity and quality of permanent pastures in Adisols of temperate regions. *Journal of Soil Science and Plant Nutrition,* 15**,** 111-128.

DERMODY, J. 2013. *Fodder crisis sees cows starve to death* [Online]. Available: https://www.irishexaminer.com/news/arid-20228661.html [Accessed 23/02/21].

DEVANEY, F., MARTIN, J., O'NEILL, F. & DELANEY, A. 2013. Irish semi-natural grasslands survey. *Annual Report.* 4 ed.

DILLON, A., COX, T. & GREANEY, J. 2018a. Teagasc/Irish Farmers Journal BETTER Farm Beef Programme. *EGF Sustainable meat and milk production from grasslands*

DILLON, E., DONNELLAN, T., HANRAHAN, K., HOULIHANE, T., LOUGHREY, J., MCKEON, M., MORAN, B. & THORNE, F. 2018b. Outlook 2019: Economic prospects for agriculture. *Athenry: Teagasc*.

DILLON, E., MORAN, B., LENNON, J. & DONNELLAN, T. 2018c. Teagasc National Farm Survey. Teagasc.

DILLON, P., CROSSE, S., STAKELUM, G. & FLYNN, F. 1995. The effect of calving date and stocking rate on the performance of spring-calving dairy cows. *Grass and Forage Science,* 50**,** 286-299.

DILLON, P., HENNESSY, T., SHALLOO, L., THORNE, F. & 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**,** 16-29.

DILON, P. & KENNEDY, J. 2009. Grazing notebook. *Teagasc and Irish Farmer Journal.* Teagasc.

DONOHUE, R. J., LAWES, R. A., MATA, G., GOBBETT, D. & OUZMAN, J. 2018. Towards a national, remote-sensing-based model for predicting field-scale crop yield. *Field Crops Research*, 227**,** 79-90.

DORMATEY, R., SUN, C., ALI, K., COULTER, J. A., BI, Z. & BAI, J. 2020. Gene Pyramiding for Sustainable Crop Improvement against Biotic and Abiotic Stresses. *Agronomy,* 10.

DREWNOWSKI, A. 2018. Measures and metrics of sustainable diets with a focus on milk, yogurt, and dairy products. *Nutr Rev,* 76**,** 21-28.

EDIRISINGHE, A., CLARK, D. & WAUGH, D. 2012. Spatio-temporal modelling of biomass of intensively grazed perennial dairy pastures using multispectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation,* 16**,** 5-16.

ÉIREANN, M. 2012. A summary of climate averages for ireland 1981-2010. Met Éireann, Glasnevin Hill, Dublin.

ENVIRONNEMENT, A. 2019. Evaluation of the impact of the CAP on habitats, landscapes, biodiversity.

ESTES, L., BRADLEY, B., BEUKES, H., HOLE, D., LAU, M., OPPENHEIMER, M., SCHULZE, R., TADROSS, M. & TURNER, W. 2013. Comparing mechanistic and empirical model projections of crop suitability and productivity: implications for ecological forecasting. *Global Ecology and Biogeography,* 22**,** 1007-1018.

EUROSTAT. 2015. *Land cover statistics* [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Land_cover_statistics#Land_cover_in_the_EU [Accessed 19/02/21].

EUROSTAT. 2021a. *Agri-environmental indicator - cropping patterns* [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Agri-environmental_indicator_-_cropping_patterns [Accessed 19/02/2021].

EUROSTAT. 2021b. *Number of dairy cows* [Online]. Available: https://ec.europa.eu/eurostat/databrowser/view/tag00014/default/table?lang=en [Accessed 18/02/21].

FALZOI, S., GLEESON, E., LAMBKIN, K., ZIMMERMANN, J., MARWAHA, R., GREEN, R. O. H. S. & FRATIANNI, S. 2019. Analysis of the severe drought in Ireland in 2018. *Weather,* 74.

FAO. 2021. *FAOSTAT Livestock Primary Data* [Online]. Available: fao.org/faostat/en/#data/QL [Accessed 18/02/21].

FARRAG, A., OTHMAN, S., MAHMOUD, T. & ELRAFFIEI, A. Y. 2019. Satellite swarm survey and new conceptual design for Earth observation applications. *The Egyptian Journal of Remote Sensing and Space Science*.

FAY, D., KRAMERS, G., ZHANG, C. & GRENNAN, D. M. E. 2007. Soil Geochemical Atlas of Ireland.

FEALY, R. & FEALY, R. M. 2008. The spatial variation in degree days derived from locational attributes for the 1961 to 1990 period. *Irish Journal of Agricultural and Food Research***,** 1-11.

FENG, P., WANG, B., LI LIU, D., WATERS, C., XIAO, D., SHI, L. & YU, Q. 2020. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique. *Agricultural and Forest Meteorology,* 285**,** 107922.

FITZGERALD, J., BRERETON, A. & HOLDEN, N. 2008. Simulation of the influence of poor soil drainage on grass-based dairy production systems in Ireland. *Grass and Forage Science,* 63**,** 380-389.

FRANK, A. B. & HOFMANN, L. 1989. Relationship among grazing management, growing degreedays, and morphological development for native grasses on the Northern Great Plains *Rangeland Ecology & Management/Journal of Range Management Archives,* 42**,** 199-202.

FRENCH, P., HENNESSY, D., O'DONOVAN, M. & LAIDLAW, S. 2006. Manipulation of grass supply to meet feed demand. Teagasc.

GAO, X., DONG, S., LI, S., XU, Y., LIU, S., ZHAO, H., YEOMANS, J., LI, Y., SHEN, H. & WU, S. 2020a. Using the random forest model and validated MODIS with the field spectrometer

measurement promote the accuracy of estimating aboveground biomass and coverage of alpine grasslands on the Qinghai-Tibetan Plateau. *Ecological Indicators,* 112**,** 106114.

GAO, X., DONG, S., LI, S., XU, Y., LIU, S., ZHAO, H., YEOMANS, J., LI, Y., SHEN, H., WU, S. & ZHI, Y. 2020b. Using the random forest model and validated MODIS with the field spectrometer measurement promote the accuracy of estimating aboveground biomass and coverage of alpine grasslands on the Qinghai-Tibetan Plateau. *Ecological Indicators,* 112.

GARGIULO, J., CLARK, C., LYONS, N., DE VEYRAC, G., BEALE, P. & GARCIA, S. 2020. Spatial and Temporal Pasture Biomass Estimation Integrating Electronic Plate Meter, Planet CubeSats and Sentinel-2 Satellite Data. *Remote Sensing,* 12.

GAUJOUR, E., AMIAUD, B., MIGNOLET, C. & PLANTUREUX, S. 2011. Factors and processes affecting plant biodiversity in permanent grasslands. A review. *Agronomy for Sustainable Development,* 32**,** 133-160.

GE, L., JUANLE, W., YANJIE, W. & HAISHUO, W. 2019. Estimation of Grassland Production in Central and Eastern Mongolia from 2006 to 2015 via Remote Sensing. *Journal of Resources and Ecology,* 10.

GOMEZ-ZAVAGLIA, A., MEJUTO, J. C. & SIMAL-GANDARA, J. 2020. Mitigation of emerging implications of climate change on food production systems. *Food Res Int,* 134**,** 109256.

GOURLEY, C. J. P., HANNAH, M. C. & CHIA, K. T. H. 2017. Predicting pasture yield response to nitrogenous fertiliser in Australia using a meta-analysis-derived model, with field validation. *Soil Research,* 55.

GREEN, S. 2019. *Investigation into the bio-physical constraints on farmer turn-out-date decisions using remote sensing and meteorological data.* PhD, University College Cork.

GREEN, S., CAWKWELL, F. & DWYER, E. 2018. A time-domain NDVI anomaly service for intensively managed grassland agriculture. *Remote Sensing Applications: Society and Environment,* 11**,** 282-290.

GRIFFITHS, P., NENDEL, C., PICKERT, J. & HOSTERT, P. 2020. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sensing of Environment,* 238.

GSI, G. S. O. I. 2021. *GSI geological data download* [Online]. Available: https://www.gsi.ie/en-ie/data-and-maps/Pages/Bedrock.aspx [Accessed].

GUAN, S., FUKAMI, K., MATSUNAKA, H., OKAMI, M., TANAKA, R., NAKANO, H., SAKAI, T., NAKANO, K., OHDAN, H. & TAKAHASHI, K. 2019. Assessing Correlation of High-Resolution NDVI with Fertilizer Application Level and Yield of Rice and Wheat Crops using Small UAVs. *Remote Sensing,* 11.

GUO, Y., FU, Y., HAO, F., ZHANG, X., WU, W., JIN, X., ROBIN BRYANT, C. & SENTHILNATH, J. 2021. Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecological Indicators,* 120.

HAGOLLE, O., HUC, M., PASCUAL, D. V. & DEDIEU, G. 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment,* 114**,** 1747-1755.

HAKL, J., HREVUŠOVÁ, Z., HEJCMAN, M. & FUKSA, P. 2012. The use of a rising plate meter to evaluate lucerne (Medicago sativaL.) height as an important agronomic trait enabling yield estimation. *Grass and Forage Science,* 67**,** 589-596.

HAN, D., O'KIELY, P. & SUN, D.-W. 2003. Linear models for the dry matter yield of the primary growth of a permanent grassland pasture. *Irish Journal of Agricultural and Food Research,* 42**,** 17-38.

HANRAHAN, L., GEOGHEGAN, A., O'DONOVAN, M., GRIFFITH, V., RUELLE, E., WALLACE, M. & SHALLOO, L. 2017. PastureBase Ireland: A grassland decision support system and national database. *Computers and Electronics in Agriculture,* 136**,** 193-201.

HARA, R. O., GREEN, S., MCCARTHY, T., CAHALANE, C., FENTON, O. & TUOHY, P. 2020. Identifying artificially drained pasture soils using machine learning and Earth observation imagery. *Journal of Applied Remote Sensing,* 14.

HARGY, V. T. 1997. Objectively mapping accumulated temperature for Ireland. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 17**,** 909-927.

HARMONEY, K. R., MOORE, K. J., GEORGE, J. R., BRUMMER, E. C. & RUSSELL, J. R. 1997. Determination of Pasture Biomass Using Four Indirect Methods *Agronomy journal,* 89.

HART, A. M. K., BALDOCK, D. & AUBERT, S. W. P.-M. 2020. Aligning the post-2020 CAP with the Green Deal. Institute for European Environmental Policy.

HASTINGS, A., CLIFTON-BROWN, J., WATTENBACH, M., MITCHELL, C. P. & SMITH, P. 2009. The development of MISCANFOR, a newMiscanthuscrop growth model: towards more robust yield predictions under different climatic and soil conditions. *GCB Bioenergy,* 1**,** 154-170.

HENNESSY, D., DELABY, L., VAN DEN POL-VAN DASSELAAR, A. & SHALLOO, L. 2020. Increasing Grazing in Dairy Cow Milk Production Systems in Europe. *Sustainability,* 12.

HIGGINS, S., SCHELLBERG, J. & BAILEY, J. 2019. Improving productivity and increasing the efficiency of soil nutrient management on grassland farms in the UK and Ireland using precision agriculture technology. *European Journal of Agronomy,* 106**,** 67-74.

HODGSON, J. A., THOMAS, C. D., OLIVER, T. H., ANDERSON, B. J., BRERETON, T. M. & CRONE, E. E. 2011. Predicting insect phenology across space and time. *Global Change Biology,* 17**,** 1289-1300.

HOLDEN, N. M. & BRERETON, A. J. 2002. An Assessment of the Potential Impact of Climate Change on Grass Yield in Ireland over the Next 100 Years. *Irish Journal of Agricultural and Food Research* 41**,** 213-226.

HOLLBERG, J. & SCHELLBERG, J. 2017. Distinguishing Intensity Levels of Grassland Fertilization Using Vegetation Indices. *Remote Sensing,* 9.

HOLTGRAVE, A.-K., RÖDER, N., ACKERMANN, A., ERASMI, S. & KLEINSCHMIT, B. 2020. Comparing Sentinel-1 and-2 Data and Indices for Agricultural Land Use Monitoring. *Remote Sensing,* 12**,** 2919.

HOPKINS, A. & PRADO, A. D. 2007 Implications of climate change for grassland in Europe: impacts, adaptations and mitigation options: a review. *Grass and Forage Science,* 62**,** 118–126.

HURTADO-URIA, C., HENNESSY, D., SHALLOO, L., O'CONNOR, D. & DELABY, L. 2014. Relationships between meteorological data and grass growth over time in the south of Ireland. *Irish Geography,* 46**,** 175-201.

HURTADO-URIA, C., HENNESSY, D., SHALLOO, L., SCHULTE, R. P. O., DELABY, L. & O'CONNOR, D. 2012. Evaluation of three grass growth models to predict grass growth in Ireland. *The Journal of Agricultural Science,* 151**,** 91-104.

HUTCHINSON, G. K., RICHARDS, K. & RISK, W. H. Aspects of accumulated heat patterns (growing-degree days) and pasture growth in Southland. Proceedings of the New Zealand Grassland Association, 2000. 81-85.

JAIN, R., URBAN, L., BALBACH, H. & WEBB, M. D. 2012. Contemporary Issues in Environmental Assessment. *Handbook of Environmental Engineering Assessment.*

JANITZA, S. & HORNUNG, R. 2018. On the overestimation of random forest's out-of-bag error. *PLoS One,* 13**,** e0201904.

JANSEN, B. V. S., KOLDEN, C. A., GREAVES, H. E. & EITEL, J. U. H. 2019. Lidar provides novel insights into the effect of pixel size and grazing intensity on measures of spatial heterogeneity in a native bunchgrass ecosystem. *Remote Sensing of Environment,* 235.

JENAL, A., HÜGING, H., AHRENDS, H. E., BOLTEN, A., BONGARTZ, J. & BARETH, G. 2021. Investigating the Potential of a Newly Developed UAV-Mounted VNIR/SWIR Imaging System for Monitoring Crop Traits—A Case Study for Winter Wheat. *Remote Sensing,* 13.

JEONG, J. H., RESOP, J. P., MUELLER, N. D., FLEISHER, D. H., YUN, K., BUTLER, E. E., TIMLIN, D. J., SHIM, K. M., GERBER, J. S., REDDY, V. R. & KIM, S. H. 2016. Random Forests for Global and Regional Crop Yield Predictions. *PLoS One,* 11**,** e0156571.

JOÃO PEDRO SILVA JUSTIN TOLAND, W. J. & O'HARA, J. E. E. T. E. 2008. LIFE and Europe´s grasslands Restoring a forgotten habitat.

JOHNSON, I., CHAPMAN, D., PARSONS, A., ECKARD, R. & FULKERSON, W. DairyMod: a biophysical simulation model of the Australian dairy system. Australian Farming Systems Conference, 2003a. Citeseer.

JOHNSON, I., LODGE, G. & WHITE, R. 2003b. The sustainable grazing systems pasture model: description, philosophy and application to the SGS national experiment. *Australian Journal of Experimental Agriculture,* 43**,** 711-728.

JOHNSON, I. R., CHAPMAN, D. F., SNOW, V. O., ECKARD, R. J., PARSONS, A. J., LAMBER, M. G. & CULLEN, B. R. 2008. DairyMod and EcoMod: biophysical pasture-simulation models for Australia and New Zealand. *Australian Journal of Experimental Agriculture,* 48**,** 621–631.

JOHNSON, I. R. & THORNLEY, J. 1983. Vegetative crop growth model incorporating leaf area expansion and senescence, and applied to grass. *Plant, Cell and Environment* 6**,** 721-729.

JOUVEN, M., CARRÈRE, P. & 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**,** 112-124.

JUST, A. C., ARFER, K. B., RUSH, J., DORMAN, M., SHTEIN, A., LYAPUSTIN, A. & KLOOG, I. 2020. Advancing methodologies for applying machine learning and evaluating spatiotemporal models of fine particulate matter (PM2. 5) using satellite data over large regions. *Atmospheric Environment,* 239**,** 117649.

KASAMPALIS, D., ALEXANDRIDIS, T., DEVA, C., CHALLINOR, A., MOSHOU, D. & ZALIDIS, G. 2018. Contribution of Remote Sensing on Crop Models: A Review. *Journal of Imaging,* 4.

KEANE, T. 1986. *Climate, weather and Irish agriculture,* Meteorological Service.

KEANE, T. & COLLINS, J. 1986. Climate, weather and Irish agriculture, joint working group on applied agricultural meteorology (AGMET), c/o Met Éireann. Dublin.

KESSLER, A., ARCHONTOULIS, S. V. & LICHT, M. A. 2020. Soybean yield and crop stage response to planting date and cultivar maturity in Iowa, USA. *Agronomy Journal,* 112**,** 382-394.

KHOSRAVI, I., JOUYBARI-MOGHADDAM, Y. & SARAJIAN, M. R. 2017. The comparison of NN, SVR, LSSVR and ANFIS at modeling meteorological and remotely sensed drought indices over the eastern district of Isfahan, Iran. *Natural Hazards,* 87**,** 1507-1522.

KOLECKA, N., GINZLER, C., PAZUR, R., PRICE, B. & VERBURG, P. 2018. Regional Scale Mapping of Grassland Mowing Frequency with Sentinel-2 Time Series. *Remote Sensing,* 10.

KONG, B., YU, H., DU, R. & WANG, Q. 2019. Quantitative estimation of biomass of alpine grasslands using hyperspectral remote sensing. *Rangeland Ecology & Management,* 72**,** 336-346.

KUKAL, M. S. & IRMAK, S. 2018. U.S. Agro-Climate in 20(th) Century: Growing Degree Days, First and Last Frost, Growing Season Length, and Impacts on Crop Yields. *Sci Rep,* 8**,** 6977.

LAI, Y. R., PRINGLE, M. J., KOPITTKE, P. M., MENZIES, N. W., ORTON, T. G. & DANG, Y. P. 2018. An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. *International Journal of Applied Earth Observation and Geoinformation,* 72**,** 99-108.

LAPPLE, D., HENNESSY, T. & O'DONOVAN, M. 2012. Extended grazing: a detailed analysis of Irish dairy farms. *J Dairy Sci,* 95**,** 188-95.

LEE, T.-H., ULLAH, A. & WANG, R. 2020. Bootstrap aggregating and random forest. *Macroeconomic Forecasting in the Era of Big Data.* Springer.

LEGG, M. & BRADLEY, S. 2019. Ultrasonic Proximal Sensing of Pasture Biomass. *Remote Sensing,* 11.

LI, F. Y., SNOW, V. O. & HOLZWORTH, D. P. 2011. Modelling the seasonal and geographical pattern of pasture production in New Zealand. *New Zealand Journal of Agricultural Research,* 54**,** 331-352.

LI, J. & CHEN, B. 2020. Global Revisit Interval Analysis of Landsat-8 -9 and Sentinel-2A -2B Data for Terrestrial Monitoring. *Sensors (Basel),* 20.

LI, Q.-Y., YIN, J., LIU, W.-D., ZHOU, S.-M., LI, L., NIU, J.-S., NIU, H.-B. & MA, Y. 2012. Determination of Optimum Growing Degree-Days (GDD) Range Before Winter for Wheat Cultivars with Different Growth Characteristics in North China Plain. *Journal of Integrative Agriculture,* 11**,** 405-415.

LI, Y., LIANG, S., ZHAO, Y., LI, W. & WANG, Y. 2017. Machine learning for the prediction of L. chinensis carbon, nitrogen and phosphorus contents and understanding of mechanisms underlying grassland degradation. 192**,** 116-123.

LIANG, T., FENG, Q., GE, J., XIE, H. & LIANG, T. 2019. Assessment of Machine Learning Methods for Modeling Alpine Grassland Biomass in Southern Qinghai Province, China. *Proceedings of the 3rd International Conference on Computer Science and Application Engineering - CSAE 2019.*

LIAO, C., WANG, J. & SHAN, B. Corn Biomass Estimation Using Sentinel-2 and VENµS Data Based on A Simple Light Use Efficiency Method. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, 2019. IEEE, 7294-7297.

LOUIS, J., DEBAECKER, V., PFLUG, B., MAIN-KNORN, M., BIENIARZ, J., MUELLER-WILM, U., CADAU, E. & GASCON, F. Sentinel-2 sen2cor: L2a processor for users. Proceedings Living Planet Symposium 2016, 2016. Spacebooks Online, 1-8.

LUEDELING, E., STEINMANN, K. P., ZHANG, M., BROWN, P. H., GRANT, J. & GIRVETZ, E. H. 2011. Climate change effects on walnut pests in California. *Global Change Biology,* 17**,** 228-238.

LYU, X., LI, X., DANG, D., DOU, H., XUAN, X., LIU, S., LI, M. & GONG, J. 2020. A new method for grassland degradation monitoring by vegetation species composition using hyperspectral remote sensing. *Ecological Indicators,* 114.

MA, Q., CHAI, L., HOU, F., CHANG, S., MA, Y., TSUNEKAWA, A. & CHENG, Y. 2019. Quantifying Grazing Intensity Using Remote Sensing in Alpine Meadows on Qinghai-Tibetan Plateau. *Sustainability,* 11.

MA, Y., SHEN, Y. & LIU, Y. 2020. Food Waste to Biofertilizer: A Potential Game Changer of Global Circular Agricultural Economy. *J Agric Food Chem,* 68**,** 5021-5023.

MAHER, J., BOGUE, F. & DOUGLAS, J. 2019. Grass10 campaign. *Moorepark '19 - the Irish Dairying Growing Sustainably.* Teagasc.

MANDANICI, E. & BITELLI, G. 2016. Preliminary comparison of sentinel-2 and landsat 8 imagery for a combined use. *Remote Sensing,* 8**,** 1014.

MARKHAM, B. L., JENSTROM, D., MASEK, J. G., DABNEY, P., PEDELTY, J. A., BARSI, J. A. & MONTANARO, M. Landsat 9: status and plans. Earth Observing Systems XXI, 2016. International Society for Optics and Photonics, 99720G.

MARKHAM, B. L., STOREY, J. C., WILLIAMS, D. L. & IRONS, J. R. 2004. Landsat sensor performance: history and current status. *IEEE Transactions on Geoscience and Remote Sensing,* 42**,** 2691-2694.

MARSHALL, M., TU, K. & BROWN, J. 2018. Optimizing a remote sensing production efficiency model for macro-scale GPP and yield estimation in agroecosystems. *Remote Sensing of Environment,* 217**,** 258-271.

MARTI, J., BORT, J., SLAFER, G. A. & ARAUS, J. L. 2007. Can wheat yield be assessed by early measurements of Normalized Difference Vegetation Index? *Annals of Applied Biology,* 150**,** 253-257.

MASEK, J. G., WULDER, M. A., MARKHAM, B., MCCORKEL, J., CRAWFORD, C. J., STOREY, J. & JENSTROM, D. T. 2020. Landsat 9: Empowering open science and applications through continuity. *Remote Sensing of Environment,* 248**,** 111968.

MATTHEWS, A. 2017. Brexit Impacts on Irish Agri-food Exports to the UK. *EuroChoices,* 16**,** 26-32.

MCCARTHY, B., SHALLOO, L. & GEARY, U. 2011. The Grass €alculator. *In:* TEAGASC (ed.). Fermoy, Co. Cork: Moorepark Animal & Grassland Research and Innovation Centre, Teagasc.

MCDONNELL, J., BROPHY, C., RUELLE, E., SHALLOO, L., LAMBKIN, K. & HENNESSY, D. 2019. Weather forecasts to enhance an Irish grass growth model. *European Journal of Agronomy,* 105**,** 168-175.

MCMASTER, G. & SMIKA, D. 1988. Estimation and evaluation of winter wheat phenology in the central Great Plains. *Agricultural and Forest Meteorology,* 43.

MCMASTER, G. S. & WILHELM, W. W. 1997. Growing degree-days: one equation, two interpretations *Agricultural and Forest Meteorology,* 87**,** 291-300.

MCSWEENEY, D., COUGHLAN, N., CUTHBERT, R., HALTON, P. & IVANOV, S. 2019. Micro-sonic sensor technology enables enhanced grass height measurement by a Rising Plate Meter. *Information Processing in Agriculture,* 6**,** 279-284.

MENG, B., GAO, J., LIANG, T., CUI, X., GE, J., YIN, J., FENG, Q. & XIE, H. 2018. Modeling of Alpine Grassland Cover Based on Unmanned Aerial Vehicle Technology and Multi-Factor Methods: A Case Study in the East of Tibetan Plateau, China. *Remote Sensing,* 10.

MICHELS, M., FECKE, W., FEIL, J.-H., MUSSHOFF, O., PIGISCH, J. & KRONE, S. 2019. Smartphone adoption and use in agriculture: empirical evidence from Germany. *Precision Agriculture,* 21**,** 403-425.

MICHEZ, A., LEJEUNE, P., BAUWENS, S., HERINAINA, A. A. L., BLAISE, Y., CASTRO MUÑOZ, E., LEBEAU, F. & BINDELLE, J. 2019. Mapping and monitoring of biomass and grazing in pasture with an unmanned aerial system. *Remote Sensing,* 11**,** 473.

MOECKEL, T., DAYANANDA, S., NIDAMANURI, R., NAUTIYAL, S., HANUMAIAH, N., BUERKERT, A. & WACHENDORF, M. 2018. Estimation of Vegetable Crop Parameter by Multi-temporal UAV-Borne Images. *Remote Sensing,* 10**,** 805.

MÜLLER, L., LIPIEC, J., KORNECKI, T. S. & GEBHARDT, S. 2011. Trafficability and Workability of Soils. *Encyclopedia of Earth Sciences Series*.

MURPHY, A. 2020. Drought Summary. Met Éireann.

MURPHY, D., SHINE, P. & MURPHY, M. 2021. Utilising grassland management and climate data for more accurate prediction of herbage mass using the rising plate meter. *Precision Agriculture***,** 1-28.

MURPHY, D. J., O' BRIEN, B. & MURPHY, M. D. 2018. Development of a Labour Utilisation Decision Support Tool to Efficiently Measure Grass Herbage Mass Using a Rising Plate Meter</i&gt. *2018 Detroit, Michigan July 29 - August 1, 2018.*

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. 2019 ASABE Annual International Meeting, 2019. American Society of Agricultural and Biological Engineers, 1.

MURRAY, J. & HENRY, T. 2018. Waulsortian limestone: geology and hydrogeology. *Groundwater Matters: science and practice.* Tullamore, Co. Offaly, Ireland.

NAIDOO, L., VAN DEVENTER, H., RAMOELO, A., MATHIEU, R., NONDLAZI, B. & GANGAT, R. 2019. Estimating above ground biomass as an indicator of carbon storage in vegetated wetlands of the grassland biome of South Africa. *International Journal of Applied Earth Observation and Geoinformation,* 78**,** 118-129.

NELSON, C. J., MOORE, K. J. & COLLINS, M. 2017. *Forages and grasslands in a changing world*.

NGUYEN, P., BADENHORST, P. E., SHI, F., SPANGENBERG, G. C., SMITH, K. F. & DAETWYLER, H. D. 2020. Design of an Unmanned Ground Vehicle and LiDAR Pipeline for the High-Throughput Phenotyping of Biomass in Perennial Ryegrass. *Remote Sensing,* 13.

NICKMILDER, C., TEDDE, A., DUFRASNE, I., LESSIRE, F., TYCHON, B., CURNEL, Y., BINDELLE, J. & 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.

NOLAN, P. & FLANAGAN, J. 2020. High-resolution climate projections for Ireland–A multi-model ensemble approach. *Environmental Protection Agency*.

O'BRIEN, D., HENNESSY, T., MORAN, B. & SHALLOO, L. 2015. Relating the carbon footprint of milk from Irish dairy farms to economic performance. *J Dairy Sci,* 98**,** 7394-407.

O'DONOVAN, M., DILLON, P., RATH, M. & STAKELUM, G. 2002. A comparison of four methods of herbage mass estimation. *Irish Journal of Agricultural and Food Research***,** 17-27.

O'MARA, F. P. 2012. The role of grasslands in food security and climate change. *Ann Bot,* 110**,** 1263-70.

O'DONOVAN, M., HENNESSY, D. & CREIGHTON, P. 2020. Ruminant grassland production systems in Ireland. *Irish Journal of Agricultural and Food Research*.

O'DONOVAN, M. O. L. A. G. M. 2017. PastureBase Ireland – Capturing Grassland Data on Commercial Irish Farms. *Irish Grassland Association IGA*

OECD. 2003. *Glossary of Statistical Terms–Sustainable Agriculture* [Online]. Paris, France: The Organisation for Economic Co-operation and Development (OECD). Available: https://stats.oecd.org/glossary/detail.asp?ID=2624 [Accessed 19/02/2021].

PARENTE, G. & BOVOLENTS, S. 2012. The role of grassland in rural tourism and recreation in Europe. *EGF*. Poland: Polish Grassland Society.

PENG, Y., FAN, M., BAI, L., SANG, W., FENG, J., ZHAO, Z. & TAO, Z. 2019. Identification of the best hyperspectral indices in estimating plant species richness in sandy grasslands. *Remote Sensing,* 11**,** 588.

PUNALEKAR, S. M., VERHOEF, A., QUAIFE, T. L., HUMPHRIES, D., BERMINGHAM, L. & 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**,** 207-220.

QUAN, X., HE, B., YEBRA, M., YIN, C., LIAO, Z., ZHANG, X. & LI, X. 2017. A radiative transfer model-based method for the estimation of grassland aboveground biomass. *International Journal of Applied Earth Observation and Geoinformation,* 54**,** 159-168.

RADKOWSKI, A., RADKOWSKA, I., DRZEWIECKI, W., PIROWSKI, T. & SZEWCZYK, W. Initial evaluation of PlanetScope nanosatellite images applicability for identification of grazed plant communities. *In:* ASTOR, T. & DZENE, I., eds. EGF 2021 Sensing – New Insights into Grassland Science and Practice 2021 Germany.

RAMANKUTTY, N., EVAN, A. T., MONFREDA, C. & FOLEY, J. A. 2008. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global biogeochemical cycles,* 22.

RAMOELO, A., CHO, M., MATHIEU, R. & SKIDMORE, A. K. 2015. Potential of Sentinel-2 spectral configuration to assess rangeland quality. *Journal of applied remote sensing,* 9**,** 094096.

RASOOLIMANESH, S. M., ALI, F. & JAAFAR, M. 2018. Modeling residents' perceptions of tourism development: Linear versus nonlinear models. *Journal of Destination Marketing & Management,* 10**,** 1-9.

RÉAUMUR, R.-A. F. D. 1735. Observation du thérmomére, faitesa Paris pedant l'année 1735, compares avec celles qui ont été faites sous la ligne,a l'Isle de France,a Alger et quelques-unes de nos isles de l'Amerique. *Mémoires de l'Académie Royal des Sciences***,** 545–576.

REICHSTEIN, M., CAMPS-VALLS, G., STEVENS, B., JUNG, M., DENZLER, J., CARVALHAIS, N. & PRABHAT 2019. Deep learning and process understanding for data-driven Earth system science. *Nature,* 566**,** 195-204.

REINERMANN, S., ASAM, S. & KUENZER, C. 2020. Remote Sensing of Grassland Production and Management—A Review. *Remote Sensing,* 12.

RENAULT, T. 2019. Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. *Digital Finance,* 2**,** 1-13.

RENOU-WILSON, F., WILSON, D., BARRY, C., FOY, B. & MÜLLER, C. 2015. Carbon Loss from Drained Organic Soils under Grassland–CALISTO. *Environmental Protection Agency, Johnstown Castle, Wexford, Ireland*.

ROCHE, J. R., BERRY, D. P., BRYANT, A. M., BURKE, C. R., BUTLER, S. T., DILLON, P. G., DONAGHY, D. J., HORAN, B., MACDONALD, K. A. & MACMILLAN, K. L. 2017. A 100-Year Review: A century of change in temperate grazing dairy systems. *J Dairy Sci,* 100**,** 10189-10233.

ROSSITER, D. 2009. Biophysical models in land evaluation. *Encyclopedia of land use, land cover and soil sciences: Land evaluation,* 2**,** 181-195.

RUELLE, E., DELABY, L., WALLACE, M. & SHALLOO, L. 2018a. Using models to establish the financially optimum strategy for Irish dairy farms. *J Dairy Sci,* 101**,** 614-623.

RUELLE, E., HENNESSY, D. & DELABY, L. 2018b. Development of the Moorepark St Gilles grass growth model (MoSt GG model): A predictive model for grass growth for pasture based systems. *European Journal of Agronomy,* 99**,** 80-91.

SAKOWSKA, K., JUSZCZAK, R. & GIANELLE, D. 2016. Remote Sensing of Grassland Biophysical Parameters in the Context of the Sentinel-2 Satellite Mission. *Journal of Sensors,* 2016**,** 1-16.

SALAM, A. 2020. Internet of things in agricultural innovation and security. *Internet of Things for Sustainable Community Development.* Springer.

SALSKI, A. & HOLSTEN, B. 2006. A fuzzy and neuro-fuzzy approach to modelling cattle grazing on pastures with low stocking rates in Central Europe. *Ecological Informatics,* 1**,** 269-276.

SANDERSON, M. A., ROTZ, C. A., FULTZ, S. W. & RAYBURN, E. B. 2001. Estimating Forage Mass with a commercial Capacitance Meter, Rising Plate Meter, and Pasture Ruler. *Agronomy Journal,* 93**,** 1281-1286.

SANJEEVI, P., PRASANNA, S., SIVA KUMAR, B., GUNASEKARAN, G., ALAGIRI, I. & VIJAY ANAND, R. 2020. Precision agriculture and farming using Internet of Things based on wireless sensor network. *Transactions on Emerging Telecommunications Technologies,* 31**,** e3978.

SCHAPENDONK, A., STOL, W., VAN KRAALINGEN, D. & BOUMAN, B. 1998. LINGRA, a sink/source model to simulate grassland productivity in Europe. *European Journal of Agronomy,* 9**,** 87-100.

SCHIRMEL, J. 2021. COVID-19 Pandemic Turns Life-Science Students into "Citizen Scientists": Data Indicate Multiple Negative Effects of Urbanization on Biota. *Sustainability,* 13.

SCHIRPKE, U., KOHLER, M., LEITINGER, G., FONTANA, V., TASSER, E. & TAPPEINER, U. 2019. Future impacts of changing land-use and climate on ecosystem services of mountain grassland and their resilience. *Ecosyst Serv,* 26**,** 79-94.

SCHMIDHUBER, J. & TUBIELLO, F. N. 2007. Global food security under climate change. *Proceedings of the National Academy of Sciences,* 104**,** 19703-19708.

SCHULTE, R., FEALY, R., CREAMER, R., TOWERS, W., HARTY, T. & JONES, R. 2012. A review of the role of excess soil moisture conditions in constraining farm practices under Atlantic conditions. *Soil Use and Management,* 28**,** 580-589.

SCHULTE, R. P. O., DIAMOND, J., FINKELE, K., HOLDEN, N. M. & BRERETON, A. J. 2005. Predicting the soil moisture conditions of Irish grasslands. *Irish Journal of Agricultural and Food Research,* 44**,** 95-110.

SCHWALBERT, R. A., AMADO, T., CORASSA, G., POTT, L. P., PRASAD, P. V. & CIAMPITTI, I. A. 2020. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agricultural and Forest Meteorology,* 284**,** 107886.

SCHWIEDER, M., BUDDEBERG, M., KOWALSKI, K., PFOCH, K., BARTSCH, J., BACH, H., PICKERT, J. & HOSTERT, P. 2020. Estimating Grassland Parameters from Sentinel-2: A Model Comparison Study. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science,* 88**,** 379-390.

SERRANO, J., SHAHIDIAN, S., MORAL, F., CARVAJAL-RAMIREZ, F. & MARQUES DA SILVA, J. 2020. Estimation of Productivity in Dryland Mediterranean Pastures: Long-Term

Field Tests to Calibration and Validation of the Grassmaster II Probe. *AgriEngineering,* 2**,** 240-255.

SHALLOO, L. Milk production costs–Can we compete. Teagasc National Dairy Conference, 2009. 19-38.

SHALLOO, L., BYRNE, T., LESO, L., RUELLE, E., STARSMORE, K., GEOGHEGAN, A., WERNER, J. & O'LEARY, N. 2021. A review of precision technologies in pasture-based dairying systems. *Irish Journal of Agricultural and Food Research*.

SHALLOO, L., CREIGHTON, P. & O'DONOVAN, M. 2011. The economics of reseeding on a dairy farm. *Irish Journal of Agricultural and Food Research,* 50**,** 113–122.

SHALLOO, L., DILLON, P., O'LOUGHLIN, J., RATH, M. & WALLACE, M. 2004. Comparison of a pasture-based system of milk production on a high rainfall, heavy-clay soil with that on a lower rainfall, free-draining soil. *Grass and Forage Science,* 59**,** 157-168.

SHALLOO, L., M, O. D., LESO, L., WERNER, J., RUELLE, E., GEOGHEGAN, A., DELABY, L. & O'LEARY, N. 2018. Review: Grass-based dairy systems, data and precision technologies. *Animal,* 12**,** s262-s271.

SHAMMI, S. A. & MENG, Q. 2021. Use time series NDVI and EVI to develop dynamic crop growth metrics for yield modeling. *Ecological Indicators,* 121.

SHENDRYK, Y., DAVY, R. & THORBURN, P. 2021. Integrating satellite imagery and environmental data to predict field-level cane and sugar yields in Australia using machine learning. *Field Crops Research,* 260.

SHOKO, C., MUTANGA, O., DUBE, T. & SLOTOW, R. 2018. Characterizing the spatio-temporal variations of C3 and C4 dominated grasslands aboveground biomass in the Drakensberg, South Africa. *International Journal of Applied Earth Observation and Geoinformation,* 68**,** 51-60.

SIBANDA, M., MUTANGA, O., DUBE, T., MOTHAPO, M. C. & MAFONGOYA, P. L. 2019. Remote sensing equivalent water thickness of grass treated with different fertiliser regimes using resample HyspIRI and EnMAP data. *Physics and Chemistry of the Earth, Parts A/B/C,* 112**,** 246-254.

SIBANDA, M., MUTANGA, O. & ROUGET, M. 2015. Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments. *ISPRS Journal of Photogrammetry and Remote Sensing,* 110**,** 55-65.

SKOVSEN, S., DYRMANN, M., MORTENSEN, A. K., LAURSEN, M. S., GISLUM, R., ERIKSEN, J., FARKHANI, S., KARSTOFT, H. & JORGENSEN, R. N. The GrassClover image dataset for semantic and hierarchical species understanding in agriculture. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019. 0-0.

SNIP, K.-J. 2016. A successful pasture management in spring and autumn in the Netherlands. Van Hall Larenstein

SOLYALI, D. 2020. A Comparative Analysis of Machine Learning Approaches for Short-/Long-Term Electricity Load Forecasting in Cyprus. *Sustainability,* 12.

SPEISER, J. L., MILLER, M. E., TOOZE, J. & IP, E. 2019. A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling. *Expert Syst Appl,* 134**,** 93-101.

STATISTA. 2021. *Cow milk production worldwide from 2015 to 2020* [Online]. Available: https://www.statista.com/statistics/263952/production-of-milk-worldwide/ [Accessed 18/02/21].

STURM, A., DRECHSLER, M., JOHST, K., MEWES, M. & WÄTZOLD, F. 2018. DSS-Ecopay – A decision support software for designing ecologically effective and cost-effective agri-environment schemes to conserve endangered grassland biodiversity. *Agricultural Systems,* 161**,** 113-116.

TARAVAT, A., WAGNER, M. & OPPELT, N. 2019. Automatic Grassland Cutting Status Detection in the Context of Spatiotemporal Sentinel-1 Imagery Analysis and Artificial Neural Networks. *Remote Sensing,* 11.

TARRIO, K., TANG, X., MASEK, J. G., CLAVERIE, M., JU, J., QIU, S., ZHU, Z. & WOODCOCK, C. E. 2020. Comparison of cloud detection algorithms for Sentinel-2 imagery. *Science of Remote Sensing,* 2**,** 100010.

TEAGASC 2017. How to measure grass. *In:* IRELAND, P. (ed.).

TIEN BUI, D., KHOSRAVI, K., LI, S., SHAHABI, H., PANAHI, M., SINGH, V., CHAPI, K., SHIRZADI, A., PANAHI, S., CHEN, W. & BIN AHMAD, B. 2018. New Hybrids of ANFIS with Several Optimization Algorithms for Flood Susceptibility Modeling. *Water,* 10.

TIMSINA, J. 2018. Can Organic Sources of Nutrients Increase Crop Yields to Meet Global Food Demand? *Agronomy,* 8.

TODD, S. W., HOFFER, R. M. & MILCHUNAS, D. G. 2010. Biomass estimation on grazed and ungrazed rangelands using spectral indices. *International Journal of Remote Sensing,* 19**,** 427-438.

TWARAKAVI, N. K. C., SAKAI, M. & ŠIMŮNEK, J. 2009. An objective analysis of the dynamic nature of field capacity. *Water Resources Research*, 45.

URBAZAEV, M., CREMER, F., MIGLIAVACCA, M., REICHSTEIN, M., SCHMULLIUS, C. & THIEL, C. 2018. Potential of Multi-Temporal ALOS-2 PALSAR-2 ScanSAR Data for Vegetation Height Estimation in Tropical Forests of Mexico. *Remote Sensing,* 10**,** 1277.

VALENZUELA, I. C., BALDOVINO, R. G., BANDALA, A. A. & DADIOS, E. P. Optimization of Photosynthetic Rate Parameters using Adaptive Neuro-Fuzzy Inference System (ANFIS). International Conference on Computer and Applications (ICCA), 2017. 129-134.

VAN DEN POL-VAN DASSELAAR, A., HENNESSY, D. & ISSELSTEIN, J. 2020. Grazing of Dairy Cows in Europe—An In-Depth Analysis Based on the Perception of Grassland Experts. *Sustainability,* 12.

VAN ZONNEVELD, M., RAKHA, M., TAN, S. Y., CHOU, Y. Y., CHANG, C. H., YEN, J. Y., SCHAFLEITNER, R., NAIR, R., NAITO, K. & SOLBERG, S. O. 2020. Mapping patterns of abiotic and biotic stress resilience uncovers conservation gaps and breeding potential of Vigna wild relatives. *Sci Rep,* 10**,** 2111.

VELTHOF, G., LESSCHEN, J., SCHILS, R., SMIT, A., ELBERSEN, B., HAZEU, G., MUCHER, C. & OENEMA, O. 2014. Grassland areas, production and use. Lot 2. Methodological studies in the field of Agro-Environmental Indicators.

VERDUGO-VÁSQUEZ, N., PAÑITRUR-DE LA FUENTE, C. & ORTEGA-FARÍAS, S. 2017. Model Development to Predict Phenological scale of Table Grapes (cvs. Thompson, Crimson and Superior Seedless and Red Globe) using Growing Degree Days. *OENO One,* 51.

VINOGRADOVS, I., VILLOSLADA, M., NIKODEMUS, O., RUSKULE, A., VEIDEMANE, K., GULBINAS, J., MORKVENAS, Ž., KASPARINSKIS, R., SEPP, K., JÄRV, H., KLIMASK, J., ZARIŅA, A., BRŪMELIS, G., DOTAS, A. & KRYŽANAUSKAS, A. 2020. Integrating ecosystem services into decision support for management of agroecosystems: Viva Grass tool. *One Ecosystem,* 5.

WACHENDORF, M., FRICKE, T. & MÖCKEL, T. 2018. Remote sensing as a tool to assess botanical composition, structure, quantity and quality of temperate grasslands. *Grass and Forage Science,* 73**,** 1-14.

WALSH, J. & HORNER, A. 1984. Regional aspects of agricultural production in Ireland 1970–1980. *Irish Geography,* 17**,** 95-101.

WANG, E. & ENGEL, T. 1998. Simulation of phenological development of wheat crops. *Agricultural systems,* 58**,** 1-24.

WANG, J., XIAO, X., BAJGAIN, R., STARKS, P., STEINER, J., DOUGHTY, R. B. & 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**,** 189-201.

WANG, Y., WU, G., DENG, L., TANG, Z., WANG, K., SUN, W. & SHANGGUAN, Z. 2017. Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm. *Sci Rep,* 7**,** 6940.

WHITE, H. J., GAUL, W., SADYKOVA, D., LEON-SANCHEZ, L., CAPLAT, P., EMMERSON, M. C. & YEARSLEY, J. M. 2020. Quantifying large-scale ecosystem stability with remote sensing data. *Remote Sens Ecol Conserv,* 6**,** 354-365.

WHITE, R. & ROHWEDER, S. M. M. 2000. Pilot analysis of global ecosystems: Grassland ecosystems.

WHITE, T. & SNOW, V. 2012. A modelling analysis to identify plant traits for enhanced water-use efficiency of pasture. *Crop and Pasture Science,* 63**,** 63-76.

WINGLER, A. & HENNESSY, D. 2016. Limitation of grassland productivity by low temperature and seasonality of growth. *Frontiers in plant science,* 7**,** 1130.

WU, J. S. & FU, G. 2018. Modelling aboveground biomass using MODIS FPAR/LAI data in alpine grasslands of the Northern Tibetan Plateau. *Remote Sensing Letters,* 9**,** 150-159.

WULDER, M. A., WHITE, J. C., LOVELAND, T. R., WOODCOCK, C. E., BELWARD, A. S., COHEN, W. B., FOSNIGHT, E. A., SHAW, J., MASEK, J. G. & ROY, D. P. 2016. The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment,* 185**,** 271-283.

XIA, J., MA, M., LIANG, T., WU, C., YANG, Y., ZHANG, L., ZHANG, Y. & YUAN, W. 2018. Estimates of grassland biomass and turnover time on the Tibetan Plateau. *Environmental Research Letters,* 13**,** 014020.

XIE, Q., DASH, J., HUANG, W., PENG, D., QIN, Q., MORTIMER, H., CASA, R., PIGNATTI, S., LANEVE, G., PASCUCCI, S., DONG, Y. & YE, H. 2018. Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 11**,** 1482-1493.

YANG, S., FENG, Q., LIANG, T., LIU, B., ZHANG, W. & XIE, H. 2018. Modeling grassland above-ground biomass based on artificial neural network and remote sensing in the Three-River Headwaters Region. *Remote Sensing of Environment,* 204**,** 448-455.

YIN, J., WANG, W., YU, H., LIU, B., FENG, Q., LIANG, T., MENG, B., YANG, S., GAO, J., GE, J., HOU, M. & LIU, J. 2020. Estimation of Grassland Height Based on the Random Forest Algorithm and Remote Sensing in the Tibetan Plateau. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 13**,** 178-186.

YIN, X., WANG, C., ZONG, Z., WANG, H., ZHANG, H. & ZHANG, W. 2018. Biomass estimation of desert steppe based on spectral indices along a precipitation gradient. *Spectroscopy Letters,* 51**,** 324-331.

YUE, S., MUNIR, I. U., HYDER, S., NASSANI, A. A., QAZI ABRO, M. M. & ZAMAN, K. 2020. Sustainable food production, forest biodiversity and mineral pricing: Interconnected global issues. *Resources Policy,* 65.

ZAREI, A., ASADI, E., EBRAHIMI, A., JAFARI, M., MALEKIAN, A., MOHAMMADI NASRABADI, H., CHEMURA, A. & MASKELL, G. 2020. Prediction of future grassland vegetation cover fluctuation under climate change scenarios. *Ecological Indicators,* 119.

ZEKIĆ-SUŠAC, M., PFEIFER, S. & ŠARLIJA, N. 2014. A Comparison of Machine Learning Methods in a High-Dimensional Classification Problem. *Business Systems Research Journal,* 5**,** 82-96.

ZENG, L. & CHEN, C. 2018. Using remote sensing to estimate forage biomass and nutrient contents at different growth stages. *Biomass and Bioenergy,* 115**,** 74-81.

ZENG, N., REN, X., HE, H., ZHANG, L., ZHAO, D., GE, R., LI, P. & NIU, Z. 2019. Estimating grassland aboveground biomass on the Tibetan Plateau using a random forest algorithm. *Ecological Indicators,* 102**,** 479-487.

ZHANG, C., DENKA, S., COOPER, H. & MISHRA, D. R. 2018. Quantification of sawgrass marsh aboveground biomass in the coastal Everglades using object-based ensemble analysis and Landsat data. *Remote Sensing of Environment,* 204**,** 366-379.

ZHANG, S., BAI, Y., ZHANG, J.-H. & ALI, S. 2021a. Developing a process-based and remote sensing driven crop yield model for maize (PRYM–Maize) and its validation over the Northeast China Plain. *Journal of Integrative Agriculture,* 20**,** 408-423.

ZHANG, X., BAO, Y., WANG, D., XIN, X., DING, L., XU, D., HOU, L. & SHEN, J. 2021b. Using UAV LiDAR to Extract Vegetation Parameters of Inner Mongolian Grassland. *Remote Sensing,* 13**,** 656.

ZHU, Y., LIU, K., LIU, L., MYINT, S. W., WANG, S., LIU, H. & HE, Z. 2017. Exploring the potential of world view-2 red-edge band-based vegetation indices for estimation of mangrove leaf area index with machine learning algorithms. 9.

ZUMO, I. M. & HASHIM, M. 2020. Mapping Seasonal Variations of Grazing Land Above-ground Biomass with Sentinel 2A Satellite Data. *IOP Conference Series: Earth and Environmental Science.*

ZUMO, I. M., HASHIM, M. & HASSAN, N. 2021. Mapping grass above-ground biomass of grazing-lands using satellite remote sensing. *Geocarto International***,** 1-14.

5.1 Matching grass growth rate from PBI to image acquisition dates

```
library(dplyr)
library(data.table)

NDVI <-read.csv(file.choose(), header=T)
PB <-read.csv(file.choose(), header=T)
PB<-PB %>% mutate(Growth_2 = Growth)
PB<-PB %>% mutate(Farm_2 = Farm)
PB<-PB %>% mutate(Date_2 = Date)

joined<-NDVI %>% left_join(PB, by = c("Growth","Date"))
joined<-NDVI[PB,roll = TRUE, on = .(Date, Farm)]
write.csv(gf,"H:/National_model_new/test.csv", row.names = TRUE)

dplr <-left_join(NDVI,PB , by=c("Date","Farm"))
setDT(NDVI)
setDT(PB)

NDVI$Date <- as.Date(NDVI$Date, format = "%d-%b-%y")
PB$Date <- as.Date(PB$Date, format = "%d-%b-%y")
gf<-PB[NDVI, on = .(Farm = Farm, Date =Date), roll = 7]
gf<-na.omit(gf)
write.csv(gf,"PB_NDVI.csv", row.names = TRUE)

mingf<-PB[NDVI, on = .(Farm = Farm, Date = Date), roll = TRUE]
mingf<-na.omit(mingf)
write.csv(mingf,"H:/National_model_new/2020.csv", row.names = TRUE)
```

5.2 Script for Brereton model

```
from scipy.stats import pearsonr
from sklearn.metrics import mean_squared_error
import sklearn.metrics as sk
import scipy as sp
from scipy import stats
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import string

data=pd.read_csv('path to your folder')
data_g=pd.read_csv('path to your folder')


k=1.26
alpha=0.23
Q=18.81
G_biomass=np.asarray(list(data_g.ground))
''' Extract date'''
date=list(data.date)
l=len(date)
R=list(data.radiation)
Rs=np.asarray(R)
T=np.asarray(list(data.mean_temp))
E_p=np.asarray(list(data.pe))
#E_a=np.asarray(list(data.evap))
smd=np.asarray(list(data.smd_md))
Sa=smd[0]
SMD_max=110
SMD_c=0
E_a=np.zeros(l)
for i in range(l):
    E_a[i]=E_p[i]* ((SMD_max-smd[[i-1]])/(SMD_max-SMD_c))
e=np.zeros(l)
for i in range (l):
    if T[i]<5.5:
        e[i]=0
    elif T[i]>5.5 and T[i]<9.5:
        e[i]=0.00893+(0.00204*T[i])
    elif T[i]>5.5:
        e[i]=0.00349+(0.00070*T[i])
x=np.zeros(l)
for i in range(l):
    if i<=167:
        x[i]=1.0
    elif i>167:
        x[i]=0.4

Y_p=np.zeros(l)
for i in range(l):
    Y_p[i]=(e[i]*Rs_mj[i]*x[i])/Q

Ya_Yp=np.zeros(l)
```

```
for i in range(l):
    Ya_Yp[i]=0.20+0.80*(E_a[i]/E_p[i])

Y_a=np.zeros(l)
for i in range(l):
    Y_a[i]=Y_p[i]*Ya_Yp[i]

df = pd.DataFrame(Y_a)
cb3=df.rolling(window=3).mean()
```

5.3 Script for ANFIS model

```
## Load package
library(caret)
library(ggplot2)
library(ggpubr)
library(ggpmisc)
library(frbs)
library(GGally)
library(dplyr)
library(data.table)
library(ggcorrplot)
library(Boruta)

## Reading data and standardization
dataset <- read.csv(file.choose(), header=T)
pp_boxcox <-preProcess(dataset[,-1],method=c("BoxCox","center","scale"))
AllFeatures<-predict(pp_boxcox,dataset[,-1])

## Select predictors using RFE algorithm
boruta_output <- Boruta(Growth ~ .,data=(AllFeatures[,6:17]),  doTrace=2)
boruta_signif <- names(boruta_output$finalDecision[boruta_output$finalDecision %in%
c("Confirmed", "Tentative")])
plot(boruta_output, cex.axis=0.8, cex.lab=1,las=2, xlab="Variables", main="Variable
Importance")
Dataset1<-dplyr::select(dataset[,],variables)
set.seed(76418)
as.data.frame(table(unlist(Dataset1)))
row_count <- nrow(Dataset1)
shuffled_rows <- sample(row_count)
trainSets1 <- Dataset1[head(shuffled_rows,floor(row_count*0.60)),]
tests1<- Dataset1[tail(shuffled_rows,floor(row_count*0.40)),]
trainSets<-trainSets1[,3:14]
testSets<-tests1[,3:13]
gr<-tests1[,14]
range.data<-apply(Dataset1[,3:14], 2, range) #2 indicate columns

## ANFIS
## set the method and its parameters
method.type <- "ANFIS"
control <- list(num.labels = 10, max.iter =100, step.size = 0.1,  type.tnorm = "YAGER",
type.snorm = "YAGER", type.implication.func = "ZADEH", name = "biomass")
## generate fuzzy model
object <- frbs.learn(trainSets,range.data, method.type, control)

## Fit model using data training.
anfisTrainY <- predict(object, trainSets[,-12])
anfisTestY<- predict(object, testSets[,])
anfis.train <- cbind.data.frame(trainSets$Growth, anfisTrainY)
anfis.test <- cbind.data.frame(gr, anfisTestY)
residuals <- (gr - anfisTestY)
MSE <- mean(residuals^2)
MAE <- mean(abs(residuals))
RMSE <- sqrt(mean(residuals^2))
SMAPE <- mean(abs(residuals)/(abs(gr) + abs(anfisTestY))/2)*100
```

5.4 Script for RF model

```
## Load package
library(caret)
library(ggplot2)
library(ggpubr)
library(dplyr)
library(GGally)
library(ggcorrplot)
library(randomForest)

## Reading data and standardization
dataset <- read.csv(file.choose(), header=T)
dataset<-subset(dataset, Growth >0)
pp_boxcox <-preProcess(dataset[,-1],method=c("center"))
AllFeatures<-predict(pp_boxcox,dataset[,-1])

# Select predictors using RFE algorithm
rfeCtrl<-rfeControl(method="repeatedcv",
                    repeats=10,number=10,
                    verbos=TRUE,
                    functions=rfFuncs)

Dataset1<-dplyr::select(dataset[,],Date,Farm,NDVI,NDVI_re1,evap,gdd,Growth)
set.seed(76418)
pre<-preProcess(Dataset1[,-10],method = "center")
Dataset<-predict(pre,newdata=Dataset1[,])
as.data.frame(table(unlist(Dataset)))
row_count <- nrow(Dataset)
shuffled_rows <- sample(row_count)
trainSets <- Dataset[head(shuffled_rows,floor(row_count*0.60)),]
tests<- Dataset[tail(shuffled_rows,floor(row_count*0.40)),]
testSets<-tests[,-10]
gr<-tests[,10]
ctrKFCV<-trainControl(method="repeatedcv",repeats=10,number=10,search="grid")

## Hypertuning parameter
set.seed(76418)
m1 <- randomForest(
  formula = Growth ~ .,
  data    = trainSets
)
m1
# number of trees with lowest MSE
which.min(m1$mse)
# RMSE of this optimal random forest
sqrt(m1$mse[which.min(m1$mse)])
#################
set.seed(76418)
rfModel<-train(Growth~., data=trainSets,
               method="rf",
               tuneLength=5,
               n.trees=130,
               importance=TRUE,
               trControl=ctrKFCV)
```

```
rfModel
rfTestY <- predict(rfModel, testSets[,])
rfTrainY <- predict(rfModel, trainSets[,-10])
rf.train <- cbind.data.frame(trainSets$Growth, rfTrainY)
rf.test <- cbind.data.frame(gr, rfTestY)

residuals <- (gr - rfTestY)
MSE <- mean(residuals^2)
MAE <- mean(abs(residuals))
RMSE <- sqrt(mean(residuals^2))
SMAPE <- mean(abs(residuals)/(abs(gr) + abs(rfTestY))/2)*100
err <- c(MSE,MAE, RMSE, SMAPE)
```

5.5 RMSE for different values of 'k' for cross-validation

Table 5.5A The error metrics- RMSE (kg DM $ha^{-1}day^{-1}$) for all the models 1-12 from k=1-15.

k	*RMSE for model 7*	*RMSE for model 8*	*RMSE for model 9*	*RMSE for model 10*	*RMSE for model 11*	*RMSE for model 12*
1	*14.88*	*16.01*	*15.71*	*15.04*	*15.24*	*15.03*
2	*14.86*	*15.80*	*15.04*	*14.78*	*15.32*	*15.01*
3	*14.80*	*15.85*	*15.73*	*14.76*	*15.63*	*15.00*
4	*14.82*	*15.74*	*15.63*	*14.80*	*15.00*	*14.97*
5	*14.77*	*15.71*	*15.60*	*14.94*	*15.98*	*14.94*
6	*14.80*	*15.80*	*15.84*	*14.77*	*15.90*	*14.92*
7	*14.78*	*15.93*	*15.53*	*14.75*	*15.56*	*14.90*
8	*14.76*	*15.40*	*14.83*	*14.89*	*15.84*	*14.85*
9	*14.73*	*15.45*	*14.72*	*14.70*	*15.04*	*14.85*
10	*14.70*	*15.33*	*14.68*	*14.65*	*14.94*	*14.83*
11	*14.72*	*15.90*	*14.93*	*15.58*	*15.23*	*15.00*
12	*14.80*	*15.80*	*15.01*	*15.43*	*15.59*	*15.05*
13	*15.20*	*15.66*	*15.33*	*15.46*	*15.67*	*15.57*
14	*15.55*	*15.68*	*15.38*	*15.80*	*15.69*	*15.70*
15	*15.58*	*15.79*	*15.04*	*15.85*	*15.06*	*15.29*

5.6 Plot of OOB error with many trees with the 'mtry' tables

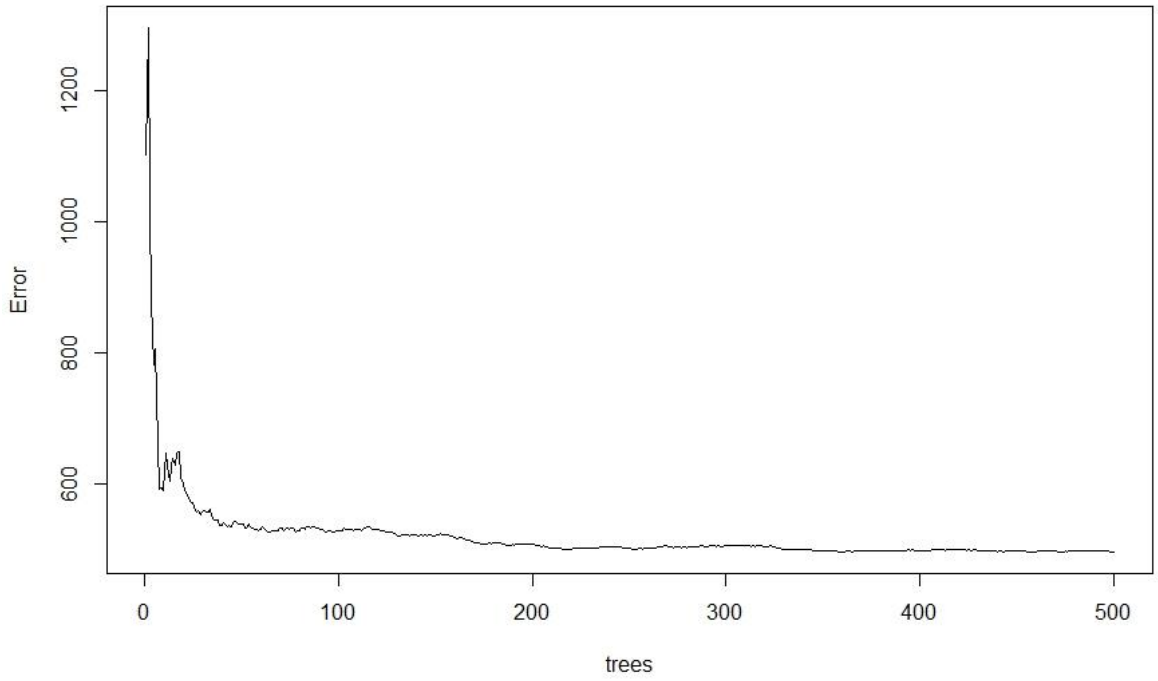


Figure 5.6A Plot of out-of-bag (OOB) error vs many trees (ntree) for ntree in random forest. Out-of-bag error is the mean prediction error on the sample that was not used in the random forest training. The tree with the lowest RMSE is chosen for the final model. At 219 trees RMSE is the lowest.

Table 5.6A mtry values with their respective RMSE.
mtry of 10 has the lowest RMSE of 20.60 kg DM $ha^{-1}day^{-1}$

mtry	*RMSE (kg DM $ha^{-1}day^{-1}$)*
2	*21.77*
4	*21.04*
6	*20.73*
8	*20.63*
10	*20.60*

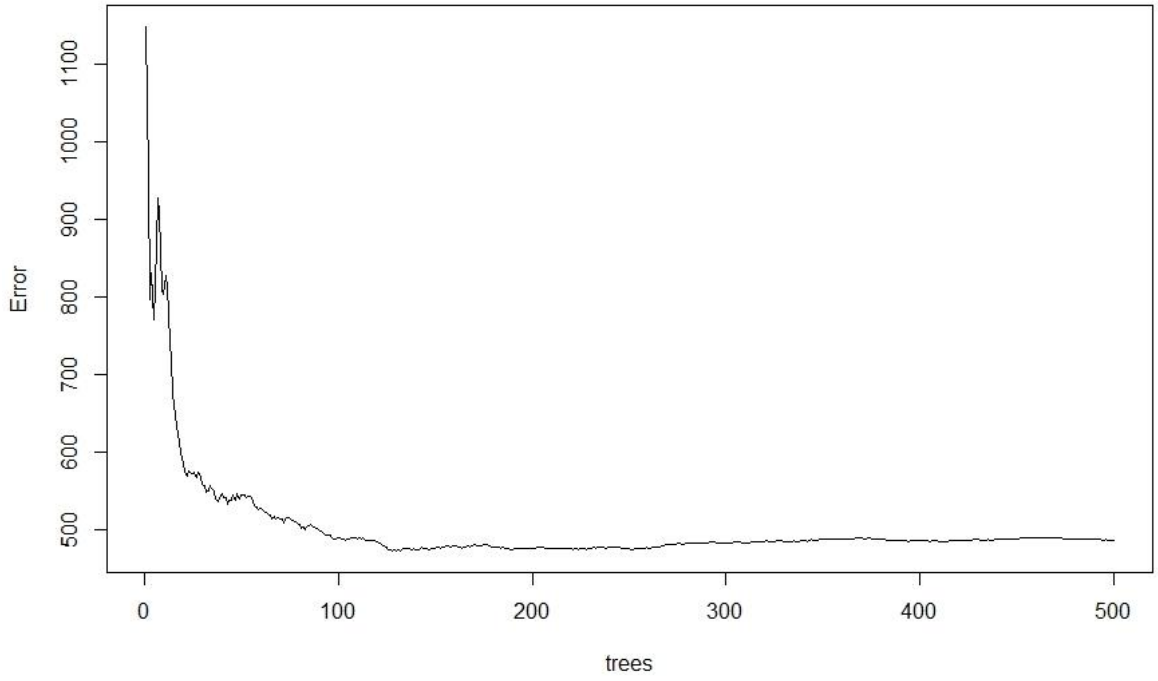


Figure 5.6B Plot of out-of-bag (OOB) error vs several trees (ntree) for ntree in random forest. Out-of-bag error is the mean prediction error on the sample that was not used in the random forest training. The tree with the lowest RMSE is chosen for the final model.

Table 5.6B mtry values with their respective RMSE.
mtry of 9 had lowest RMSE of 20.49 kg DM $ha^{-1}day^{-1}$

mtry	*RMSE (kg DM $ha^{-1}day^{-1}$)*
2	*21.36*
3	*20.96*
5	*20.60*
7	*20.49*
9	*20.49*

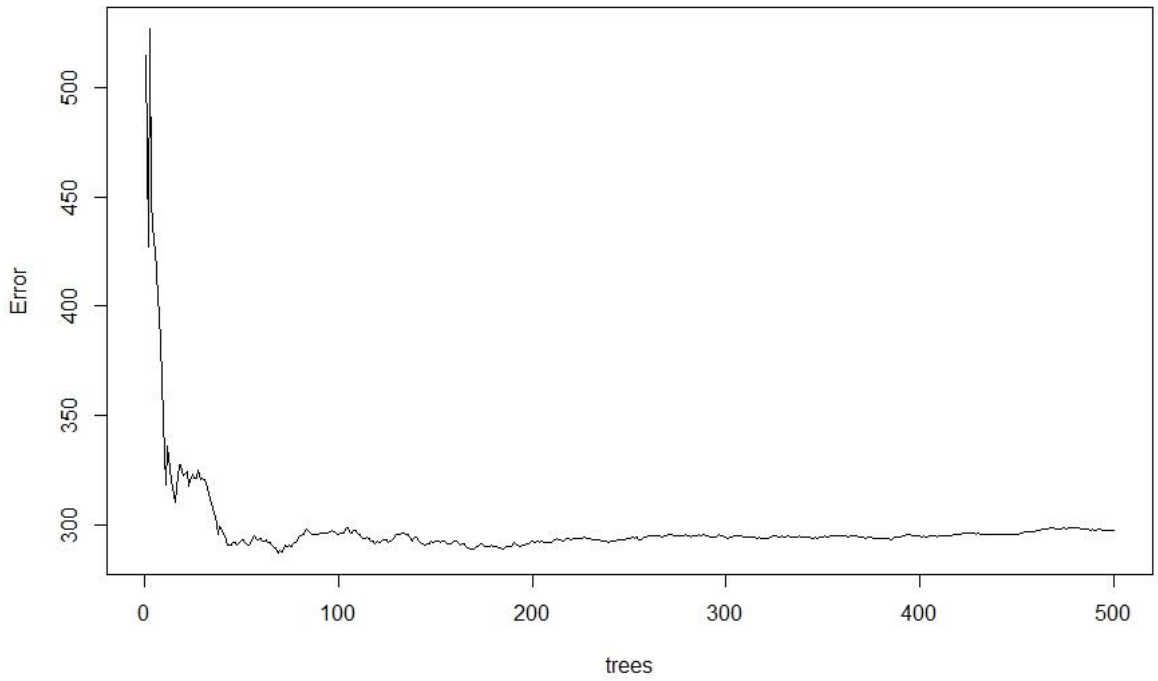


Figure 5.6C Plot of out-of-bag (OOB) error vs a number of trees (ntree) for ntree in random forest. Out-of-bag error is the mean prediction error on the sample that was not used in the random forest training. The tree with the lowest RMSE is chosen for the final model.

Table 5.6C mtry values with their respective RMSE. mtry of 2 has the lowest RMSE of 16.46 kg DM $ha^{-1}day^{-1}$

mtry	*RMSE (kg DM $ha^{-1}day^{-1}$)*
2	*16.46*
4	*16.60*
6	*16.74*
8	*16.87*
11	*17.01*

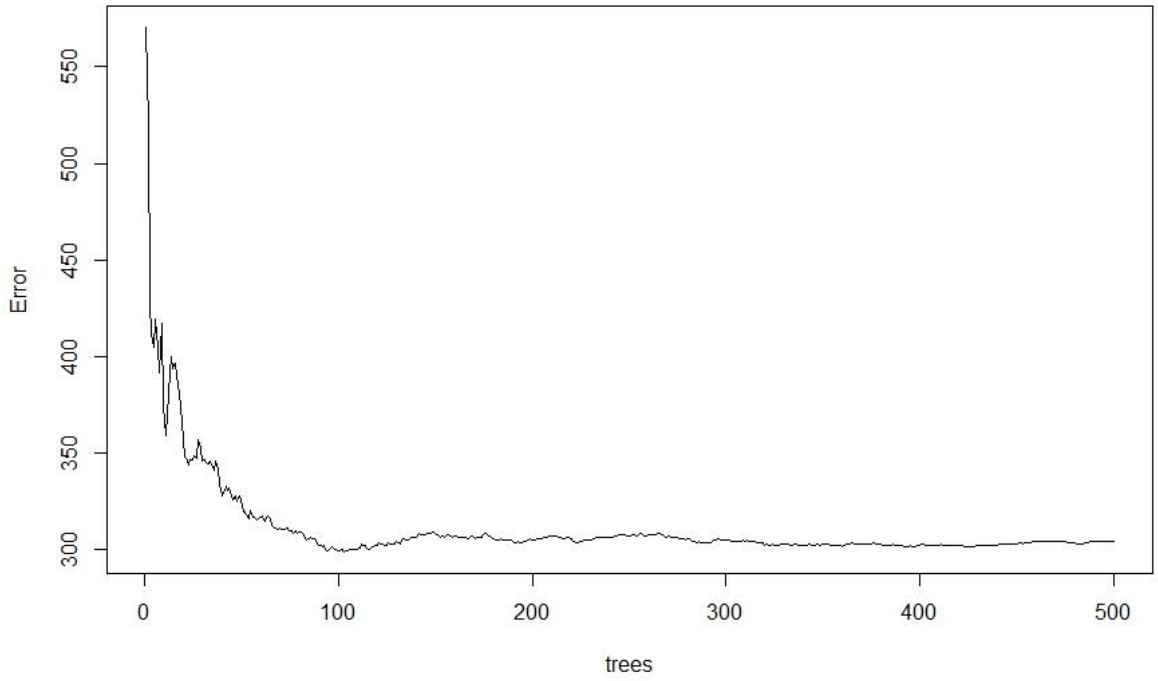


Figure 5.6D Plot of out-of-bag (OOB) error vs a number of trees (ntree) for ntree in random forest. Out-of-bag error is the mean prediction error on the sample that was not used in the training of the random forest. The tree with the lowest RMSE is chosen for the final model. At ntree=103, the RMSE is the lowest and is chosen for the RF model in the plot.

Table 5.5D mtry values with their respective RMSE. mtry of 2 has the lowest RMSE of 16.47 kg DM $ha^{-1}day^{-1}$

mtry	*RMSE (kg DM $ha^{-1}day^{-1}$)*
2	*16.47*
4	*16.60*
6	*16.79*
8	*16.87*
10	*16.98*

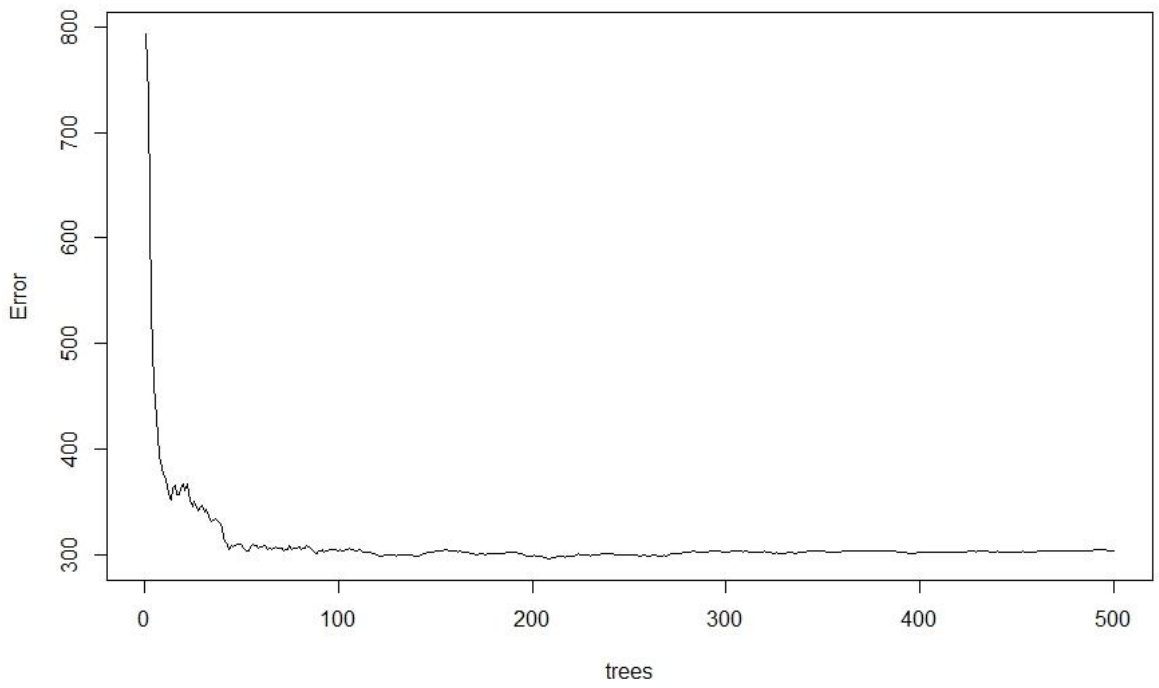


Figure 5.6E Plot of out-of-bag (OOB) error vs a number of trees (ntree) for ntree in random forest. Out-of-bag error is the mean prediction error on the sample that was not used in the training of the random forest. The tree with the lowest RMSE is chosen for the final model. In the plot, at ntree=209, the RMSE is the lowest and is chosen for the RF model.

Table 5.6E mtry values with their respective RMSE. mtry of 2 has the lowest RMSE of 16.51 kg DM $ha^{-1}day^{-1}$

mtry	*RMSE*
2	*16.51*
3	*16.60*
5	*16.69*
7	*16.76*
9	*16.89*

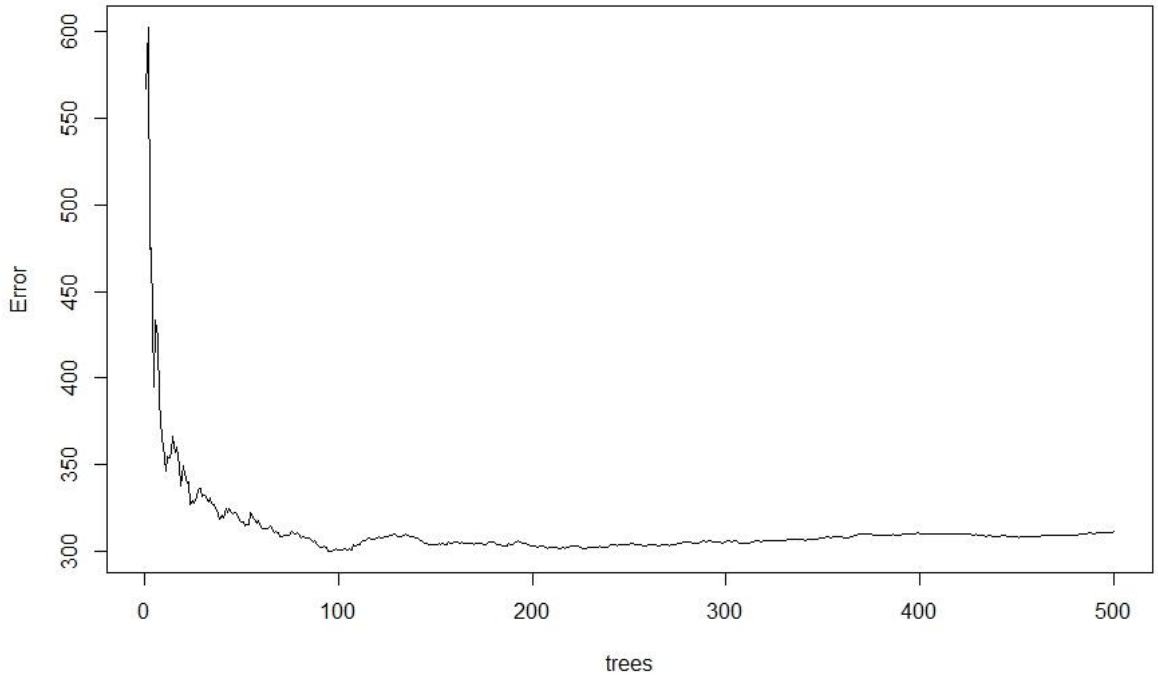


Figure 5.6F Plot of out-of-bag (OOB) error vs a number of trees (ntree) for ntree in random forest. Out-of-bag error is the mean prediction error on the sample that was not used in the training of the random forest. The tree with the lowest RMSE is chosen for the final model.

Table 5.6F mtry values with their respective RMSE. mtry of 2 has the lowest RMSE of 16.54 kg DM $ha^{-1}day^{-1}$

mtry	*RMSE*
2	*16.54*
3	*16.62*
5	*16.71*
7	*16.81*
9	*17.00*

Table 5.6G Comparing four models with R^2 for training and testing, optimal trees, grass growth rate value at optimal tree and mtry are shown. Model 9 contains all the variables; model 10 contains all the variables except rainfall; model 11 contains all the variables except NDRE and rainfall; model 12 contains all the variables except NDRE and rainfall.

Model	*Tree*	*Value*	*mtry*
9	*69*	*16.94*	*2*
10	*103*	*17.28*	*2*
11	*209*	*17.19*	*2*
12	*96*	*17.31*	*2*

5.7 Grass growth curve plots from model 10 (RF) (Chapter 5)

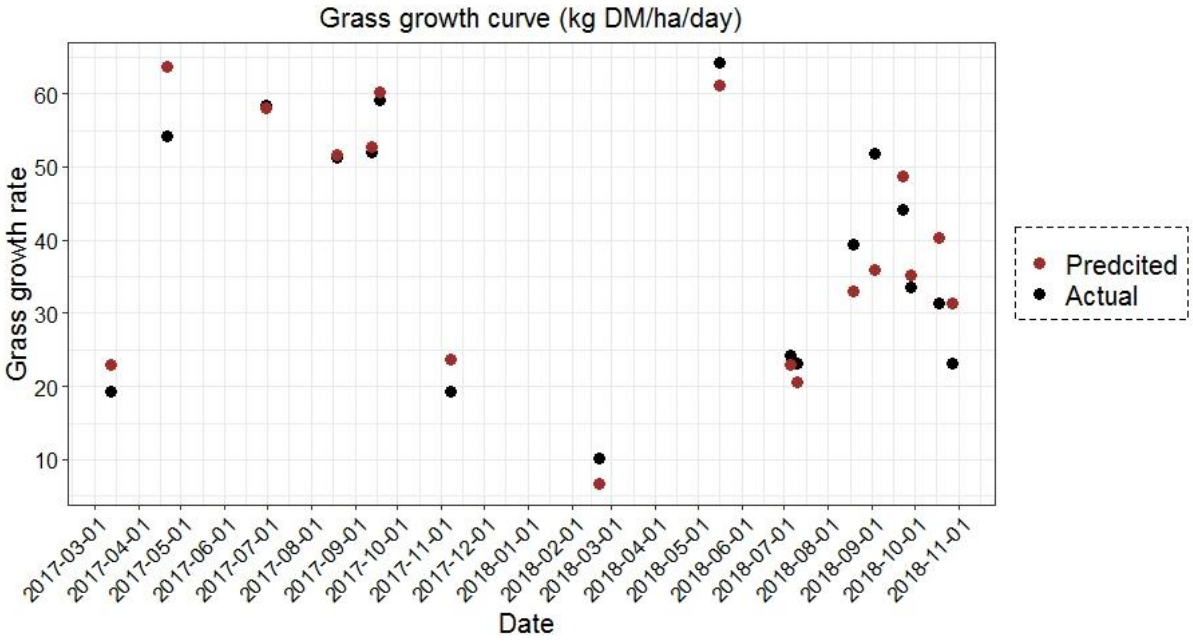


Figure 5.7A Grass growth curve from PBI (Black) and predictions from RF model 10 (red) for 2017 and 2018 for Curtin farm (Cork cluster)

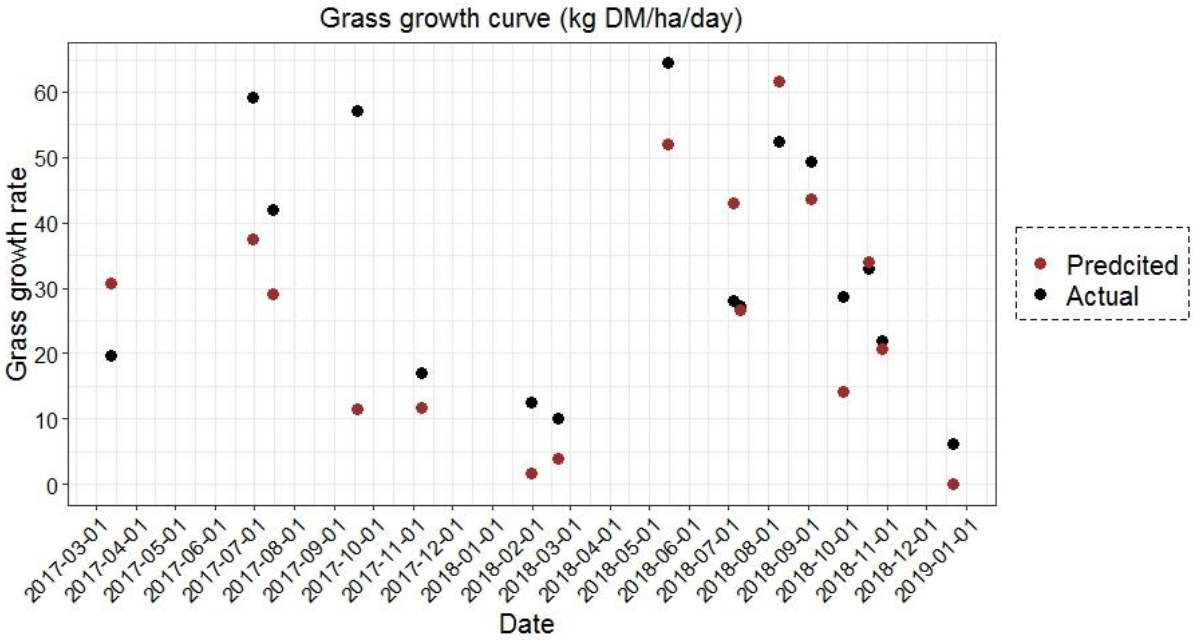


Figure 5.7B Grass growth curve from PBI (Black) and predictions from RF model 10 (red) for 2017 and 2018 for Kilworth farm (Cork cluster)

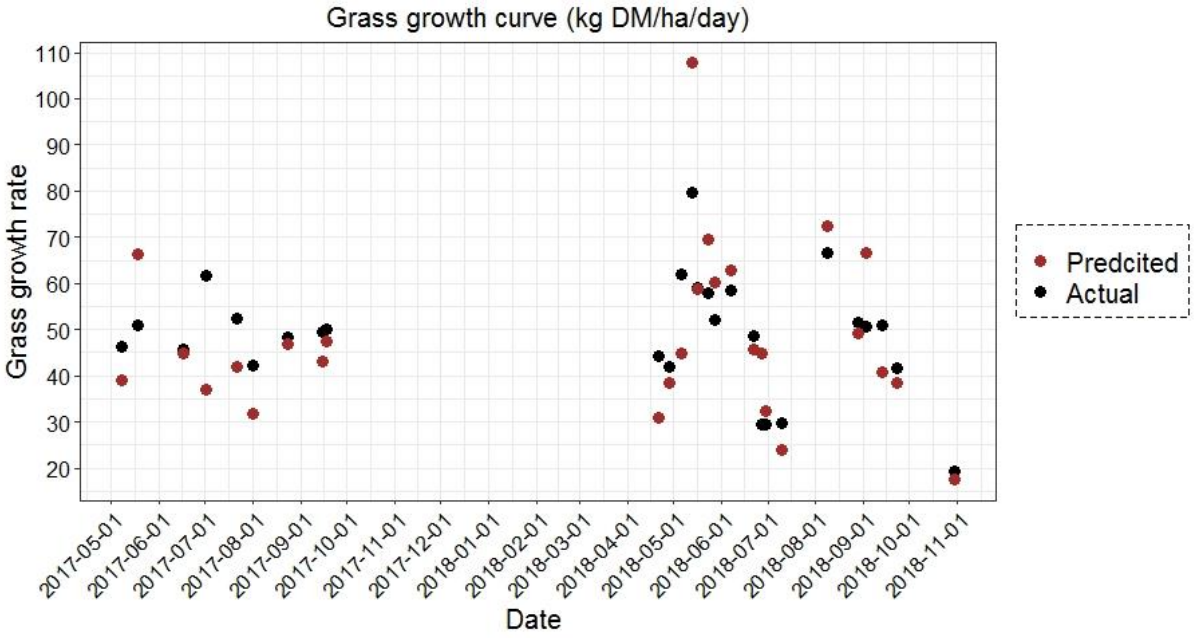


Figure 5.7C Grass growth curve from PBI (Black) and predictions from RF model 10 (red) for 2017 and 2018 for Kildavin farm (Wexford cluster)

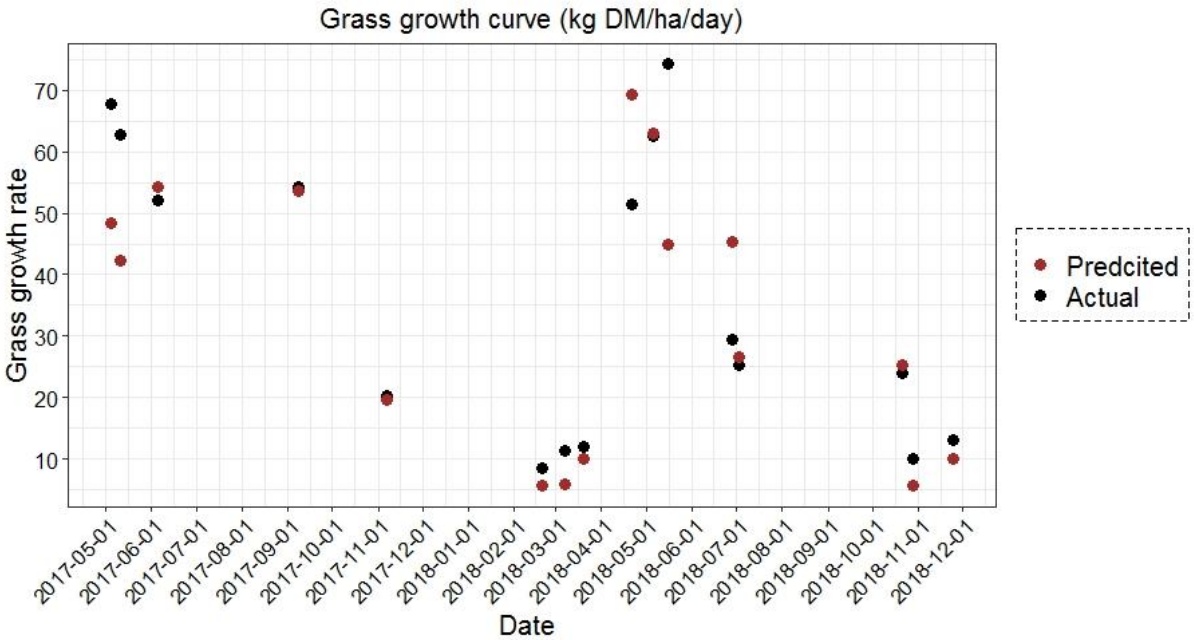


Figure 5.7D Grass growth curve from PBI (Black) and predictions from RF model 10 (red) for 2017 and 2018 for INZAC farm (Galway cluster)

6.1 Residuals of grass growth rate from PBI & model predictions (Chapter 6)

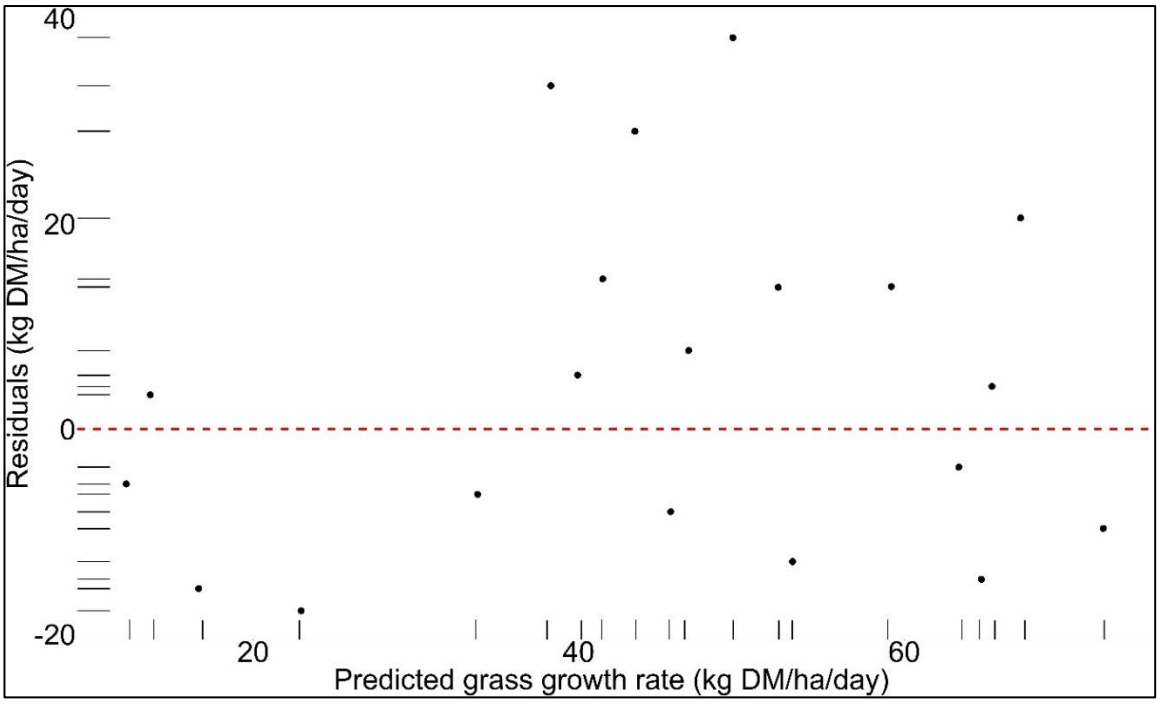


Figure 6.1A Residual plot between residual and predicted grass growth rate values from random forest (Commercial model)

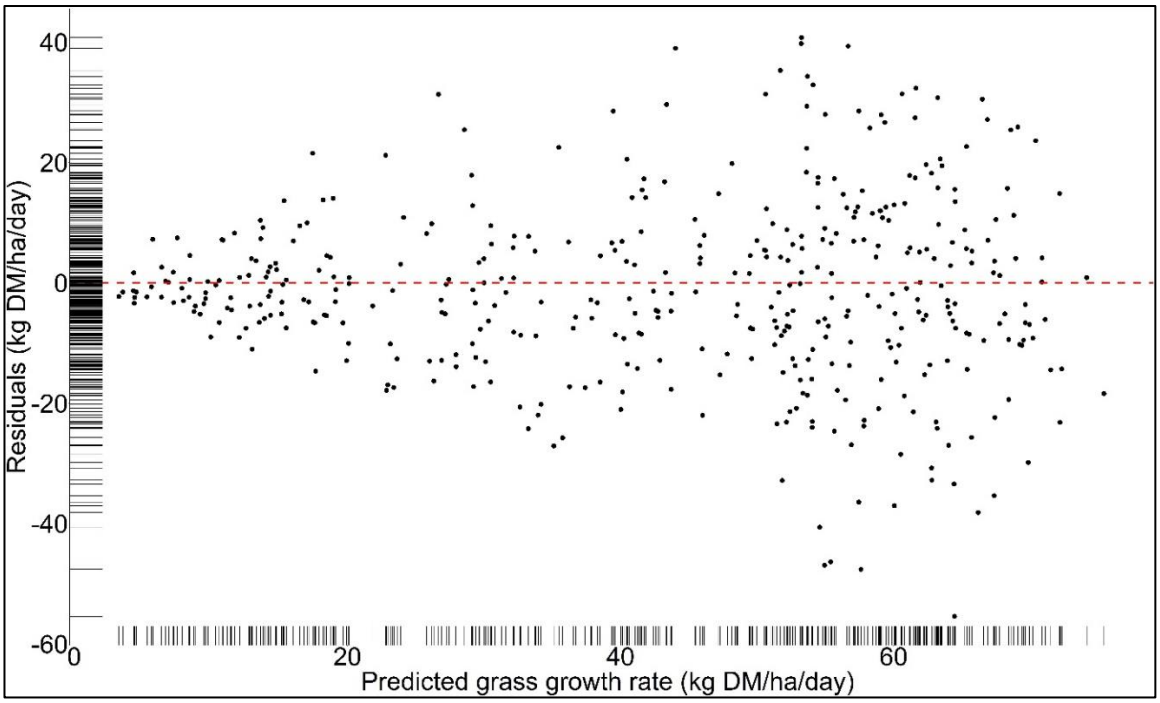


Figure 6.1B Residual plot for the model with lowest RMSE and highest R^2 for Agmodel (test data)

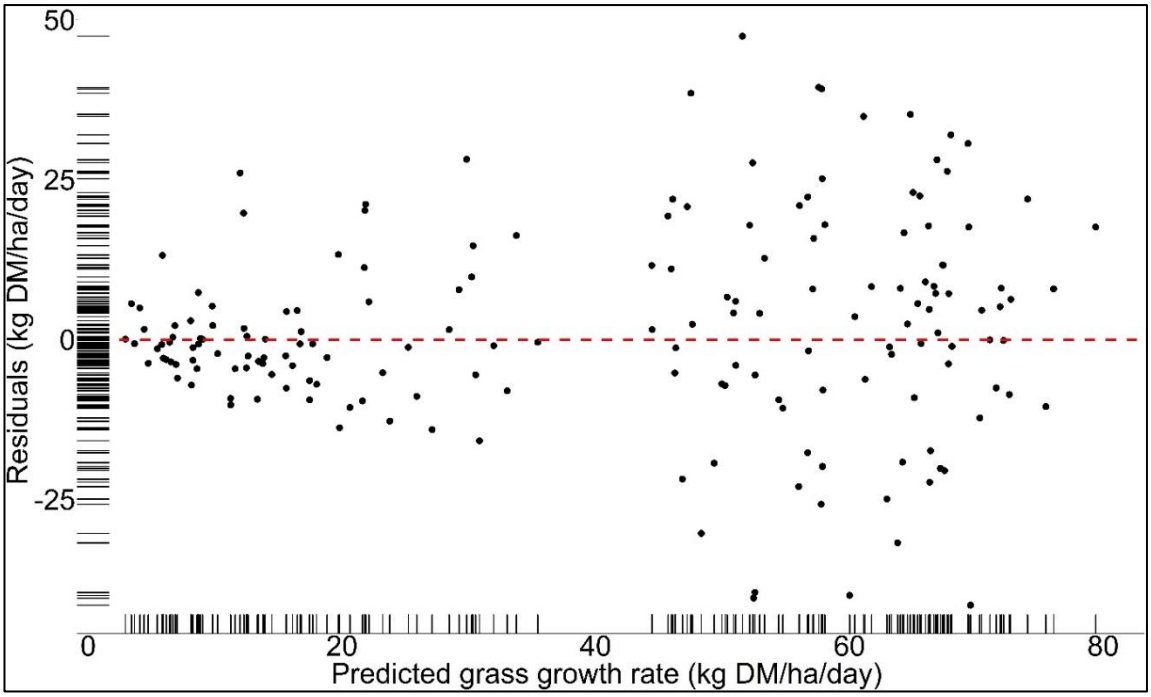


Figure 6.1C Residual plot for predicted and actual grass growth rate for test data for Agmodel1

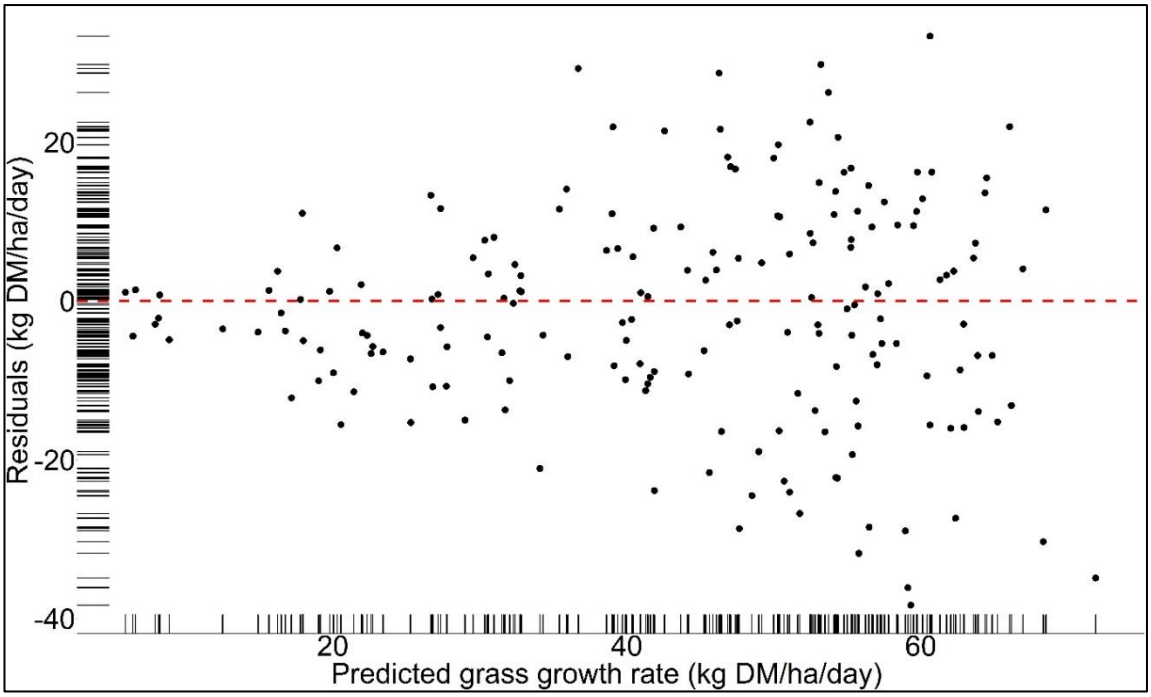


Figure 6.1D Residual plot for predicted grass growth rate from Agmodel 2 for test data

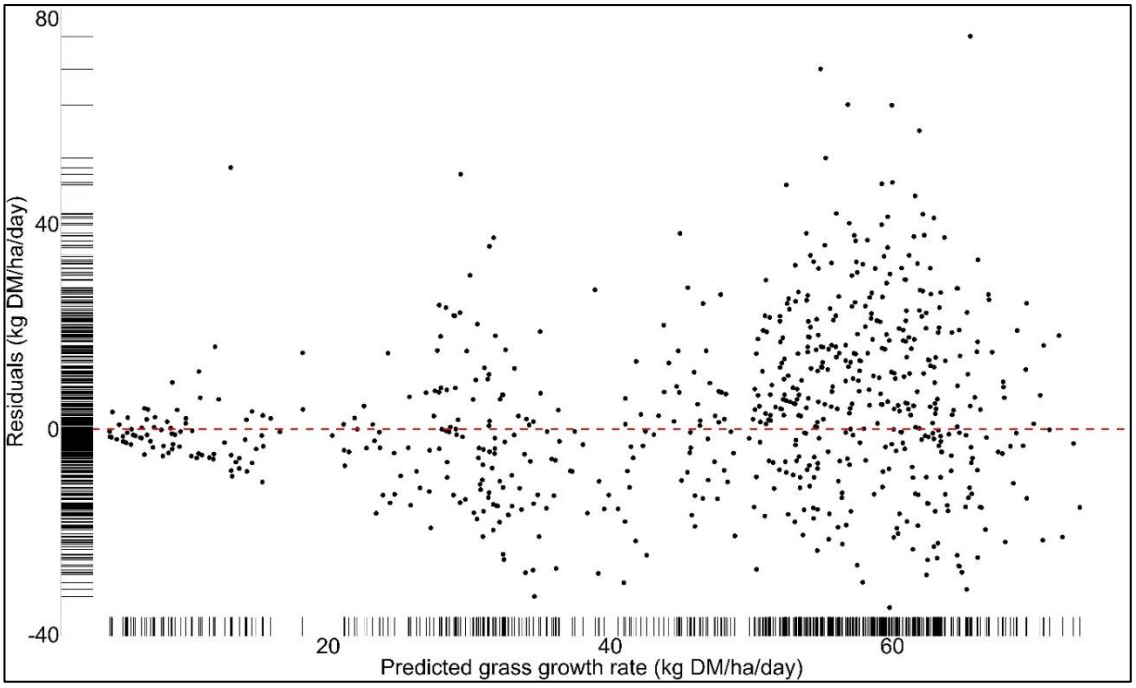


Figure 6.1E Residual plot for prediction from Agmodel for 2020 data. The x-axis represents the predicted grass growth rate from Agmodel, and the y-axis is the residuals (kg DM ha^{-1}day^{-1})

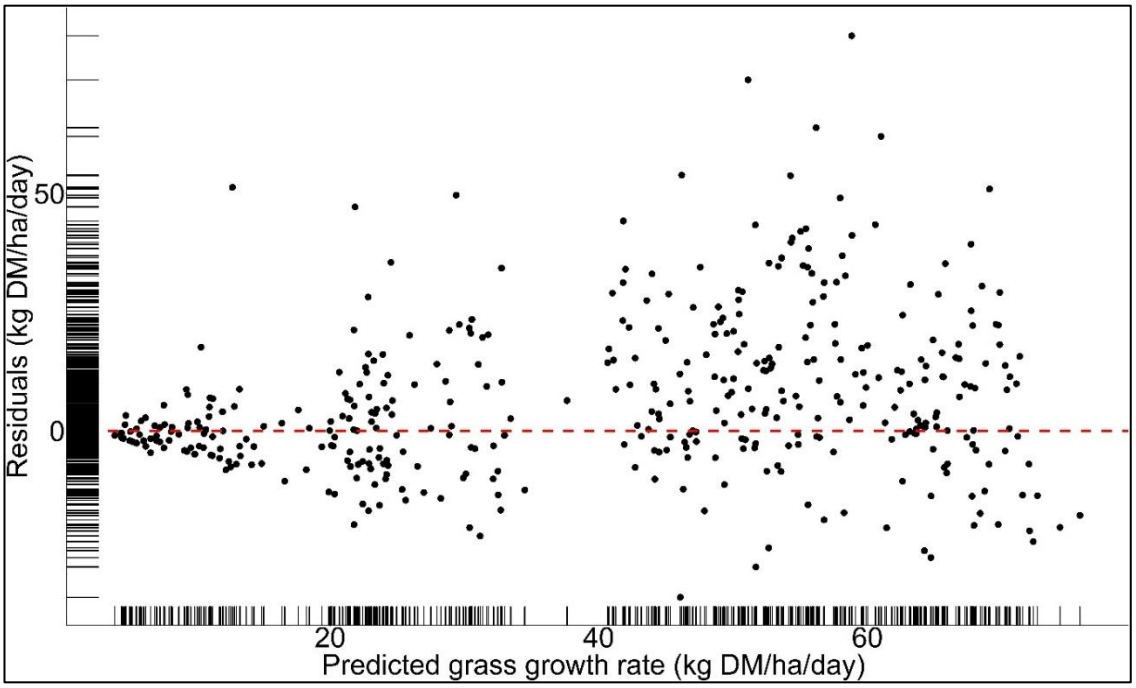


Figure 6.1F Residual plot for predicted grass growth rates values for 2020 data using Agmodel 1

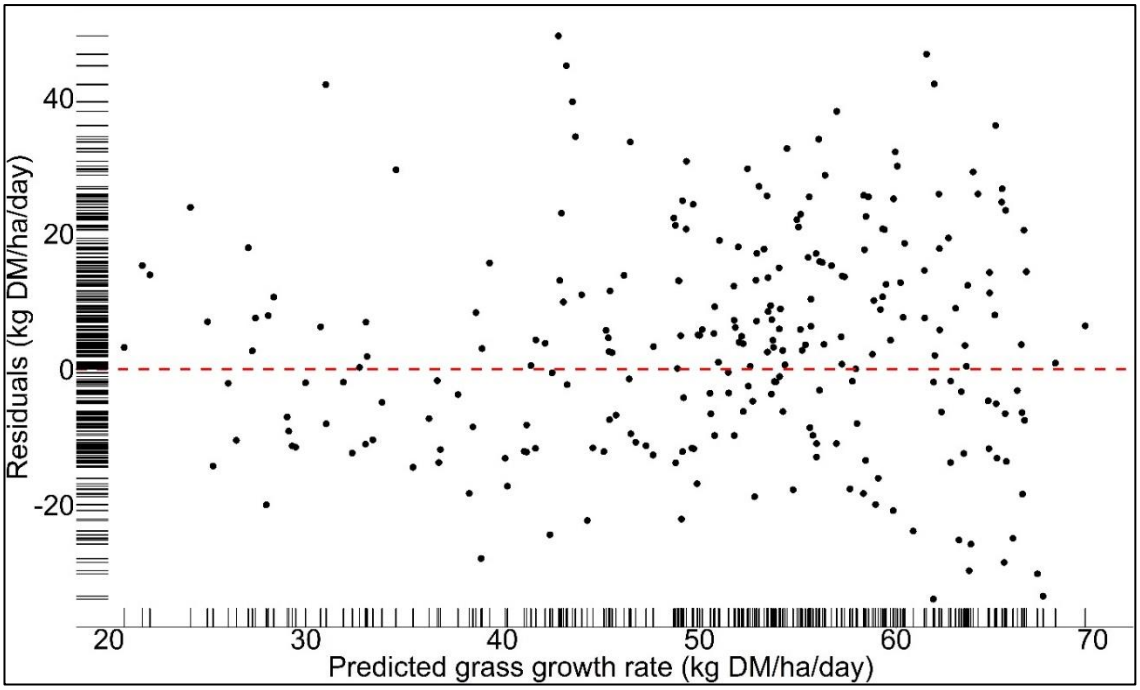


Figure 6.1G Residual plot for predicted grass growth rate (kg DM $ha^{-1}day^{-1}$) from Agmodel 2 for 2020 data.

6.2 Random Forest for national model

```
## Load package
library(ranger)
library(ggplot2)
library(ggpubr)
library(dplyr)
library(GGally)
library(ggcorrplot)
library(ggpmisc)

## Reading data and standardization
dataset <- read.csv(file=Path to the file)
dataset<-subset(dataset, Growth >0 & Growth <=100)
pp_boxcox <-preProcess(dataset[,-1],method=c("BoxCox","center","scale"))
AllFeatures<-predict(pp_boxcox,dataset[,-1])

Dataset1<-dplyr::select(dataset[,],Variables)

nrFolds <- 5
set.seed(76418)
# generate array containing fold-number for each sample (row)
folds <- rep_len(1:nrFolds, nrow(Dataset1))
# actual cross validation
for(k in 1:nrFolds) {
  # actual split of the data
  fold <- which(folds == k)
  trainSets <- Dataset1[-fold,]
  tests <- Dataset1[fold,]
  # train and test your model with data.train and data.test
}
testSets<-tests[,-20]
gr<-tests[,20]

hyper_grid <- expand.grid(
  mtry       = 1:19,
  node_size  = 1:4,
  num.trees = seq(50,500,50),
  OOB_RMSE   = 0
)

system.time(
  for(i in 1:nrow(hyper_grid)) {
    # train model
    rf <- ranger(
      formula = Growth ~ .,
      data= trainSets,
      num.trees= hyper_grid$num.trees[i],
      mtry = hyper_grid$mtry[i],
      min.node.size  = hyper_grid$node_size[i],
      importance = 'impurity')
    # add OOB error to grid
    hyper_grid$OOB_RMSE[i] <- sqrt(rf$prediction.error)
  })
```

```
nrow(hyper_grid)
position = which.min(hyper_grid$OOB_RMSE)
head(hyper_grid[order(hyper_grid$OOB_RMSE),],5)

# fit best model
rf.model <- ranger(Growth ~ .,data = trainSets, num.trees = hyper_grid$num.trees[position],
importance = 'impurity', probability = FALSE, min.node.size =
hyper_grid$node_size[position], mtry = hyper_grid$mtry[position])

rf.model
b<-rf.model$variable.importance
trainSets$prediction<- predict(rf.model,trainSets)$predictions

testSets$prediction<- predict(rf.model,testSets,se.method ='infjack' )$predictions
d_test = gr-testSets$prediction
mse_test = mean((d_test)^2)
mae_test = mean(abs(d_test))
rmse_test = sqrt(mse_test)
df <- data.frame(gr,testSets$prediction)
R2_test =cor(x = df$gr, y = df$testSets.prediction, method = "pearson")^2
set.seed(76418)

residuals <- (gr - df$testSets.prediction)
```