Management of barley for improving yield and profitability in Northern Sweden

Skötsel av korn för att förbättra avkastning och lönsamhet i norra Sverige

RJN Project 5/2022

Final report

Junxiang Peng

SLU

Summary

The project aim was to develop remote sensing-based models to map barley yield estimation and identify limiting factors. A pre-experiment was conducted in 2022 and concluded that plant nitrogen (N) uptake (*PNU*, kg ha^{-1}) and soil water content (*SWC*, %) are two main limiting factors for barley growth at the Röbäcksdalen field station. Field experiments in 2023 and 2024 focusing on N treatments and *SWC* were implemented in several barley fields. Different N topdressing strategies were applied to the field and several crop and environmental variables including *SWC* were measured. Remote sensing data including drone and satellite data were also collected. The experiment results confirmed that *PNU* and *SWC* are among the main factors affecting barley yield. Models were developed to estimate *PNU* and *SWC*. A parallel work was performed to build a barley yield estimation model using remote sensing data. Even though model accuracy needs to be improved, the remote sensing approach shows potential to track barley crop growth status and give suggestions for farmers to manage their fields more profitably.

1. Introduction

Barley is an important crop in Sweden, as it represents approximately 30 percent of the total grain harvest of the country[1]. It is especially important in Northern Sweden, where it is the major cereal crop. It is likely that the agronomic potential of barley will increase in Northern Sweden in the coming decades with the current global warming trend, reinforcing the strategic importance of this crop. However, there is still a large variability in yield, between farms, between the fields of the same farm, and within fields. The yield of barley can be largely affected by various factors due to specific weather conditions found at high latitudes (e.g. short growing seasons with very long days, low temperatures, early or late frost risks, etc.), along with more generic limiting factors such as inappropriate nitrogen levels, waterlogging, etc. Remote sensing techniques (drone and satellite), relying on the interaction of light with vegetation features, such as the canopy chlorophyll content or the accumulated biomass, could be used to monitor crop status in a quick, accurate and low-cost manner. In this project, we focused on the use of remote sensing digital data and various modelling approaches to map yield variations and understand the dynamics of the limiting factors. This knowledge was then applied to test in-field adaptation strategies, validating the assumptions developed with remote sensing.

2. Material and methods

2.1 Field experiment

A preliminary experiment, focusing on identifying the barley growth limited factors, was conducted in 2022. Two fields with undersown grass were utilized in the experiment. In each field, three spots identified as high, medium and low yielding places determined by using a yield map from the previous year were chosen to collect the data. Two more spots with contrasting yield potential identified by using in-season NDVI data were selected for data collection as well. According to the results, nitrogen (N) and waterlogging were recognized as two main limiting factors for barley growth.

Therefore, the field experiments in 2023 (3 fields) and 2024 (2 fields) focused on these two factors. In addition to the typical fertilization rate (manure plus 20 kg N/ha fertiliser nitrogen at sowing), three additional chemical fertilization rates were applied in different strips in the fields:

- Treatment1 (T1): 30 kg N/ha topdressing directly after sowing.
- Treatment2 (T2): 30 kg N/ha topdressing at growth stage 31-32 (under Stem elongation), approximately at the end of June.
- Treatment3 (T3): 15 kg N/ha topdressing directly after sowing, plus 15 kg N/ha topdressing at growth stage 31-32, approximately at the end of June.

Strips were approximately 24 m wide, and the length was the width of each field. In each treatment strip, two points with relatively higher and lower elevations, identified by drone-based digital surface model (DSM), were chosen to install tubes to measure relative soil water content (*SWC*, %) using the Diviner 2000 sensor (Sentek Sensor Tech., Stepney, Australia) (Figure 1).

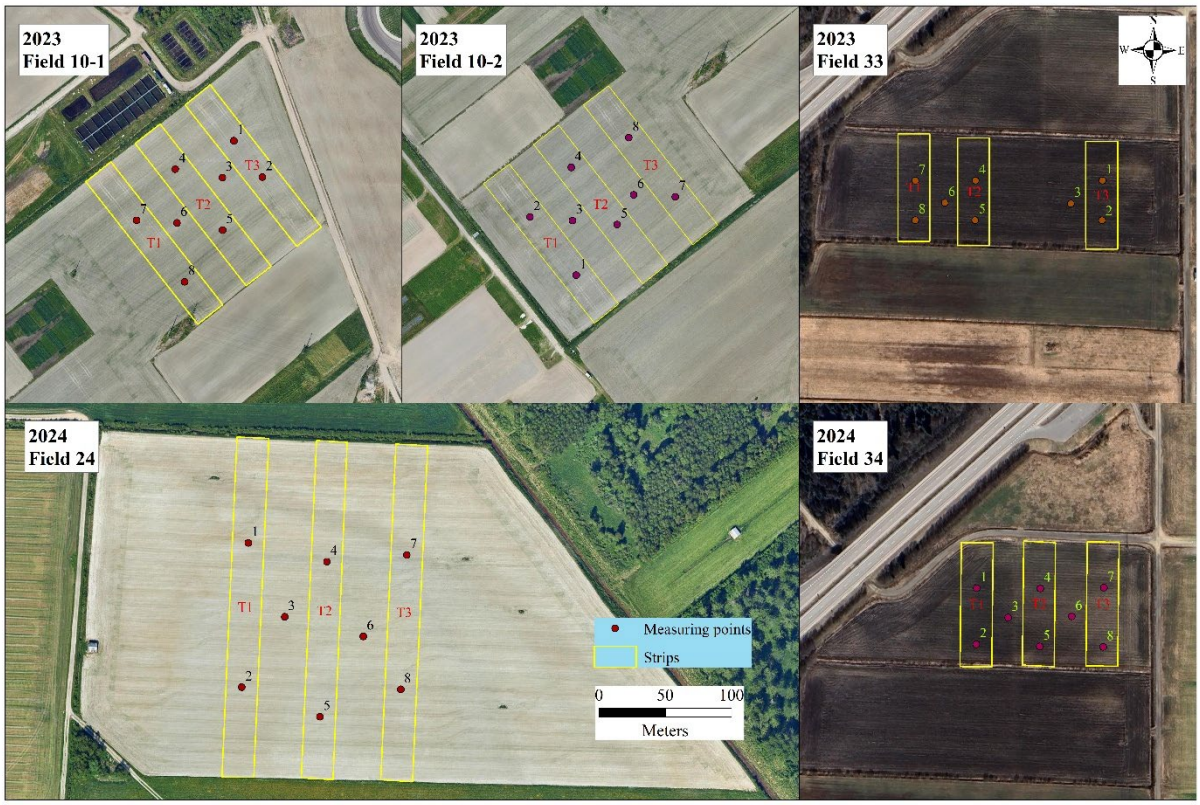


Figure 1. Layout of the nitrogen topdressing treatment strips and soil water content (SWC) measurement locations in 2023 and 2024.

In each year, crop sampling was conducted to collect barley samples within a 50×50 cm^2 quadrat, four times (end of June, stem elongation; middle of July, flowering; middle of August, ripening; and once just before harvest, senescence). The coordinates of the four corners of the quadrat were measured using Emlid Reach RS3 RTK (Emlid Tech Kft, Budapest, Hungary). During the sampling, the leaf area index and fraction of intercepted photosynthetic active radiation were measured using a Sunscan SS1 (Delta-T Devices Ltd, Cambridge, UK). The barley fresh matter was measured in the lab directly after the sampling, and then the samples were oven-dried at 60 °C for 48 hours. In each season, the samples from the last two sampling instances were threshed and the grain fresh and dry matter measured. At the end of each season, the barley crop samples from the first two samplings and grain samples from the last two samplings were analysed to determine the N concentration.
The harvest yield data for each field were collected using the Trimble FMX precision harvesting system (Trimble Inc., Westminster, Colorado, USA). The extreme yield outliers were removed using Interquartile Range (IQR). The yield data were then interpolated from point shapefile to raster file using a Kriging approach in Arcmap 10.8.2 (Esri, Redlands, California, USA).

2.2 Remote sensing data collection

A MAIA S2 camera (SAL Engineering, Russi, Italy) with nine spectral bands (Table 1) was mounted on a DJI Matrice 210 RTK drone (DJI, Shenzhen, China) to collect multispectral images every week over the experimental fields. The MAIA S2 raw images were pre-processed by radiometric correction using the irradiance light sensor (ILS) and by cropping the "No data" border. The pre-processed images were then aligned and mosaiced using Agisoft Metashape (Agisoft LLC, St. Petersburg, Russia), following the radiometric calibration using the professional calibration panel and geographic calibration using the black-white panels placed at the field corners, where the exact coordinates were measured. A DJI Mavic 3T drone (DJI, Shenzhen, China) with installed RGB camera and thermal camera was used to collected thermal images weekly over the experimental fields. The thermal raw images were pre-processed to convert pixel values from digital numbers to temperature in Celsius unit using Atygeo thermal software (Atyges store, Málaga, Spain). The thermal images were then aligned and mosaiced using Agisoft Metashape. Five hand-made panels (NRA, Umeå, Sweden) with colours of white, 25% black, 50% black, 75 % black and black were used for the radiometric calibration. During the drone flights, the panels were placed within the field, and the instantaneous land surface temperature (*LST*) was measured using a hand-held thermometer (IR thermometer optris MSpro LT, Optris GmbH & Co. KG, Berlin Germany). A linear regression model was built based on the *LST* data measured by the thermometer and extracted from the mosaic thermal image and the linear regression model was used for the radiometric calibration. The geographic calibration was implemented using the same approach as for the MAIA S2 data.

Table 1 Spectral band information for the MAIA S2 camera

Band	Start wavelength (nm)	Stop wavelength (nm)	Central wavelength (nm)	Wavelength width (nm)
S1 - Violet	433	453	443	20
S2 - Blue	457.5	522.5	490	65
S3 - Green	542.5	577.5	560	35
S4 - Red	650	680	665	30
S5 – Red edge 1	697.5	712.5	705	15
S6 – Red edge 2	732.5	747.5	740	15
S7 - NIR 1	773	793	783	20
S8 - NIR 2	784.5	899.5	842	115
S9 - NIR 3	855	875	865	20

In addition to collection of the drone data, Sentinel-1 and Sentinel-2 satellite data in 2023 and 2024 over the experimental fields were downloaded from Google earth engine platform.

2.3 Barley yield estimation model

The barley yield was estimated using the CASA (Carnegie-Ames-Stanford Approach) model. The model calibration was implemented by linking barley yield and intercepted photosynthetically active radiation (*Ipar*, MJ m^{-2} d^{-1}) via radiation use efficiency (*RUE*, g MJ^{-1}). The relationship between barley yield and *Ipar* was affected by environmental constraints, which were also developed. In this study, this part of work was implemented by a thesis project named "A study of radiation interception and related barley yield estimation using Sentinel-2 imagery in Northern Sweden" which was funded by RJN and related to this project. For the details, readers are referred to the project report of the thesis project.

2.4 Early-season nitrogen content estimation using MAIA data

Given the hypothesis that N is one of the main limiting factors of barley growth in Northern Sweden, accurate early-season detection of the nitrogen content is crucial for precision N management, e.g., topdressing. In this study, we used drone images obtained from a MAIA S2 camera to estimate barley plant nitrogen uptake (*PNU*, kg m^{-2}), which was calculated by multiplying crop dry matter by the measured N concentration. Data collected from the first sampling instances in 2023 and 2024 (26th June 2023 and 25th June 2024) were used to calibrate the model. Data from field 24 on 25th June 2024 were not included due to missing drone data. Several vegetation indices were calculated (see table 4 in Peng et al. (2023)[2]). The sampling quadrat was used as the region of interest (ROI) to extract the values of each band's reflectance and vegetation index. A random forest (RF) algorithm was used to link the *PNU* and remote sensing data (reflectance from individual bands and values from vegetation indices); 75% of the dataset was used for model calibration and 25% was used for validation. The algorithm was implemented using "*randomForest*" package[3] in R environment[4].

2.5 Surface soil water content estimation using Sentinel-1 SAR and Sentinel-2 data

For exploring the effect from *SWC* on the final yield, the yield was extracted from the interpolated yield map using a circle buffer zone as the region of interest (ROI), which is a 50 cm radius around each *SWC* measurement tube. The measured *SWCs* were averaged at depths of 10 cm, 50 cm and 100 cm for each field across the season. The linear regression models

that linked the averaged *SWC* at each depth and the extracted yields were developed to test the effect of *SWC* on the final yield.

Sentinel-1 SAR data and Sentinel-2 multispectral data were used to estimate surface *SWC* (at 10 cm depth) following the method of Ma et al. (2020)[5], using the following equations:

$$\sigma_T^0 = \sigma_{veg}^0 + \tau^2 \sigma_{soil}^0 \quad (1)$$

$$\sigma_{veg}^0 = Am_V \cos(\theta)(1-\tau^2)(1-e^{-\alpha}) \quad (2)$$

$$\tau^2 = e^{-2Bm_V/\cos(\theta)} \quad (3)$$

$$\sigma_{soil}^0 = \frac{0.11SM^{0.7}[\cos(\theta)]^{2.2}[1-e^{-0.32(ks)^{1.8}}]}{0.095[0.13+\sin(1.5\theta)]^{1.4}[1-e^{-1.3(ks)^{0.9}}]} \quad (4)$$

Where, *SM* is the *SWC*, *k* is the wave number, *s* is the root mean of the surface height, θ is the incidence angle, *A* and *B* are empirical parameters depending on the canopy type, α is radar shadow coefficient depending on vegetation type and land use, m_V is the vegetation water content computed from NDVI which was calculated from Sentinel-2 data ($m_V = 0.078e^{3.51NDVI}$)[6]. σ_T^0 is the total backscatter (in this study, VV polarized backscatters) received by the Sentinel-1 satellite, which is the sum of the canopy and soil scatterings (σ_{veg}^0 and σ_{soil}^0). In this study, *SWC* from field 10 and 33 in 2023 and field 24 in 2024 were used to optimize the parameters (*ks*, *A*, *B*, and α) using "Solver" function in Excel and the data from field 34 in 2024 were used for validation.

3. Results and discussions

3.1 Final yield distributions

The yield maps interpolated from the point data downloaded from the Trimble FMX harvest machine using a Kriging approach are illustrated in Figure 2. It is apparent that there were yield differences within the field and among fields. The mean values of yield in each field in each year were 4056 kg ha^{-1}, 4368 kg ha^{-1}, and 4979 kg ha^{-1} for fields 10-1, 10-2 and 33 in 2023, and 3637 kg ha^{-1}, 3425 kg ha^{-1} for fields 24 and 34 in 2024. The average yield in 2023 was higher than 2024, likely a result of the environmental effects (e.g., weather differences and *SWC*).

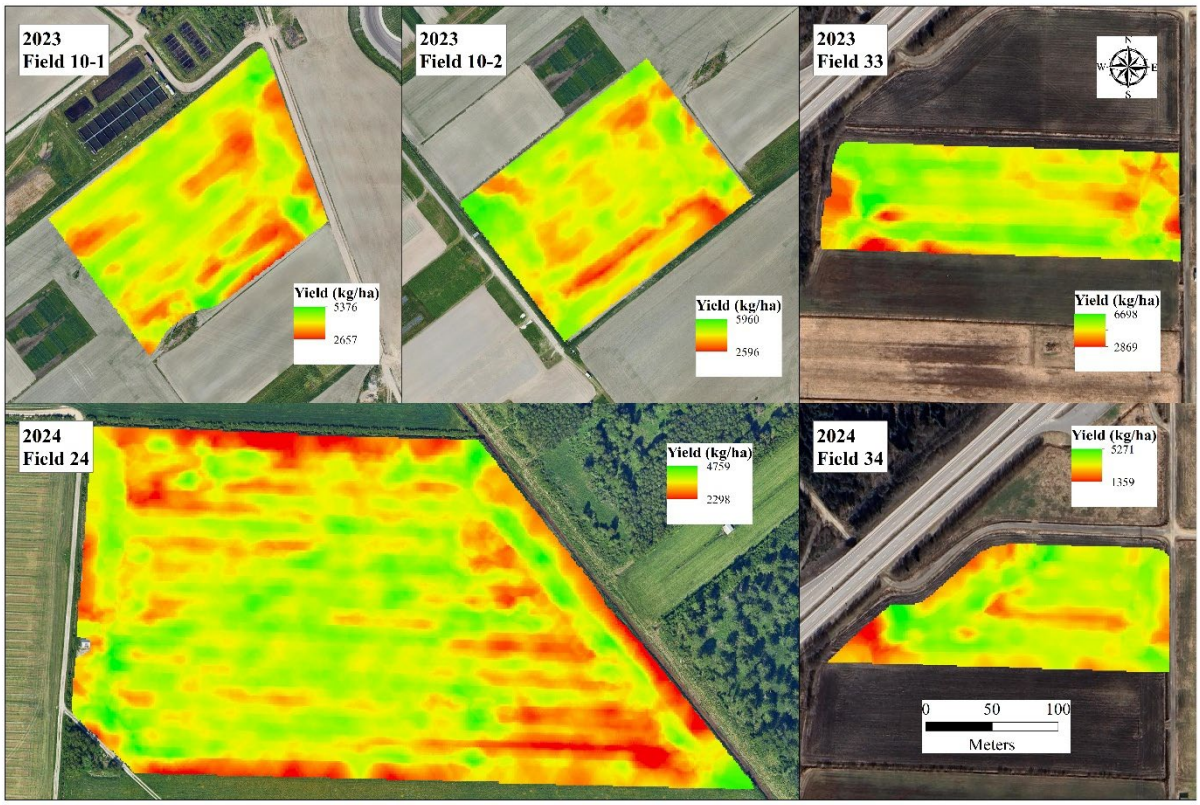


Figure 2. Layout of the yield distributions in different experimental fields in 2023 and 2024. It should be noted that the legend for each field is not identical, thus there are clear differences between different fields.

Figure 3 shows the differences in final yield for different treatment strips in different fields and years. The yields from Control (C) in each field were extracted from the areas between the main treatments in which the *SWC* measurement tubes were also installed (Figure 1). The yields from topdressing treatment strips were in general higher than the yield from the control (no extra N top-dressed), except the field 10-2 in 2023. The result clearly shows that extra topdressing could in general increase the barley yield in these fields. However, it is worthwhile to note that the responses of the yield to the topdressing treatments in each field were different. For example, in field 10-1 in 2023, the biggest improvement in yield was from T1 (topdressing at sowing), which showed that adding more N at the seed sowing time would increase the yield to a large degree; whereas, in field 34 in 2024, the biggest improvement of the yield occurred at T2, which indicated that topdressing at the stage of Stem elongation increased the yield to the largest extent. Future studies could explore the detailed mechanisms which influence the effects from topdressing on the yields.

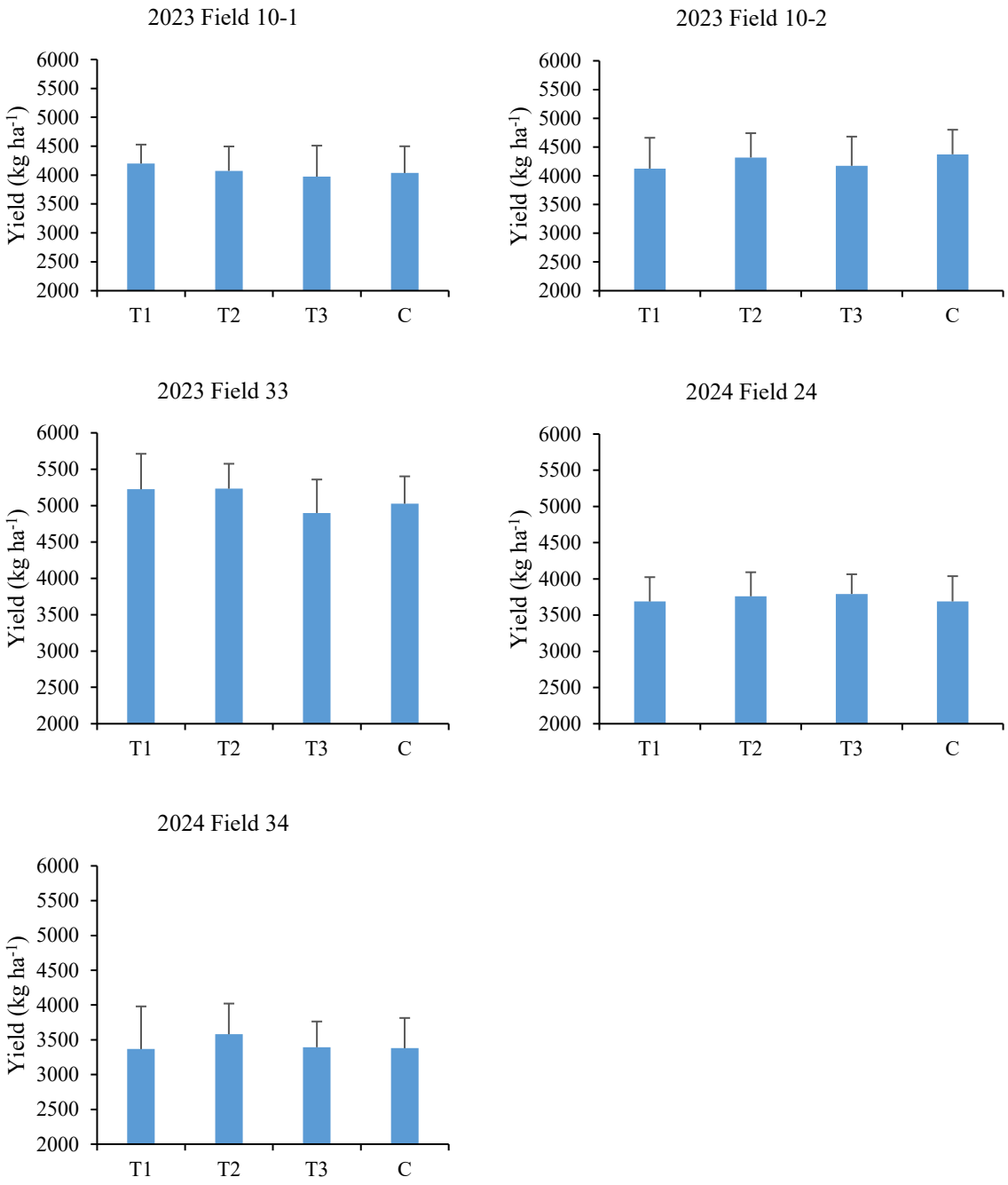


Figure 3. Final yield for different treatment strips in different fields and years. The details of the treatments (T1, T2 and T3) can be seen from section 2.1. The yields from Control (C) in each field were extracted from the areas between the main strip treatments in which the soil wate content (SWC) measurement tubes were installed as well (see figure 1). The error bars indicate standard errors of yield for each treatment.

3.2 Barley yield estimation based on remote sensing data

A CASA model was developed based on the historical barley yield data, weather data and downloaded Sentinel-2 data:

$$Yield = HI \times RUE_{opt} \sum_{i=1}^{n} Ipar_i \times f_{CI_i} \times f_{Tmax_i} \times f_{Ws_i} \times f_{moisture_i} \quad (5)$$

Where, yield is dry matter grain yield, HI is harvest index, which is set as 0.62, RUE_{opt} is the optimal radiation use efficiency, f_{CI}, f_{Tmax}, f_{Ws}, and $f_{moisture}$ are functions of cloudiness index, daily maximum temperature, water stress and soil moisture, respectively. Regarding the model accuracy, R^2 values were 0.51 and 0.30 for model calibration and validation, respectively. The model could potentially in terms of accuracies, and efforts should be put on this direction in the future. For the details of the model development and validation, readers are referred to the internship report of the project named "A study of radiation interception and related barley yield estimation using Sentinel-2 imagery in Northern Sweden".

3.3 Barley nitrogen content estimation based on remote sensing data

A RF algorithm regression model was developed based on the data (MAIA multispectral data and *PNU*) collected at the end of June in 2023 and 2024. The linear fits of the calibration and validation indicate the high robustness of the regression model; hence it is reliable enough to use this model to predict barley *PNU* early in the season (Figure 4).

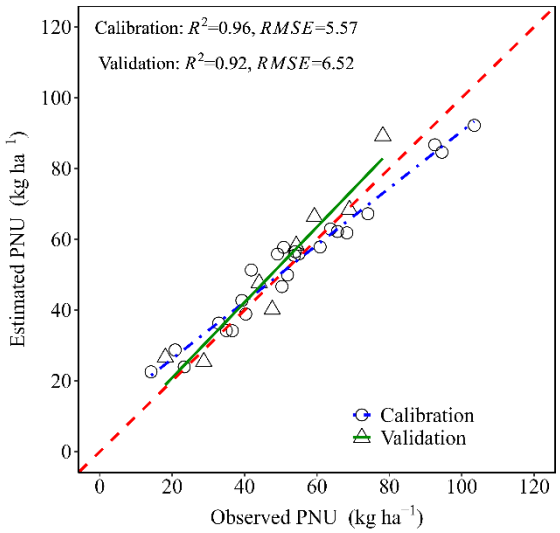


Figure 4. The linear fits between observed plant nitrogen uptake (PNU, kg ha^{-1}) and estimated PNU derived from the random forest (RF) regression model which was developed for PNU estimation using drone-based data for calibration and validation. The dotted red line is the 1:1 line.

Figure 5 illustrates the distributions of estimated *PNU* early in the season (end of June) across experimental fields in 2023 and 2024. The distributions of the estimated *PNU* only partially reflected the distribution of the final yields (Figure 2). Future studies could focus on the nitrogen dilution curve construction and nitrogen requirement calculation based on the estimated *PNU*[7].

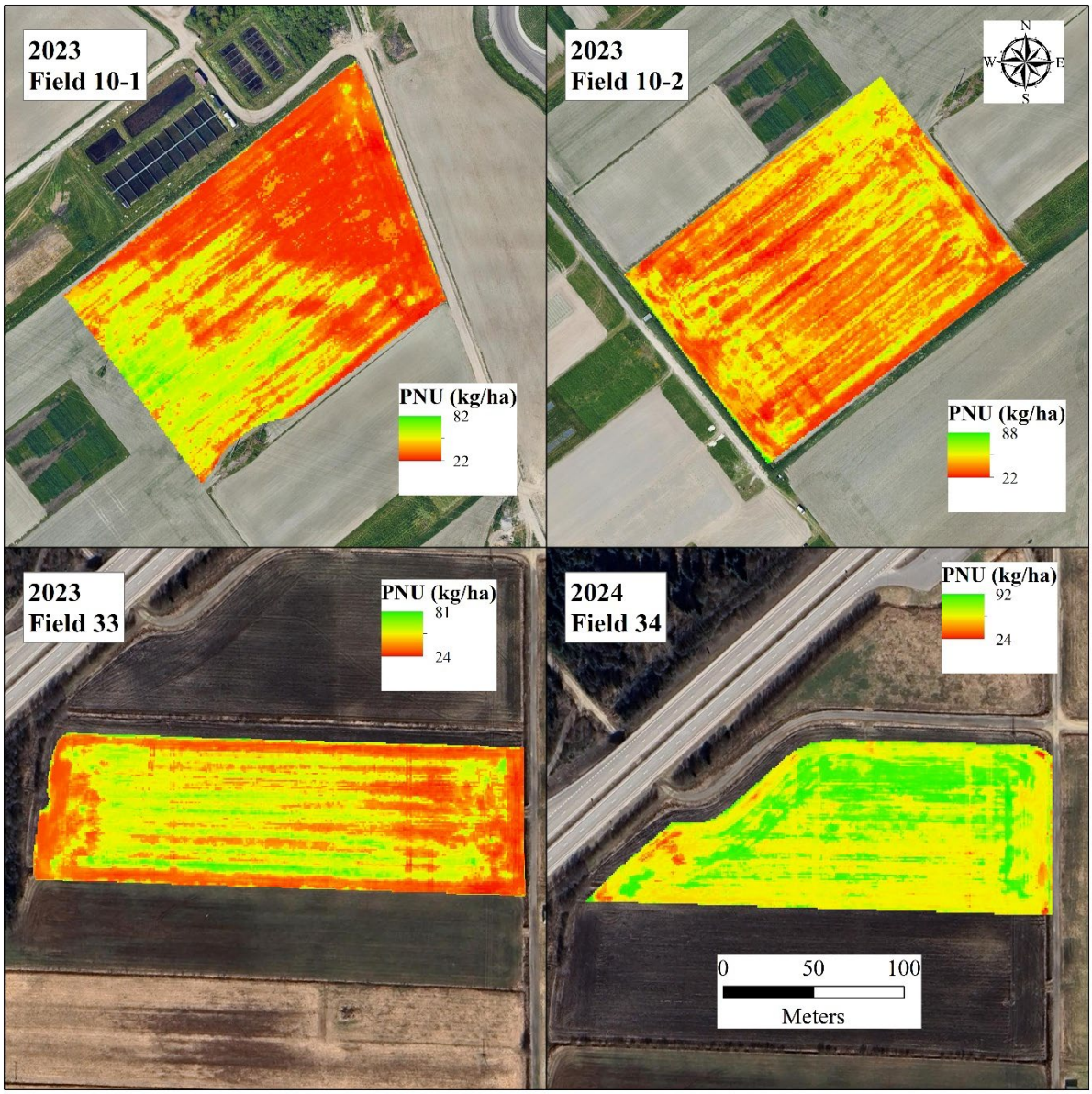


Figure 5. Distributions of estimated plant nitrogen uptake (PNU, kg ha^{-1}) in different experimental fields in 2023 and 2024. It should be noted that the legend for each field is not identical.

3.4 *SWC* estimation based on remote sensing data

The linear regressions between final yield and averaged seasonal *SWCs* (at depths of 10 cm, 50 cm and 100 cm) are shown in Figure 6. The linear fit between the averaged seasonal volumetric *SWC* and the final yield were not significant ($p > 0.05$) at the depth of 10 cm. However, the linear regressions between final yield and the averaged seasonal *SWC* were significant at 50 cm ($p<0.01$) and at 100 cm ($p<0.001$). The linear fits show apparent year difference, i.e. higher volumetric *SWC* and lower yield in 2024, and *vice versa* for 2023

(Figure 6). It was therefore confirmed that volumetric *SWC* is a factor affecting yield in these particular fields.

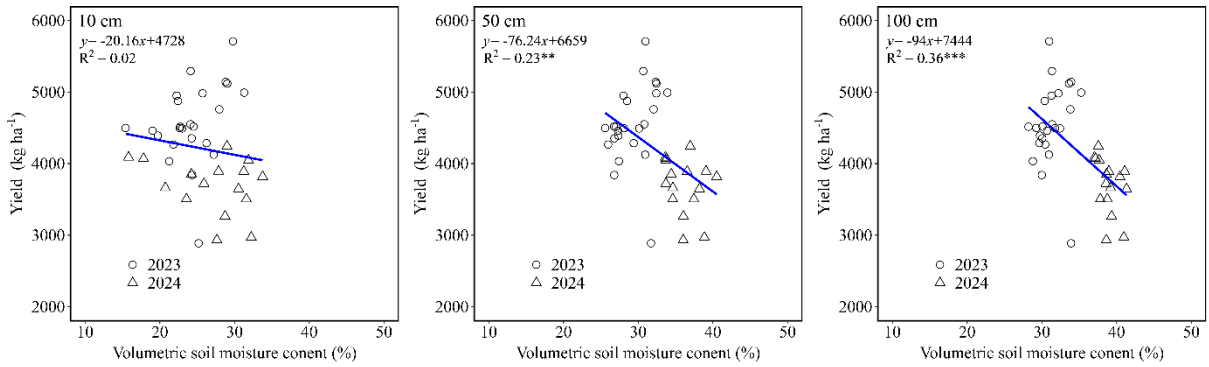


Figure 6. Linear regressions between the final yield (kg ha^{-1}) and the seasonal-average volumetric soil water content (*SWC*, %) at depths of 10 cm, 50 cm and 100 cm. The volumetric *SWC* was measured using a Sentek Diviner 2000 sensor.

Although the linear fit between the average seasonal *SWC* measured at 10 cm depth and the final yield was not significant, a remote sensing based surface *SWC* estimation model and parameters were trained and optimized using the *SWC* measured at 10 cm depth, since the Synthetic Aperture Radar (SAR) data from Sentinel-1 satellite has been proven to be sensitive to the surface *SWC* (from 3-10 cm)[5]. Following the equations 1-4 and the optimization using "Solver" function in Excel, the parameters were optimized (*ks*, A, B, and α are 1, -1.14, 0.007, and 10.59, respectively). The model validation shows that even though the linear fit between the observed and predicted *SWCs* was significant ($p<0.01$), the accuracy was low ($R^2 = 0.23$) (Figure 7), and the reason might be that the equations utilized in this study were calibrated using data from other locations rather than Northern Sweden[5]. The model parameters should be more carefully optimized based on the local specific conditions, especially for these particular fields where *SWC* is normally high. Another issue might be that the time overlapping between the Sentinel-1 SAR data and the measured *SWC* data was not exact. The passing time of the Sentinel-1 satellite over the experimental field was either between 5-6 am (Descending orbit) or between 16-17 pm (Ascending orbit) during a day, but the measuring time of the instantaneous *SWC* was mainly between 10-16 during a day. The Sentinel-1 SAR band (VV) values and incident angle are different at different passing times. A continuous measurement of the *SWC* might be helpful to address this issue.

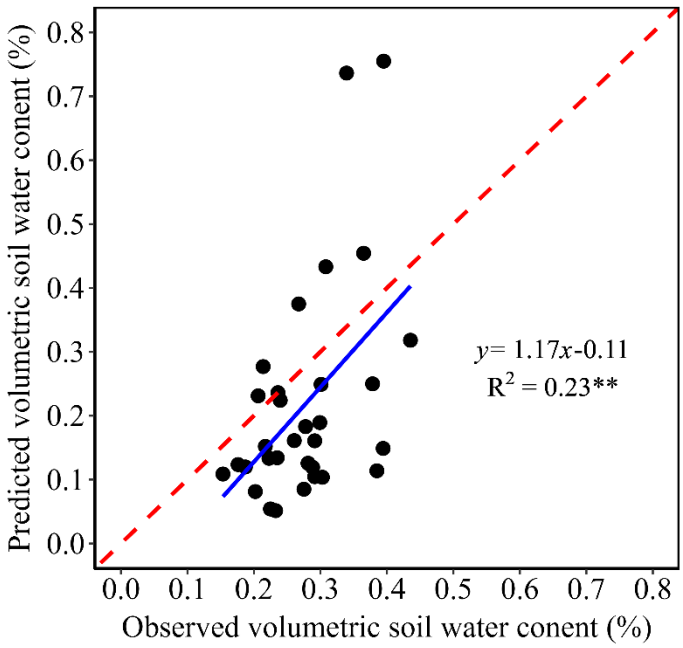


Figure 7. Validation of the volumetric soil water content (*SWC*, %) estimation in this study by linear fit between observed and predicted *SWC*.

The high *SWC* could result in lower barley yield; thus having properly functioning underground drainage pipes are important. Thermal images can be used to detect the function of existing drainage pipes, as the flowing water in the drainage pipes can make the land surface temperature lower than surrounding and this temperature difference can be detected in the thermal images. Thermal images obtained from drone mounted with a thermal camera have potential for assessing the function of the drainage pipes, such as presented in Kratt et al. (2020)[8]. The thermal images obtained in this study also can be used to detect the drainage pipes. In the example shown in Figure 8, the black line across the field might indicate the drainage pipe; however, this methods needs further development. Processing drone thermal images contains a lot of uncertainties such as the internal non-uniformity correction (NUC) and temperature drift, and environmental factors (e.g., air humidity and surface emissivity property). These uncertainties could affect the quality of the drone images processing, thus special attention should be given to the data collection and processing in the future[9].

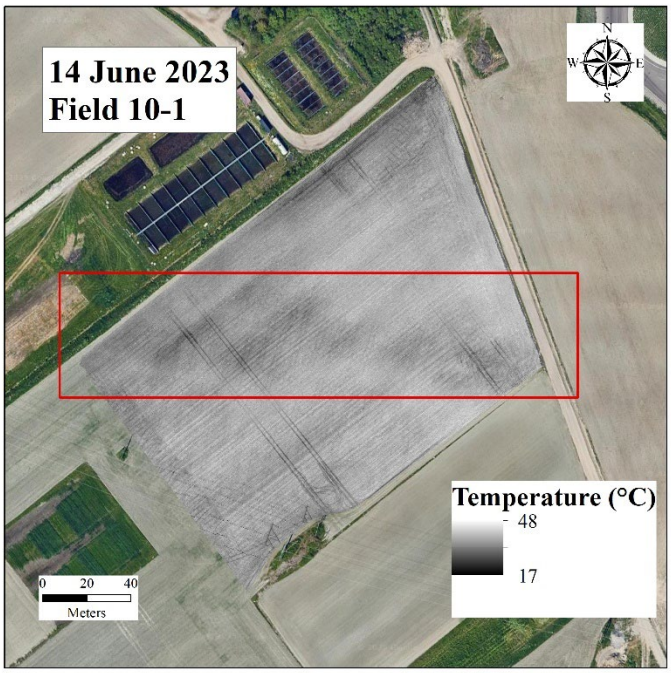


Figure 8. Example of the distribution of the land surface temperature obtained from the drone mounted with a thermal camera in this study. The red rectangle shows the potential location of the underground drainage pipe.

4. Summary

- For the fields studied, it was confirmed that *PNU* and *SWC* are two factors that influence patterns of in- and between-field yield differences.
- Drone data were used to build an accurate *PNU* estimation model. Satellite data also has potential to estimate *PNU* as well, which should be tested in the future. Existing platforms, such as Cropsat, which use satellite data to monitor crops could supply a similar service.
- It is possible to use satellite SAR data to estimate surface *SWC*, but the model parameters should be further optimized, and the validation accuracy needs to be improved.
- Future studies could focus on detailed strategies for topdressing and measures for waterlogging, to potentially improve barley yield in Northern Sweden.

5. Potential application of this research

- Two main limiting factors of barley growth in the study fields were confirmed by this study.

- Remote sensing data obtained from satellites and drones show potential to retrieve crop N status. A nitrogen fertilization strategy can be developed through using the Cropsat.se platform which utilizes satellite data. However, drone data could provide much detailed information, which could help farmers to manage nitrogen more precisely.
- With the development of more accurate remote sensing-based surface *SWC* estimation models in the future, it could be possible to provide information about potential waterlogging risks.

6. Scientific outputs and collaborations.

The following papers are planned:

- Machine-learning based estimation of the nitrogen content of barley from drone and Sentinel-2 multispectral data (drafted).
- Surface soil water content estimation using Synthetic Aperture Radar data from Sentinel-1 satellite (planned).
- Limiting factors of barley growth in Northern Sweden and potential measures to alleviate these factors (planned).

Implementation of the project was supported from SLU Röbäcksdalen Field Research Station, the Röbäcksdalen dairy research facility, and SITES-Röbäcksdalen. More specifically help was provided by Reija Danielsson, Viktor Brodin, Boel Sandström, Olav Tjernström, Håkan Nilsson, Victor Manabe, Sanna Bergqvist, Florentin Foucault, Camilla Destefanis, Weicai Yang, and Qihua Yu for field experiment operations, data collections and analysis.

References

1. https://www.lantmannen.com/good-food/four-kinds-of-cereal/barley/
2. Peng, J., Zeiner, N., Parsons, D., Féret, J. B., Söderström, M., & Morel, J. (2023). Forage biomass estimation using Sentinel-2 Imagery at high latitudes. Remote Sensing, 15(9), 2350.
3. Liaw, A., & Wiener, M. (2018). Classification and regression based on a forest of trees using random inputs, based on Breiman (2001). R Doc. Package Randomforest, 4, 14.
4. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
5. Ma, C., Li, X., & McCabe, M. F. (2020). Retrieval of high-resolution soil moisture through combination of Sentinel-1 and Sentinel-2 data. Remote Sensing, 12(14), 2303.
6. Gao, Y., Walker, J. P., Allahmoradi, M., Monerris, A., Ryu, D., & Jackson, T. J. (2015). Optical sensing of vegetation water content: A synthesis study. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(4), 1456-1464.

7. Peng, J., Manevski, K., Kørup, K., Larsen, R., & Andersen, M. N. (2021). Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. Field Crops Research, 268, 108158.
8. Kratt, C. B., Woo, D. K., Johnson, K. N., Haagsma, M., Kumar, P., Selker, J., & Tyler, S. (2020). Field trials to detect drainage pipe networks using thermal and RGB data from unmanned aircraft. *Agricultural Water Management*, *229*, 105895.
9. Peng, J., Nieto, H., Andersen, M. N., Kørup, K., Larsen, R., Morel, J., ... & Manevski, K. (2023). Accurate estimates of land surface energy fluxes and irrigation requirements from UAV-based thermal and multispectral sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, *198*, 238-254.

Agronomy Specialty

INTERNSHIP IN A PROFESSIONAL ORGANIZATION ABROAD

Engineering training Institut Agro Dijon

Initial training

Construction of a reliable barley yield estimation model based on remote sensing data

(Internship from 03/25/2024 to 08/09/2024)

Florentin FOUCAULT

Pr. Christelle GEE

Referent teacher

Pr. David PARSONS

Junxiang PENG

Internship tutors

Swedish University of Agricultural Sciences

(Sveriges lantbruksuniversitet, SLU)

Department of Crop Production Ecology

Skogsmarksgränd, 90183 Umeå, Suède

2023-2024

Table of contents

List of figures

List of abbreviations

F_{Ipar}: fraction of radiation useful for photosynthesis intercepted

GPS: Global Positioning System

Ipar: radiation useful for photosynthesis intercepted

NIR: Near-Infrared

NDMI: Normalized Difference Moisture Index

NDVI: Normalized Difference Vegetation Index

NMDI: Normalized Multi-band Drought Index

RTK: Real Time Kinematic

RVI: Ratio Vegetation Index

SWIR: Short-Wave Infrared

Introduction

The ‘management of barley for improving yield and profitability in Northern Sweden' project aims to understand and reduce the heterogeneity of barley yields in northern Sweden. Indeed, there is very little data to explain these heterogeneous yields in this region. The project is divided into three major phases.

In the first phase, the main objective is to collect and analyze aerial images to find relationships between yield heterogeneities and these images. Yields are derived from the combine harvester's controller on previous years' harvests. Models will thus be created to estimate barley yield throughout its growth cycle, and to determine when limiting factors appear.

The second phase aims to measure the factors that can influence yield directly in the field. It is based on remote sensing and physical and chemical measurements during barley growth.

Finally, the third phase aims to find and evaluate solutions to reduce yield heterogeneity in barley.

Ultimately, this project aims not only to explain the variations in barley yield in northern Sweden, but also to propose concrete solutions for improving crop productivity and profitability in this region (Peng et Parsons, 2022).

Here, we will focus on the yield estimation part of the study, using remote sensing from Sentinel 2 satellites.

I. Materials and methods

To estimate barley yield using Sentinel 2 satellite data, the first step is to collect satellite reflectance data from barley plots. Next, we'll look at the field data collection required to run and validate the model.

Once the data have been collected, the Ipar will be calculated over the entire barley growth period. This will be used directly to calculate barley yield, but before that, yield data from previous years will be processed.

A. Download satellite data

The first step is to retrieve RVI (Ratio Vegetation Index) data from barley plots harvested in previous years (2018-2023). RVI data are calculated from red and near infrared reflectances. These are retrieved from the Sentinel 2 constellation satellites, which pass over the Röbäcksdalen station every 2 days. Here, the project focuses on the B4 and B8 bands, corresponding to the red and near infrared respectively, as shown in Figure 1. Bands B8A, B11 and B12 will also be used in the future. Google Earth Engine is used to extract plot reflectances while filtering out pixels where clouds are present. This is because clouds have much higher

red and near infrared reflectances than vegetation and the ground, which distorts the data when they are present.

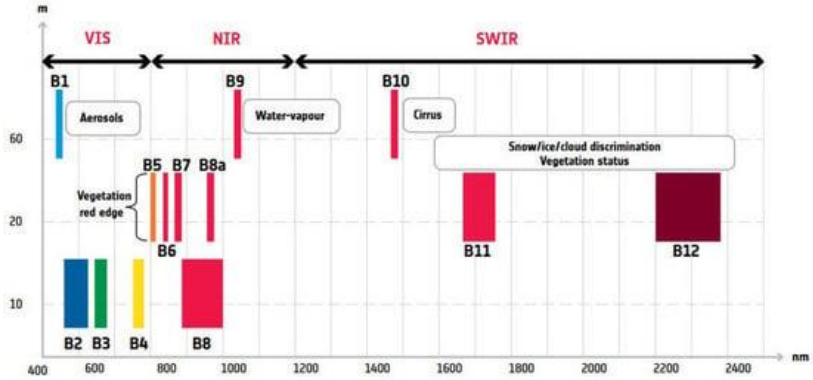


Figure 1 : Wavelengths visible from Sentinel 2 (Alparone et al.*, 2024)*

The reflectances are then imported into R, to calculate the RVI of each microplot via equation 1, where R corresponds to red reflectance and NIR (Near-Infrared) to near-infrared reflectance.

$$RVI = \frac{R}{NIR} (1)$$

In parallel to RVI, NMDI (Normalized Multi-band Drought Index) is also downloaded in the same way and in the same microplots. This is an index of soil and vegetation moisture, derived from reflectance at wavelengths of 860nm, 1640nm and 2130 nm (Wang et Qu, 2007), which can be associated respectively with bands B8A, B11 and B12 of the Sentinel 2 constellation (Nguyen et Binh, 2016). The B8A band is similar to the B8 band, but has a narrower bandwidth and a resolution of 20 m, just like the B11 and B12 bands, as shown in Figure 2. The NMDI calculation follows equation 2.

$$NMDI = \frac{B8A-(B11-B12)}{B8A+(B11-B12)} (2)$$

As shown in equation 2, NMDI is an index using 3 bands to determine vegetation and soil moisture. In fact, the difference between B11 and B12 varies according to humidity, being high when the soil is wet and low when it's dry. For vegetation, it's the other way around: when it's under water stress, the difference is high, and when it's wet, the difference is low (Wang et Qu, 2007). Figure 2 shows this difference in the dotted box, where band B11 corresponds to B6 and band B12 to B7.

Band B8A serves as a reference, since it varies little with humidity, as shown in Figure 2.

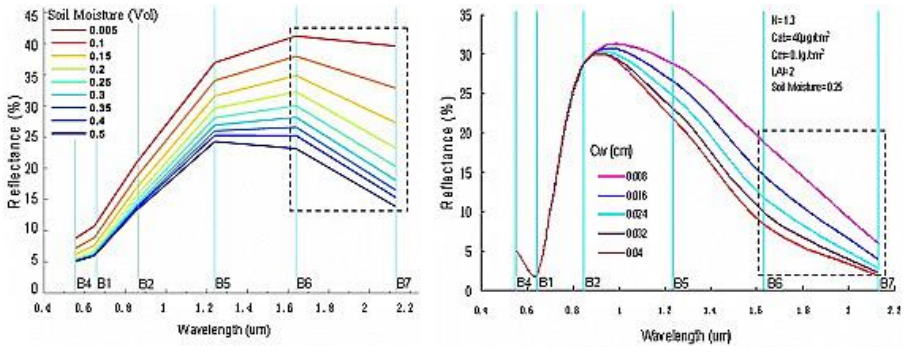


Figure 2 : Simulation of soil reflectance on the left and vegetation on the right as a function of wavelength and humidity (Wang et Qu, 2007)

NMDI is not very effective for average vegetation, so NDMI (Normalized Difference Moisture Index) is also downloaded in the same way in the same microplots to solve this problem. This is derived from bands B8A and B11 (Viehweger, 2023) and its calculation follows equation 3.

$$NDMI = \frac{B8A-B11}{B8A+B11} \quad (3)$$

B. Field data collection

There are many different types of field data, some of which are used in the model, such as meteorological data, while others are used to test the model.

The meteorological data came from the Röbäcksdalen weather station and were downloaded from the SLU website (Lundkvist, 2024). This station is in the middle of the experimental station's plots, and this made it possible to retrieve global radiation and temperature data.

Barley yield data are collected via a Trimble FmX yield controller mounted on the harvester. The harvester is also equipped with GPS (Global Positioning System) geolocation and autosteering with RTK (Real Time Kinematic) correction, which increases the accuracy of harvest data collection. This controller is also used to recover grain moisture to calculate dry weight. However, processing of the controller's data is necessary, which will be explained in another section.

Other data are collected during barley growth, notably the fraction of radiation useful for photosynthesis intercepted (f_{Ipar}) by the barley. However, these data are only being collected from the 2023 season onwards. The f_{Ipar} is measured using a SunScan, for which a measurement of the incident radiation is first made by placing this device above the vegetation and a repetition of 3 radiation measurements below the vegetation is carried out, the average of these 3 values gives the reflected radiation. The f_{Ipar} is the relative difference between these 2 measurements, as shown in equation 4.

$$f_{Ipar} = \frac{Incident-Transmis}{Incident} \quad (4)$$

These f_{Ipar} measurements are carried out at the same time as other measurements from the 'Management of barley for improving yield and profitability in Northern Sweden' project, such as biomass, soil moisture, proxi-sensing, ...

C. Data analysis equipment

Data analysis for model design is conducted in R studio. The automap, caret, dypyr, ggplot2, plot3D, raster, sf, sirad, sp and stringr libraries were also added to R.

Part of the geographic data processing was conducted in QGIS.

And as mentioned above, Google Earth Engine was used to pre-process Sentinel 2 images via a Java-Script interface.

D. Method for calculating Ipar over the entire barley growth period

1. Model for calculating f_{Ipar} from RVI

f_{Ipar} can be determined in a number of ways, one of which is to measure it directly in the field using probes that measure incident and transmitted radiation, as seen in part I.B with equation 4.

This method will be used to verify the model created from a second method.

A model can be used to find the f_{Ipar} from the RVI, which can be extracted from a Sentinel 2 image or, more simply, using a multispectral camera with the equivalent of bands B4 (red) and B11 (NIR). This model was described by (Christensen et Goudriaan, 1993)) in equations 5, 6 and 7.

$$RVI = \frac{\left(\rho_{NIR,\infty}+\left(\frac{\eta_{NIR}}{\rho_{NIR,\infty}}\right)(1-f_{Ipar})\right)\left(1+\eta_R(1-f_{Ipar})^2\right)}{\left(\rho_{R,\infty}+\left(\frac{\eta_R}{\rho_{R,\infty}}\right)(1-f_{Ipar})^2\right)\left(1+\eta_{NIR}(1-f_{Ipar})\right)} \quad (5)$$

$$\eta_{NIR} = \frac{\rho_{R,\infty}-\rho_{R,S}}{\rho_{R,S}-\frac{1}{\rho_{R,\infty}}} \quad (6) \qquad \eta_{NIR} = \frac{\rho_{NIR,\infty}-\rho_{NIR,S}}{\rho_{NIR,S}-\frac{1}{\rho_{NIR,\infty}}} \quad (7)$$

Equation 5 links RVI to f_{Ipar} with various parameters: near-infrared reflectance for maximum plant cover ($\rho_{NIR,\infty}$), red reflectance for maximum plant cover ($\rho_{R,\infty}$)and 2 other parameters that are calculated via equations 6 and 7. These 2 equations use the first 2 parameters described but add the near-infrared reflectance for soil ($\rho_{NIR,S}$) and the red reflectance for soil ($\rho_{R,S}$).

So $\rho_{NIR,\infty}$ and $\rho_{R,\infty}$ will correspond to the NIR and red reflectances respectively when the RVI is at its maximum in the dataset. And this will be similar for $\rho_{NIR,S}$ and $\rho_{R,S}$ but this is the minimum of the RVI. To determine these maximum and minimum values for the dataset that combines all the plots over different years, it's not enough just to take the maximum and minimum value of the RVI, since these can be biased. To do this, the first x maximum and y minimum reflectance values are averaged.

The values of x and y will be strategically chosen so that the model for determining f_{Ipar} by RVI has results that are as close as possible to field measurements. To achieve this, linear regressions between the values measured in the field and the f_{Ipar} values predicted from the Sentinel 2 RVI will be performed for x and y values ranging from 1 to 300 000, given that the database contains around 1.1 million values. The x and y will be chosen so that the r^2 is as close as possible to 1 and the parameters a and b of the regression line (measurement = a model + b) are as close as possible to 1 for a and 0 for b. These parameters are shown in Figure 3.

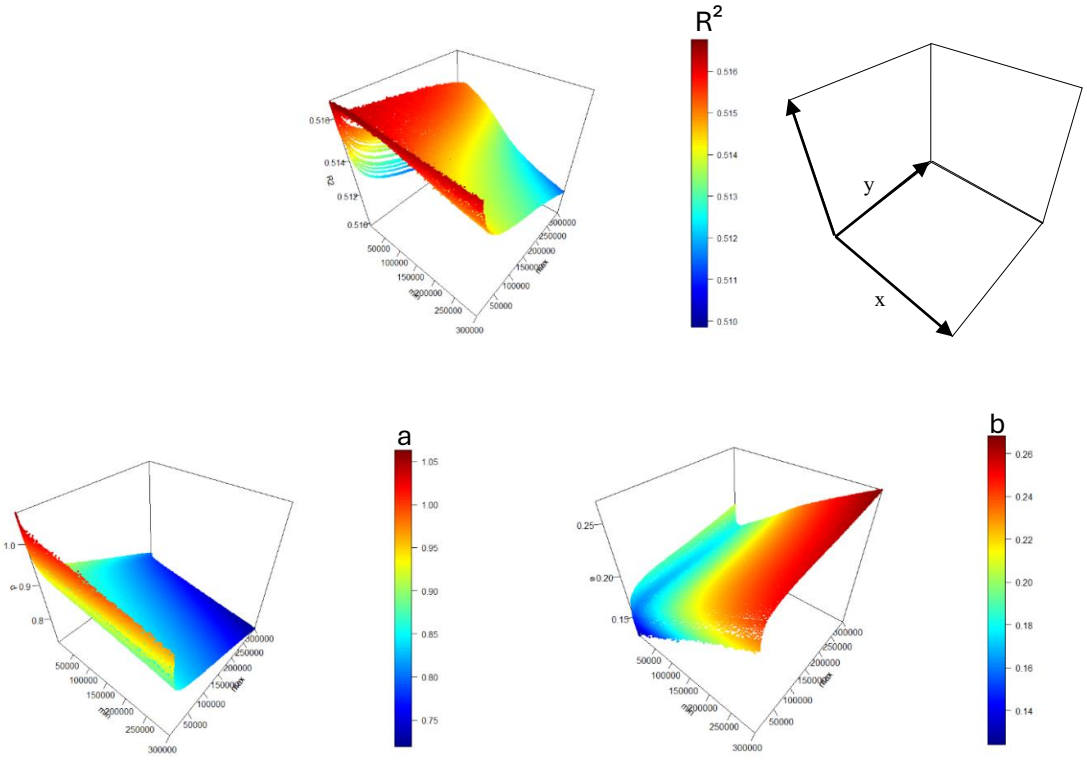


Figure 3 : Optimization of x and y to obtain the most reliable model for estimating f_{Ipar}, with the R^2 at the top, the slope of the linear regression line on the left (a) and the y-intercept on the right (b).

The pair of x and y values with the best result is x = 27 and y = 14000. The results of the linear regression will be discussed in in part **Erreur ! Source du renvoi introuvable.**.

This allows us to determine the value of the parameters $\rho_{NIR,\infty}$, $\rho_{R,\infty}$, $\rho_{NIR,S}$ and $\rho_{R,S}$. Thus, using R $\rho_{NIR,\infty} \approx 0{,}487$, $\rho_{R,\infty} \approx 0{,}012$, $\rho_{NIR,S} \approx 0{,}245$ et $\rho_{R,S} \approx 0{,}203$.

Having determined these 4 parameters ($\rho_{NIR,\infty}$, $\rho_{R,\infty}$, $\rho_{NIR,S}$ et $\rho_{R,S}$) it is possible to determine RVI values as a function of f_{Ipar} (corresponding to the red dots in Figure 4).

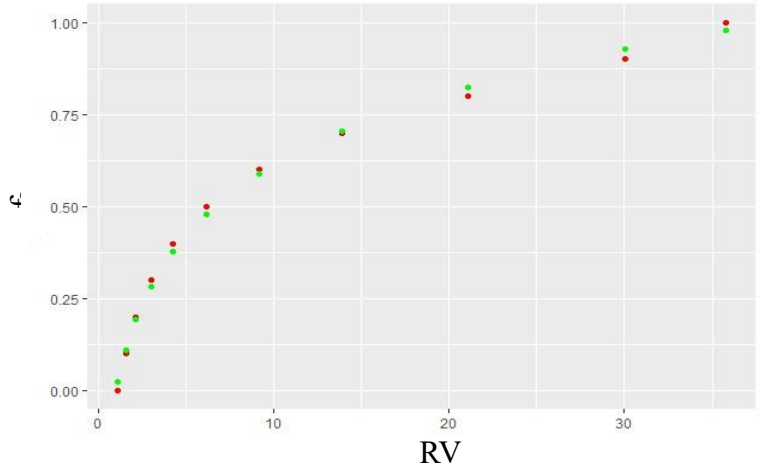


Figure 4 : f_{Ipar} calculated (red) and modeled (green) as a function of RVI with parameters set previously

The f_{Ipar} as a function of RVI can be modeled by a function, as shown in equation 7. (Christensen et Goudriaan, 1993).

$$f_{Ipar}(RVI) = a + b\,RVI^{c}\ (7)$$

The 3 constants a, b and c must be determined to enable this model for determining f_{Ipar} to be used. To do this, a regression of the values calculated via equation 4 by equation 7 will be performed. Thus, using R, it is determined that the calculated f_{Ipar} values follow equation 7 with a ≈ -11,01922, b ≈ 11,00463 et c ≈ 0,02354765.

The red dots in Figure 4 correspond to the modeling of f_{Ipar} by equation 7 and the above parameters.

Thus, with the RVI values from Sentinel 2, it is possible to calculate the f_{Ipar} of barley.

2. Interpolation of f_{Ipar} and calculation of Ipar over the entire growth period

Having determined the f_{Ipar} for the dates of passage of one of the Sentinel 2 satellites without the presence of clouds, it is necessary to interpolate these data over the entire barley growth period in order to obtain a daily f_{Ipar} value.

The f_{Ipar} for sowing and harvesting dates are set to 0, as barley at these stages does not absorb light energy, so the f_{Ipar} is zero.

A simple linear interpolation is not ideal, as f_{Ipar} values can be slightly erroneous, as can be seen in Figure 5, where the black curve corresponds to a linear interpolation between the various points, and abnormal variations in f_{Ipar} can be observed. To eliminate this problem, a first smoothing is performed on R using the smooth function, which corresponds to the blue curve in Figure 5.

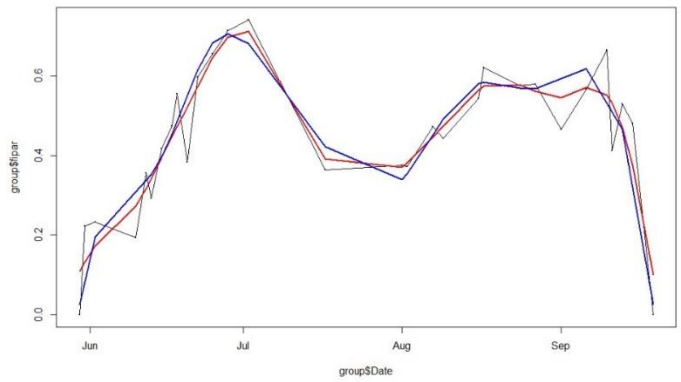


Figure 5 : f_{Ipar} of a barley micro plot throughout the growth cycle with linear interpolation in black, 1st smoothing in red and 2nd smoothing in blue.

Calculating the difference between the smoothing curve (red) and the modeled f_{Ipar} values (black) eliminates the most extreme values, so that the interpolation is as close to reality as possible. The difference between these 2 curves can be seen in Figure 6.

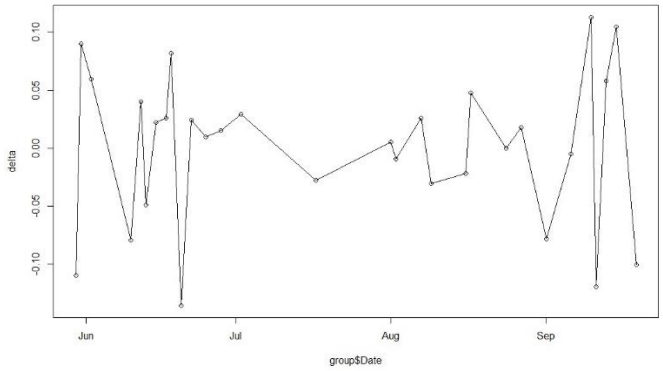


Figure 6 : Difference between smoothed curve and linear interpolation

A threshold of 0,07 positive or negative difference was chosen to eliminate erroneous f_{Ipar} values.

Once these values have been eliminated, a second smoothing step enables interpolation over the entire growth period. Linear interpolation is used not only between sowing and the first value, but also between the last value and harvest, in order to respect the value of 0 at the beginning and end of the cycle.

After estimating f_{Ipar} over the entire barley growth period, calculating Ipar is easier. All that's needed is the global irradiance, which can be measured at a weather station close to the plots. Ipar is thus determined by equation 8 (Peng *et al.*, 2021).

$$Ipar = 0{,}5\ G_0\ f_{Ipar} \quad (8)$$

Thus, Ipar was calculated on each of the microplots over the entire barley growth period.

E. Yield data processing

Yields are not usable as such, since measurements are biased by various factors.

These factors can be grouped into different categories, the first of which is harvester dynamics. These include latency errors and the time taken for the harvester to load and unload (Lyle *et al.*, 2014). The latency time corresponds to the time between the moment when the plant is cut and the moment when the grains of the plant pass in front of the yield sensor. During this time, the harvester continues to move at a certain speed; this average speed multiplied by the latency time corresponds to the distance between the true position of the yield and the position indicated on the map. In the data processing for this project, this error has been neglected.

The up- and down-load time errors correspond to the time taken for the harvester to go from a state where its threshing chain is empty to one where it is under load, and vice versa for the down-load time. During this time, there is an underestimation of output during the ramp-up and an overestimation during the ramp-down. (Leroux, 2020). This can easily be corrected by deleting points during these 2 phases. Here, 24s of ramp-up time and 10s of ramp-down time were chosen as recommended by (Lyle *et al.*, 2014). These times also correspond to data acquired in barley as shown in Figure 7.

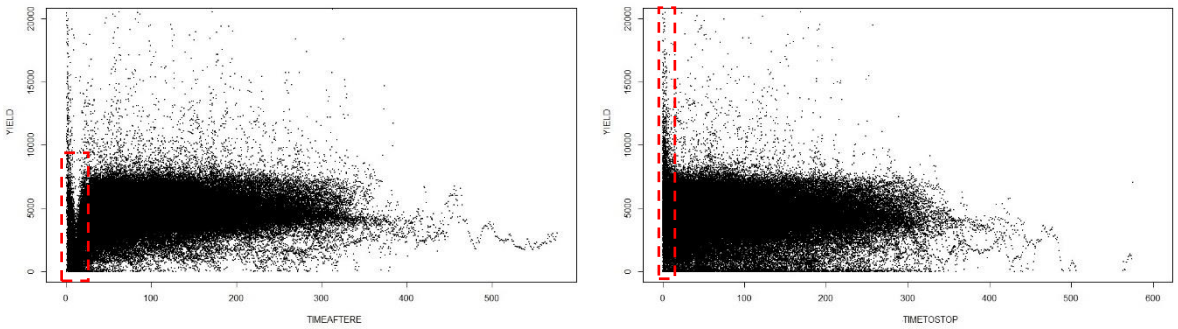


Figure 7 : Visualization of harvester loading (left) and unloading (right)

In fact, Figure 7 provides an easy visualization of the load descent time, which materializes as a yield peak just before detection of the cutter lift. This is due to the counting of a smaller area at the end of the line, but with a similar grain flow rate, resulting in a yield peak. The actual downhill run is not counted by the yield controller, since the cutter lift stops yield recording.

As far as the rise in load is concerned, a yield peak can be seen at the start, but the reason for this is the same as for the descent in load time. A dip can also be seen immediately afterwards, corresponding to the loading of the harvester. This loading time corresponds to the recommended 24s.

Another type of error corresponds directly to the error of the moisture and yield measuring instruments. This type of error is difficult to detect, particularly in the case of humidity. Nevertheless, some humidity errors were easily corrected. Indeed, in some plots the humidity is equal to 40%, which can be corrected by excluding these points. This sensor problem can be seen in Figure 8, with a peak in the frequency of humidity levels equal to 40%.

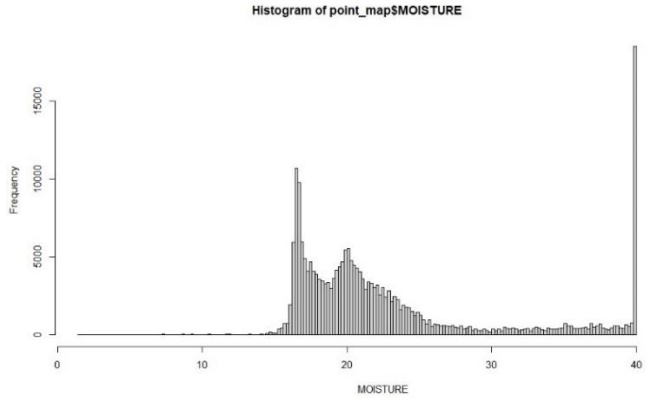


Figure 8 : Frequencies of humidity values before filtering on all yield maps

Yield measurement errors are derived from several sensors (grain flow and speed), but the accuracy of these sensors is not known, since it differs according to the type of sensor. For example, speed can be measured using pulse sensors, radar sensors, GPS signals, etc.

The third type of error concerns the accuracy of the positioning system. In this case, the positioning system uses GPS with RTK correction, giving centimeter-level accuracy (Trimble, 2024). This accuracy is satisfactory given that the yields will be averages over the micro-parcels, which have an order of magnitude of several meters.

And the last type of error, according to (Lyle *et al.*, 2014), concerns harvester driving errors. Some of these errors are negligible in these data, such as recrossing between two passes, which changes the width harvested, but here the precision of RTK guidance limits this. However, crossover at the end of a row when the field ends in a bevel is not limited by RTK guidance. But the previous correction of loading and unloading times helps to limit this. In fact, the previous correction eliminates the points at the beginning and end of the line, which removes the points where the cut is only partially filled. However, there is still a driving problem: the harvester's speed is always controlled by the operator. So when the operator causes sudden changes in speed, the yield is skewed. To correct this, speed variations between points were calculated. By comparing speed variations with yield (Figure 9), a threshold of 0,6 km.h-1 of speed variation between 2 points was set for this to be eliminated.

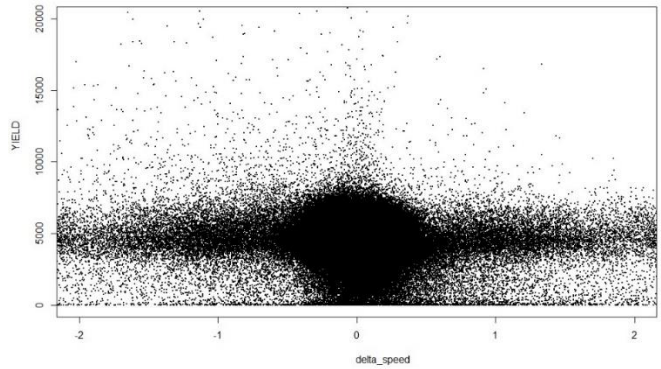


Figure 9 : Grain yields in kg.ha^{-1} as a function of speed variation

And more simply, as far as speed is concerned, points whose speed does not lie between 4 and 10 km.h^{-1} have been excluded, since they would lead to yield errors, as shown in Figure 10.

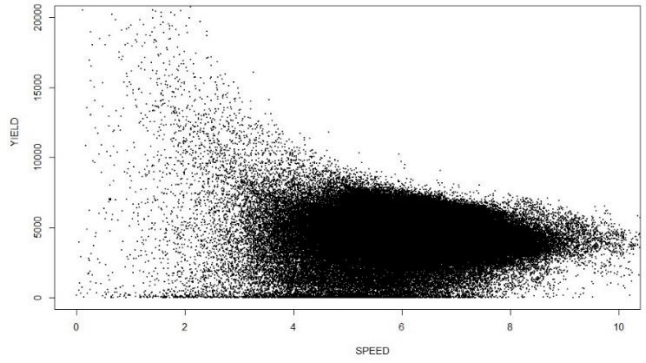


Figure 10 : Grain yields in kg.ha^{-1} as a function of working speed

This is more striking at low speeds, where yields reach improbable values for barley. In Figure 10, the scale is limited to 20 T.ha^{-1} but some values were much higher.

Another factor, which is difficult to consider, is machine setting, which can lead to varying degrees of grain loss. This was not used, as the machine does not record these parameters and it would be difficult to estimate losses without actual field measurements.

Now, once many points have been eliminated, kriging is carried out. This has 2 objectives: to recover the yield data where the points have been removed, and to correct the last erroneous points through the smoothing effect of block kriging. Once the kriging has been carried out, yields are extracted for each of the microplots. The microplots are then filtered according to the standard deviation of the kriging to eliminate certain errors.

F. Calculation of barley biomass at harvest

Barley biomass at harvest is the sum of the biomass accumulated daily. This accumulated biomass will be found in the different parts of the barley. But to determine grain yield, we need to know the proportion of seed biomass in relation to the rest of the plant. The harvest index is the ratio of seed dry matter to total plant dry matter. But first, we need to calculate the biomass accumulated daily.

1. Daily biomass gain at potential

The daily biomass gain (NPP) can be calculated from the Ipar. At potential NPP is calculated via equation 9 (Peng *et al.*, 2021).

$$NPP = RUE_{opt} \times Ipar \quad (9)$$

Thus, the daily biomass gain in g.m-2.day -1 is derived from the radiation intercepted by barley (Ipar) in MJ.m-2.day -1, calculated in a previous section. The RUEopt corresponds to the optimum efficiency of radiation use to produce biomass. RUEopt is expressed in g.MJ -1.

By definition, RUEopt is the optimum efficiency, so it is constant throughout the barley growth cycle.

Having determined the NPP, the accumulated biomass (DMtot) over the entire barley growth cycle, from sowing (i = 1) to harvesting (i = n), is simply the daily sum of biomass gains as summarized by equation 10.

$$DM_{tot} = RUE_{opt} \sum_{i=1}^{n} Ipar_i \quad (10)$$

As seen above, RUEopt is constant over the entire growth cycle. So to determine its value, we simply determine the sums of Ipar for a set of plots and compare them with the actual biomasses. A simple linear regression through 0 is then used to determine RUEopt.

But since barley growth is never at its full potential due to abiotic and biotic stress, the sum of Ipar is not sufficient.

2. Taking environmental factors into account

To improve the CASA model, it is essential to take environmental factors into account. Biotic factors are more difficult to model than abiotic factors, which can be modeled using measurements from weather stations near the plots.

As a result, this model will only take abiotic factors into account. These factors will be integrated into the model via multiplier coefficients that will modulate daily biomass gains.

a) Temperature limits growth

One of the first factors we could think of would be temperature. In a publication on the same model but for potatoes (Peng *et al.*, 2021), the influence of maximum daily temperature is modeled by a second-degree polynomial with a maximum value of 1 for the optimal growth

temperature of potatoes and 0 when the same polynomial is assumed to be negative, thus modeling the maximum and minimum growth temperatures of the plant. For barley, the base temperature is considered to be 5°C (Peltonen-Sainio *et al.*, 2008), which sets the minimum temperature threshold at which barley growth is possible. For an optimum temperature of 16.5 to 18°C (Karsai *et al.*, 2013), so the maximum of the polynomial function will be set at 1 between these two values, i.e. at around 17.5°C.

Thus, the function to take into account the influence of temperature will follow equation 11.

$$f_{Tmax} = \begin{cases} 0;\ si\ Tmax\ \leq 5\ °C \\ -0{,}0064\ Tmax + 0{,}224\ Tmax - 0{,}96;\ si\ 5\ °C \leq Tmax \leq 30\ °C \\ 0;\ si\ Tmax\ \geq 30\ °C \end{cases} (11)$$

And this function will follow the form shown in Figure 11.

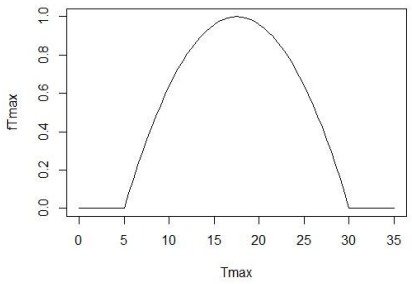


Figure 11 : Viewing the model's temperature function

b) Considering the best efficiency of diffuse radiation

It is also possible to include diffuse radiation in the model. This is due to clouds scattering direct radiation, which is more efficient for photosynthesis than direct radiation (Peng *et al.*, 2021)

A cloud index is therefore calculated, using the daily global radiation (GR_{obs}) measured at the Röbäcksdalen weather station and the daily global radiation at the top of the atmosphere (GR_{TOA}), which is calculated using the "extract" R function in the "sirad" library, with date and latitude (63.808347014401065) as input variables (Jedrzej S. Bojanowski, 2012). The index is calculated via equation 12.

$$CI = 1 - \frac{GR_{obs}}{GR_{TOA}} (12)$$

After calculating this index, the function to take diffuse radiation into account uses the minimum and maximum values of the index and its value at the desired moment. The function follows equation 13.

$$f_{CI} = \frac{CI - CI_{min}}{CI_{max} - CI_{min}} (13)$$

And this function applies in the form ($1-\mu * f_{CI}$) with $\mu = 0{,}46$ (Peng *et al.*, 2021). This ensures that this function does not have too great an impact on the model.

c) Growth limitation due to stomatal opening

Biomass accumulation can also be affected by stomatal activity in barley. When the air is dry, the stomata are closed, preventing exchanges between the atmosphere and the photosynthetic

cells. These exchanges are necessary for the photosynthetic activity of the plant, which enables the creation of biomass.

Function 14 takes this activity into account using the vapor pressure deficit (VPD) (Peng *et al.*, 2021).

$$f_{VPD} = \frac{1}{1+\frac{VPD}{D_0}} \text{ with } D_0 = 1{,}5 \text{ KPa} \quad (14)$$

Calculating VPD requires relative humidity, which is readily available from the weather station. But it is also necessary to find the saturated vapor pressure of the air. This is a function of temperature. Tetens' formula can be used to find this pressure as a function of temperature. The formula differs according to whether temperatures are negative or positive, but during barley's growing season the temperature is practically always positive. Thus equation 15 gives the saturated vapor pressure of air in KPa as a function of temperature in °C (Monteith et Unsworth, 2013).

$$P_{sat} = 0{,}61078\, e^{\frac{17{,}27\,T}{T+237{,}3}} \quad (15)$$

Thus, we can calculate the saturated vapor pressure of air using equation 16.

$$VPD = P_{sat} \times \left(1 - \frac{\text{relative humidity}}{100}\right) \quad (16)$$

d) Adding the impact of evapotranspiration to the model

Evapotranspiration can have a significant impact on biomass accumulation during the growth cycle. Thus it can be integrated into the model via function 17 (Zhang *et al.*, 2024).

$$f_{WS} = 0{,}5 + 0{,}5 \times \frac{ET_a}{ET_p} \quad (17)$$

Actual evapotranspiration (ETa) is a function of potential evapotranspiration (ETp) multiplied by a crop index (Brouwer et Heibloem, 2024). Thus $\frac{ET_a}{ET_p}$ can be replaced by the crop index. This is a function of the barley stage, so 4 stage ranges will be defined which will vary according to sowing and harvest dates and a coefficient is defined for each stage (Brouwer et Heibloem, 2024). The first stage corresponds to the first 15 days after sowing, with a coefficient of 0,35. The other 3 stages correspond to the total duration of the crop cycle minus the first 15 days. The 2nd stage corresponds to 24%, the 3rd to 48% and the last to 28%. The coefficients are 0.75, 1.15 and 0.45 respectively. Figure 12 summarizes the different phases.

Figure 12 : Coefficients according to the 4 growth stages which are related to the cultivation time

e) Estimating water stress in barley

To estimate water stress in barley, the NMDI will mainly be used. As we saw in part I.A, this index allows us to assess soil and vegetation moisture. But NMDI is not effective on moderately dense vegetation (0,5 < LAI < 1,5) (Wang et Qu, 2007). So, for LAI (Leaf Area Index) values within this range, NDMI is used, which is less effective but possible within this range.

LAI values are calculated via the f_{Ipar}, in fact using an exponential model, it is possible to relate the f_{Ipar} to the LAI. This relationship follows equation 18.

$$LAI = e^{3,42397 \times f_{Ipar} - 1,94404} \text{ (18)}$$

Using the values measured in the field by the Sunscan, the relationship is quite correct, with a p-value: < 2,2e-16 and an $R^2 = 0.9583$. What's more, the relationship is all the stronger for small LAI values, so for LAI values between 0,5 and 1,5 it's correct.

For LAI below 0,5, the higher the NMDI, the drier the soil, and an NMDI below 0,6 is considered wet (Wang et Qu, 2007). Function 19 is used to apply this, and its representation follows the blue curve in Figure 13.

$$f_{moisture}\ (LAI \leq 0{,}5) = \begin{cases} 1\ ;\ si\ NMDI \leq 0{,}6 \\ 2{,}5 - 2{,}5\ NMDI\ ;\ si\ 0{,}6 < NMDI < 1 \\ 0\ ;\ si\ NMDI \geq 1 \end{cases} \text{ (19)}$$

For LAI greater than 1.5, NMDI interpretation is the reverse. The higher the NMDI, the lower the water stress (Wang et Qu, 2007). Function 20 therefore calculates $f_{moisture}$ for LAI above 1,5, and its representation corresponds to the orange curve in Figure 13.

$$f_{moisture}\ (LAI > 1{,}5) = \begin{cases} 0\ ;\ si\ NMDI \leq 0{,}5 \\ 4\ NMDI - 2\ ;\ si\ 0{,}5 < NMDI < 0{,}75 \\ 1\ ;\ si\ NMDI \geq 0{,}75 \end{cases} \text{ (20)}$$

In the last case, where LAI are between the 2 previous values, NMDI is not applicable. To estimate water stress, the NDMI (Normalized Difference Moisture Index) is used. It uses 2 of the 3 NMDI bands, (bands B8A and B11), which have already been downloaded (Viehweger, 2023). Its calculation follows equation 21.

$$NDMI = \frac{B8A - B11}{B8A + B11} \text{ (21)}$$

NDMI values range from -1 to 1 and are related to water stress, but also to plant cover. To compensate for this, NDMI is divided by NDVI (Normalized Difference Vegetation Index). NDVI is calculated using the same bands as RVI via equation 22.

$$NDMI = \frac{B8 - B4}{B8 + B4} \text{ (22)}$$

Thus, using function 23, $f_{moisture}$ can be calculated, and this function can be visualized by the green curve in Figure 13.

$$f_{moisture}\ (0{,}5 < LAI \leq 1{,}5) = \begin{cases} 0\ ;\ si\ \frac{NMDI}{NDVI} \leq 0{,}1 \\ 2{,}5\ \frac{NMDI}{NDVI} - 0{,}25\ ;\ si\ 0{,}1 < \frac{NMDI}{NDVI} < 0{,}5 \\ 0\ ;\ si\ \frac{NMDI}{NDVI} \geq 0{,}5 \end{cases} \text{ (23)}$$

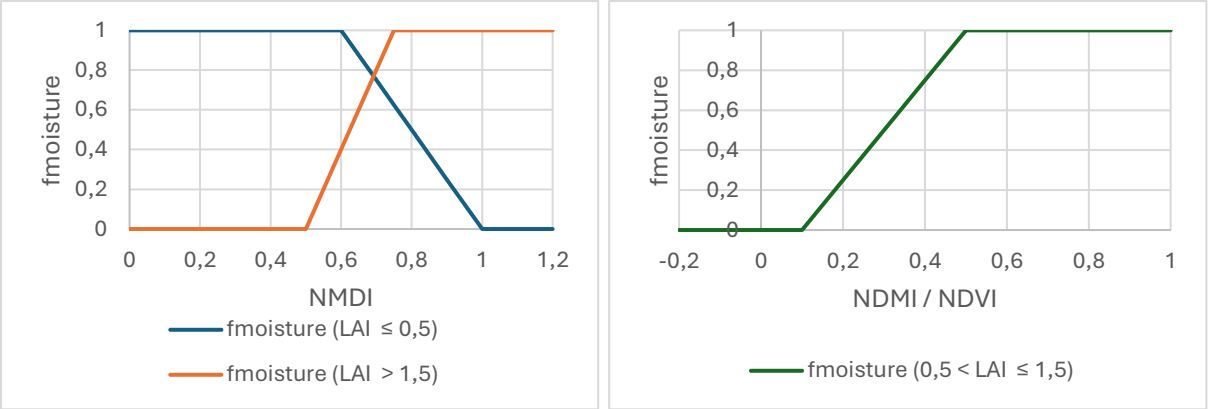


Figure 13 : $f_{moisture}$ as a function of NDMI for IPL outside the interval]0.5; 1.5] on the left and as a function of NDMI/NDVI for IPL in this interval on the right

These 3 functions can be used to model water stress over the entire barley growth period.

II. Results

First, we'll look back at the fIpar calculation model to analyze the results, then we'll study the model's relevant environmental factors. Finally, the model will be applied to data for the year 2023 to verify its validity.

A. Analysis of the f_{Ipar} determination model

To verify the model's predictions of fIpar as a function of RVI, field surveys were carried out. These were compared with the model results, as shown in Figure 14.

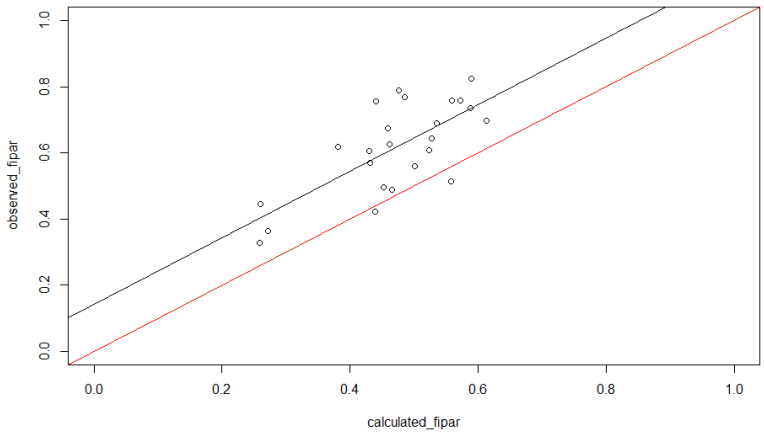


Figure 14 : Comparison of calculated fipar with measured fIpar with linear regression in black and x=y line in red

Linear regression between the calculated and observed results in the equation y = 1,00588 x + 0,14139, with a coefficient of determination of 0,52. There is therefore a slight overestimation of the model compared with the measurements. The coefficient of determination is rather low, but this may also be due to the choice of using Sentinel 2 for the RVI measurement, which gives poorer results than proxi detection in another study (Peng *et al.*, 2021).

B. Selection of relevant environmental factors

The various environmental factors are applied to the Ipar values to find those that best match the model results. To do this, the observed yields are converted into biomass yields via the harvest index, which is 0,62. These are then compared with the calculated yield. Figure 15 shows the different results based on R^2 and RRMSE.

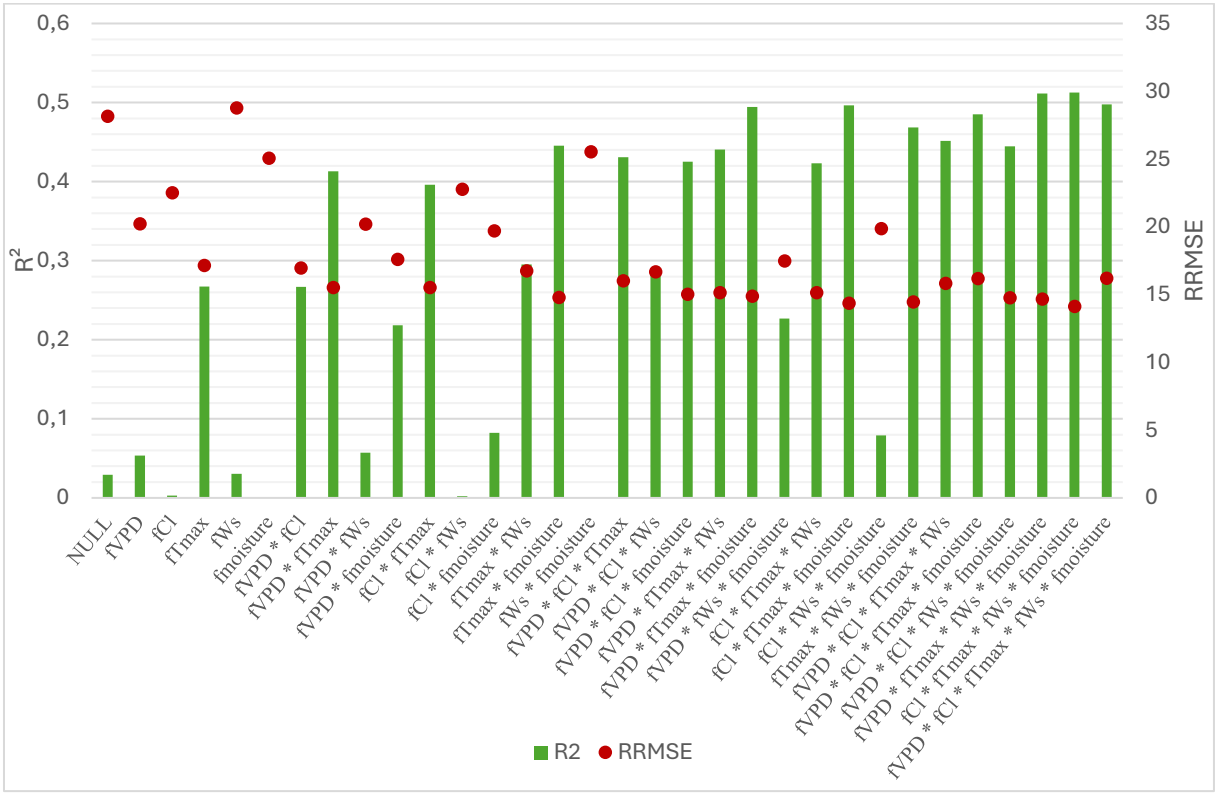


Figure 15 : Model results as a function of applied environmental factors

Taking these results, the best combination of environmental factors corresponds to $f_{Cl} * f_{Tmax} * f_{Ws} * f_{moisture}$, giving a result of $R^2 = 0,513$ and RRMSE = 14,1. To find the RUE_{opt}, linear regression through zero is used. Figure 16 shows this linear regression.

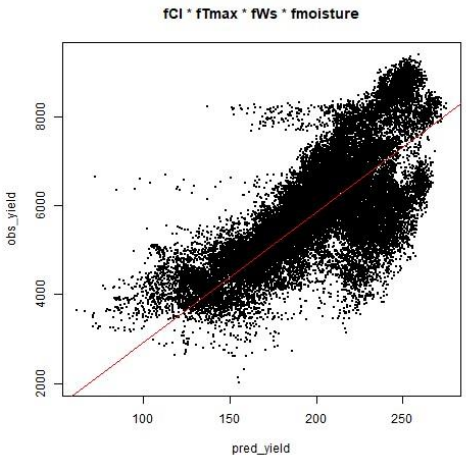


Figure 16 : Visualization of optimal linear regression with observed biomass yield (kg / ha) on the ordinate and predicted yield without RUE_{opt} (MJ / m^2) on the abscissa

Linear regression gives $RUE_{opt} = 2,92$ g.MJ^{-1}. The optimal model thus follows equation 24.

$$DM_{tot} = RUE_{opt} \sum_{i=1}^{n} Ipar_i \times f_{Cl_i} \times f_{Tmax_i} \times f_{Ws_i} \times f_{moisture_i} \quad (24)$$

And barley grain yield follows equation 25 with HI = 0.62, which corresponds to the harvest index, which is the ratio of grain mass to total mass.

$$Yield = RUE_{opt} \times HI \times \sum_{i=1}^{n} Ipar_i \times f_{Cl_i} \times f_{Tmax_i} \times f_{Ws_i} \times f_{moisture_i} \quad (25)$$

C. Analysis of the barley yield model

Data from 2023 that have not been used up to now will be used to check the validity of the model. Thus, using equation 24, the theoretical biomass that barley would have made is calculated and compared by linear regression to the actual biomass obtained via the yield controller and the harvest index. This can be seen in Figure 17.

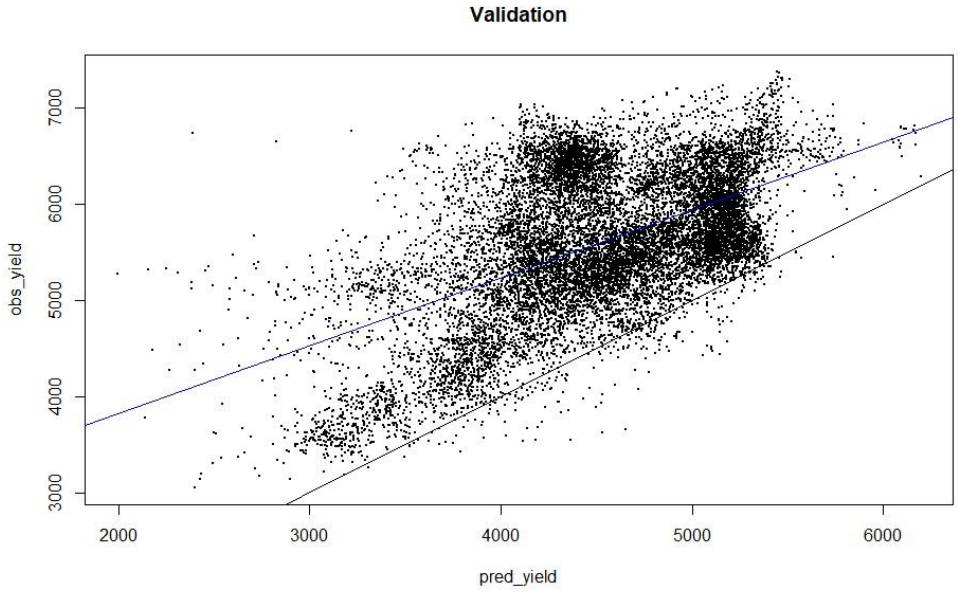


Figure 17 : Linear regression of actual biomass against calculated biomass in 2023 (blue line) and the black line corresponds to the observed yield, which is equal to the calculated yield

Linear regression gives an R^2 = 0.30, which is low. The RRMSE is 22,4 and the bias is +1078 kg. This means that the model underestimates reality for 2023.

The low R^2 is probably due to environmental factors that have not been taken into account, or have been poorly taken into account, in the model. Some factors, such as humidity, were built directly into the model. These indicators need to be verified by field measurements.

Conclusion

The model for estimating barley yield using the Sentinel 2 constellation mainly uses 2 bands, one red and one near-infrared, to calculate the RVI needed to calculate the Ipar via the fIpar. In this way, the sum of the Ipars corresponds to the energy used by photosynthesis. But this energy is not only used for biomass production, there are also losses that can be modeled by functions. Here, 4 functions proved useful: a function modulating biomass production as a function of temperature (f_{Tmax}), a function favoring diffuse radiation over direct radiation (f_{Cl}), a function simulating evapotranspiration (f_{WS}) and a function modeling water stress ($f_{moisture}$). The $f_{moisture}$

function requires the use of 2 additional Sentinel 2 bands in the mid-infrared, while the other functions mainly use data from the weather station.

The model still gives some yield errors, probably due to missing or poorly optimized functions. For example, the water stress function could be calibrated using field measurements.

References

Alparone L., **Arienzo A.**, et **Garzelli A.**, 2024. Spatial Resolution Enhancement of Vegetation Indexes via Fusion of Hyperspectral and Multispectral Satellite Data. *Remote Sensing*, 16 (5), p. 875. DOI : 10.3390/rs16050875

Brouwer C. et **Heibloem M.**, 2024. *CHAPTER 3: CROP WATER NEEDS*. https://www.fao.org/4/s2022e/s2022e07.htm (Consulté le 30 mai 2024).

Christensen S. et **Goudriaan J.**, 1993. Deriving light interception and biomass from spectral reflectance ratio. *Remote Sensing of Environment*, 43 (1), p. 87-95. DOI : 10.1016/0034-4257(93)90066-7

Jedrzej S. Bojanowski, 2012. *sirad: Functions for Calculating Daily Solar Radiation and Evapotranspiration*. , p. 2.3-3. DOI : 10.32614/CRAN.package.sirad

Karsai I., **Igartua E.**, **Casas A. m.**, **Kiss T.**, **Soós V.**, **Balla K.**, **Bedő Z.**, et **Veisz O.**, 2013. Developmental patterns of a large set of barley (Hordeum vulgare) cultivars in response to ambient temperature. *Annals of Applied Biology*, 162 (3), p. 309-323. DOI : 10.1111/aab.12023

Leroux C., 2020. *Filtering - Cleaning Yield Maps*. https://www.aspexit.com/filtering-cleaning-yield-maps/ (Consulté le 12 juillet 2024).

Lundkvist A., 2024. *Weather stations*. https://www.slu.se/en/faculties/nj/this-is-the-nj-faculty/collaborative-centres-and-major-research-platforms/faltforsk-field-research-unit/weather/weather-stations/ (Consulté le 5 mai 2024).

Lyle G., **Bryan B.A.**, et **Ostendorf B.**, 2014. Post-processing methods to eliminate erroneous grain yield measurements: review and directions for future development. *Precision Agriculture*, 15 (4), p. 377-402. DOI : 10.1007/s11119-013-9336-3

Monteith J.L. et **Unsworth M.H.**, 2013. *Principles of environmental physics: plants, animals, and the atmosphere*. 4th edition. Amsterdam Boston : Elsevier/Academic Press, 401 p. ISBN 978-0-12-386910-4

Nguyen N. et **Binh N.**, 2016. Potential of drought monitoring using Sentinel-2 data. Dans : Hanoi, Viêt Nam : https://www.researchgate.net/publication/330401106_Potential_of_drought_monitoring_using_Sentinel-2_data

Peltonen-Sainio P., **Jauhiainen L.**, et **Hakala K.**, 2008. Climate change and prolongation of growing season: changes in regional potential for field crop production in Finland. *Agricultural and Food Science*, 18 (3-4), p. 171. DOI : 10.2137/145960609790059479

Peng J., **Manevski K.**, **Kørup K.**, **Larsen R.**, **Zhou Z.**, et **Andersen M.N.**, 2021. Environmental constraints to net primary productivity at northern latitudes: A study across scales of radiation interception and biomass production of potato. *International Journal of Applied Earth Observation and Geoinformation*, 94, p. 102232. DOI : 10.1016/j.jag.2020.102232

Peng J. et **Parsons D.**, 2022. *Management of barley for improving yield and profitability in Northern Sweden*.

Trimble, 2024. *Real-time kinematic (RTK) GPS technology*. https://www.trimble.com/en/solutions/technologies/rtk-real-time-kinematic (Consulté le 15 juillet 2024).

Viehweger J., 2023. *Normalized Difference Moisture Index (NDMI)*. https://custom-scripts.sentinel-hub.com/custom-scripts/sentinel-2/ndmi/ (Consulté le 4 juillet 2024).

Wang L. et **Qu J.J.**, 2007. NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical Research Letters*, 34 (20). DOI : 10.1029/2007GL031021

Zhang S., **Lærke P.E.**, **Andersen M.N.**, **Peng J.**, **Mortensen E.Ø.**, **Pullens J.W.M.**, **Larsen K.S.**, **Jørgensen U.**, et **Manevski K.**, 2024. Validating and improving the Carnegie-Ames-Stanford Approach for remote sensing of perennial grass biomass. *ISPRS Journal of Photogrammetry and Remote Sensing*,

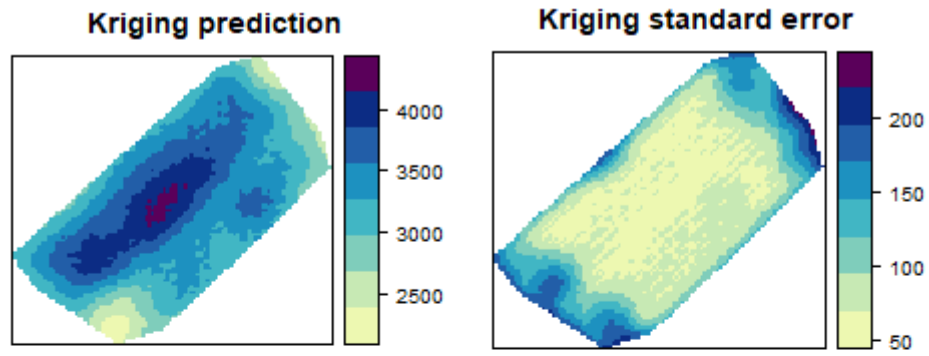
Appendix 1: Regression results for the model used to determine f_{Ipar} from RVI

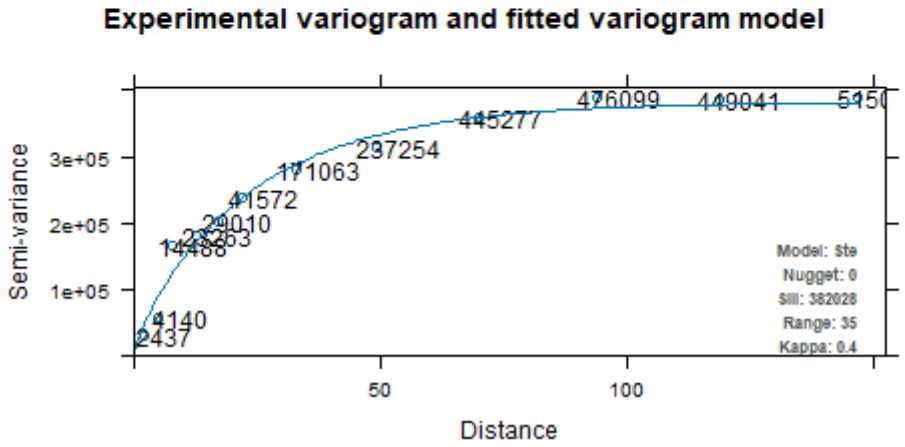
```
> reg<-lm(Fpar_origin ~ predicted_Fpar)
> summary(reg)

Call:
lm(formula = Fpar_origin ~ predicted_Fpar)

Residuals:
     Min       1Q   Median       3Q      Max
-0.03602 -0.01932  0.00336  0.01824  0.02729

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.001665   0.013874    0.12    0.907
predicted_Fpar 1.000884   0.023545   42.51  1.1e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.02461 on 9 degrees of freedom
Multiple R-squared:  0.995,     Adjusted R-squared:  0.9945
F-statistic:  1807 on 1 and 9 DF,  p-value: 1.101e-11
```

Appendix 2: Linear regression comparing the f_{Ipar} estimation model with the measured f_{Ipar}

```
>     reg2 <- lm(observed.fipar ~ calculated.fipar)
> summary(reg2)

Call:
lm(formula = observed.fipar ~ calculated.fipar)

Residuals:
      Min        1Q    Median        3Q       Max
-0.188019 -0.063989  0.006377  0.058656  0.171695

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.14139    0.09954   1.421    0.169
calculated.fipar  1.00588    0.20743   4.849 7.59e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09914 on 22 degrees of freedom
Multiple R-squared:  0.5166,    Adjusted R-squared:  0.4947
F-statistic: 23.52 on 1 and 22 DF,  p-value: 7.595e-05
```

Appendix 3: Kriging result for one of the plots after point filtering





Appendix 4: Results for each function combination

Fonction(s)	R^2	RMSE (kg/ha)	RRMSE	RUE_{opt} (g/MJ)
NULL	0,029	1648	28,1	1,50
fVPD	0,054	1183	20,2	1,95
fCl	0,003	1317	22,5	2,04
fTmax	0,267	1003	17,1	1,93
fWs	0,030	1684	28,8	1,53
fmoisture	0,000	1467	25,1	1,76
fVPD * fCl	0,267	993	17,0	2,57
fVPD * fTmax	0,413	909	15,5	2,35
fVPD * fWs	0,057	1182	20,2	2,00
fVPD * fmoisture	0,218	1030	17,6	2,25
fCl * fTmax	0,396	908	15,5	2,51
fCl * fWs	0,002	1333	22,8	2,08
fCl * fmoisture	0,082	1154	19,7	2,36
fTmax * fWs	0,295	980	16,7	1,98
fTmax * fmoisture	0,445	866	14,8	2,22
fWs * fmoisture	0,001	1495	25,5	1,78
fVPD * fCl * fTmax	0,431	937	16,0	3,02
fVPD * fCl * fWs	0,290	977	16,7	2,63
fVPD * fCl * fmoisture	0,425	880	15,0	2,95
fVPD * fTmax * fWs	0,441	887	15,2	2,42
fVPD * fTmax * fmoisture	0,494	871	14,9	2,68
fVPD * fWs * fmoisture	0,227	1024	17,5	2,29
fCl * fTmax * fWs	0,423	886	15,1	2,58
fCl * fTmax * fmoisture	0,496	840	14,4	2,87
fCl * fWs * fmoisture	0,079	1164	19,9	2,40
fTmax * fWs * fmoisture	0,468	847	14,5	2,26
fVPD * fCl * fTmax * fWs	0,452	926	15,8	3,10
fVPD * fCl * fTmax * fmoisture	0,485	948	16,2	3,43
fVPD * fCl * fWs * fmoisture	0,445	865	14,8	3,00
fVPD * fTmax * fWs * fmoisture	0,512	860	14,7	2,73
fCl * fTmax * fWs * fmoisture	0,513	828	14,1	2,92
fVPD * fCl * fTmax * fWs * fmoisture	0,498	948	16,2	3,49

Appendix 5: Linear regression to find optimal environmental factors

```
> lin_mod <- lm(obs_yield ~ pred_yield)
> summary(lin_mod)

Call:
lm(formula = obs_yield ~ pred_yield)

Residuals:
    Min      1Q  Median      3Q     Max
-3161.5  -436.3    -3.7   443.8  3871.1

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1025.110     24.039   42.64   <2e-16 ***
pred_yield    24.248      0.119  203.81   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 809.3 on 39495 degrees of freedom
Multiple R-squared:  0.5126,    Adjusted R-squared:  0.5126
F-statistic: 4.154e+04 on 1 and 39495 DF,  p-value: < 2.2e-16
```

Appendix 6: Linear regression of model validation in 2023

```
> lin_mod <- lm(obs_yield ~ pred_yield)
> summary(lin_mod)

Call:
lm(formula = obs_yield ~ pred_yield)

Residuals:
     Min       1Q   Median       3Q      Max
-2038.34  -448.77   -64.02   447.04  2646.70

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.412e+03  4.042e+01   59.66   <2e-16 ***
pred_yield  7.054e-01  8.862e-03   79.60   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 617.8 on 15080 degrees of freedom
Multiple R-squared:  0.2959,    Adjusted R-squared:  0.2958
F-statistic:  6337 on 1 and 15080 DF,  p-value: < 2.2e-16
```