Nicolas Vuille-dit-Bille
Avenue de Morges 20
1004 Lausanne
nicolas.vuille-dit-bille@agroscope.admin.ch; vuillenicolas@gmail.com

Hyperspectral data to characterize multi-environment response of winter wheat varieties in nitrogen management context

Conceptual Design Report

10 October 2024

Abstract

Hyperspectral instruments can give insights concerning ecophysiological plant parameters by using reflectance spectroscopy. This technology can be difficult to handle to extract meaningful information because it measures one reflectance value linked to one specific wavelength in a range for example of 350 to 2500 nm. These reflectance values can also be highly inter-correlated considering narrow wavelength range. However, some parts of the spectrum are linked to specific plant parameter and can be interesting to investigate in order to evaluate for example crops needs (nutrient or water deficit) and other biotic stresses like disease or insect damage. In this study, three field trials across Switzerland (Changins, Goumoens and Reckenholz) have been conducted over two years testing five top Swiss varieties of winter wheat (CH Camedo, CH Claro, Montalbano, CH Nara and Runal) with three main nitrogen treatments (none, reduced amount and conventional amount). In a context of intrants reduction in agriculture to protect the environment and reduce cost, it is important to characterize varieties and their performance under reduced amount of nitrogen in interaction with their environment while optimizing grain yield and quality. To reach this goal, agronomic and hyperspectral measurements have been collected over two winter wheat growth seasons in order to estimate the variety performance in interaction with nitrogen amount, experimental site and year.

Table of Contents

1 Project Objectives

The first objective of this study is to be able to extract relevant information from the hyperspectral data which should be correlated to conventional agronomic measurements like for example, leaf chlorophyll content (LCC), leaf area index (LAI), grain and straw yield. In the literature, there are already some insights about which part of the spectrum can be related to which ecophysiological components depending on crop species (Figure 1).

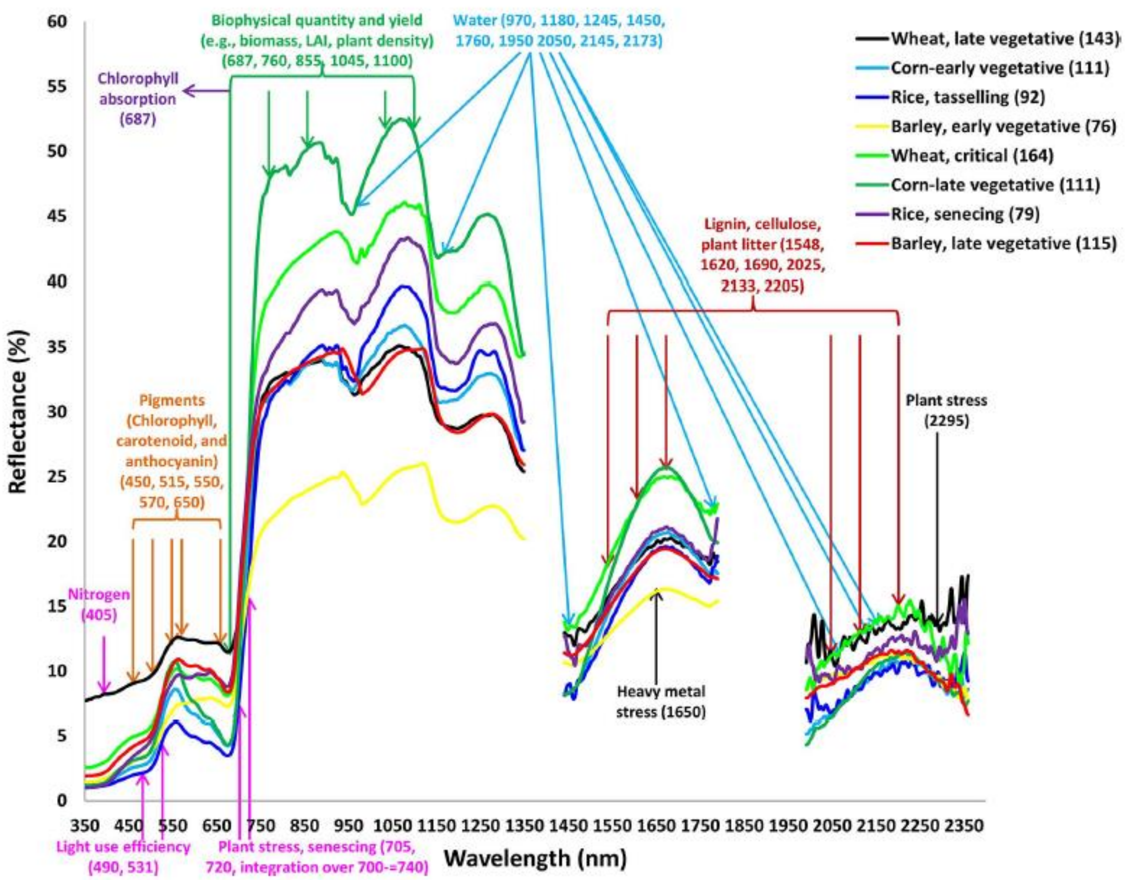


Figure 1. Reflectance spectrum according to crop species and their link to ecophysiological parameters [1].

Similar studies have already been performed using multispectral sensors mounted on unmanned aerial vehicle (UAV) in a nitrogen management context [2] and can be compared to the results of this study. However, before starting to interpret this data, it would be needed to find the best way to clean the data in order that misleading information is excluded from the analysis. Indeed, data acquisition on the field with a spectroradiometer instrument can be difficult and is really sensitive to climatic conditions and the measurement time. Reference measurements with white calibrated panel are performed at specific time interval to improve the comparison between measurements but it is not the perfect solution and sometimes for some reasons a lot of noise is captured when measuring crop canopy reflectance.

After ensuring that the data quality is sufficient to be considered in the analysis, the second objective will be to find pattern of specific wavelength range linked to agronomic data and compare the results with the literature.

Then, another approach will be tested with the computation of vegetations indices (VI) by combining specific wavelengths. There are a lot of literatures using these VIs to estimate crop parameters but they are not always performing the same depending on specific environment. For example, it has been found that the Near-Infrared Red (NIR) region is sensitive to water content in plant tissues and this will therefore have an effect on plant N content estimation [3]. An investigation should also be performed to evaluate the robustness of these spectral indicators through a diversity of climatic conditions depending on sites and years.

Finally, once correlation between wavelengths or VIs and agronomic parameters are found, heritability on these indicators can be assessed taking into account spatial variability on the field. This will allow to know if spectral information is a good tool to characterize winter wheat varieties interacting with their environment compared to conventional agronomic measurements.

2 Methods

The field trials were conducted at three different sites across Switzerland in Changins, Goumoens and Reckenholz over two crop seasons from 2021 to 2022.

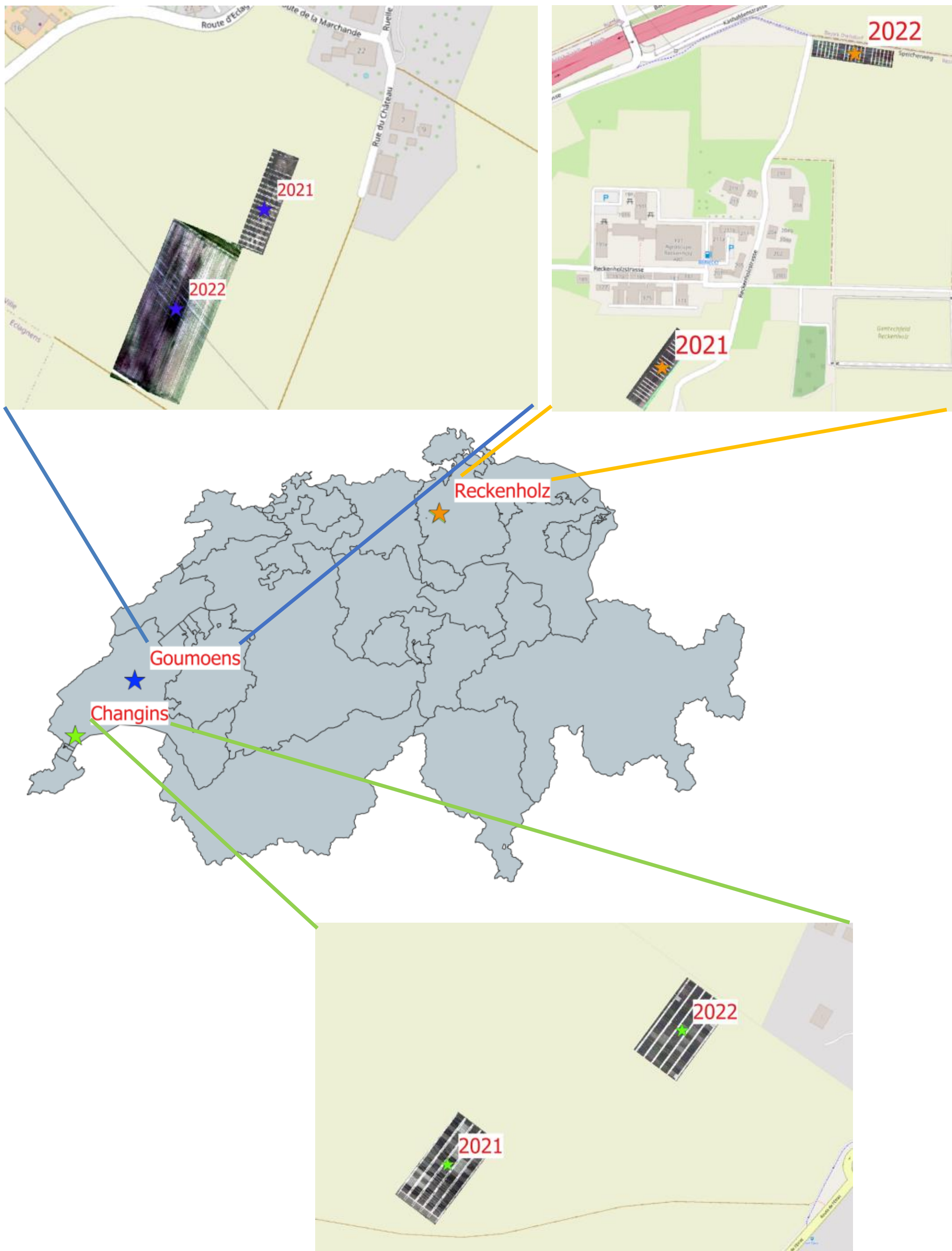
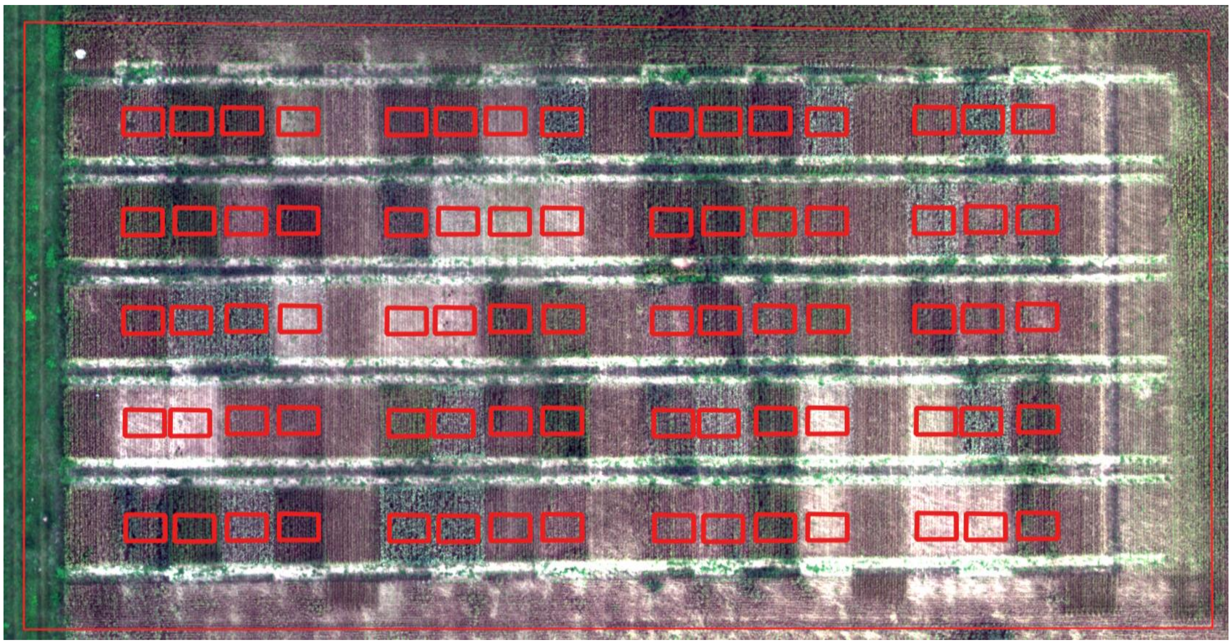


Figure 2. Basemap illustrating Switzerland with three experimental locations: Changins (green), Goumoens (blue), Reckenholz (orange). On each experimental location, trials have been conducted each year on different fields. RGB illustration of field trials according to years are showed in the zoomed area.

The field trials were organized as a complete randomized block design (Figure 3; Table 1) meaning that each block included one repetition (repetition=3). The exception was in Reckenholz experimental site with a split block design for logistic reason [4]. The nitrogen treatments and the varieties have been allocated randomly within each repetition block. Concerning, the nitrogen treatments, they were applied at three different dates corresponding to specific crop growth stage to optimize nitrogen use efficiency (NUE). For that reason, for example, the treatment $N_{20\text{-}40\text{-}20}$ means that 20 kgN/ha^{-1}, 40 kgN/ha^{-1}, 20 kgN/ha^{-1} have been applied respectively for the first, second and third fertilizer application.



Repetition 1						Repetition 2						Repetition 3					
7	**25**	**24**	**16**		**2**	**23**	**1**	**15**		**10**	**17**	**4**	**13**		**22**	**14**	**8**
N2	N5	N4	N0		N2	N3	N0	N5		N5	N2	N4	N3		N2	N4	N3
17	**19**	**8**	**5**		**9**	**11**	**16**	**21**		**25**	**20**	**18**	**23**		**12**	**7**	**10**
N2	N4	N3	N5		N4	N0	N0	N0		N5	N5	N3	N3		N2	N2	N5
22	**12**	**14**	**11**		**21**	**6**	**24**	**18**		**8**	**4**	**9**	**17**		**5**	**2**	**3**
N2	N2	N4	N0		N0	N0	N4	N3		N3	N4	N4	N2		N5	N2	N3
1	**6**	**23**	**3**		**18**	**13**	**5**	**19**		**9**	**12**	**20**	**11**		**16**	**15**	**24**
N0	N0	N3	N3		N3	N3	N5	N4		N4	N2	N5	N0		N0	N5	N4
10	**20**	**13**	**4**		**15**	**14**	**2**	**3**		**22**	**7**	**19**	**21**		**6**	**1**	**25**
N5	N5	N3	N4		N5	N4	N2	N3		N2	N2	N4	N0		N0	N0	N5

Figure 3. Field experimental design in Changins in 2021 with RGB (A) and schematic (B) representation. On the schema, colors represent the nitrogen treatments: $N_{0\text{-}0\text{-}0}$ (white), $N_{20\text{-}40\text{-}20}$ (green), $N_{20\text{-}60\text{-}0}$ (blue), $N_{40\text{-}80\text{-}40}$ (yellow), $N_{40\text{-}120\text{-}0}$ (red). Repetition blocks are mentioned above the schematic figure. Numbers in bold from 1 to 25 in each experimental unit correspond to specific variety with specific nitrogen treatment as indicated in the table 1 below.

Table 1. Treatment numbers including variety and nitrogen treatment information in Changins in 2021. Colors represent the different nitrogen treatments: $N_{0\text{-}0\text{-}0}$ (white), $N_{20\text{-}40\text{-}20}$ (green), $N_{20\text{-}60\text{-}0}$ (blue), $N_{40\text{-}80\text{-}40}$ (yellow), $N_{40\text{-}120\text{-}0}$ (red).

Treatment ID	Variety	N treatment	Total N applied	N rate
1	CH Claro	N0	0 kgN/ha-1	0-0-0
2	CH Claro	N1	80 kgN/ha-1	20-40-20
3	CH Claro	N2	80 kgN/ha-1	20-60-0
4	CH Claro	N3	160 kgN/ha-1	40-80-40
5	CH Claro	N4	160 kgN/ha-1	40-120-0
6	CH Nara	N0	0 kgN/ha-1	0-0-0
7	CH Nara	N1	80 kgN/ha-1	20-40-20
8	CH Nara	N2	80 kgN/ha-1	20-60-0
9	CH Nara	N3	160 kgN/ha-1	40-80-40
10	CH Nara	N4	160 kgN/ha-1	40-120-0
11	CH Camedo	N0	0 kgN/ha-1	0-0-0
12	CH Camedo	N1	80 kgN/ha-1	20-40-20
13	CH Camedo	N2	80 kgN/ha-1	20-60-0
14	CH Camedo	N3	160 kgN/ha-1	40-80-40
15	CH Camedo	N4	160 kgN/ha-1	40-120-0
16	Montalbano	N0	0 kgN/ha-1	0-0-0
17	Montalbano	N1	80 kgN/ha-1	20-40-20
18	Montalbano	N2	80 kgN/ha-1	20-60-0
19	Montalbano	N3	160 kgN/ha-1	40-80-40
20	Montalbano	N4	160 kgN/ha-1	40-120-0
21	Runal	N0	0 kgN/ha-1	0-0-0
22	Runal	N1	80 kgN/ha-1	20-40-20
23	Runal	N2	80 kgN/ha-1	20-60-0
24	Runal	N3	160 kgN/ha-1	40-80-40
25	Runal	N4	160 kgN/ha-1	40-120-0

Over season, agronomic parameters were collected to evaluate variety performance at specific N treatment level for each experimental units mentioned above. There were time point measurements at specific crop growth stage (before heading, at heading, after heading and at anthesis) such as chlorophyll N content, biomass, N content in biomass and leaf area index (LAI) and more dynamic parameters that have been collected at the end of the crop growth during the harvest such as grain and straw yield. These parameters will be called ground truth data (GTD) and will be linked to hyperspectral data in order to validate the prediction of the models.

In parallel, crop canopy reflectance has been measured with a hyperspectral instrument as close to agronomic parameter measurement as possible. After ten measurements, the instrument was recalibrated with a white reference panel in order to take into account light intensity variance during measurement time. For each measurement, the whole experimental unit canopy was scanned and only one reflectance value for each wavelength from 350 to 2500 nm was stored (averaging ten measurements). All of this spectral information can be used as explanatory variables in order to find the parts of the light spectrum that could be used to predict and characterize winter wheat variety performance in interaction with specific environment.

Before to be able to use the raw spectral data, pre-processing steps are needed to convert raw data to csv readable format. Then exploratory analysis should be done to avoid including spectrum part with aberrant values in the analysis. To reach this goal, some visualization tools can be useful to show unexpected results like plotting wavelength bands (x-axis) versus reflectance values (y-axis). Unexpected peaks and drops will indicate noise in the data and it might not be suitable to include it to build the prediction models. The spectral signature can also be compared between samples in order to see if there was some issue in the data acquisition. Principal component analysis (PCA) can also help to investigate the variance in the data and highlight aberrant patterns. After the data exploration, some parts of the spectrum can be used as explanatory variables in supervised learning models with agronomic parameters as response variables. Random forest method could also be useful in order to highlight the most interesting features in order to exclude unnecessary features in the models according to specific agronomic variables.

Agronomic data concerning specific experimental site and year was stored in single csv files. For hyperspectral time series data, final readable csv files were stored separately according to collection date, site and year. Data management steps needed to be performed in order to make the data available for exploration, visualization and analysis. Raw data from the hyperspectral instrument was stored with ".sed" format. Additionally, a csv file including experimental unit information identity was needed in order to match the hyperspectral raw file. R script was then used to extract reflectance values from hyperspectral file and organize them with one reflectance wavelength by column. This step was assessed by using R software with "asdreader" [5] and "tidyverse" [6] libraries. Data management processes to link the different datasets and merge them for analysis and visualization were assessed by using pandas library [7] in Python. Data visualization was also done in Python by using matplotlib [8] and seaborn libraries [9].

3 Data

The data was collected in a research project called "Wheat advisor" focus on nitrogen management according to variety response in interaction with specific environment. In this context, innovative methods like UAV, satellite and hyperspectral data (ground level) were investigated in order to know if they provide relevant information to characterize variety response to its environment.

For this study, two main types of data could be distinguished: agronomic data and time series hyperspectral data. Agronomic data (Table 2) included all the observations and measurements for each crop season and for each site). In total, there were six csv files (three sites x two years) that could be merged together to have an overview during data visualization.

Table 2. Dataframes printed from Jupyter notebook including agronomic information for the three sites (Changins, Goumoens and Reckenholz) in 2021. These dataframes contain categorical information describing specific experimental unit (ID), crop season (Year), experimental location (site), variety name (Variety) and nitrogen treatment (NTrt) but also quantivative variables like grain yield (YieldGrain0percKgM2), biomass at anthesis (BmflowDry) and total N in grain (TotNGrainGround) to measure crop variety performance.

Changins

	ID	Experiment	Year	site	Variety	NTrt	YieldGrain0PercKgM2
364	1	WheatAdvisor_exact	2021	CH	CH_Nara	N4	0.708323
365	2	WheatAdvisor_exact	2021	CH	CH_Claro	N0	0.319911
366	3	WheatAdvisor_exact	2021	CH	Runal	N1	0.665302
367	4	WheatAdvisor_exact	2021	CH	Montalbano	N1	0.586376
369	6	WheatAdvisor_exact	2021	CH	Runal	N4	0.771605

	BMflowDry	BMflowDry	TotNGrainGround
364	1.444444	1.444444	2.341919
365	0.577778	0.577778	1.974734
366	1.377778	1.377778	2.110578
367	1.011111	1.011111	1.887535
369	1.522222	1.522222	2.646264

Goumoens

	ID	Experiment	Year	site	Variety	NTrt	YieldGrain0PercKgM2
64	1	WheatAdvisor_exact	2021	GO	Montalbano	N4	0.620691
65	2	WheatAdvisor_exact	2021	GO	CH_Claro	N0	0.414331
66	3	WheatAdvisor_exact	2021	GO	CH_Camedo	N4	0.676671
67	4	WheatAdvisor_exact	2021	GO	CH_Nara	N4	0.668456
68	5	WheatAdvisor_exact	2021	GO	Runal	N0	0.375471

	BMflowDry	BMflowDry	TotNGrainGround
64	1.755556	1.755556	2.359133
65	1.177778	1.177778	1.945523
66	2.022222	2.022222	2.588087
67	1.711111	1.711111	2.425046
68	1.233333	1.233333	1.939422

Reckenholz

	ID	Experiment	Year	site	Variety	NTrt	YieldGrain0PercKgM2
139	1	WheatAdvisor_exact	2021	RE	CH_Claro	N0	0.378502
140	2	WheatAdvisor_exact	2021	RE	Montalbano	N0	0.208430
141	3	WheatAdvisor_exact	2021	RE	Runal	N0	0.178234
142	4	WheatAdvisor_exact	2021	RE	CH_Camedo	N0	0.278872
143	5	WheatAdvisor_exact	2021	RE	CH_Nara	N0	0.254449

	BMflowDry	BMflowDry	TotNGrainGround
139	0.612403	0.612403	2.070925
140	0.472868	0.472868	2.018524
141	0.682171	0.682171	2.233012
142	0.503876	0.503876	1.909466
143	0.581395	0.581395	2.048005

Hyperspectral data included mainly wavelength reflectance information and categorical variables to make the link with agronomic data like site and year (Table 3). As it was time series data, several csv files corresponding to different data collection dates could be link to one specific agronomic csv dataset by year and site. Frequency of time series data depended on site for logistic reason.

Table 3. Dataframes printed Jupyter notebook including wavelengths reflectance for a specific data collection date (crop stage at heading).

	ID	CropStage	rflt_359	rflt_360	rflt_361	rflt_362	rflt_363	rflt_364	rflt_365	rflt_366	...
450	1	Heading	0.018496	0.018435	0.018400	0.018405	0.018403	0.018397	0.018387	0.018440	...
451	2	Heading	0.020588	0.020687	0.020785	0.020865	0.020946	0.021021	0.021073	0.021132	...
452	3	Heading	0.011134	0.011124	0.011104	0.011091	0.011060	0.011027	0.011021	0.011047	...
453	4	Heading	0.014384	0.014365	0.014328	0.014269	0.014220	0.014181	0.014157	0.014164	...
454	5	Heading	0.031535	0.031565	0.031607	0.031671	0.031738	0.031801	0.031855	0.031862	...

5 rows × 2285 columns

Time series hyperspectral data could provide complex information concerning ecophysiogical crop status over time and this information variability could already been observed with histograms (Figure 4). The idea would be to link this spectrum dynamic to time point measurement like grain yield or total N in grain (Figure 5) which have been affected by various factors during the crop growth. Regarding the structure of the grain yield, bimodality of data distribution could already be observed. That was probably due to the fact that there were two main groups: with or without N applications.

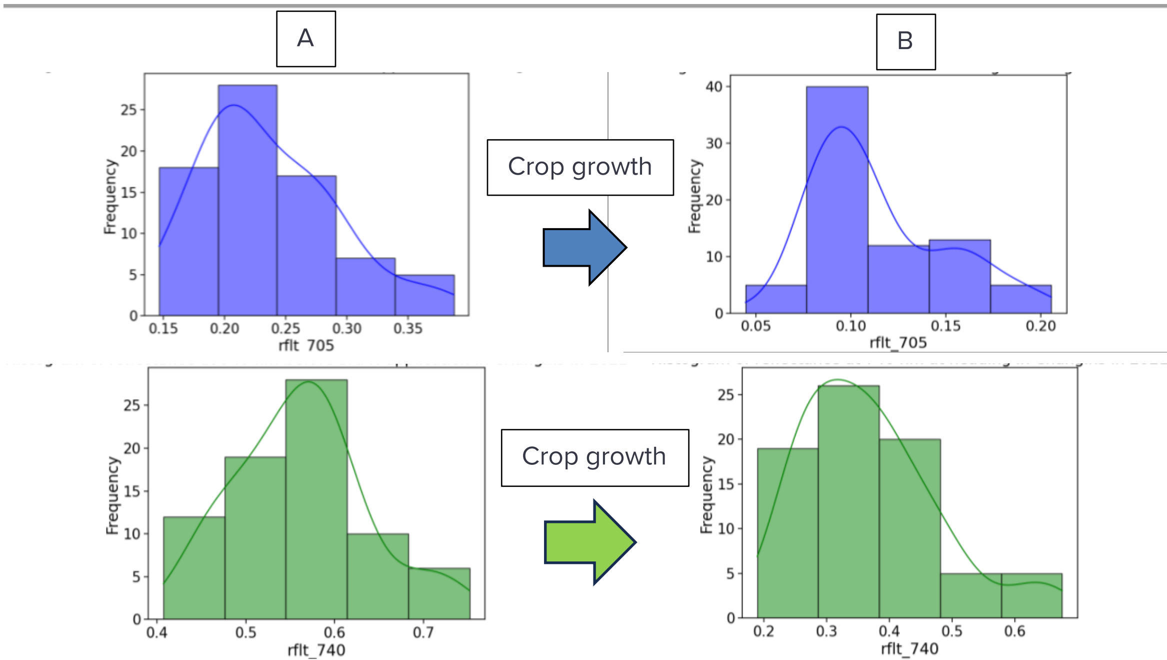


Figure 4. Histograms of reflectance at 705 nm (blue) and 740 nm (green) at two different crop stages in Changins in 2021: before second N application (A) and at heading (B).

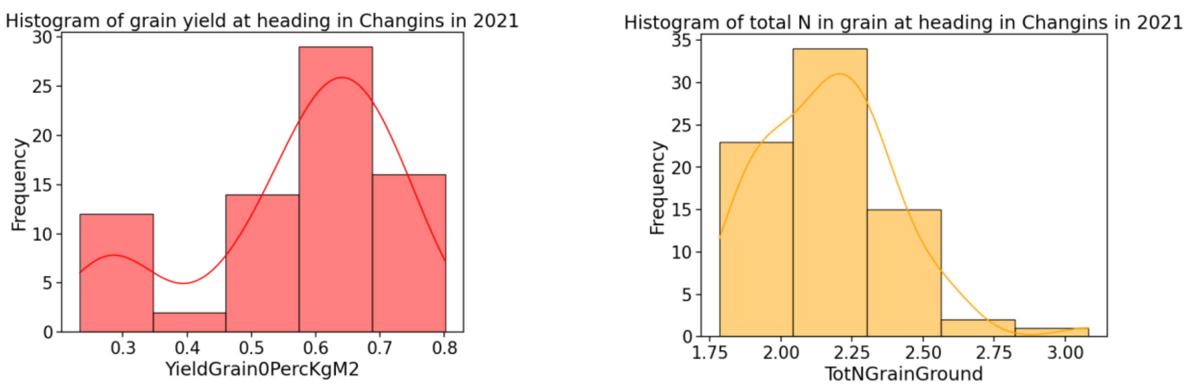


Figure 5. Histogram of two agronomic parameters in Changins 2021: grain yield (red) and total N in grain (orange).

4 Metadata

Two main types of metadata were necessary to store in order to improve the reproducibility of this study: data collection protocol and environmental condition. In the data collection protocol, the important parameters to take into consideration are summarized in the Table 4. For environmental condition, the most important parameters reported were the presence of clouds, air/soil/plant humidity and specific crop situation (ex: disease, low emergence, lodging and crop damage).

Table 4. Metadata concerning data collection parameters.

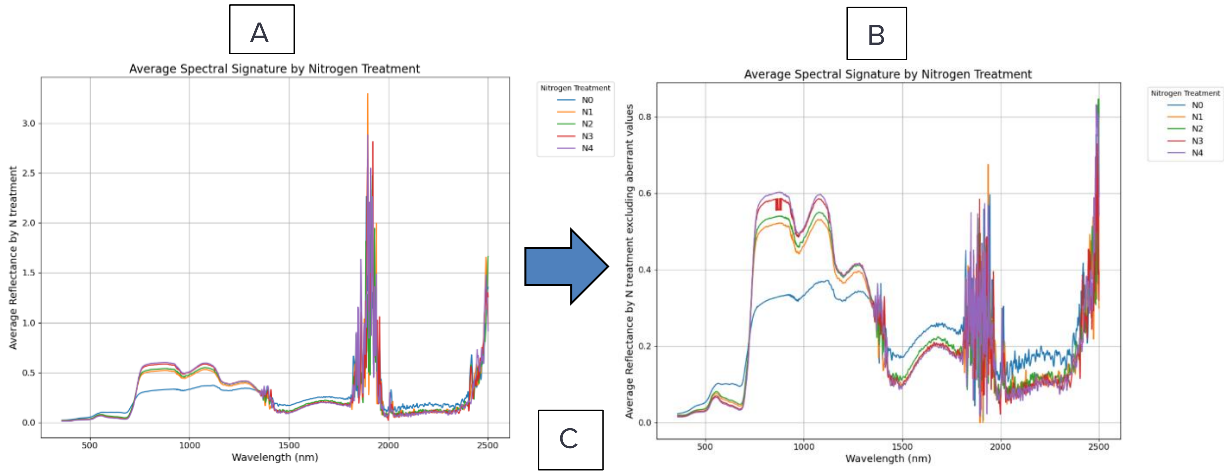
Metadata parameters	Description
Data capture orientation Scanning type	Top view; scan; contact probe Static or in movement
Number of measurements to be averaged when triggering the scan	Trade-off between measurement time and sample representativeness
Crop growth stage	Approximate because varieties can have different phenology at a specific date
Date of measurement	Specific date of data collection
Time of measurements	Interval period (from start to end)
Triggering target	Canopy; soil; ears
Number of white reference calibration between samples	Depend on environmental conditions (changing conditions like clouds involve higher frequency of calibration measurements)

This metadata could be stored in single csv file with year, site and date columns and could be linked if needed to hyperspectral data during the analysis.

5 Data Quality

Agronomic data can be very noisy and difficult to interpret because there are collected on the field with a lot of uncontrolled factors. In general, for this kind of data, some variance is expected and can be sometimes reduced with spatial correction analysis taking into consideration field heterogeneity. Concerning statistical power, for this kind of field trial, the number of replicates for each combination of treatment is limited to three (due to space availability on the field). For that reason, losing a sample can have an effect on statistical power and variance analysis and should be avoided.

Concerning hyperspectral data, they also included a lot of noise and should be pre-processed carefully before to use them in machine learning methods. This data can be very sensitive to environmental conditions especially at specific ranges of wavelength (Figure 6). This data can also contain aberrant values outside the reflectance range which should be between 0 and 1. These aberrant values should be removed before visualization and analysis. There were also some issues keeping the zero values because it could have an impact on the results when including them in statistical analysis.



rflt_1897	rflt_1898	rflt_1899	rflt_1900	rflt_1901	rflt_1902	rflt_1903	rflt_1904	rflt_1905	rflt_1906	rflt_1907	rflt_1908	rflt_1909	rflt_1910	rflt_1911	rflt_1912	rflt_1913	rflt_1914	rflt_1915
3.65109	1.57072	0.11207	0	0	0.49875	3.40531	6.31836	6.93534	5.42574	3.44232	2.08599	0.8421	0.0401	0.09552	0.68635	0.99688	0.70513	0.22659
0	0	0	0	0	0.20351	1.30842	2.54547	3.29907	3.65119	3.90309	4.32029	4.82436	5.0141	4.6142	3.82203	2.90764	1.75334	0.56256
5.4275	2.55384	0.48193	0	0	0	0	0	0.32878	1.30063	2.57771	4.41397	6.88327	8.03605	6.11016	2.43515	0	0	0
0	0	0	0	0	0.04905	0.22445	0.6276	1.59903	3.30271	4.76955	5.17186	4.84432	4.08858	2.79168	1.08563	0	0	0.01463
5.37194	2.27108	0.0801	0	0	0.3443	2.32134	4.40049	5.14362	4.61088	4.00978	4.28905	5.07373	5.92674	6.98027	8.06941	7.88266	5.33062	1.95038
0.81859	2.04315	2.59596	1.91616	0.76257	0.07337	0	0	0.27699	1.21087	1.94526	1.68752	0.6835	0.00747	0.48706	1.51293	1.99315	1.3634	0.38286
1.93819	0.74591	0	0	0	0.12372	0.73757	1.55642	2.52785	3.81587	4.84425	5.10155	4.79747	4.08797	2.79606	1.09488	0	0	0
3.27589	1.56713	0.33642	0.07524	0.17031	0.17755	0.0956	0.02367	0	0	0	0	0	0	0	0	0.0064	0.0616	0.18455
1.14922	0.42958	0	0	0	0.20863	1.36616	2.60572	3.19384	3.23986	3.05809	2.77345	2.2949	2.00312	2.3005	2.87672	2.96343	2.00092	0.77056
6.31177	2.53065	0	0	0	0.3682	2.50787	4.58384	4.73626	2.9105	1.17578	1.2578	2.49347	3.88858	5.28418	6.71098	6.92894	4.96177	2.18082
0	0	0.29913	1.14134	2.11909	2.53421	2.28112	1.75126	1.24262	0.62538	0.09923	0	0	0	0	0	0.02468	0.46145	0.99958
0.78897	0.31633	0	0	0	0.18593	1.27455	2.31343	2.34271	1.37021	0.32577	0	0	0	0.8807	2.24004	3.02623	2.84011	2.19731
0.54719	0.00091	0.05426	1.35364	2.97649	3.44338	2.39589	0.97024	0.25355	0.05448	0.01941	0	0	0	0.89127	2.24505	3.02574	2.80263	2.14106
1.8161	1.2263	0.71446	0.38403	0.12306	0	0	0	0.006	0	0.4497	2.78224	6.7009	8.98935	7.06027	2.82781	0	0	0
5.44003	3.69178	2.44344	2.3637	2.59375	2.32541	1.35329	0.31996	0	0	0	1.03008	2.87957	3.99043	3.14675	1.24665	0	0	0.10897
3.01297	1.19763	0	0	0	0.2584	1.76293	3.21643	3.25489	1.74557	0.56125	1.42127	3.72405	5.85795	7.24715	8.21736	7.8752	5.18782	1.64394
5.28126	3.29602	1.94438	1.96428	2.41014	2.42069	1.89147	1.24672	0.75402	0.21031	0.12636	1.3102	3.59699	4.9896	3.93647	1.57768	0	0	0
0	0	0.02033	0.41413	0.86717	0.97242	0.60299	0.13592	0	0.02957	0.13389	0.12389	0.05917	0.00291	0	0	0.03497	0.67788	1.44723
3.82576	1.52524	0	0	0	0	0	0	0	0	0	0.83589	2.60319	3.93772	4.10994	3.62621	2.92847	1.86369	0.70435
1.51907	0.54738	0	0.12876	0.60934	1.19642	2.14235	2.92237	2.73378	1.50242	0.31528	0	0.11688	0.88174	2.99495	5.83466	6.96641	4.94827	1.75135
9.26216	8.25811	6.86398	4.98083	2.85051	1.44894	0.91119	0.71853	0.45186	0.20684	0.03585	0	0	0	0	0	0.03759	0.59401	1.38275
0.78021	0.29783	0	0.01707	0.17226	0.26313	0.27788	0.28465	0.29655	0.24524	0.37956	1.02185	2.09111	2.95082	3.18501	3.08764	3	3	3
4.01407	10.0492	13.0255	10.5126	5.76396	2.60768	1.12044	0.29945	0	0	0	1.4921	4.64609	6.91176	6.81081	5.38885	3.86886	2.12624	0.461
0.52924	1.41811	2.03425	2.22447	2.16642	1.88139	1.22152	0.51891	0.21235	0.26766	0.37709	0.30411	0.14125	0.00744	0	0	0.09983	1.93012	4.12489
0.95261	2.37456	2.99046	2.10132	0.65867	0	0	0.23182	0.24433	0	0	0.89976	2.5685	3.92348	4.39194	4.35785	3.93777	2.66389	1.25786
1.74158	2.69089	2.97404	2.04297	0.65633	0	0	0	0	0	0	0	0	0	3.49292	8.70671	10.9645	7.66269	2.35185
2.36691	0.94899	0	0	0	0.04842	0.31588	0.60499	0.75857	0.81762	0.78215	0.60118	0.27782	0.01482	0	0	0.04005	0.78142	1.66402
0	0	0.00998	0.19301	0.41249	0.50578	0.42347	0.31326	0.33885	0.50708	0.62806	0.5209	0.24479	0.01253	0	0	0.04067	0.82827	1.73434
0.49318	1.42973	2.02426	2.03145	1.75393	1.37434	0.78361	0.19059	0	0	0	0.65621	2.0429	2.97404	2.6921	1.74414	0.94708	0.4879	0.15627
4.94584	5.88809	5.93987	4.42599	2.30051	1.0106	0.70185	0.72965	0.62906	0.51624	0.41081	0.22911	0	0	0.46711	1.27388	2.07387	2.97309	3.78122

Figure 6. Spectral signature of Changins in 2021 at heading with reflectance values (y-axis) for each wavelength (x-axis) before (A) and after removing aberrant values (B). Spectral signal is grouped by N treatments: N0 (blue), N1 (orange), N2 (green), N3 (red), N4 (purple). The table below (C) illustrates a part of the spectrum (1897 to 1915 nm) which contain noisy data with aberrant values (green>1).

6 Data Flow

All the processing steps from data collection to crop traits prediction using hyperspectral data are presented in the following diagram (Figure 7).

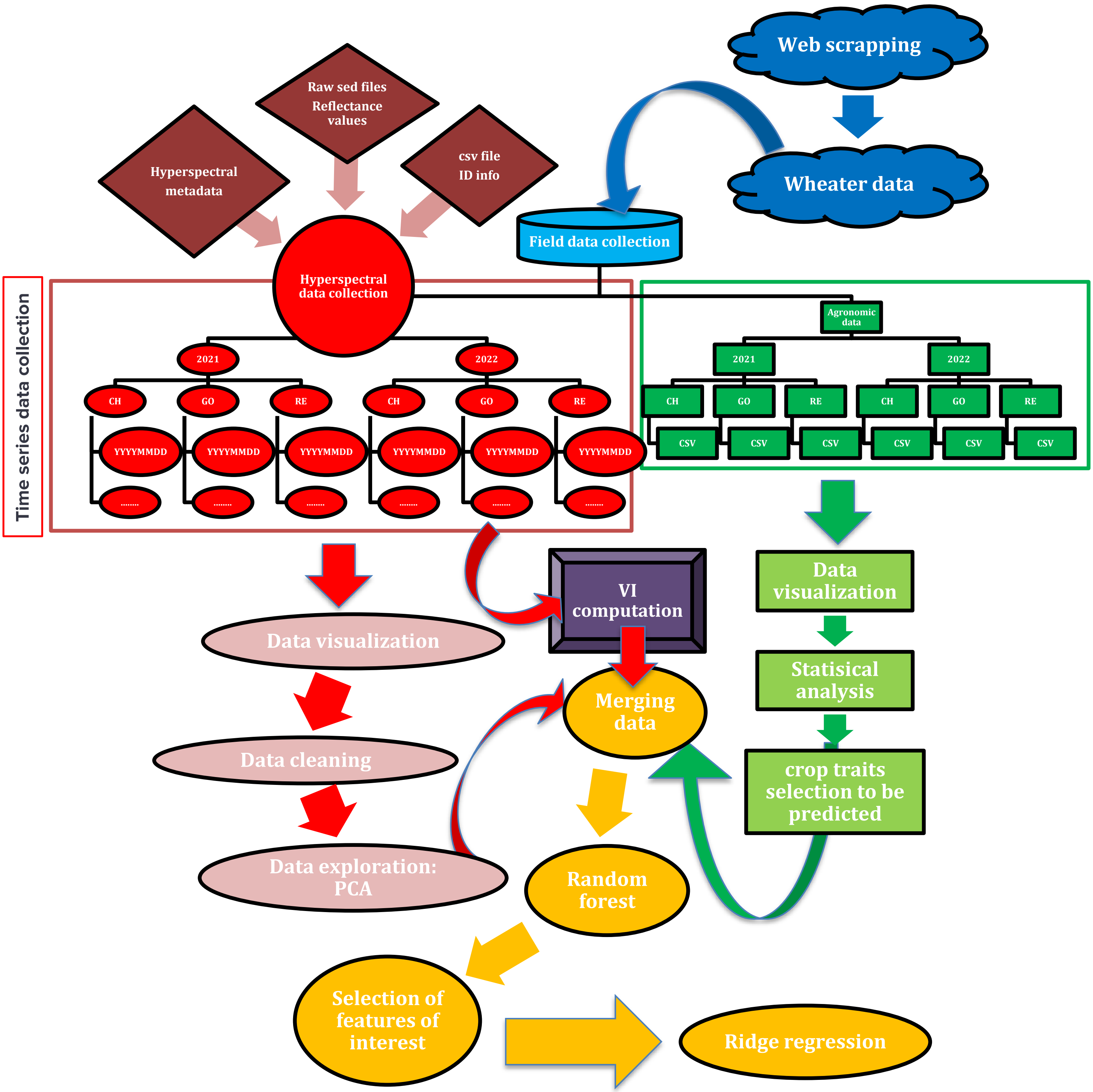


Figure 7. Data flow diagram from field data collection to crop traits prediction; CH: Changins; GO: Goumoens; RE: Reckenholz;

7 Data Model

The models investigated in this study could be used to characterize winter wheat response in nitrogen management context in order to optimize grain yield and quality while reducing fertilizer amount. Some actual companies are already using prediction models to give nitrogen recommendation to farmers using for example satellite data.

The features that will be used as inputs to the models will be in a first step reflectance values that has been filtered from data cleaning, PCA and random forest analysis. The highlighted features in previous analysis will be then used in a ridge regression model to avoid overfitting. Finally, the performance of these models will be compared to models using specific VIs already tested in literature. If needed, weather parameters could also be included as features in order to control the environment effect on reflectance values.

If the results of this study are conclusive, field trials and data collection would need to be organized in order to validate the models with more varieties and environments. To really use the models developed in this study in practice, some links still will need to be made between hyperspectral ground measurements and satellite sensors because nowadays this is the most promising tool to support farmers decision in fertilization management. At the end of the project, to be able to deploy the model in an interactive way with farmers, some application developments could be made by using for example Microsoft Power BI.

8 Documentation

The pipeline used in this study could be integrated into a github repository including the datasets and the R/Python scripts. Additional protocols will be written to guide the user through the pipeline and help him to reproduce the study. By this way, for research purpose, the observations in this study could be validated by other datasets with different environments.

9 Risks

Hyperspectral data collection is very sensitive to specific environment condition (blue sky, low wind, dry plant material and short time frame around noon) to have good quality data and all the criterions were not always respected due to logistic reasons especially for remote field experimental sites like Goumoens and Reckenholz. Without consistent good quality data over time and over sites, it might be difficult to find consistent patterns to be generalized in the model for a broader application. As the data has already been collected, the only way to deal with that is to be careful during data exploration to avoid including misleading information in the next steps with supervised learning methods which will result in bad model performance.

The other concern will be how to interpret the reflectance values at a specific time point. Indeed, reflectance value is affected by the crop stage of the wheat which can differ according to varieties. As five varieties were included in this study, it will be important to take into consideration this parameter in order to be able to interpret accurately the results.

10 Preliminary Studies

Some preliminary visualizations have already been performed on agronomic data in 2021 to observe the effect of the N treatments on varieties performance according to the three experimental sites (Figure 8, 9 and 10). Main N treatments (none, reduced and conventional application) showed already contrasted results for different crop traits like grain yield, total N in grain and biomass but this observation depended also on the experimental site (more contrast in Changins for example).

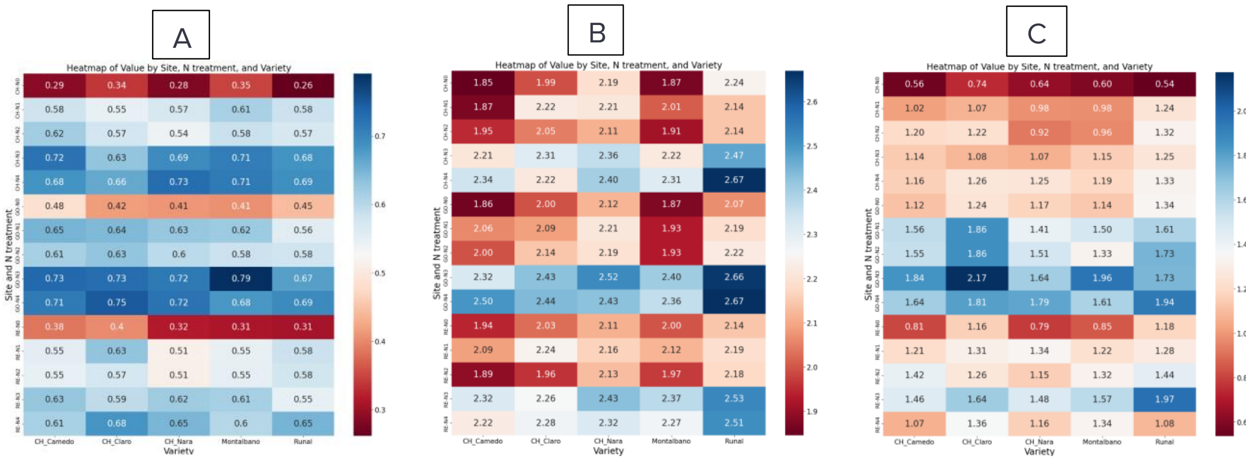


Figure 8. Heatmaps of agronomic parameters like grain yield in kg/m^2 (A) , total N in grain in % (B) and biomass in kg/m^2 (C); CH: Changins; GO: Goumoens; RE: Reckenholz; N0: $N_{0\text{-}0\text{-}0}$; N1: $_{N20\text{-}40\text{-}20}$; N2: $N_{20\text{-}60\text{-}0}$; N3: $N_{40\text{-}80\text{-}40}$; N4: $N_{40\text{-}120\text{-}0}$.

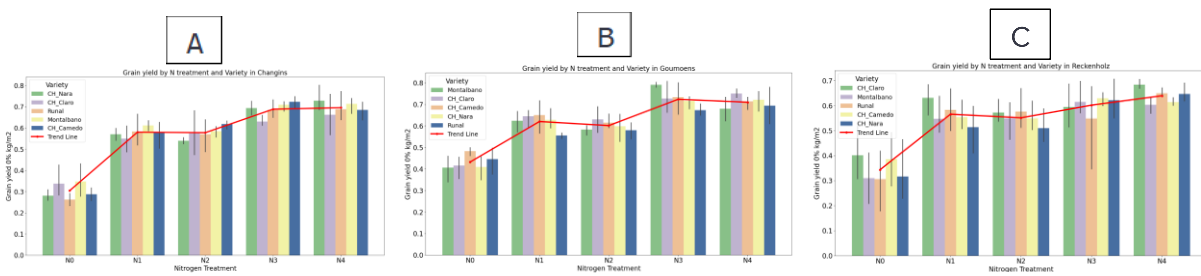


Figure 9. Barplots of grain yield in kg/m^2 at the three experimental sites: Changins (A), Goumoens (B) and Reckenholz (C). Red trend line shows the variance in grain yield according to N treatments. N0: $N_{0\text{-}0\text{-}0}$; N1: $_{N20\text{-}40\text{-}20}$; N2: $N_{20\text{-}60\text{-}0}$; N3: $N_{40\text{-}80\text{-}40}$; N4: $N_{40\text{-}120\text{-}0}$.

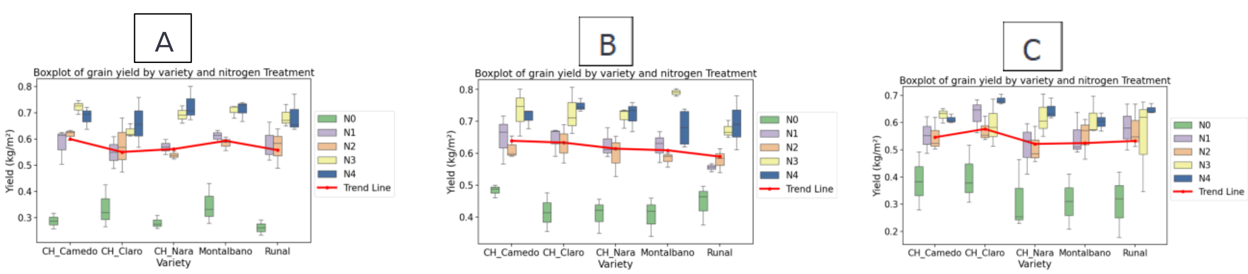


Figure 10. Boxplots of grain yield in kg/m^2 at the three experimental sites: Changins (A), Goumoens (B) and Reckenholz (C). Red trend line shows the variance in grain yield according to varieties. N0: $N_{0\text{-}0\text{-}0}$; N1: $_{N20\text{-}40\text{-}20}$; N2: $N_{20\text{-}60\text{-}0}$; N3: $N_{40\text{-}80\text{-}40}$; N4: $N_{40\text{-}120\text{-}0}$.

Some similar patterns to agronomic pattern could also been observed in spectral signature according to experimental sites in 2021 focusing on specific range of the spectrum (Figure 11).

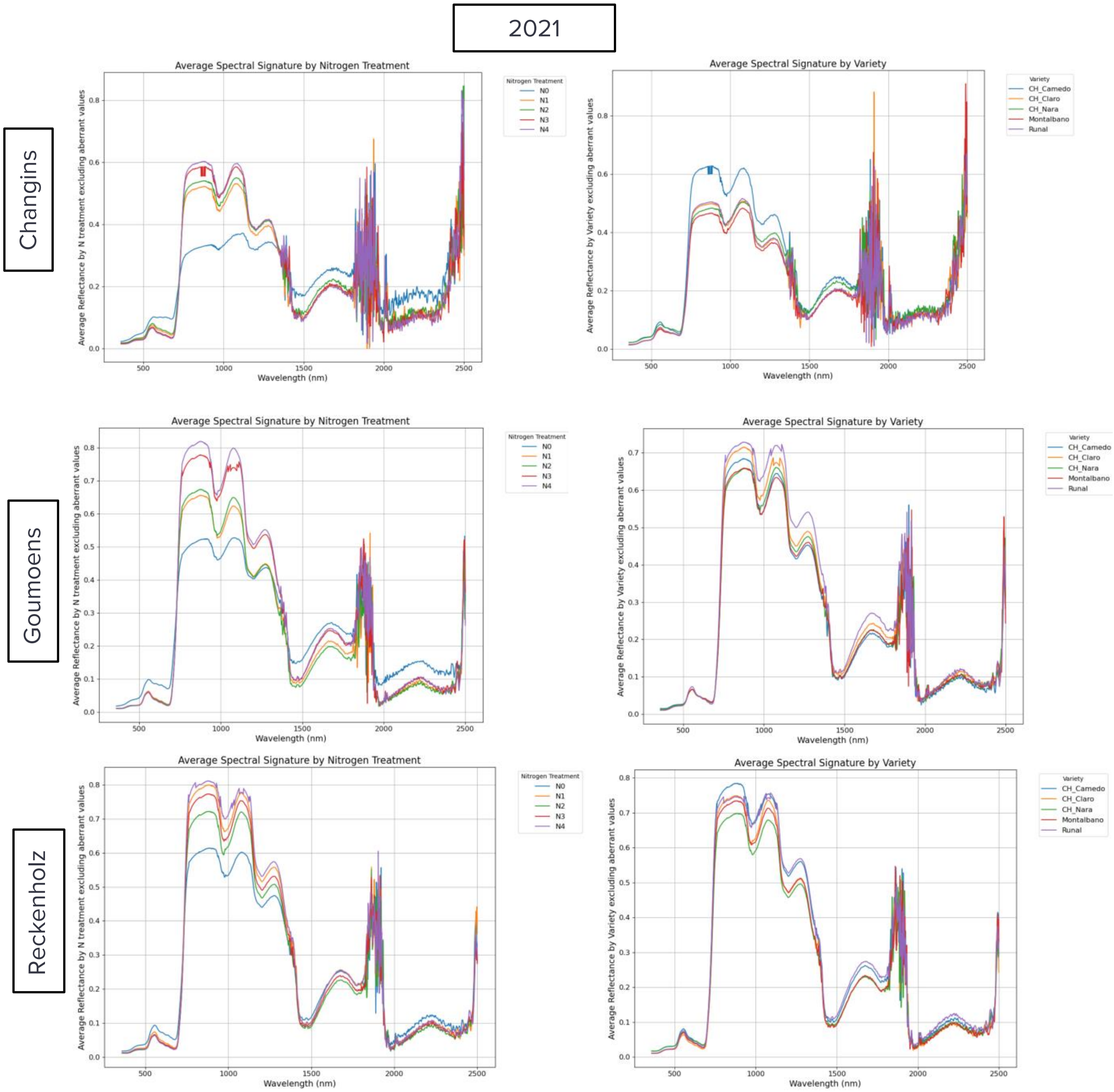


Figure 10. Spectral signature by sites grouped by N treatments (on the left) and varieties (on the right); N0: $N_{0\text{-}0\text{-}0}$; N1: $N_{20\text{-}40\text{-}20}$; N2: $N_{20\text{-}60\text{-}0}$; N3: $N_{40\text{-}80\text{-}40}$; N4: $N_{40\text{-}120\text{-}0}$.

11 Conclusions

Even if the objectives of the studies are challenging, preliminary visualizations already showed promising similarity between agronomic and hyperspectral data. Considering that, this kind of data can contain a lot of noise, it was encouraging to see similar patterns at N treatments level but also at variety level. By combining several sources of information among the spectrum, it should be possible to characterize accurately the variety response at specific N level, site and year in order to understand crop N needs.

Acknowledgements

I would like to thank the Swiss Federal Office for Agriculture [project number 19.03], Agroscope, swiss granum, Schweizerischer Getreideproduzentenverband SGPV, Timac Agro, JOWA AG, Prometerre/Proconseil, Forum Ackerbau, and Groupe Culture Romandie for supporting and funding the Wheat advisor project (data source). I would also like to thank my colleagues in Agroscope of Changins, Juan Manuel Herrera, Lilia Levy Häner, Silvan Strebel, Amanda Burton, Noémie Schaad and Didier Pellet who contributed to the project management and the data collection.

Statement

The following part is mandatory and must be signed by the author or authors.

„Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls die Arbeit als nicht erfüllt bewertet wird und dass die Universitätsleitung bzw. der Senat zum Entzug des aufgrund dieser Arbeit verliehenen Abschlusses bzw. Titels berechtigt ist. Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbstständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“

Date: 10.10.2024 Signature(s):

References and Bibliography

[1] Thenkabail, P. S., Mariotto, I., Gumma, M. K., Middleton, E. M., Landis, D. R., & Huemmrich, K. F. (2013). Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/EO-1 data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 6(2), 427-439. https://doi.org/10.1109/JSTARS.2013.2252601

[2] Argento, F., Anken, T., Abt, F., Vogelsanger, E., Walter, A., & Liebisch, F. (2021). Site-specific nitrogen management in winter wheat supported by low-altitude remote sensing and soil data. Precision Agriculture, 22, 364-386. https://doi.org/10.1007/s11119-020-09733-3

[3] Wang, J., Meyer, S., Xu, X., Weisser, W. W., & Yu, K. (2024). Drone Multispectral Imaging Captures the Effects of Soil Nmin on Canopy Structure and Nitrogen Use Efficiency in Wheat. Available at SSRN 4699313. https://dx.doi.org/10.2139/ssrn.4699313

[4] Burton, A., Häner, L. L., Schaad, N., Strebel, S., Vuille-dit-Bille, N., Bongiovani, P. D. F., Holzkämper, A., Pellet, D., & Herrera, J. M. (2024). Evaluating nitrogen fertilization strategies to optimize yield and grain nitrogen content in top winter wheat varieties across Switzerland. Field Crops Research, 307, 109251. https://doi.org/10.1016/j.fcr.2024.109251

[4] Roudier, P. (2016). asdreader: Reading ASD Binary Files in R. Retrieved from https://cran.rproject.org/package=asdreader

[5] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., & others. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686

[6] W. McKinney, pandas: a python data analysis library, http://pandas.sourceforge.net

[7] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science &Engineering, 9(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

[8] Waskom, M. L. (2021). Seaborn: Statistical data visualization. Journal of Open Source Software, 6(60), 3021. https://doi.org/10.21105/joss.03021