



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Economic Affairs,
Education and Research EAER

Agroscope

M3-Machine learning on hyperspectral data - wheat trial



Nicolas Vuille-dit-Bille



Agroscope: swiss federal institute for agricultural research

Group:

- Cultivation Techniques and Varieties in Arable Farming

Team:

- 35 People

Domains of research:

- Agronomy
- Crop science

Main areas of work:

- Characterization of genotypes of field crops
- Improvement of management practices of field crops

Crop species:


- Winter cereals (**wheat**, barley, oat, ...)
- Maize
- Potato and Sugar beet
- Oil crops and grain legumes
- Minor crops (e.g. Quinoa, Amaranth)




Agroscope

- 

Blé d'automne
Variétés


 Schweizerische Eidgenossenschaft
 Confédération suisse
 Confederazione Svizzera
 Confederaziun svizra

Département fédéral de l'économie,
 de la formation et de la recherche DFFS
 Agroscope


 swiss gramin

Variété	TOP										I					II					Fourrage				Biscuit	
	RUNAL	TITIS	CH CLARO (sem. année)	CH NARA	PIONAIR	AXEN ^a	DIABEL ^a	MONTALBANO ^b	BONAVAU ^b	BARETTA	CAOLIMO	ARINA	FOREL	SIMANO ^a	ALPVAL	HANSWIN	CAMPANILE	POSIEDA	LUDWIG	LEVIS	SPONTAN	PONCONE	CAMPESINO	SAILOR	MULAN	DILAGO
Année d'inscription	1965	1996	2009 (sem. année)	2010	2020	2022 (prov.)	2020	2018	2022 (prov.)	2018	2020	1981	2008	2012 (sem. année)	2022 (prov.)	2015	2021	2010	2004	1997	2017	2019	2022 (prov.)	2015	2007 (sem. année)	2016
Rendement (Estanso)	-	-	-	-	++	+	++	++	++(+)	++	0	0	0	+	++	++	+++	+++	+++	+	+++	+++	+++	+++	+++(+)	++
Rendement (moy ²)	++	++	++	++	++(+)	++	++	++	++	++	++	++	++	++	++(+)	++	++	++(+)	++	++	++	++	++	++	++	++
Précocité & légèreté ³	mt	1	mp	mt	mt	tp	tp	1	mt	1	0	1	mp	tp	1	mp	mt	mp	mt	mt	mt	mt	mt	mt	mt	++
Hauteur des plantes ³	m	ml	m	m	m	l	t	m	c	ml	m	tl	m	c	m	m	m	tl	tl	tl	c	m	l	c	ml	m
Verse ²	++(+)	++	+++	+++	+	++(+)	+++	+++	++	+	0	++	+++	+++	+++	+++	++	+	++	+++	++	++	++	++	+++	++
Oidium ¹	++(+)	0	+	++(+)	++(+)	++(+)	++	+	++(+)	++(+)	0	+	+	++(+)	0	+	+	+	0	++	++(+)	++(+)	++	++	+	0
Rouille jaune ¹	++	++	++	++	++	++(+)	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Rouille brune ¹	++	++	0	0	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Septoriose nodorum feuilles ^{1,3}	++	++	0	+	-	0	0	0	0	0	0	-	0	+	0	0	++	0	++	0	++	++	++	0	+	++
Septoriose nodorum épis ^{1,3}	++	0	-	0	++	0	++	++	++	++	++	++	0	-	+	0	++	++	++	++	++	++	++	++	++	++
Septoriose ténor feuilles ^{1,3}	0	0	-	0	+	++	+	++	++	++	0	-	++	+	++	++	++	++	++	++	++	++	++	++	++	++
Fusariose épis ^{1,3}	++	+	-	0	0	++	0	+	0	0	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Germénation sur pied ¹	++(+)	+	+	+	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Teneur en protéines ¹	+++	++(+)	+++	+++	+++	+++	+++	+++	+++	++(+)	++(+)	++(+)	++(+)	++(+)	++(+)	++(+)	++	++	++	++	++	++	++	++	++	++
Zélaïny ¹	+++	++	+++	+++	++(+)	+++	+++	+++	+++	++(+)	++(+)	++(+)	++(+)	++(+)	++	++	++	++	++	++	++	++	++	++	++	++
Poids & ténor ¹	++	++	0	++	++	++	++	++	++	0	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++
Poids de mille grains ¹	moyen	grand	petit	petit	petit	moyen	petit	grand	moyen	petit	très petit	moyen	très petit	moyen	grand	grand	moyen	très grand	très grand	moyen	très grand	petit	très grand	très grand	moyen	grand

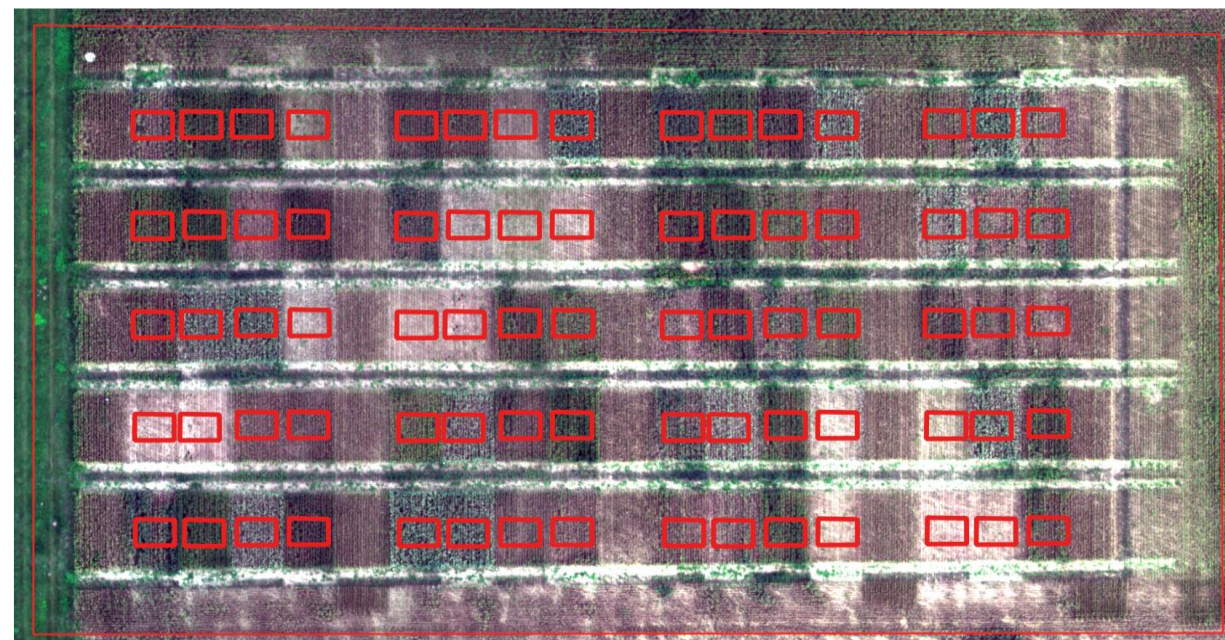
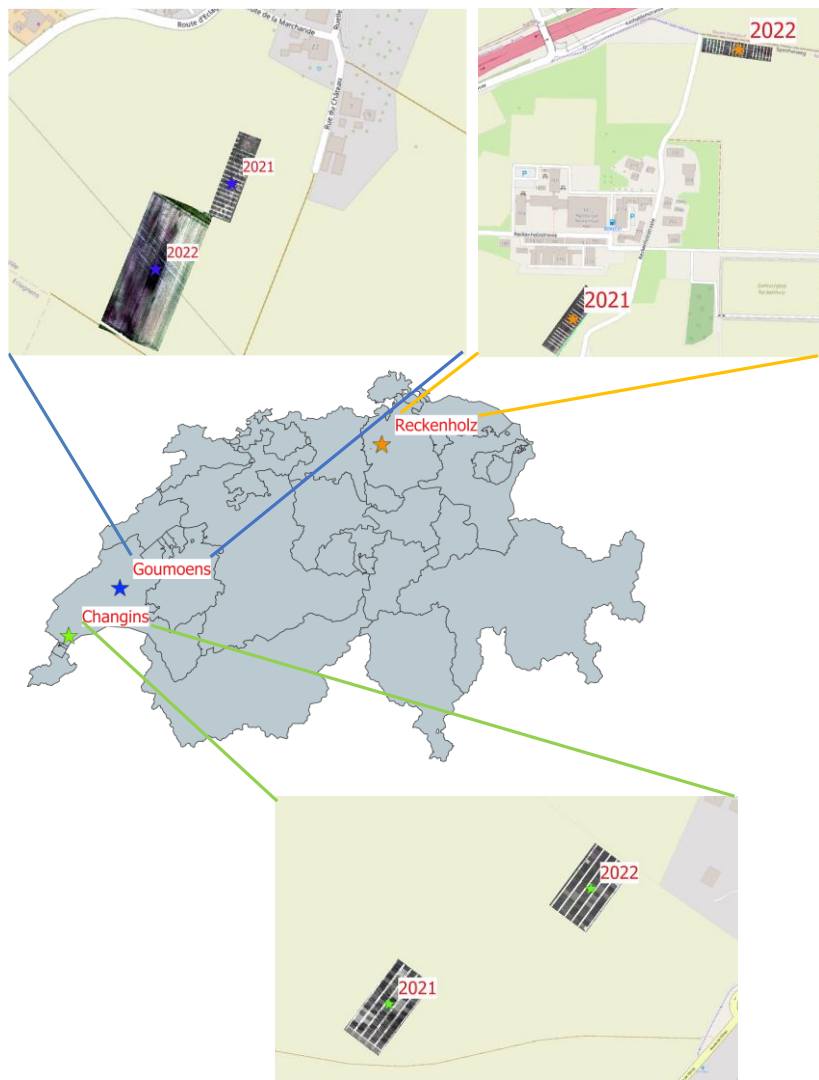


Case study: winter wheat trial with 2 treatments over 2 years on 3 sites

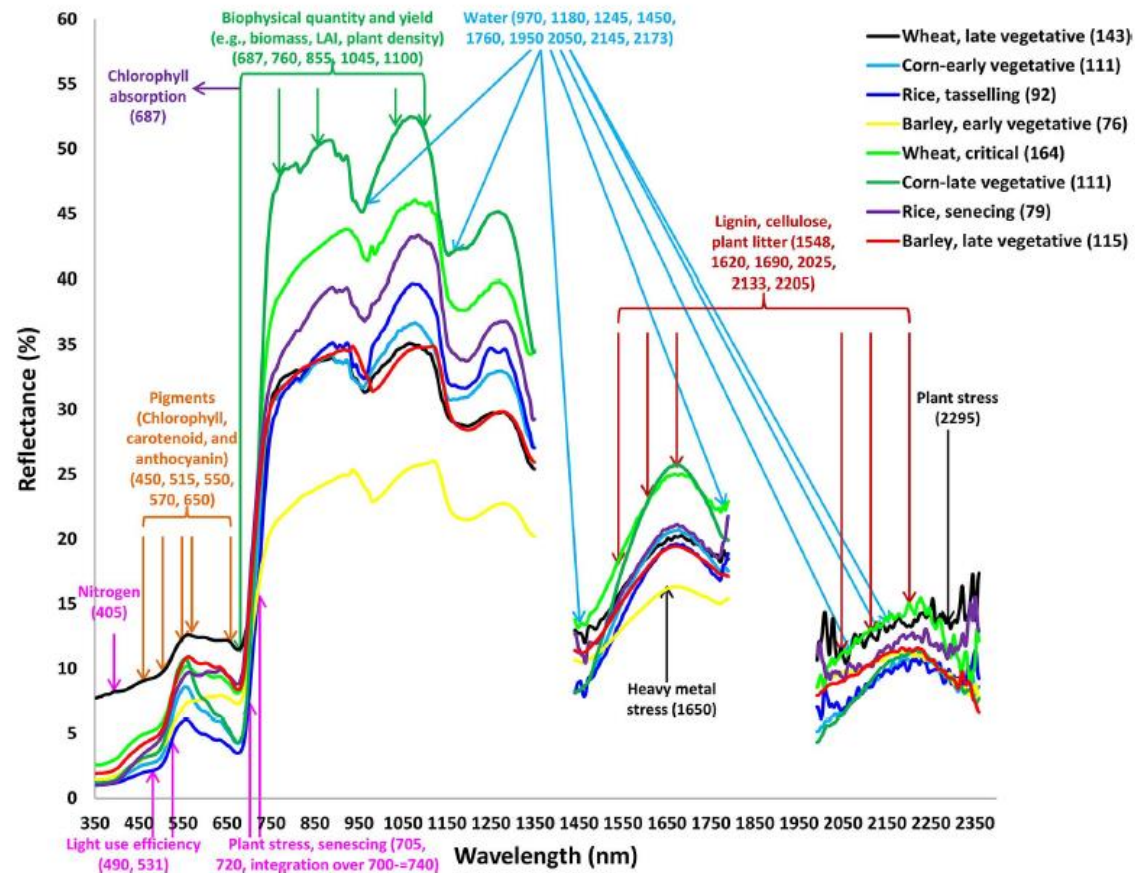
- 5 winter wheat varieties
- 2 years: 2021 to 2022
- 3 sites: Changins, Goumoens and Reckenholz
- Case study context: study variety response to different N treatment
 - Find variety that are performing well also under limited available N in order to reduce N fertilization (economic and environmental cost)
- 3 main treatments: none, reduced and conventional
- Measurement to estimate varieties performance:
 - Grain yield
 - Straw yield
 - Grain protein
 - Other physiological parameters (harvest index, leaf area index, chlorophyll content, canopy cover...)



Case study: field experimental design



Hyperspectral data: theory



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>

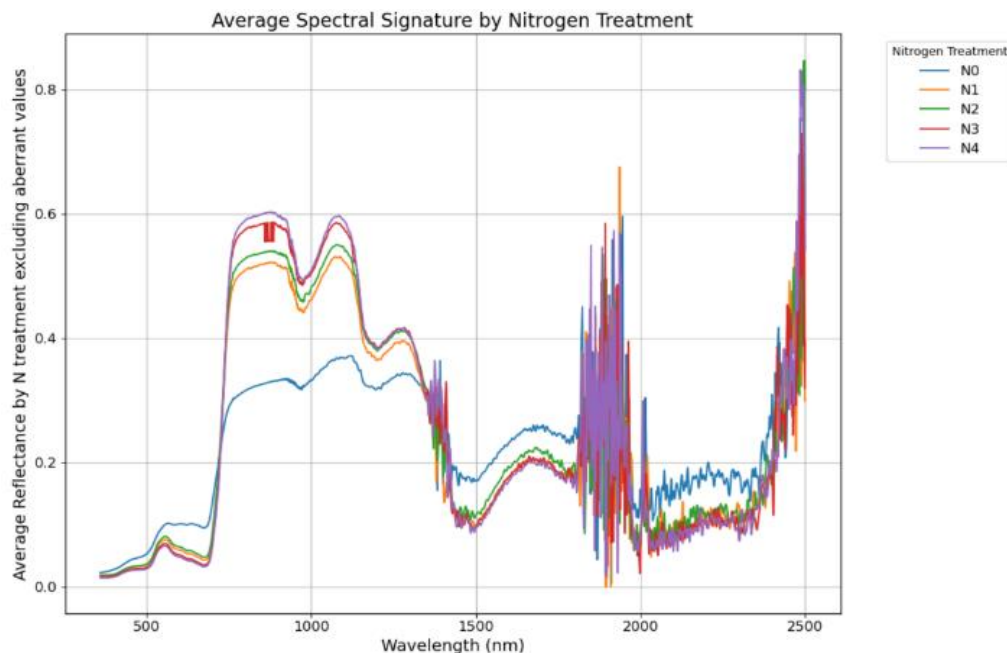


Hyperspectral data collection

	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 ...

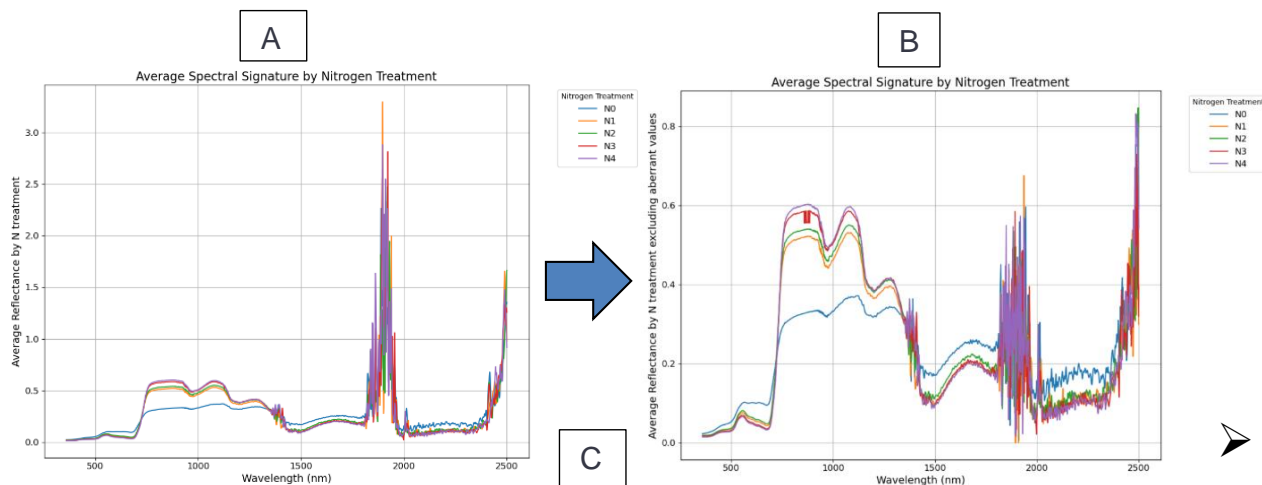
5 rows × 2285 columns

- 1) Time series data corresponding to key crop growth stage
 - 2 years: 2021-2022
 - 3 sites: Changins, Goumoens and Reckenholz





Hyperspectral data cleaning



➤ Remove bands with aberrant values : <0 or >1

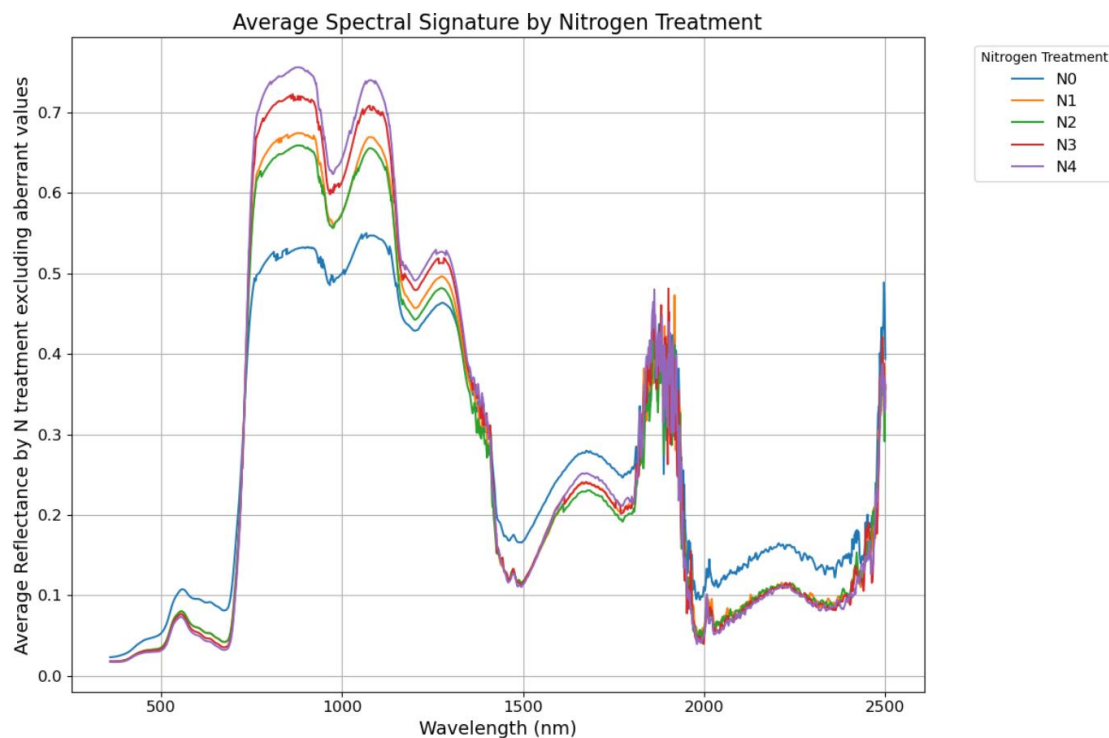
➤ Remove bands with 0 value

rft_1897	rft_1898	rft_1899	rft_1900	rft_1901	rft_1902	rft_1903	rft_1904	rft_1905	rft_1906	rft_1907	rft_1908	rft_1909	rft_1910	rft_1911	rft_1912	rft_1913	rft_1914	rft_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	0.20451	1.30842	2.54547	3.29907	3.65119	3.90309	4.37029	4.82436	5.0141	4.6142	3.82208	2.90764	1.75334
5.4275	2.55384	0.48193	0	0	0	0	0	0	0.32878	1.30063	2.57771	4.11397	6.88327	8.03605	6.11016	2.43515	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	4.78905	5.07373	5.92674	6.98027	8.06941	7.88266	3.33062	1.95048
0.81859	2.04315	2.59596	1.91616	0.76257	0.07337	0	0	0.27699	1.21087	1.91526	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	0.17091	0.17755	0.0956	0.02367	0	0	0	0	0	0	0	0.0064	0.0616	0.18445	0
1.14922	0.42958	0	0	0	0.20863	1.36616	2.60572	3.19384	3.23986	3.05809	2.77315	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	0.04468	0.46145	0.99958
0.78897	0.31633	0	0	0	0.18593	1.27455	2.31313	2.34721	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	0.12906	0	0	0.006	0	0.4497	2.78224	6.7009	8.98935	7.06027	2.82781	0	0	0	0
5.44003	3.69178	2.41434	2.3637	2.59375	2.32541	1.35329	0.31996	0	0	1.03008	2.87957	3.99043	3.14675	1.24665	0	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	8.29602	1.94438	1.96428	2.41014	2.42069	1.89147	1.74672	0.75402	0.21031	0.12636	1.3102	3.59699	4.9896	8.93647	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	0.83589	2.60319	3.93772	4.10994	3.62621	2.92847	1.86369	0.70435	0
1.51907	0.34738	0	0.1876	0.60944	1.19642	2.14235	2.92237	2.73378	1.50242	0.31528	0	0.11688	0.88174	2.99495	5.83466	6.96441	4.94827	1.75145
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	0.17126	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	5.76996	2.60768	1.12044	0.29945	0	0	1.4921	4.64609	6.91176	6.81081	5.38885	3.86886	2.17624	0.461	0
0.52924	1.41611	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	1.99046	2.10132	0.65867	0	0	0.23182	0.24433	0	0	0.89976	2.5685	3.92348	4.39184	4.35785	3.93777	2.63369	1.25786
1.74158	2.69089	2.97404	2.04297	0.65643	0	0	0	0	0	0	0	0	0	0	0	0	0	0
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	2.02476	2.08145	1.75993	1.37434	0.78361	0.19059	0	0	0.65621	2.0429	2.97404	2.6921	1.74414	0.94708	0.4879	0.15627	0
4.94584	5.88809	5.93887	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

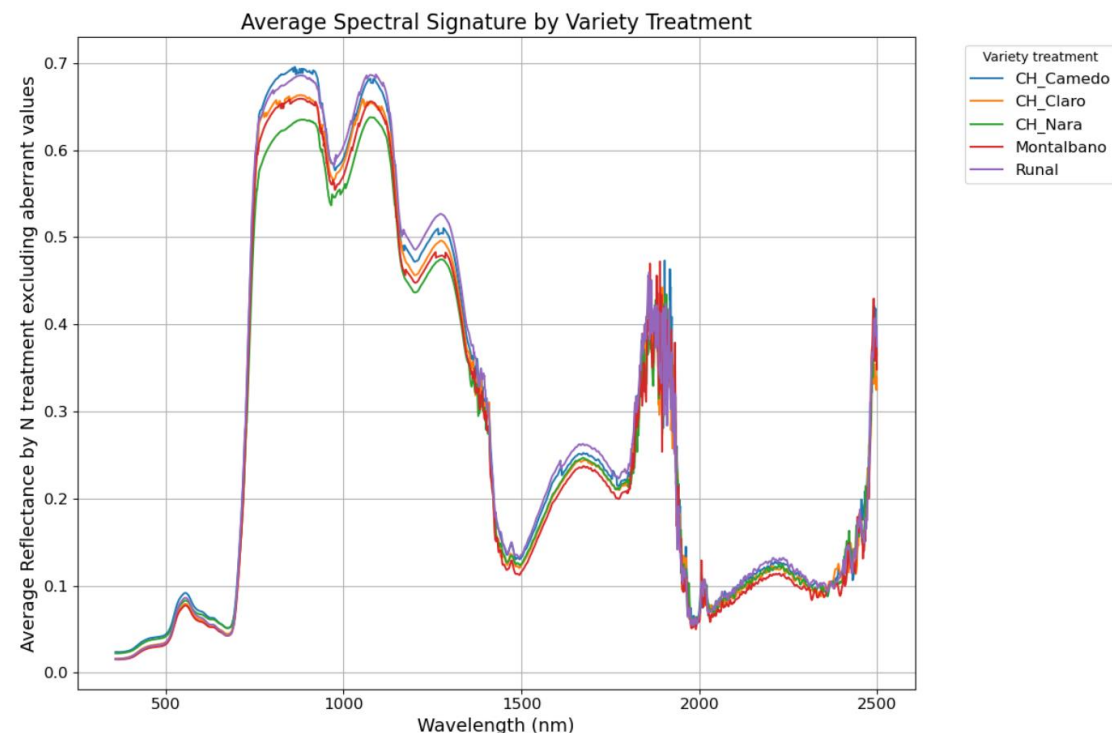


Average of spectral signature by treatments at heading

Nitrogen treatment

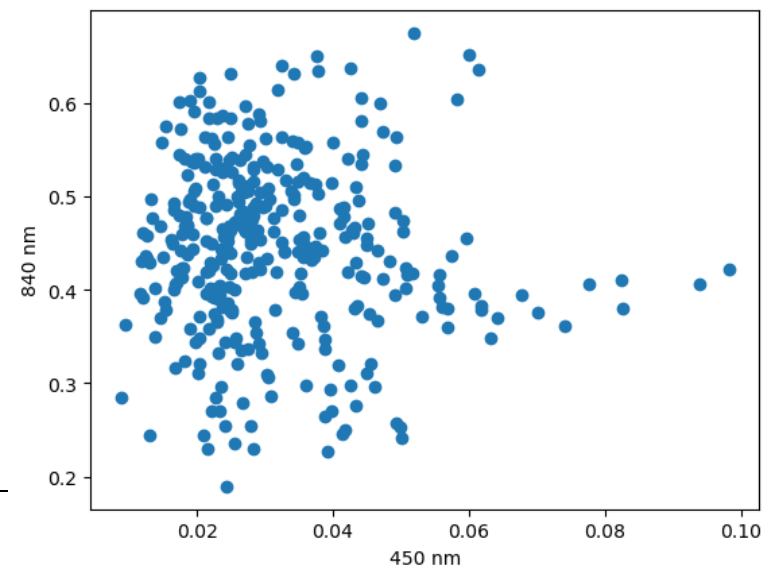
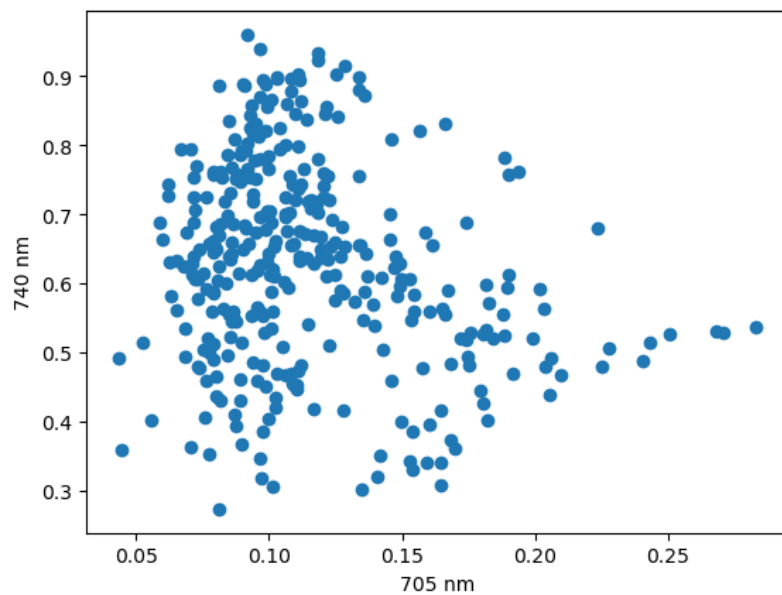
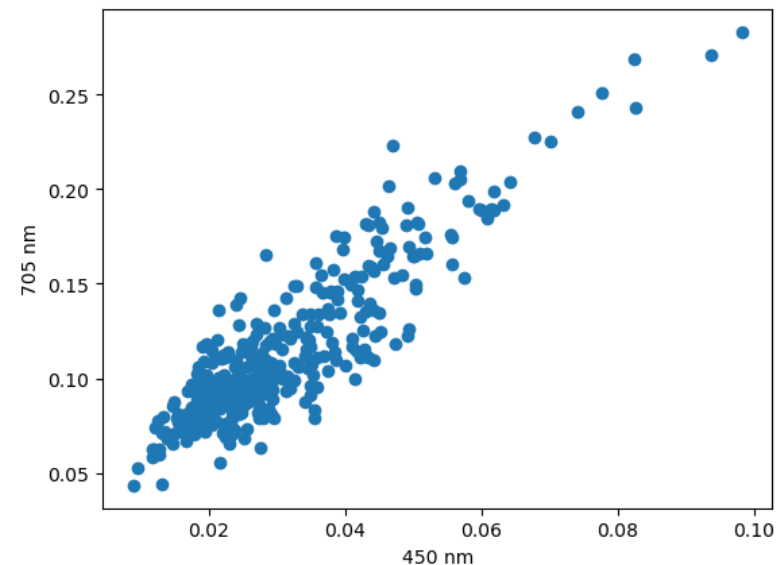
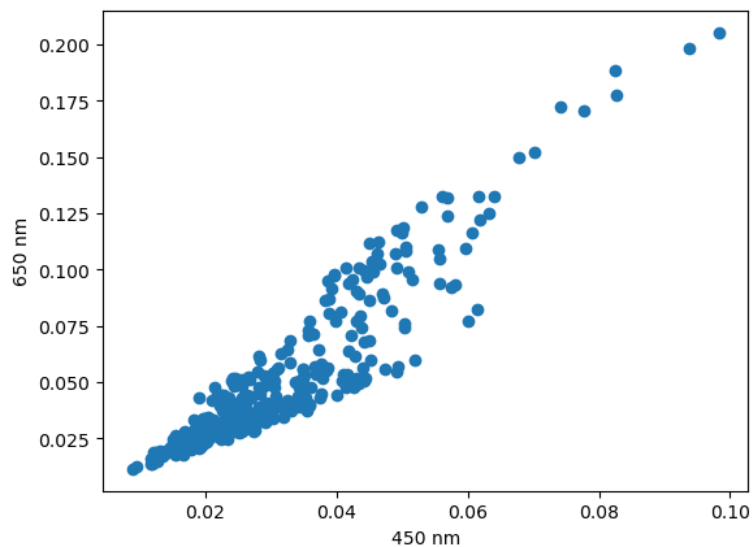


Variety treatment





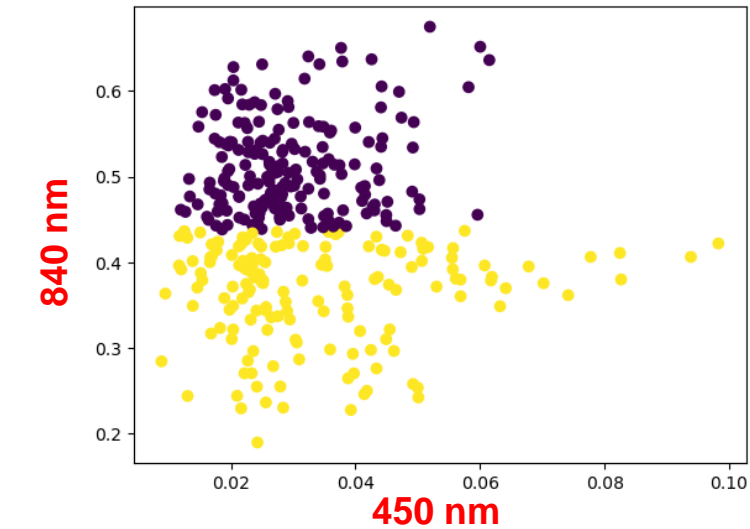
Hyperspectral: highly intercorrelated



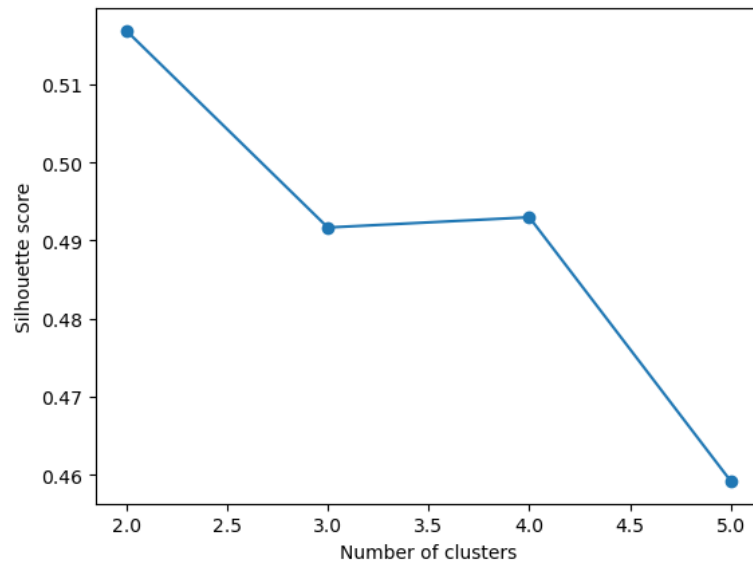
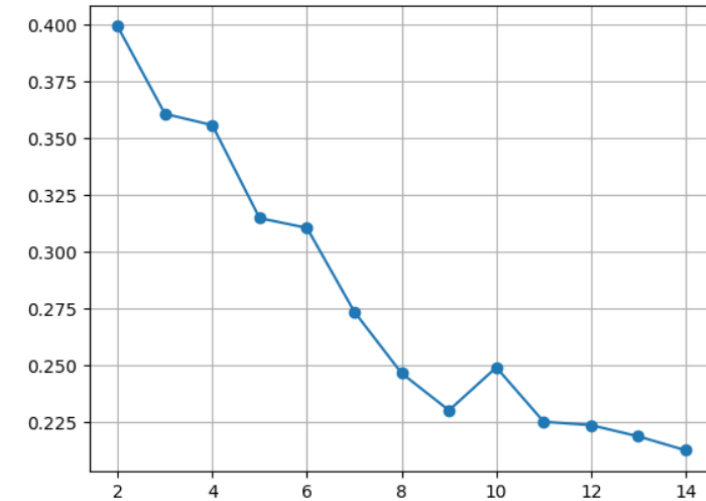


Data exploration: clustering (K-means) and umap

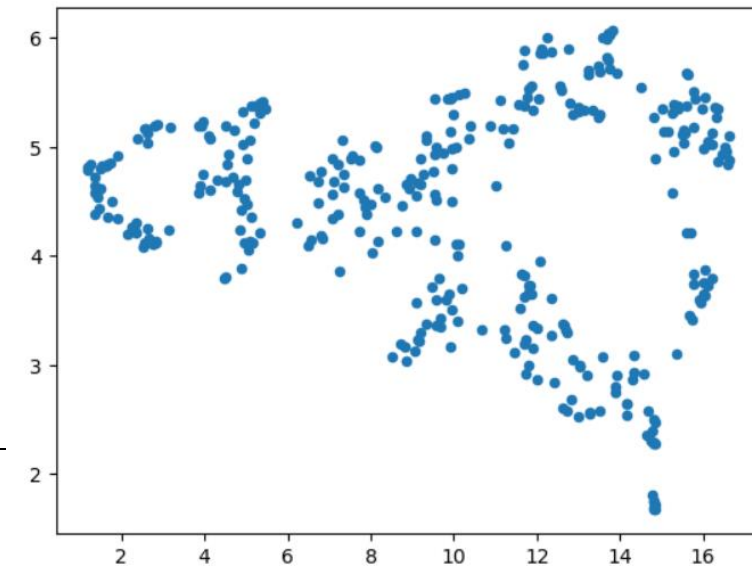
2D



High dimensional



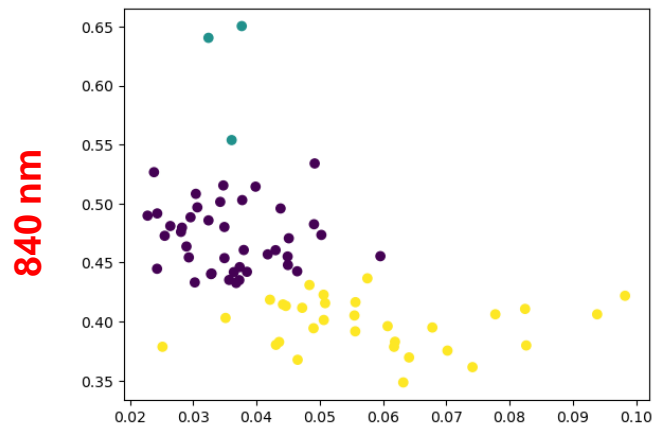
tés



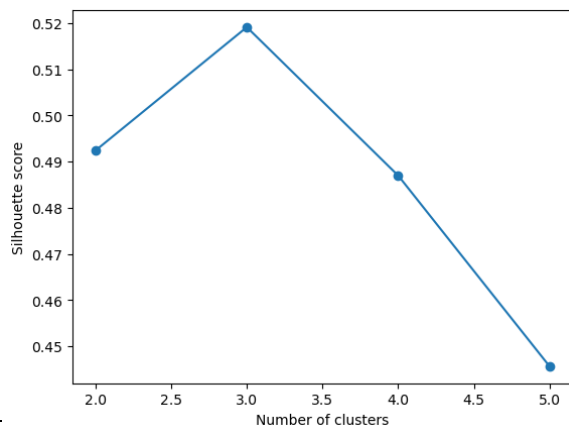


Data exploration: clustering (K-means) and umap – Focusing on Changins in 2022

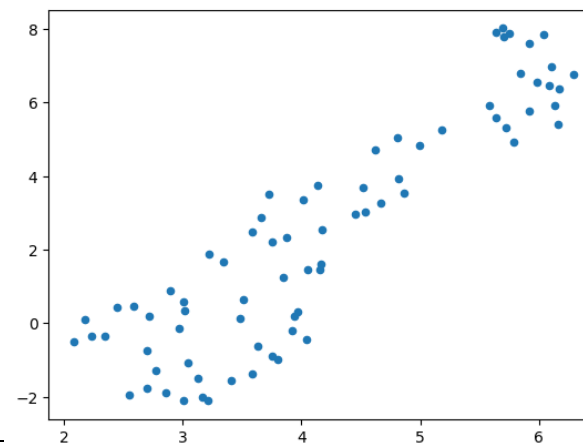
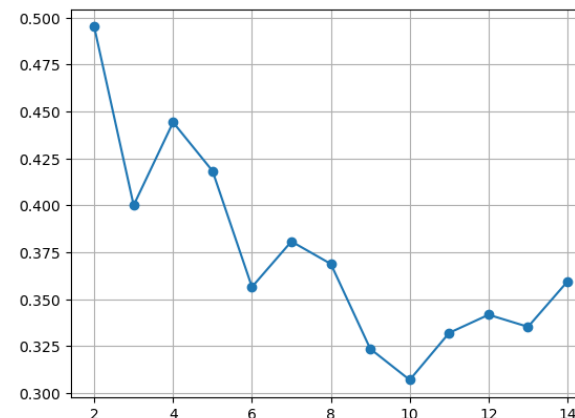
2D



450 nm

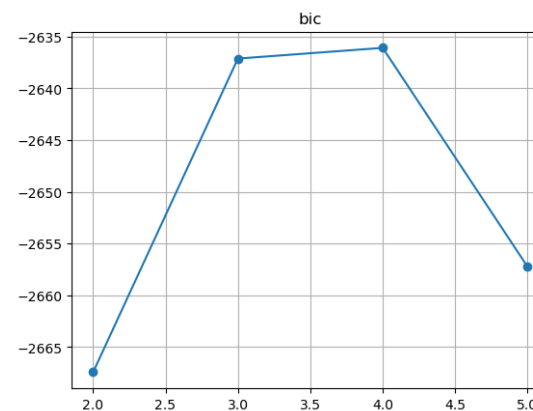
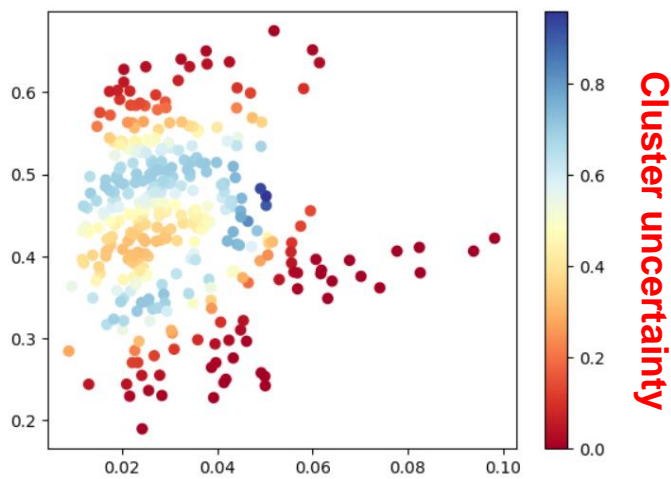
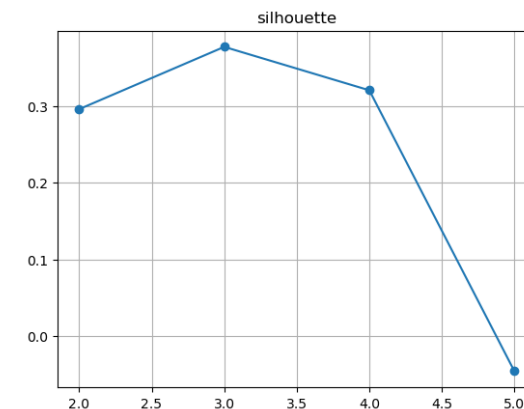
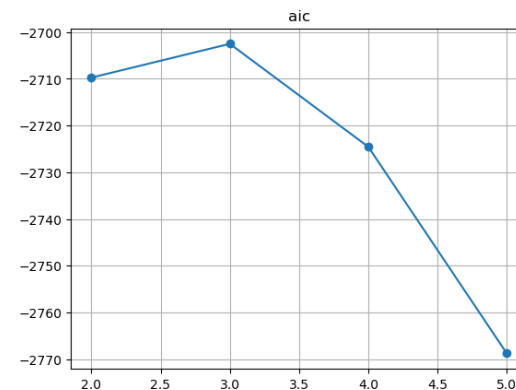
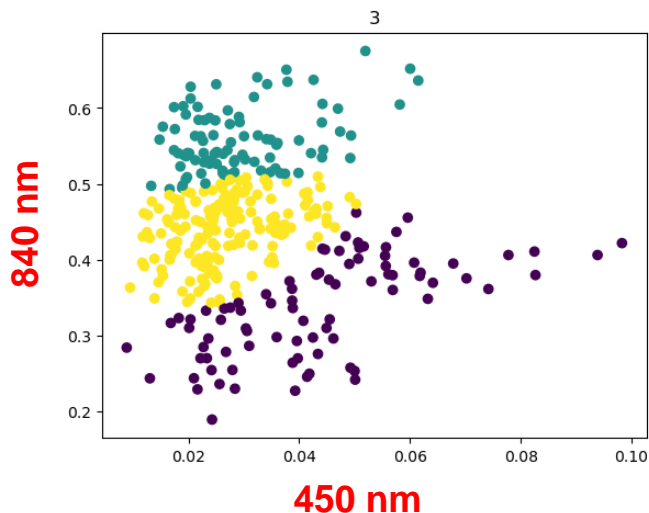


High dimensional





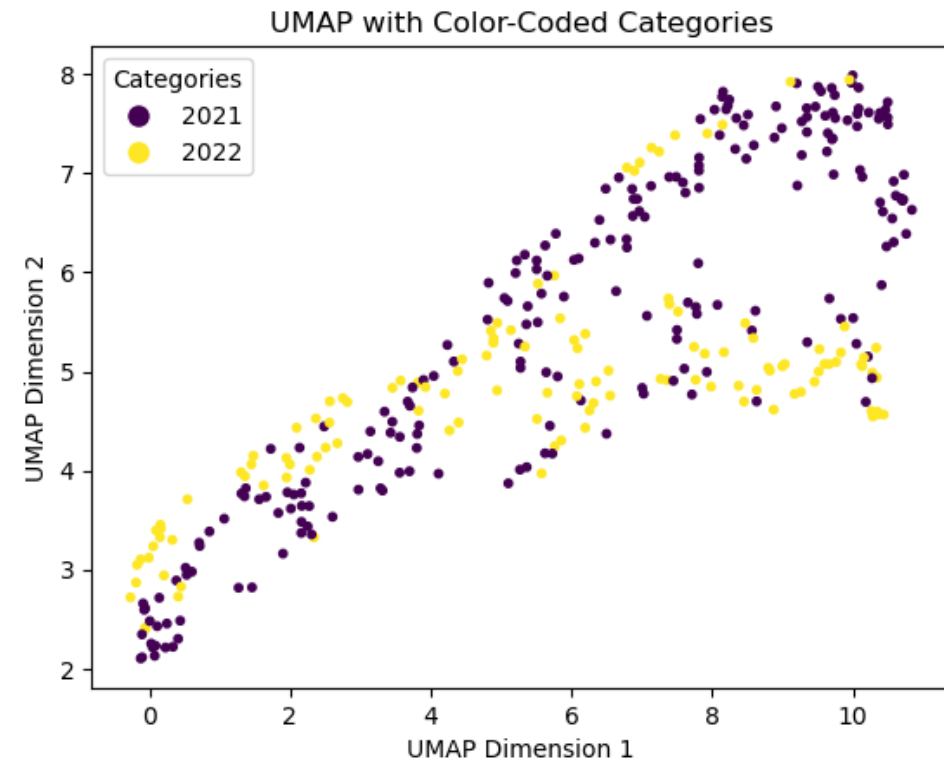
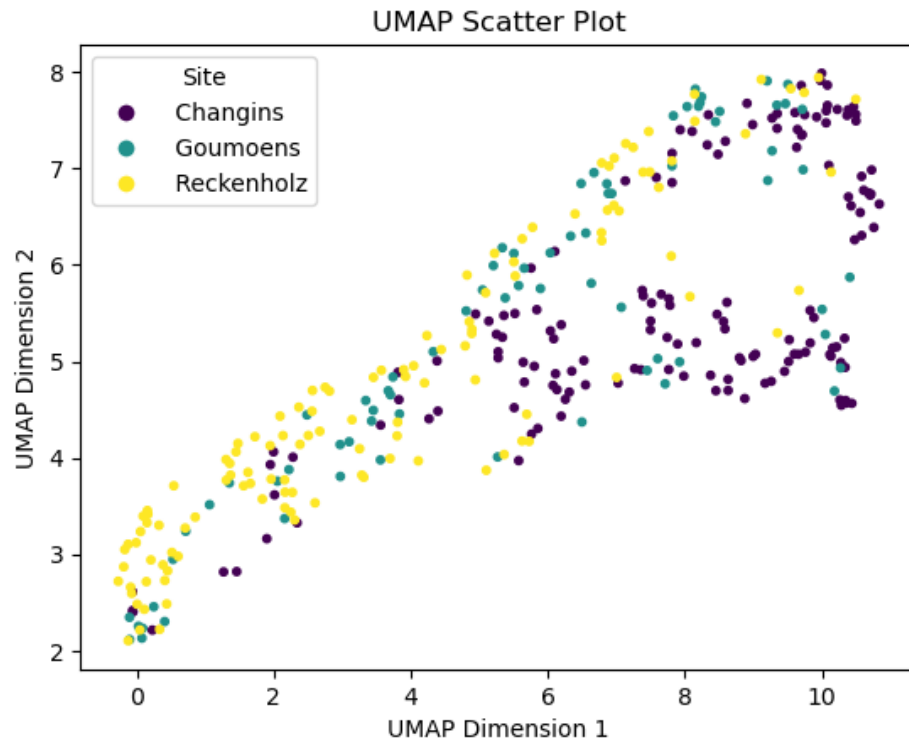
Data exploration: Gaussian mixtures (soft clustering)



➤ Take into account more complex cluster shapes than circular (K-means)

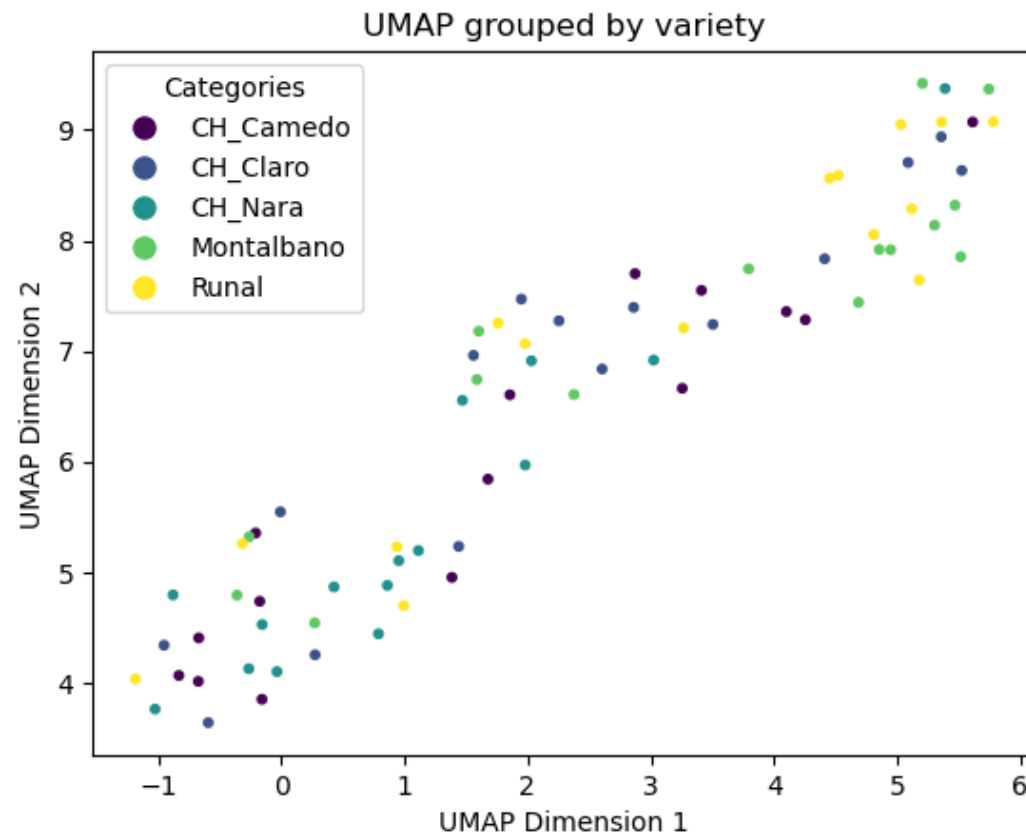
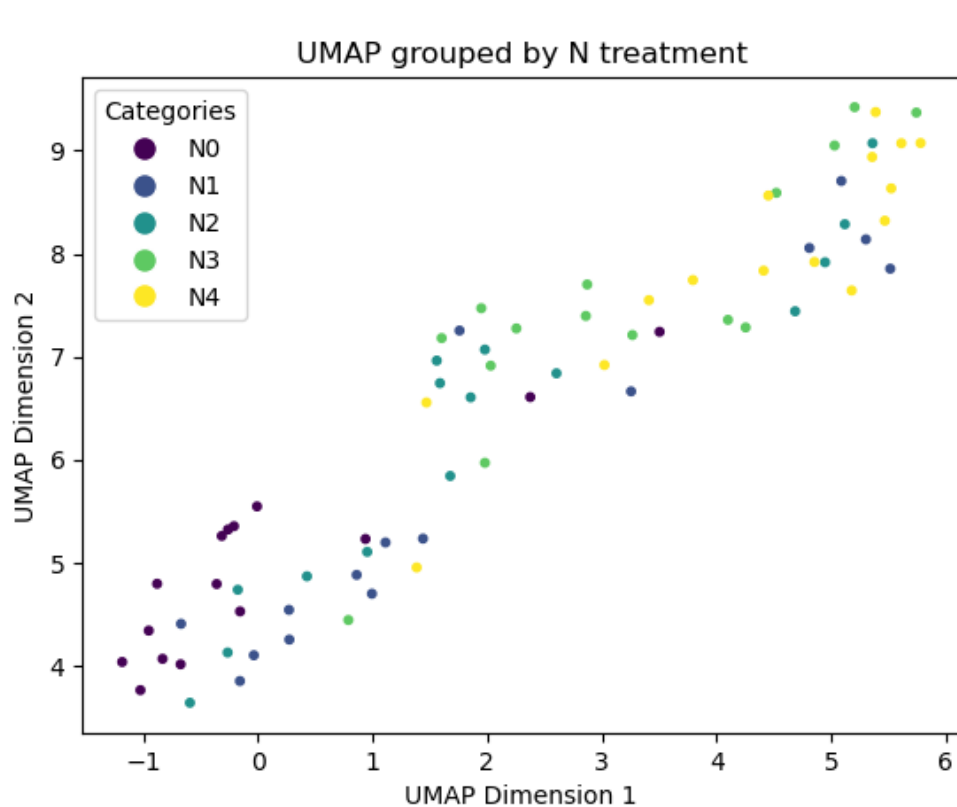


Data exploration: umap



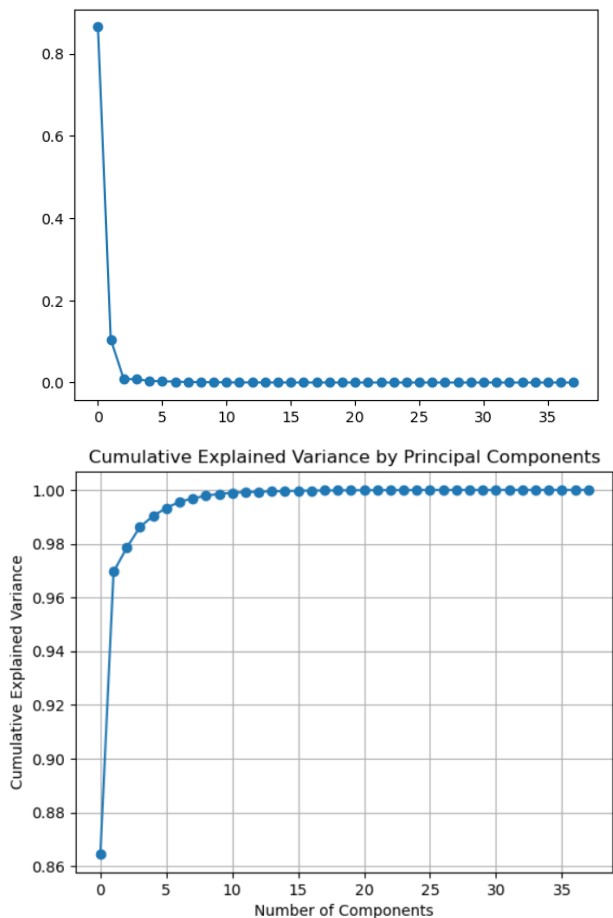


Data exploration: umap-focus on Changins 2022





Data exploration: PCA

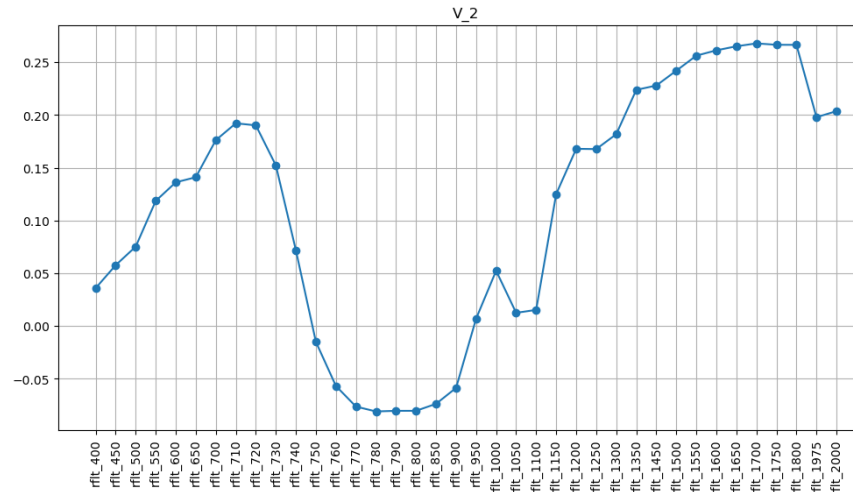
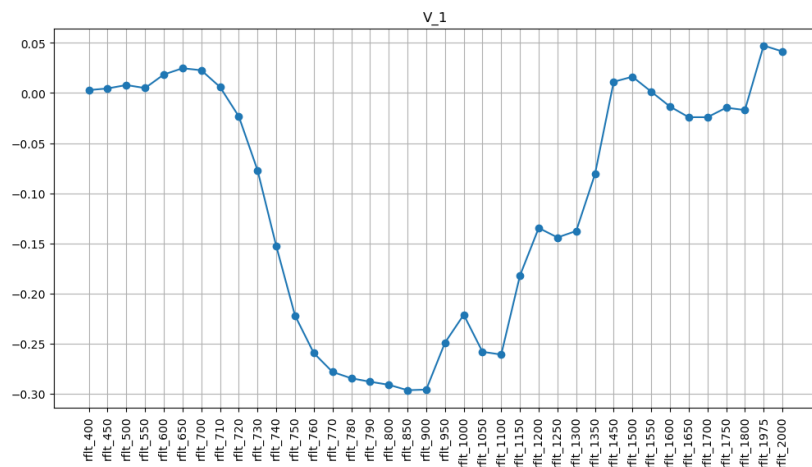


- Most of the variance is explained by the first principal components

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_10	...	V_29	V_30	V_31	V_32
rfit_400	0.002937	0.035984	-0.002985	0.036265	-0.066545	-0.077143	0.047143	0.018090	0.027047	-0.030372	...	-0.314811	0.013153	0.370203	-0.570470
rfit_450	0.004546	0.057737	0.008797	0.048449	-0.083053	-0.093928	0.058513	0.031172	0.033828	-0.038866	...	-0.097061	0.031067	-0.051995	0.439703
rfit_500	0.007968	0.075280	0.011633	0.057094	-0.102647	-0.111227	0.070736	0.038209	0.045630	-0.059384	...	0.001776	0.045085	-0.137613	0.450820
rfit_550	0.004938	0.118608	0.100272	0.041700	-0.076588	-0.132578	0.133371	0.025159	0.008326	-0.070819	...	0.464148	-0.161390	-0.025780	-0.243287
rfit_600	0.018468	0.136195	0.055064	0.063748	-0.126743	-0.139727	0.104721	0.041003	0.076547	-0.104400	...	0.177287	-0.002736	-0.130536	-0.221454
rfit_650	0.024707	0.141009	0.020273	0.078288	-0.151596	-0.132552	0.071822	0.055425	0.113306	-0.113475	...	-0.030904	0.120243	-0.309854	-0.225976
rfit_700	0.022719	0.176140	0.104292	0.057160	-0.127722	-0.154214	0.136993	0.031696	0.086213	-0.128917	...	-0.135049	-0.047044	0.202034	0.128728
rfit_710	0.005987	0.192111	0.219704	0.011086	-0.058324	-0.144916	0.194052	-0.000885	0.026911	-0.131047	...	-0.078937	-0.051935	0.400239	0.239990
rfit_720	-0.023342	0.190195	0.329054	-0.035184	0.022756	-0.115419	0.246817	-0.045449	-0.020716	-0.097410	...	-0.127544	-0.031272	-0.026472	0.001404
rfit_730	-0.077068	0.152240	0.411113	-0.068129	0.120301	-0.045753	0.256434	-0.071563	-0.090646	-0.021463	...	-0.054480	0.214311	-0.510202	-0.139891
rfit_740	-0.152834	0.071806	0.380125	-0.038900	0.173887	0.054169	0.161227	-0.040682	-0.175539	0.090713	...	0.178868	-0.147663	0.331254	0.042661
rfit_750	-0.222108	-0.014992	0.241446	0.043583	0.144238	0.120612	-0.003116	0.014278	-0.249012	0.255482	...	-0.165143	0.008243	0.025910	0.019577
rfit_760	-0.259550	-0.056748	0.115213	0.092566	0.044302	0.060705	-0.063564	0.019648	-0.047796	0.043965	...	0.085799	0.072182	0.095962	0.029375
rfit_770	-0.278402	-0.076261	0.073083	0.133530	0.008369	0.075799	-0.126264	0.056414	-0.127857	0.132030	...	0.083548	0.064112	-0.139257	0.047678
rfit_780	-0.284698	-0.080844	0.051317	0.147931	-0.025486	0.047185	-0.127509	0.068652	-0.105174	0.070686	...	0.183179	0.033072	0.105926	-0.091678
rfit_790	-0.288159	-0.080257	0.037642	0.151115	-0.061858	0.003719	-0.104245	0.053524	-0.021094	-0.036496	...	-0.215291	-0.330882	-0.070489	-0.058744
rfit_800	-0.291212	-0.080241	0.033173	0.157454	-0.080462	-0.012503	-0.093401	0.051798	-0.008328	-0.066501	...	-0.274390	-0.159398	-0.253548	0.038784
rfit_850	-0.296626	-0.073656	0.020933	0.160679	-0.134902	-0.071902	-0.069440	0.053835	0.047964	-0.159894	...	0.080215	0.707552	0.197408	0.016602
rfit_900	-0.296011	-0.058551	0.011742	0.145246	-0.168055	-0.097280	-0.043212	-0.042513	0.242654	-0.185614	...	0.148816	-0.454504	-0.026414	0.022098
rfit_950	-0.249193	0.007015	0.095887	0.063978	-0.050155	0.033271	-0.047558	-0.215876	0.580823	0.068516	...	-0.024967	0.096911	0.030991	-0.013949
rfit_1000	-0.221344	0.052433	-0.220139	-0.169310	0.070945	-0.052127	0.170324	0.011850	-0.180992	-0.034090	...	-0.045098	0.002954	0.015247	-0.014242
rfit_1050	-0.258284	0.012576	-0.313951	-0.141051	-0.033743	-0.144897	0.224923	0.084717	-0.242277	-0.161996	...	0.012214	0.006920	-0.014762	0.018567
rfit_1100	-0.261066	0.015344	-0.341905	-0.131394	-0.048759	-0.162334	0.248113	0.035208	-0.147742	-0.133742	...	-0.004369	0.022071	0.022281	-0.015644
rfit_1150	-0.182186	0.125109	-0.177318	-0.205285	0.048684	0.042876	0.155912	-0.196696	0.381576	0.250415	...	0.039289	0.045725	0.026331	0.011381
rfit_1200	-0.134697	0.167963	-0.061158	-0.194213	0.145048	0.106900	0.001842	-0.009889	0.045947	0.031876	...	-0.164195	0.045232	-0.026473	-0.007882
rfit_1250	-0.144183	0.167683	-0.070284	-0.202348	0.144352	0.106196	0.008684	0.024206	-0.016463	-0.010005	...	0.433927	-0.070000	-0.027036	0.039075
rfit_1300	-0.137737	0.181770	-0.067180	-0.201547	0.131538	0.103304	0.011670	-0.011545	0.046018	0.005243	...	-0.329916	-0.080701	-0.022178	-0.038984
rfit_1350	-0.081059	0.223784	-0.060618	-0.151056	0.067630	0.086480	-0.000640	-0.083112	0.208114	0.205121	...	0.029125	0.006407	-0.002334	0.015347
rfit_1450	0.011202	0.227878	-0.103072	0.166558	-0.268382	-0.401319	0.014654	0.149950	-0.105201	0.733861	...	-0.001113	-0.017803	-0.011077	-0.002601
rfit_1500	0.016259	0.241944	-0.042922	0.086942	-0.118142	-0.025608	-0.123851	0.219955	0.020149	-0.118426	...	-0.014608	-0.007850	0.035438	0.014579
rfit_1550	0.001391	0.256196	-0.003615	0.008483	-0.061399	0.037940	-0.191749	0.202485	-0.008351	-0.131753	...	-0.052260	0.047501	-0.000355	-0.024791
rfit_1600	-0.013472	0.261232	0.004638	-0.027502	0.010186	0.088423	-0.193448	0.185481	-0.010470	-0.086758	...	-0.029999	-0.034136	0.015457	-0.008421
rfit_1650	-0.024138	0.265059	0.012019	-0.051561	0.037006	0.103989	-0.215153	0.144908	-0.025571	-0.065104	...	0.013342	0.004593	0.012563	0.004921



Data exploration: PCA



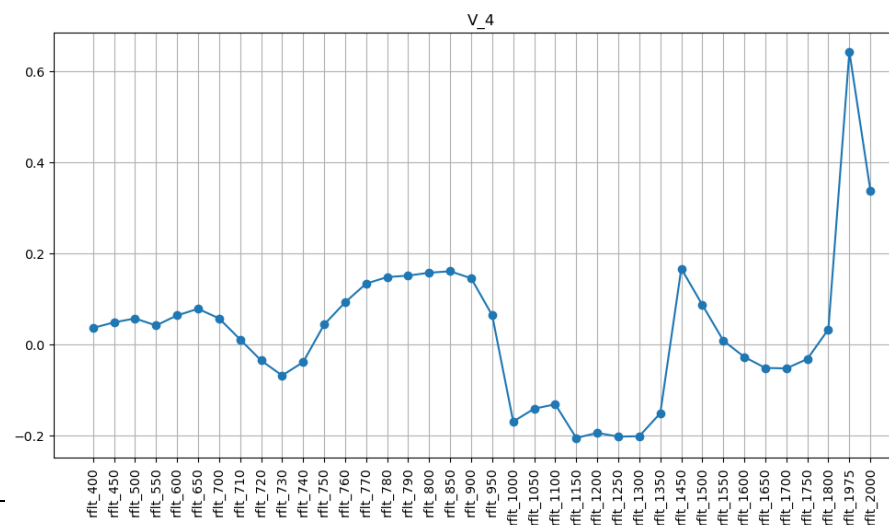
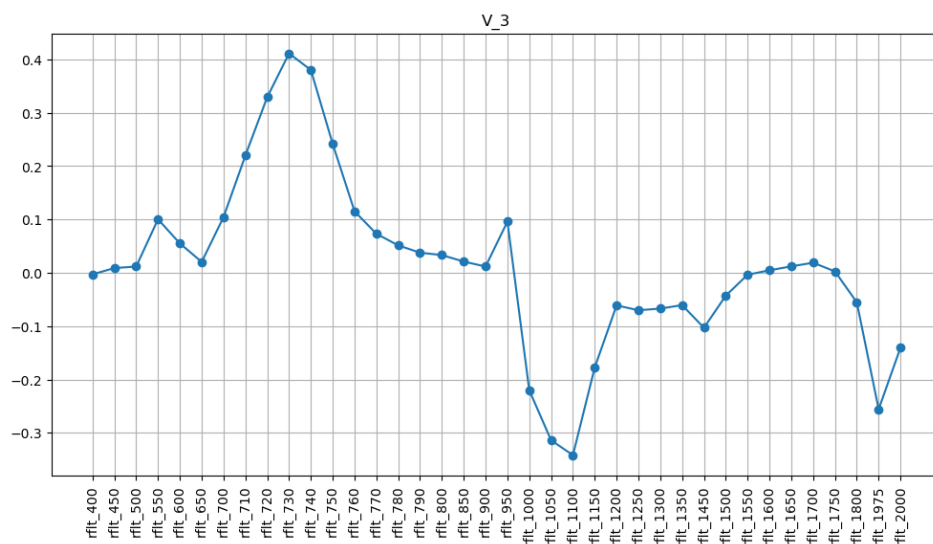
V1:

- 900 nm
- 850 nm
- 800 nm
- 790 nm
- 780 nm
- 770 nm

V2:

- 700 nm
- 710 nm
- 720 nm
- 1800 nm
- 1550 nm

38 features at the beginning → 17 features selected



V3:

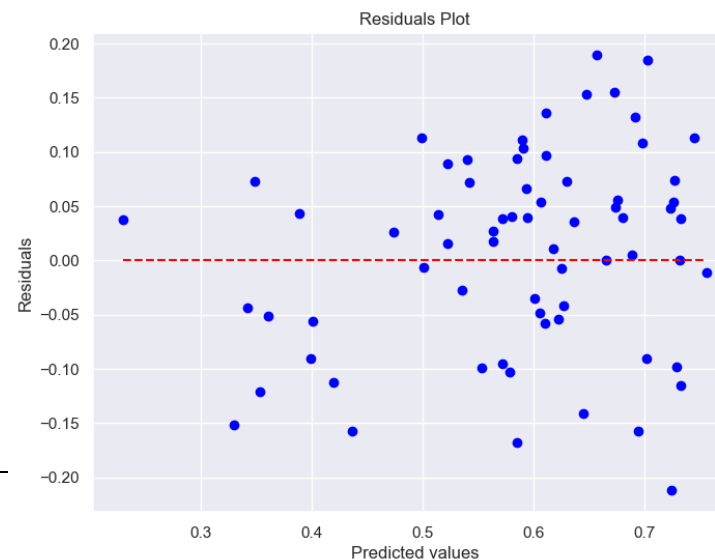
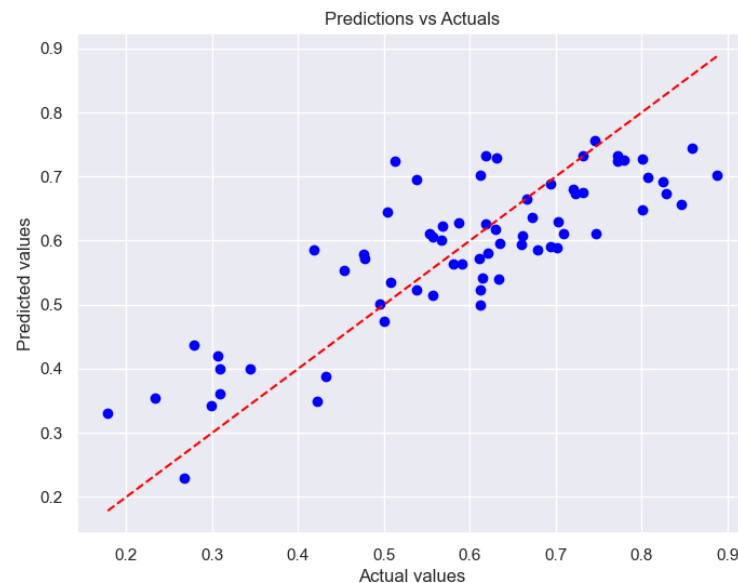
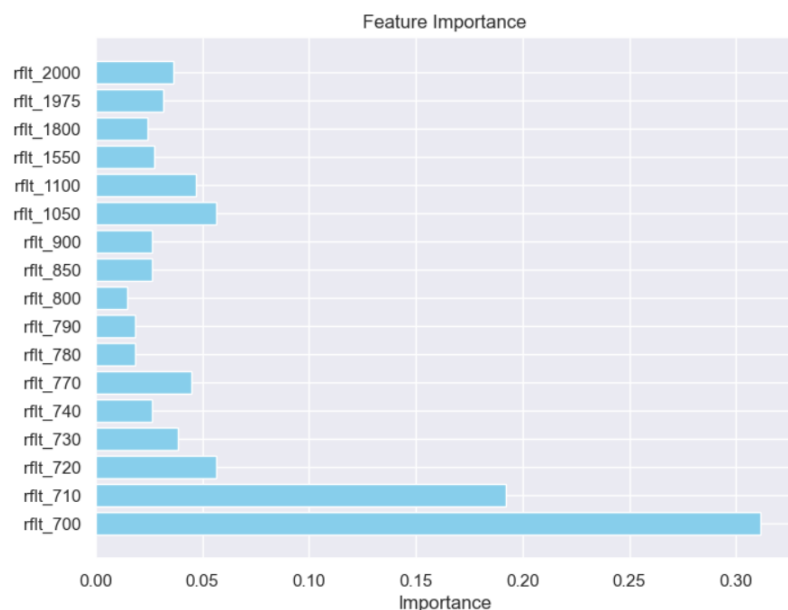
- 730 nm
- 740 nm
- 1050 nm
- 1100 nm

V4:

- 1975 nm
- 2000 nm



Random forest model with selected features



Training R-squared: 0.955

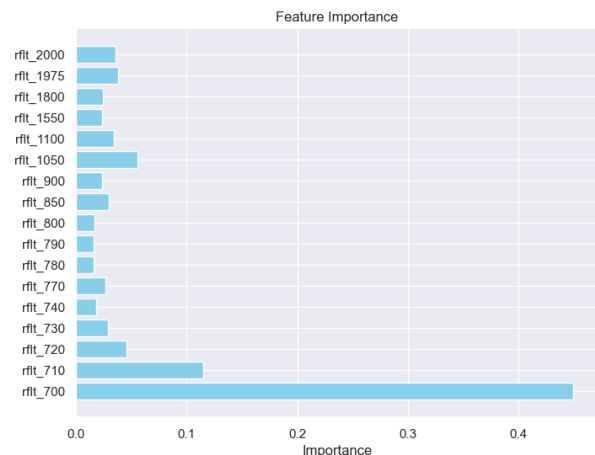
Test R-squared: 0.685



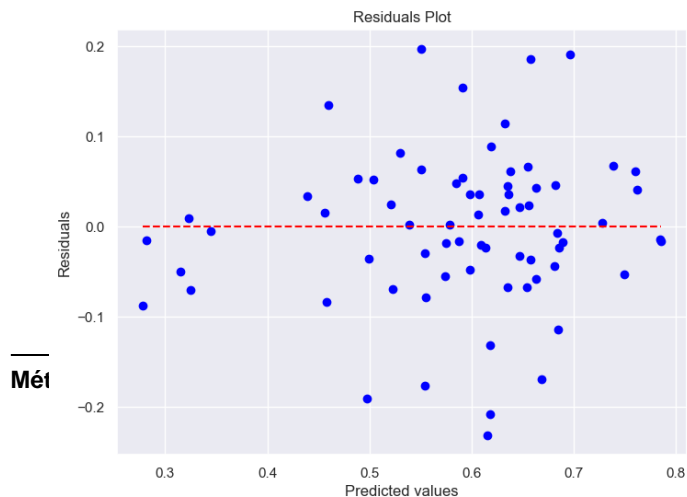
Random forest model with selected features: improvement tests

- Splitting train and test sets according to years or sites
- Reducing the tree depth (max_depth) and increasing the number of trees (n_estimators)

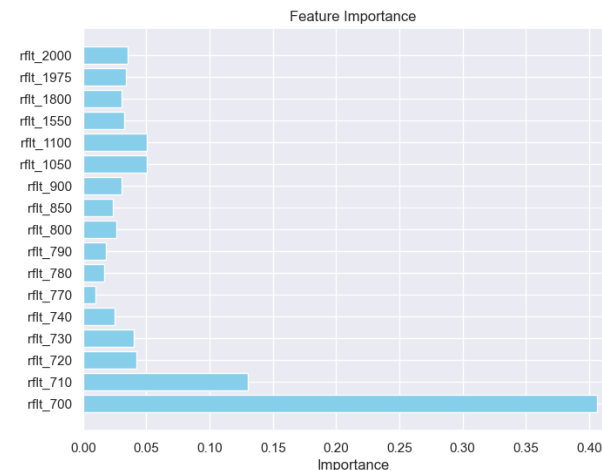
Sites



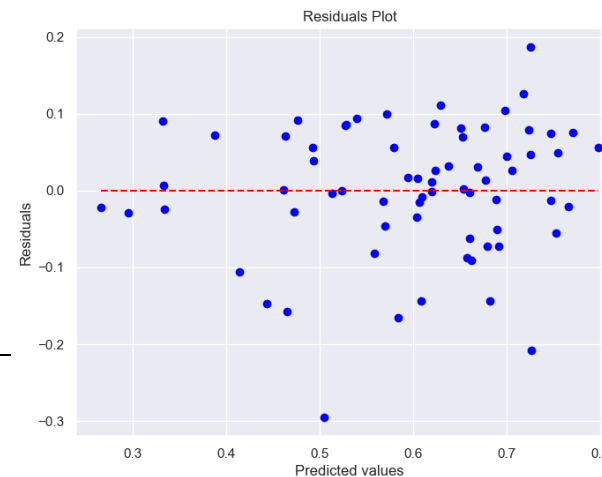
Training R-squared: 0.954
Test R-squared: 0.676



Years



Training R-squared: 0.950
Test R-squared: 0.706



Conclusion

- **Field data can be noisy:**
 - Hyperspectral data is highly sensitive to weather conditions
- **Promising results in data exploration (PCA):**
 - Highlighted the importance of red-edge bands between 700 to 800 nm to be able to characterize winter wheat performance
- **Promising results in random forest:**
 - Highlighted the 700-710 nm part of the spectrum to explain grain yield differences
 - It was expected because in theory it is linked to Nitrogen (N) chlorophyll content which is in this context (different levels of N treatments) an important indicator of grain yield
- **Low performance in the test set compared to the training set:**
 - Overfitting?
- **Other agronomic parameter could be analysed to link to hyperspectral data**
 - Straw yield (positively correlated with grain yield)
 - Harvest index (indicates the ratio of grain on total plant weight (grain + straw))
 - Protein in grain (negatively correlated with grain yield)
 - Plant height (positively correlated with grain and straw yield; lodging effect)
 - Disease (can explain unexpected low yield)



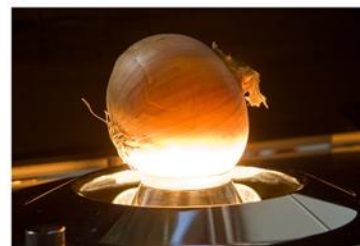
Thank you for your attention

Nicolas Vuille-dit-Bille

nicolas.vuille-dit-bille@agroscope.admin.ch

Agroscope good food, healthy environment

www.agroscope.admin.ch



Mé