Predicting functional diversity and ecological strategies: near infrared spectrometry is coming to the rescue

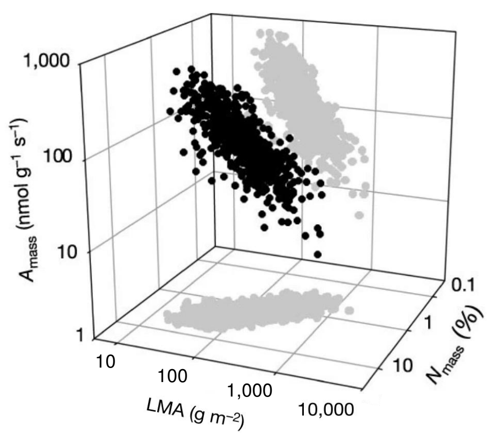
François Vasseur & Grégory Beurier
25/12/2021

Context

A long-term goal in functional ecology:
explaining and categorizing plant diversity

Describing functional diversity and ecological strategies in plants

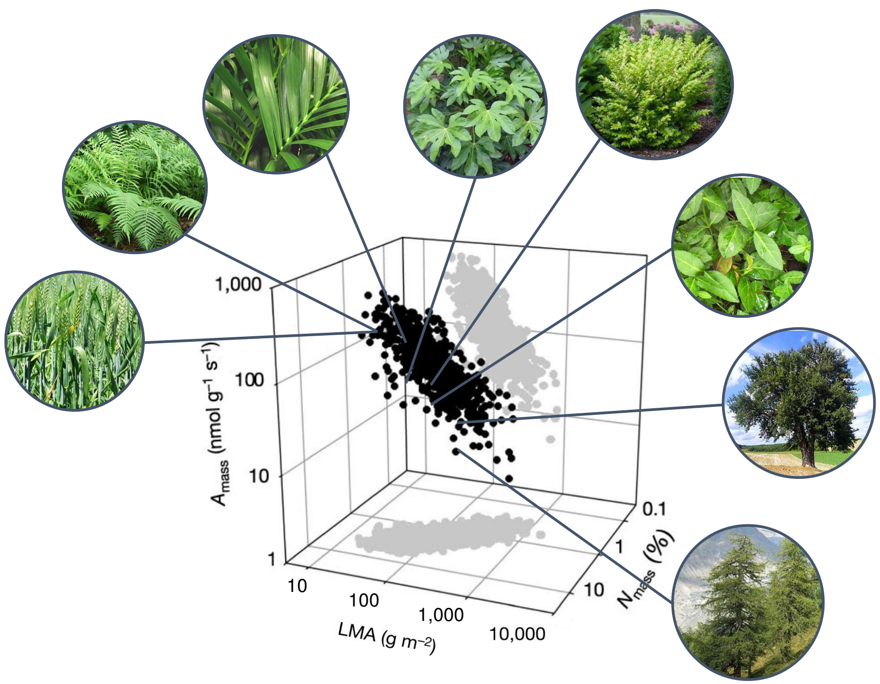
The 'worldwide leaf economics spectrum'
Wright *et al.*, Nature 2004

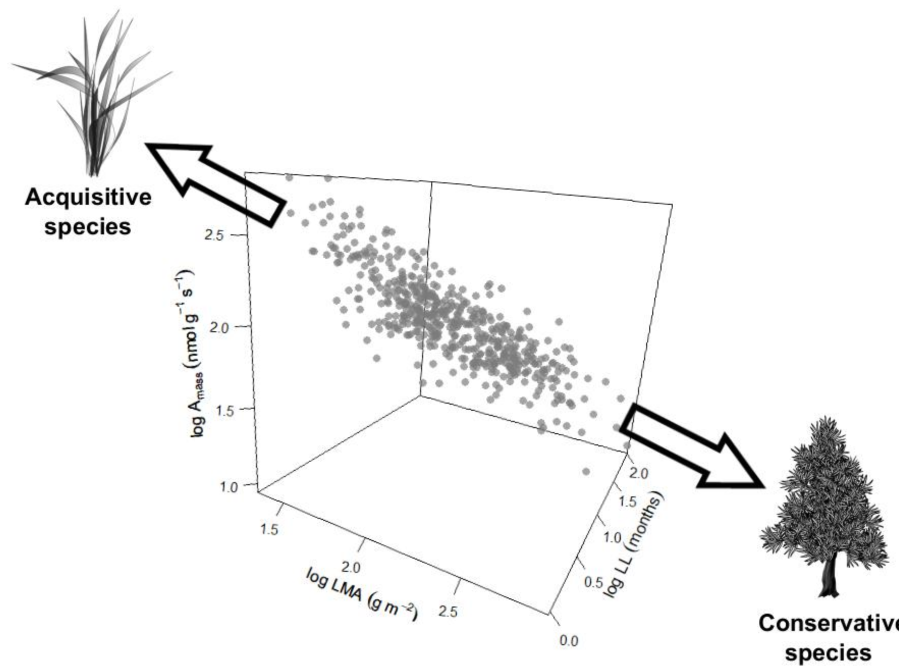


Describing functional diversity and ecological strategies in plants

The 'worldwide leaf economics spectrum'

Wright *et al.*, Nature 2004

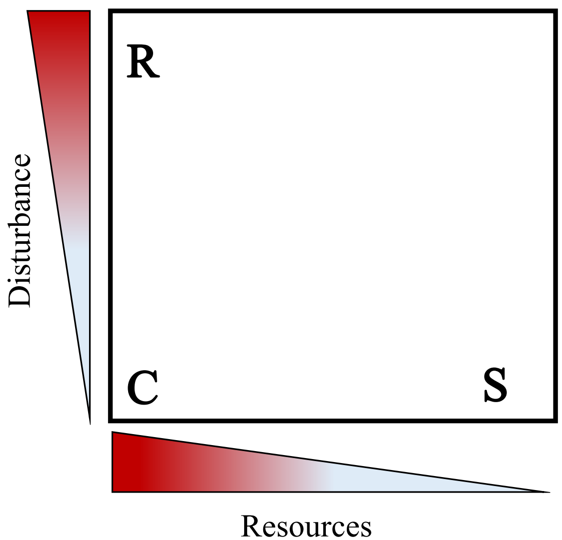




Describing functional diversity and ecological strategies in plants

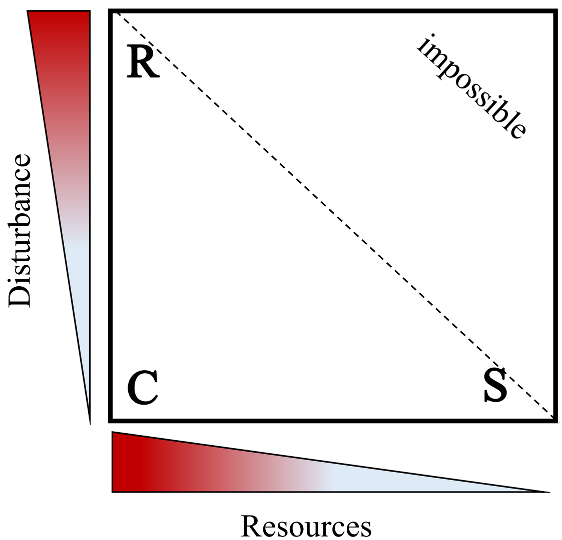
The 'C-S-R categorization' of plant ecological strategies

Grime *et al.*, Nature 1974



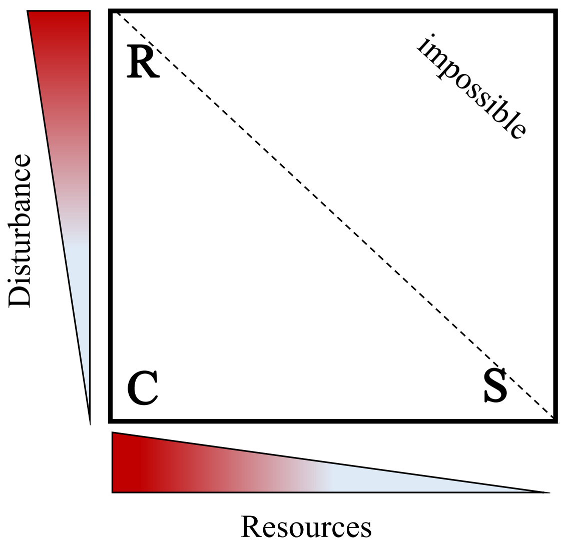
Describing functional diversity and ecological strategies in plants

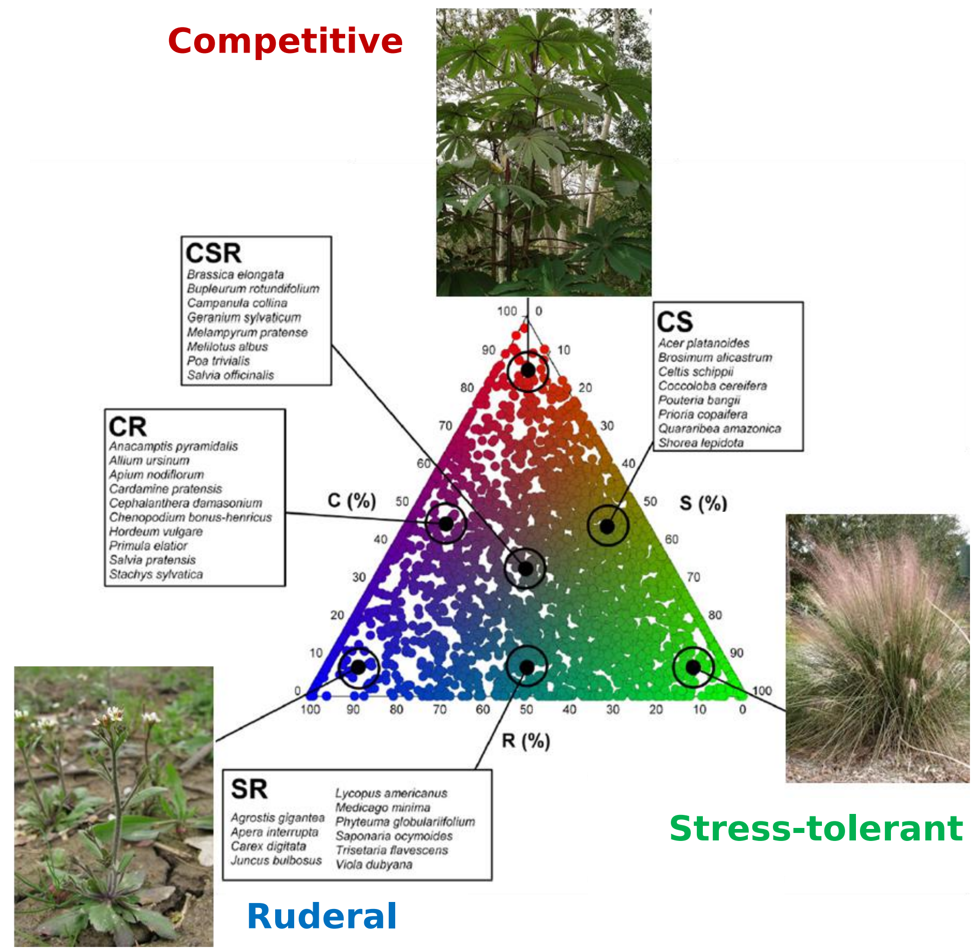
The 'C-S-R categorization' of plant ecological strategies
Grime *et al.*, Nature 1974



Describing functional diversity and ecological strategies in plants

The 'C-S-R categorization' of plant ecological strategies
Grime *et al.*, Nature 1974





Pierce *et al.*, Functional Ecology 2017

Describing functional diversity and ecological strategies in plants

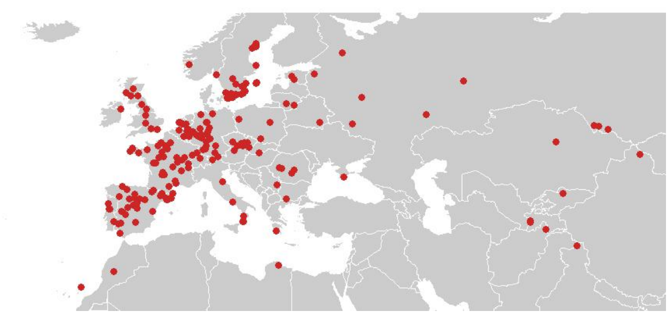
> Requires the measurement of multiple, laborious traits:

- Specific leaf area   SLA = leaf area / leaf dry mass
- Leaf dry matter content   LDMC = leaf dry mass / leaf fresh mass
- Net photosynthetic rate
- Nitrogen concentration
- etc.

> Screening many species and/or genotypes in contrasted conditions:

- In wild conditions
- In common garden
- In controlled conditions
- etc.

NIR Spectral "Collection" on *Arabidopsis thaliana*



Large genetic diversity

Growth conditions

Treatments

Herbivory
Water deficit
Hight temperature
Low temperature
Combined Stresses

Phenological stages

Bolting
Flowering
Senescence

NIR Spectral “Collection” on *Arabidopsis thaliana*

Morphological and physiological traits

Leaf nitrogen content (LNC)
Specific leaf area (SLA)
Leaf dry matter content (LDMC)
Relative water content (RWC)

Stomatal density
Vein density

Photosynthesis
Delta13C

Metabolomic traits

Sugars
Phytohormones
Phenolics
Flavonoids
Anthocyanins

Phenological traits

Flowering time

NIR Spectral "Collection" on *Arabidopsis thaliana*

Experiment	Growth Condition	Treatment	Spectrum number	Variable number	Variable
AraBreed Spring 2018	Outdoor	Herbivory Water deficit Combined stresses	1791	97	LNC, SLA, LDMC, RWC, delta13C, Metabolites
AraBreed Spring 2019	Outdoor	Herbivory Water deficit Combined stresses	227	2	SLA, LDMC
TE Outdoor	Outdoor	Wild	76	3	LNC, SLA, LDMC
Plast-Edge 2019	Greenhouse	Water deficit Temperature (Low and High) Combined stresses	1646	4	SLA, LDMC, RWC, Survival
Resorption 2017	Greenhouse	Control	11239	3	LNC, SLA, LDMC
Herbivory 2015	Greenhouse	Herbivory	5297	3	LNC, SLA, LDMC
CEFE 2018	Greenhouse	Control	114	4	LNC, SLA, LDMC, Leaf thickness
Arabreed PHENOPSIS 2018	PHENOPSIS	Control	745	5	LNC, SLA, LDMC,RWC, delta13C
AraBreed Pilot 2017	PHENOPSIS	Control	333	60	Metabolites
		Total Raw Spectra	**21468**	**100**	
		Total Spectra for prediction	**8427**		
		Total Mean Spectra	**4471**		

NIR Spectral “Collection” on *Arabidopsis thaliana*

Objectives:

1. Predict functional traits and plant strategies

2. Predict environment of origin and stress response

3. Predict leaf metabolites concentration

with a single measurement!

Statistical analyses: deep learning choice

	Spectrometer x Product x Environment	Many features / Few samples	High variability / Non linearity
Traditional approaches (PLS…)	A new model with each modification	Dimension reduction / selection	Spectra pre-treatments (SavGol, MSC, SNV, etc.) - Outliers removal
Neural Networks	A unique model One-shot hyperparametrization	Native overfitting management - Data augmentation	All pre-treatments in one model - No need for pre-treatment

Statistical analyses: deep learning choice

	Spectrometer x Product x Environment	Many features / Few samples	High variability / Non linearity
Traditional approaches (PLS…)	A new model with each modification	Dimension reduction / selection	Spectra pre-treatments (SavGol, MSC, SNV, etc.) - Outliers removal
Neural Networks	A unique model One-shot hyperparametrization	Native overfitting management - Data augmentation	All pre-treatments in one model - No need for pre-treatment

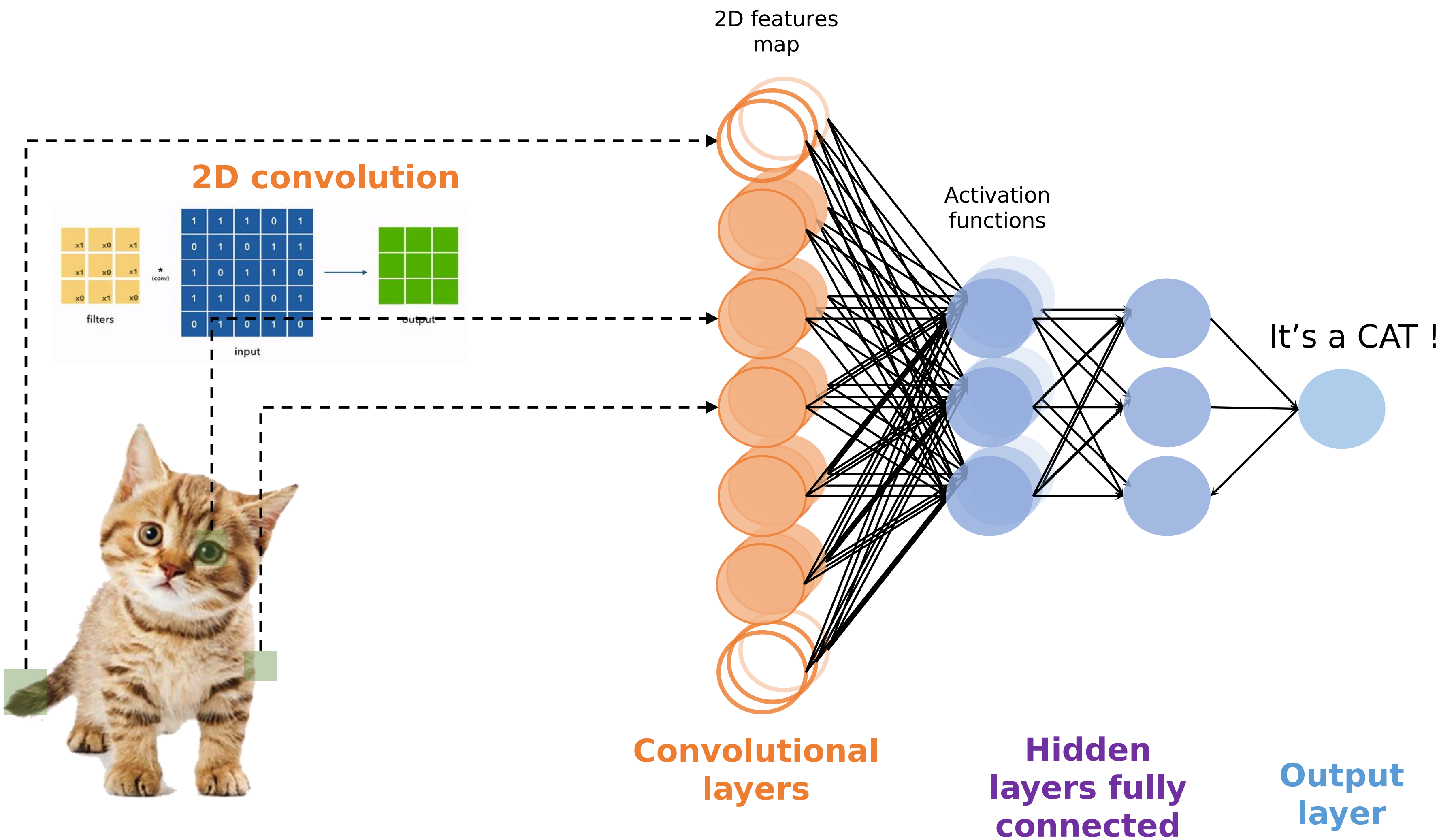
Huge Open-source community

Room for improvements

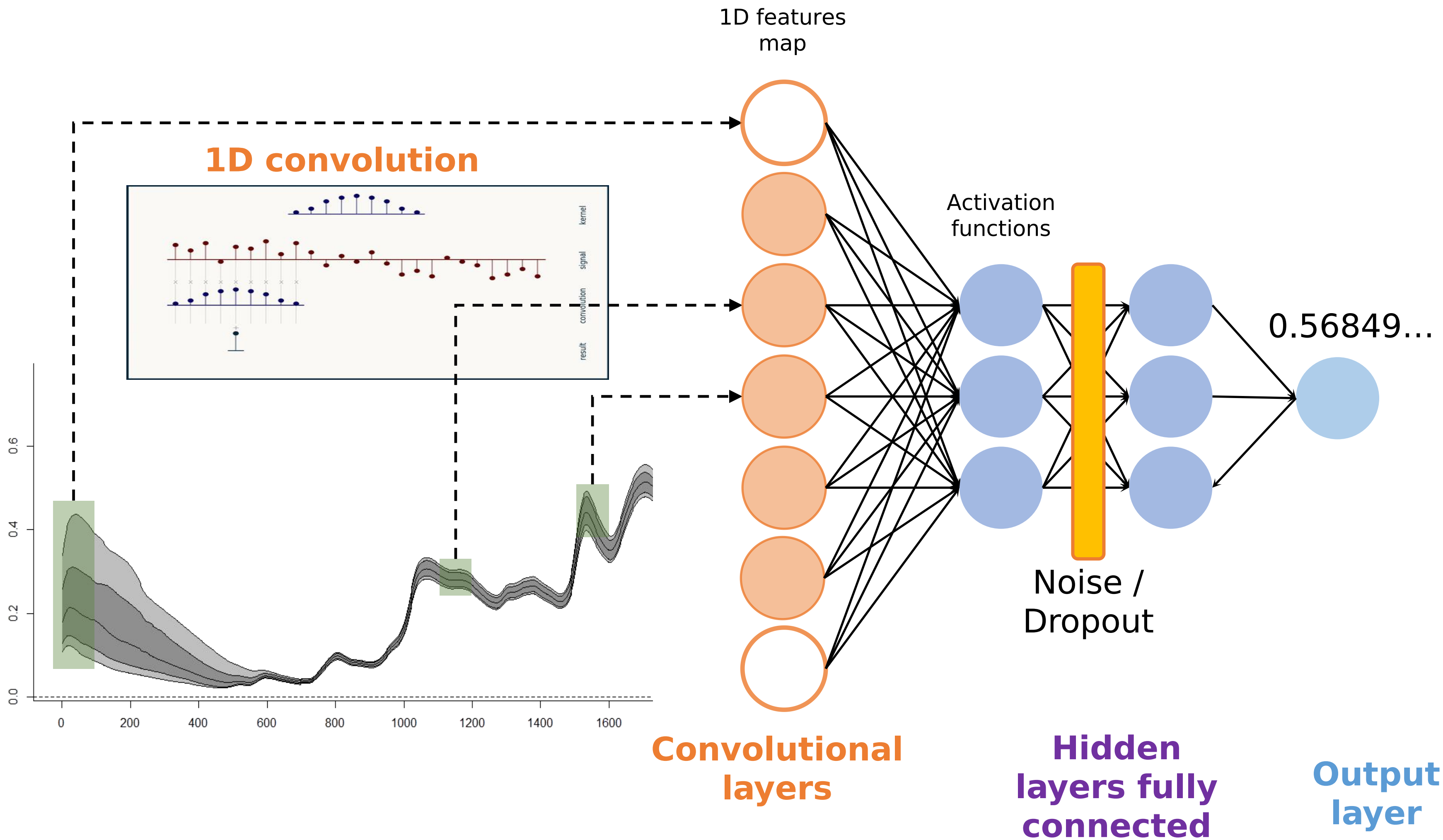
Dedicated computation

Black-box

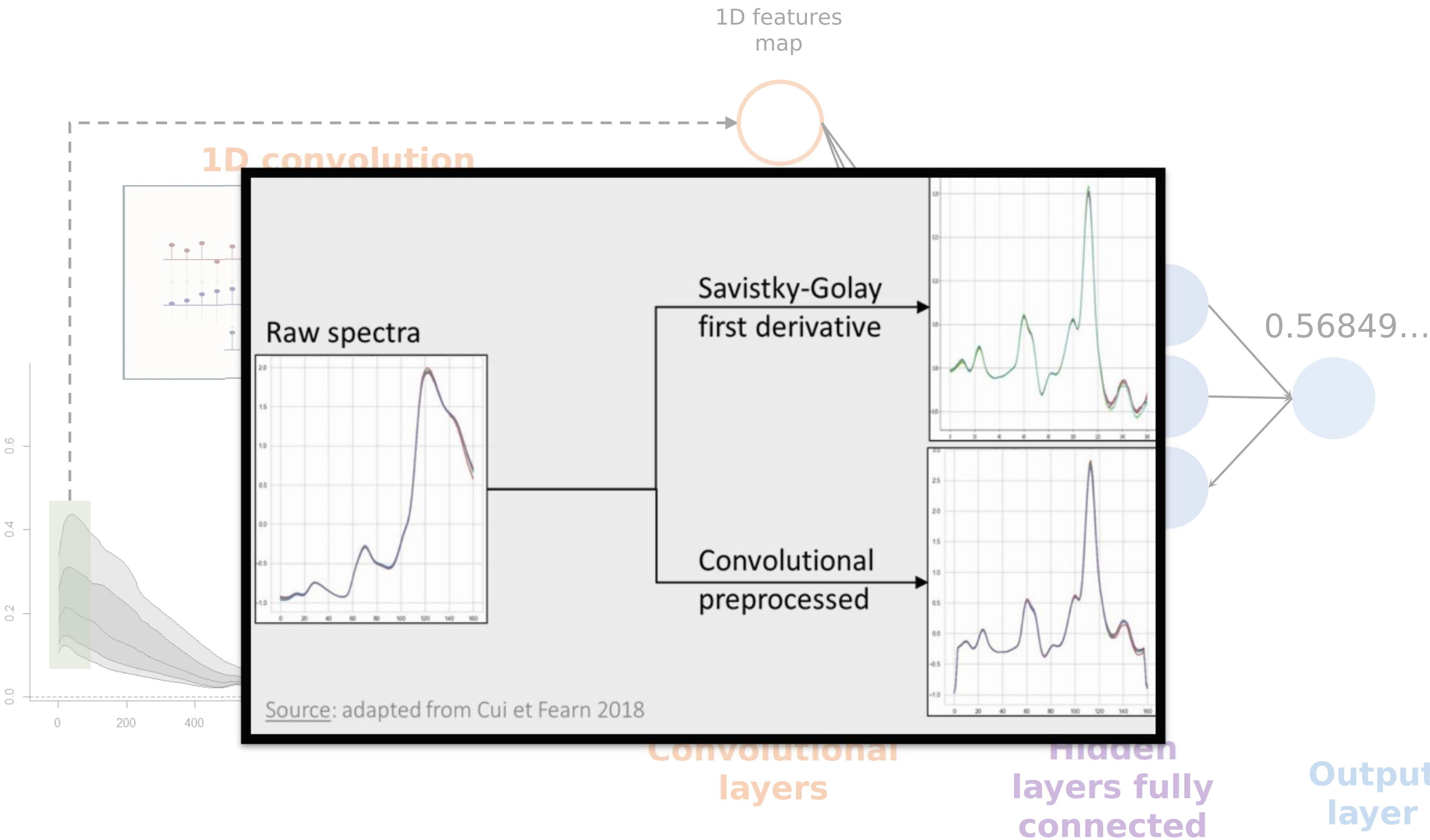
Statistical analyses: deep learning method (2D CNN)



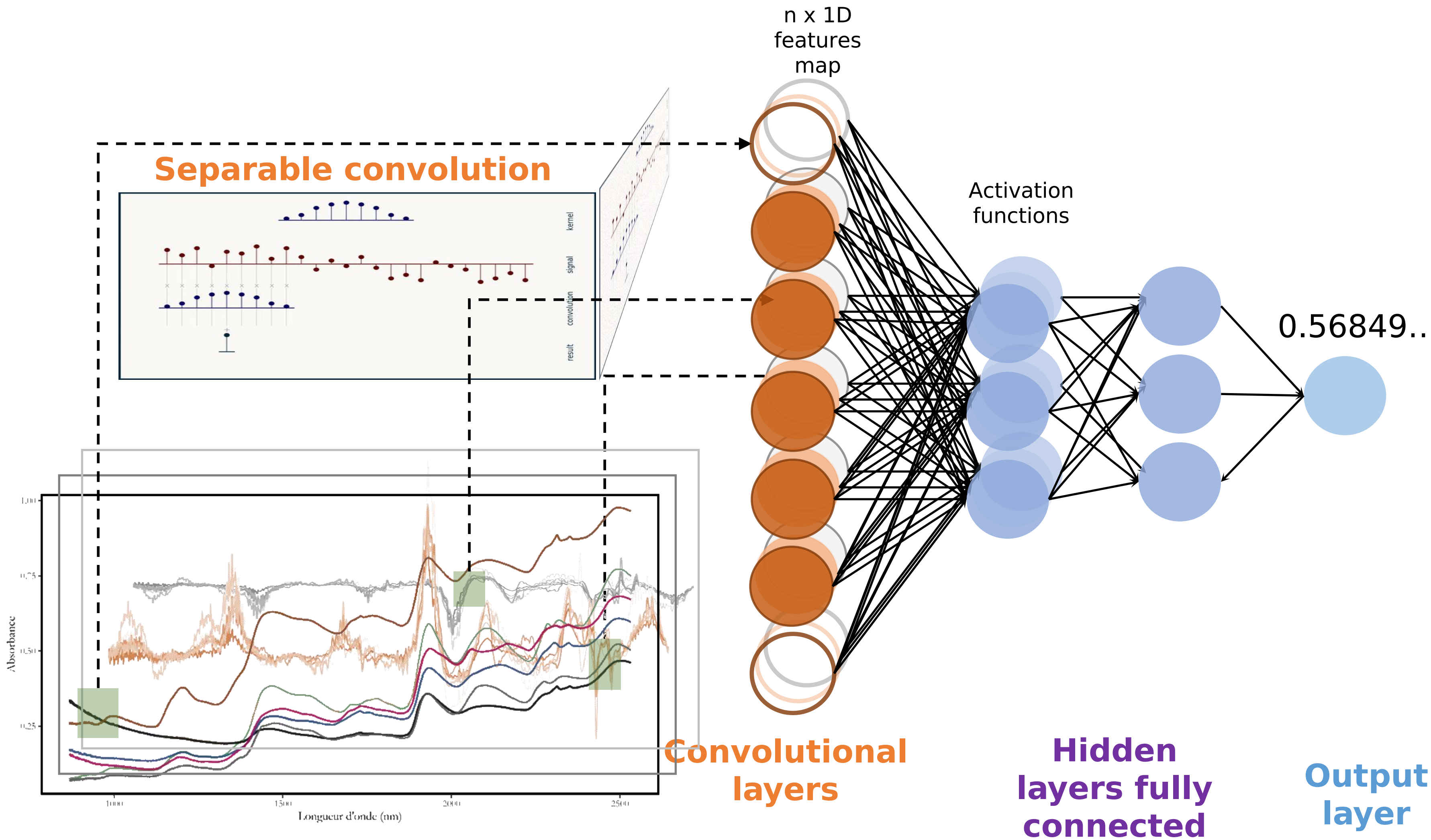
Statistical analyses: deep learning method (1D CNN NIRS)



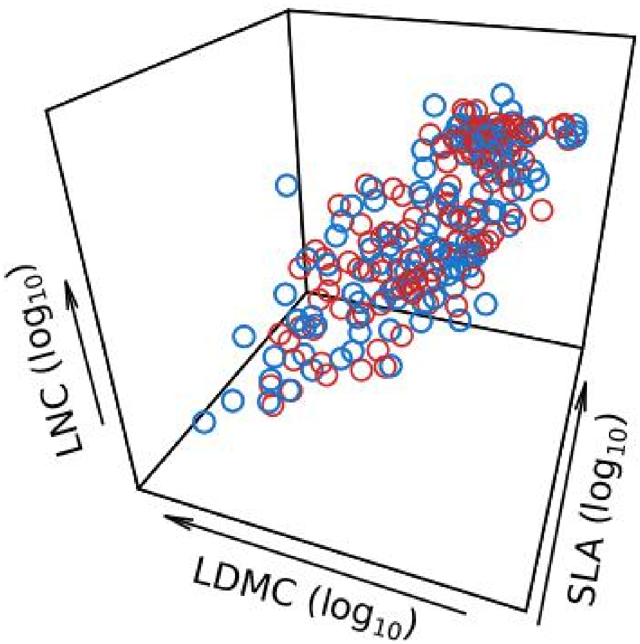
Statistical analyses: deep learning method (1D CNN NIRS)

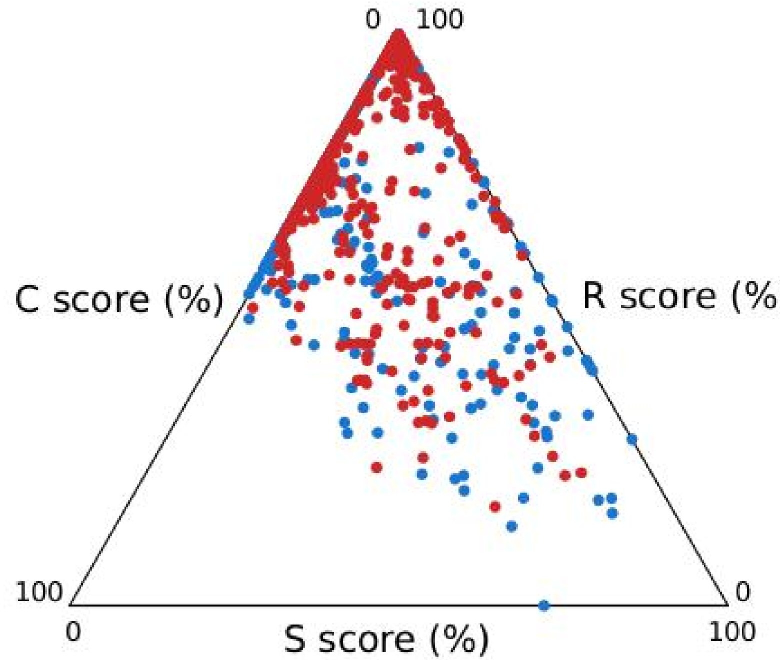


Statistical analyses: deep learning method (Separable 1D CNN NIRS)



Results: 1. Predict functional traits and ecological strategies?

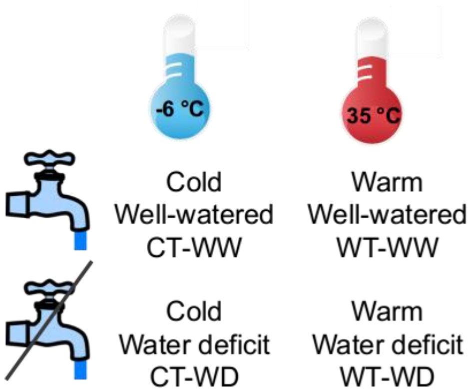


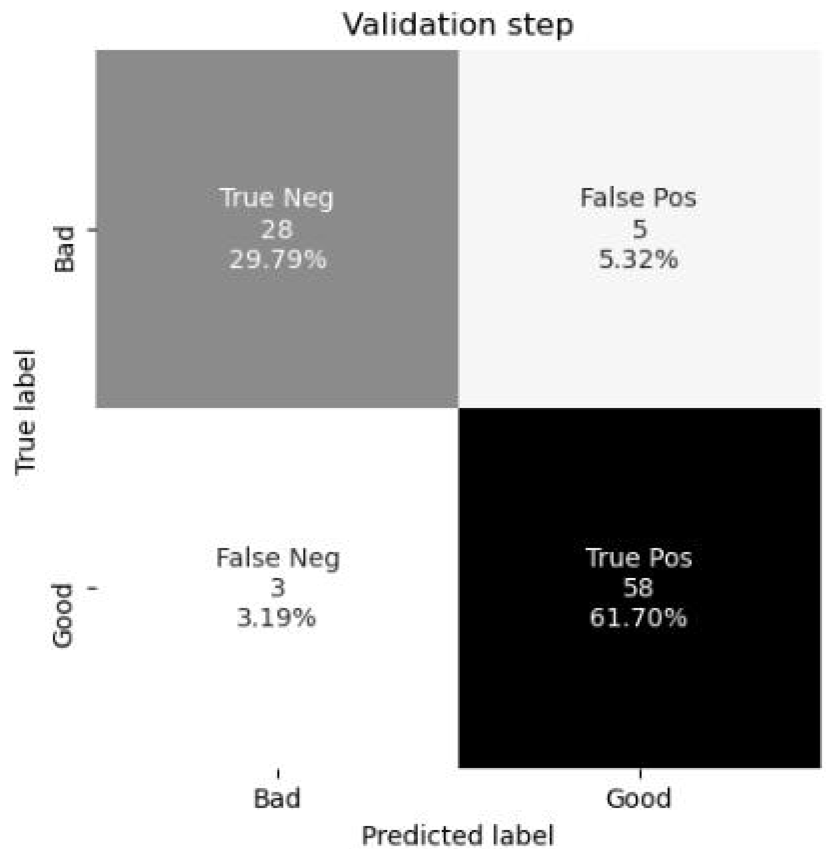


NIRS well predicts trait relationships and recapitulates plant ecological strategies

Results: 2. Predict growth environment and stress response?

NIRS predicts
plant survival to stress
(water deficit x temperature)

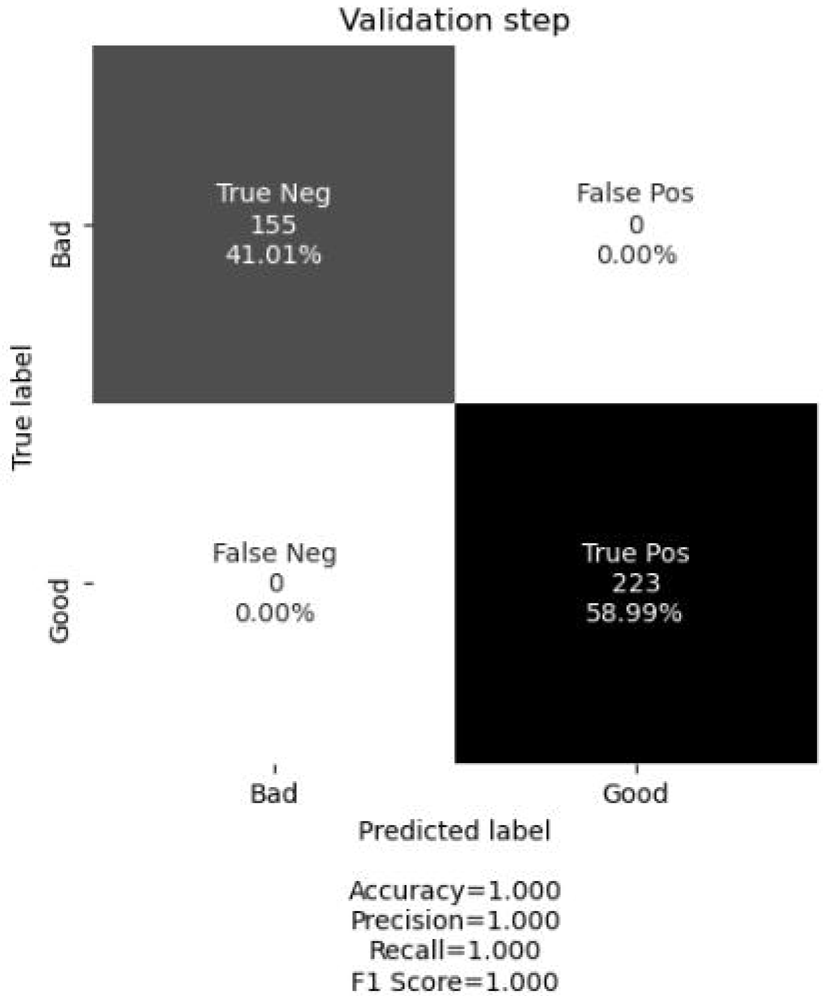




Accuracy=0.915
Precision=0.921
Recall=0.951
F1 Score=0.935

Results: 2. Predict growth conditions and stress response?

NIRS accurately predicts growth conditions (outdoor versus indoor)



Results: 3. Predict leaf secondary metabolites?

Type	Name	Unit	n	Calibration: Filter depth	Calibration: Data augm.	Calibration: RMSE	Calibration: r2	Validation: RMSE	Validation: r2
sugars	Glucose	µmol/gFW	4103	1	10	30.885	0.209	7.822	0.376
	Fructose	µmol/gFW	4037	1	10	49.419	0.132	6.618	0.666
	Sucrose	µmol/gFW	3245	1	10	34.324	0.001	5.929	0.002
	Cellobiose	µmol/gFW	3168	1	10	0.312	0.683	0.066	0.741
	Arabinose	µmol/gFW	3168	1	10	0.248	0.001	0.343	0.000
	Fucose	µmol/gFW	3168	1	10	0.113	0.333	0.011	0.037
	Galactose	µmol/gFW	4037	1	10	0.920	0.698	0.785	0.394
	Isomaltose	µmol/gFW	3168	1	10	0.059	0.728	0.011	0.279
	Maltose	µmol/gFW	746	2	1	0.190	0.000	0.064	0.002
	Mannose	µmol/gFW	3168	1	10	1.200	0.138	0.104	0.491
	Raffinose	µmol/gFW	3168	1	10	0.084	0.996	0.821	0.599
	Rhamnose	µg/gFW	3168	1	10	67.786	0.000	95.561	0.034
	Ribose	µmol/gFW	576	2	1	0.274	0.008	0.143	0.003
	Melezitose	µmol/gFW	4037	1	10	0.024	0.578	0.012	0.394
	Melibiose	µmol/gFW	3168	1	10	0.289	0.825	0.233	0.745
	Palatinose	µmol/gFW	590	2	1	0.851	0.020	0.084	0.026
	Inositol	µmol/gFW	3168	1	10	4.997	0.115	0.589	0.447
	Trehalose	µmol/gFW	576	2	1	0.583	0.003	0.085	0.003
	Xylose	µmol/gFW	3168	1	10	0.235	0.150	0.046	0.107

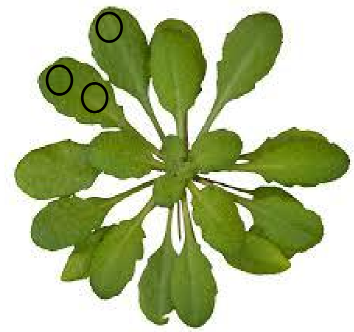
Results: 3. Predict leaf secondary metabolites?

				Calibration				Validation	
Type	Name	Unit	*n*	Filter depth	Data augm.	RMSE	*r2*	RMSE	*r2*
hormones	SA	ng/gFW	3773	1	10	834.472	0.001	495.571	0.235
	JA	ng/gFW	4191	1	10	123.582	0.898	170.864	0.457
	ABA	nmol/gFW	762	2	1	0.053	0.098	0.043	0.272
	IAA	nmol/gFW	4191	1	10	0.014	0.989	0.073	0.399
	CMLX	nmol/gFW	4191	1	10	39.066	0.012	34.951	0.089

Results: 3. Predict leaf secondary metabolites?

Type	Name	Unit	*n*	Calibration Filter depth	Data augm.	RMSE	*r2*	Validation RMSE	*r2*
other secondary metabolites	Quercetin glu	Peakarea/mg	4158	1	10	24.635	0.936	46.410	0.443
	Apigenin rutin	Peakarea/mg	4158	1	10	207.552	0.968	843.447	0.319
	Caffeic acid	Peakarea/mg	4158	1	10	28.364	0.151	0.920	0.410
	Chlorogenic a	Peakarea/mg	4158	1	10	1.282	0.998	14.293	0.740
	Cyanidin rhan	Peakarea/mg	4158	1	10	245.707	0.972	819.908	0.548
	Cyanidin sopl	Peakarea/mg	4158	1	10	50.168	0.995	373.988	0.356
	Dihydro caffe	Peakarea/mg	4158	1	10	3.291	0.987	8.470	0.866
	Citrat	Peakarea/mg	4158	1	10	309.804	0.987	1664.348	0.560
	Fumarat	Peakarea/mg	4158	1	10	56.591	0.965	163.304	0.198
	Malat	Peakarea/mg	4158	1	10	116.487	0.989	799.920	0.184
	Succinat	Peakarea/mg	4158	1	10	14.521	0.948	45.249	0.185
	Kaempherol g	Peakarea/mg	4158	1	10	218.594	0.957	508.803	0.195
	Kaempherol r	Peakarea/mg	4158	1	10	174.741	0.996	1554.556	0.618
	Kaempherol x	Peakarea/mg	4158	1	10	164.575	0.986	747.392	0.583
	mCoumaric a	Peakarea/mg	4158	1	10	111.729	0.814	143.725	0.037
	pCoumaric ac	Peakarea/mg	4158	1	10	1.352	0.917	1.318	0.492
	Pelargonidin (	Peakarea/mg	4158	1	10	7.873	0.988	33.253	0.674
	Pelargonidin s	Peakarea/mg	4158	1	10	30.145	0.990	211.416	0.518
	Prenyl naring	Peakarea/mg	4158	1	10	10.998	0.917	14.790	0.632

Conclusions



1. Predict functional traits and plant strategies?

--> yes

2. Predict environment of origin and stress response?

--> yes

3. Predict leaf metabolites concentration?

--> very variable

Acknowledgments

Cyrille Violle
Justine Bresson
Aurélien Estarague
Stefania Przybylska
Marianne Gérard
Kevin Sartori
Hans Reijnen
Amélie Emanuel
Justine Floret
Lauren Gillespie
Etienne Baron
Elena Kazakou

Denis Cornet
Lauriane Rouan

Denis Vile

Martin Ecarnot

Julie Messier

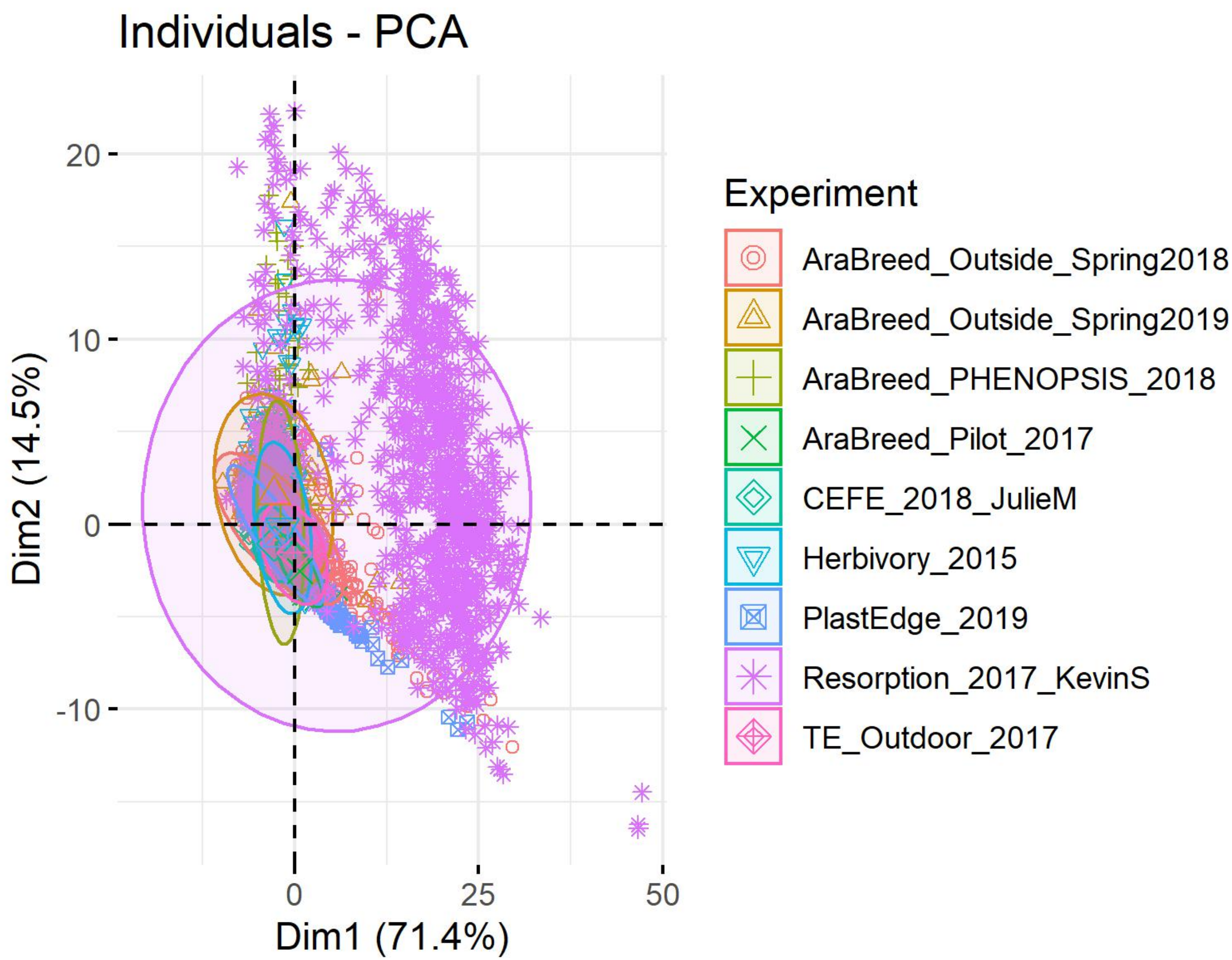
Projet AraBreed
(2017-2022)

Annexes

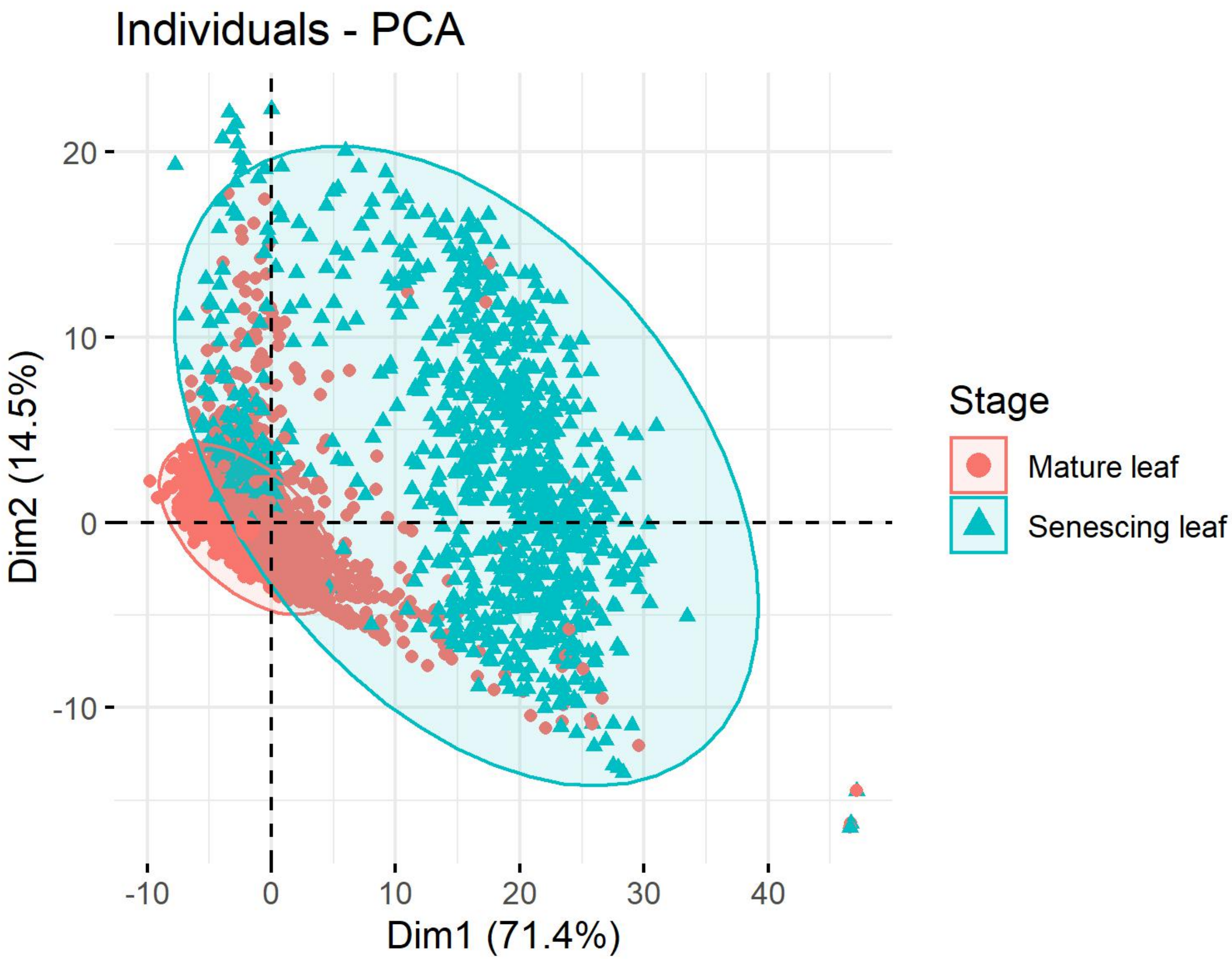
Comparing CNN and PLS approaches

	PLS				CNN	
	transfo	ncomp	RMSE	r^2	RMSE	r^2
LNC (%)	ga1msc	6	0.80	0.83	0.52	0.93
δ13C	snvga1	4	0.87	0.65	0.61	0.83
Plant growth rate (mg d-1)	ha2ha2	4	0.00	0.51	0.00	0.57
R score (%)	mscha1	5	10.08	0.59	4.79	0.87
Fructose	ga1	31	33.67	0.11	6.62	0.67
Maltose	ga1	3	53.86	0.02	0.06	0.00
Mannose	ga2sg1	10	0.16	0.19	0.10	0.49
Ribose	sg1ga3	16	38.34	0.07	0.14	0.00
Arabinose	ga1snv	3	309.23	-0.01	0.34	0.00
JA	snvga1	4	0.98	0.15	170.86	0.46
Glucoraphenin	ga2sg1	7	0.65	0.70	0.61	0.75
Hexyl glucosinolate	sg1ga3	14	54.04	-0.22	45.55	0.00
Butyl glucosinolate	ga4sg4	13	4.12	0.46	3.17	0.56
X3MTP	ga2ga1	3	41.90	-0.01	7.90	0.74
Dihydro caffeyol glucuronide	ga1snv	5	12.06	0.81	8.47	0.87

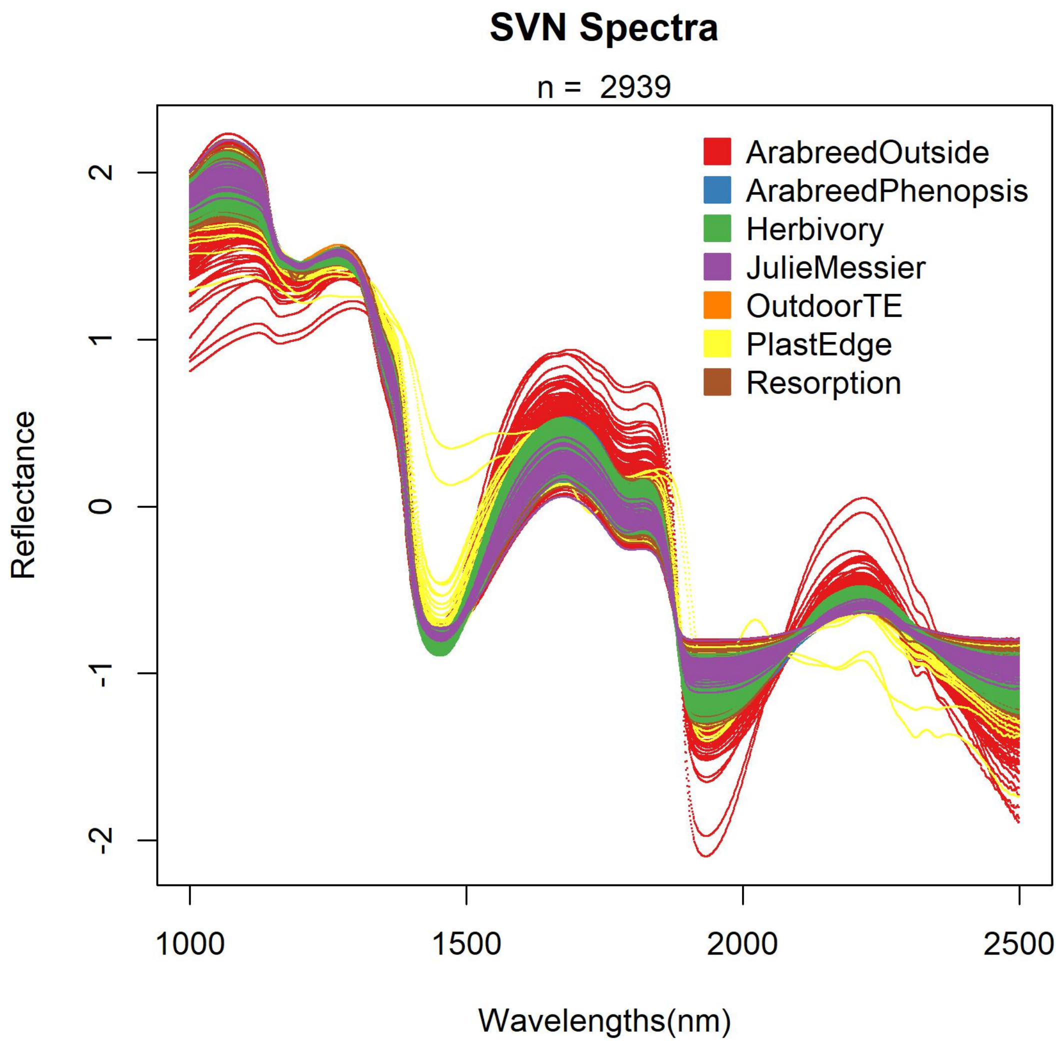
Preliminary investigation of spectra: PCA on spectrum values

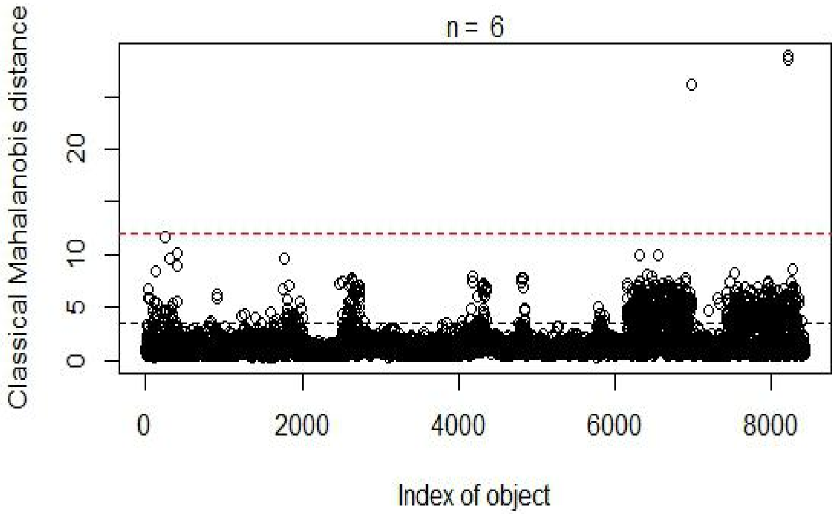


Preliminary investigation of spectra: PCA on spectrum values



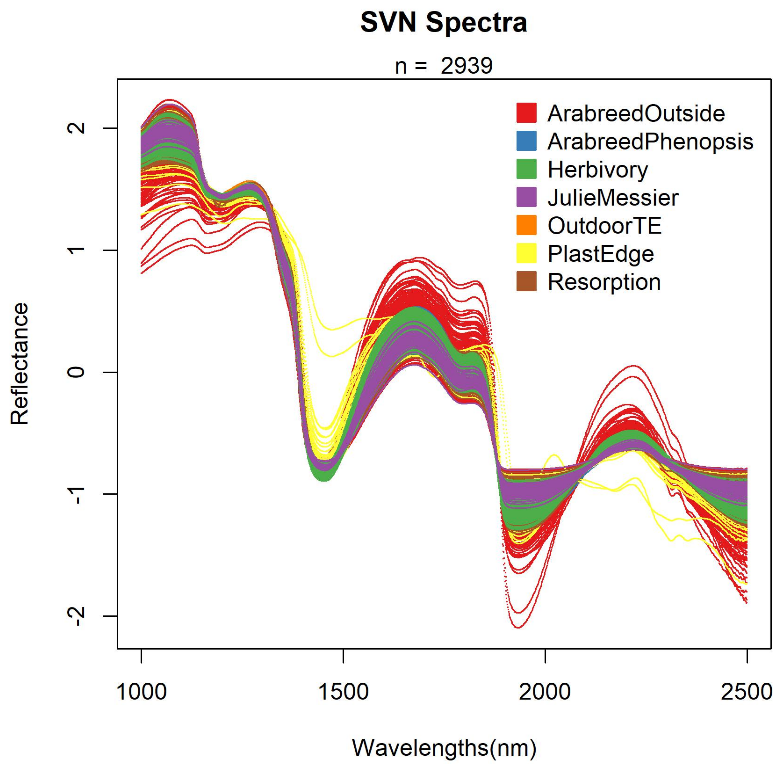
Preliminary Investigation of Spectra: Remove outliers on mean spectra





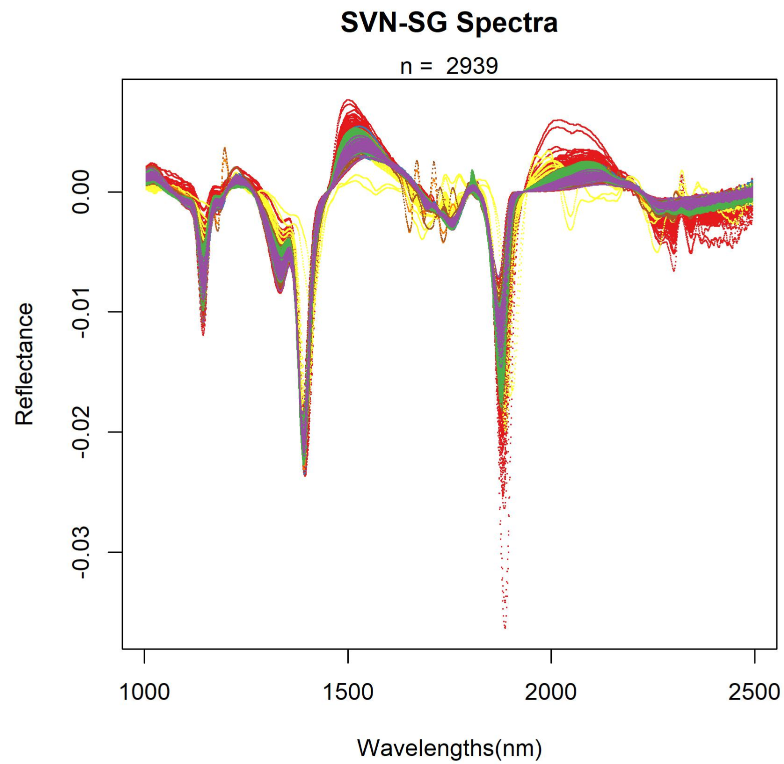
Pretreatments of NIR spectra: SNV and SG

Standard Normal Variate correction (SNV)



Every data point of the spectrum is subtracted from the mean and divided by the standard deviation: **normalization of spectra**

Savitzky-Golay smoothing filter (SG)



SG smoothes signals and calculates derivatives: **enhance signal properties, and suppress unwanted spectral features** that arise due to nonideal instrument and sample properties