Correcting measurement effects and temporal variations in NIRS models: application to the early detection of ``early blight"

HelioSPIR 2022 Nantes

Florent Abdelghafour

UMR Itap

Context : time monitoring and early detection of *alternaria solani*

ASD QualitySpec®Trek spectrometer

[350 – 2500]nm

<u>Database :</u>

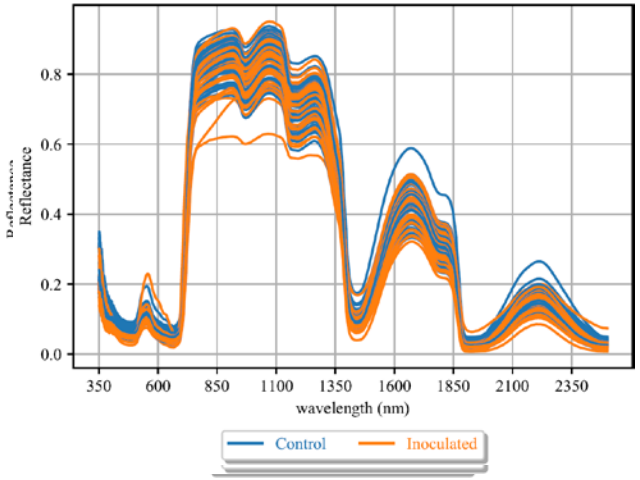
42 plants

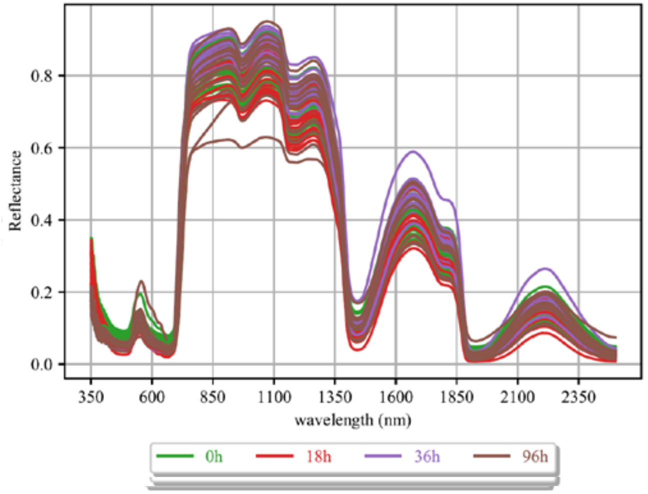
7 varieties * 3 replicates * 2 groups (Inoculated / Control)

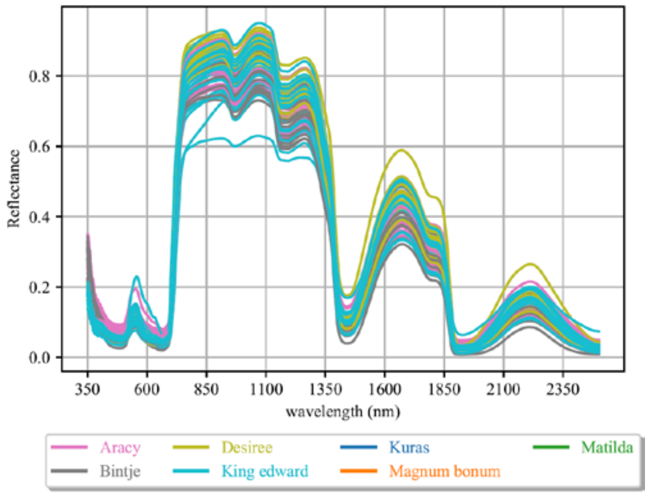
HAI 0h 18h 36h 96h

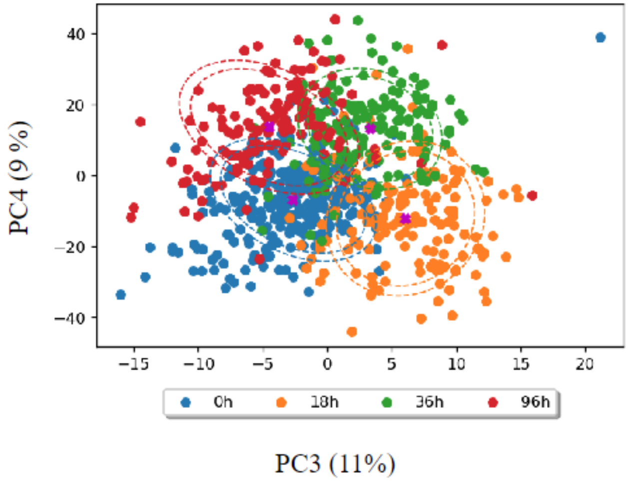
~ 2000 spectra [p = 2151]

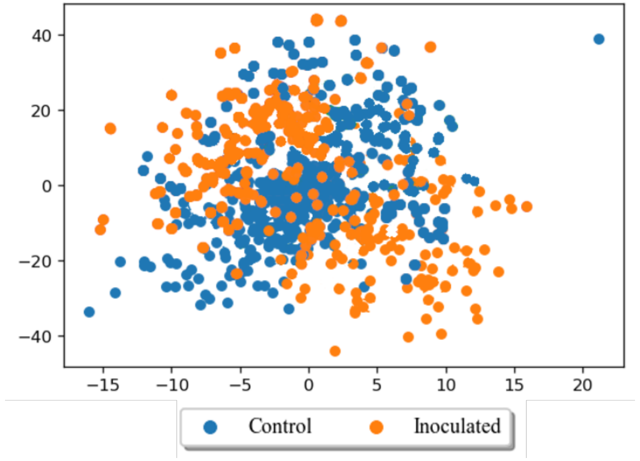
Data visualisation







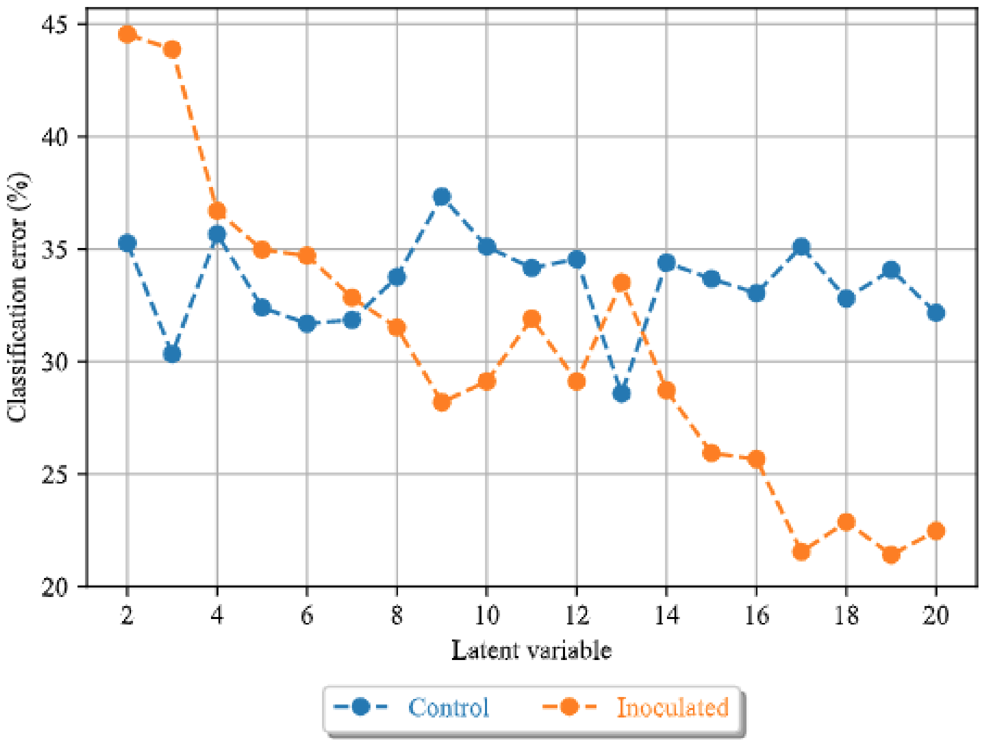




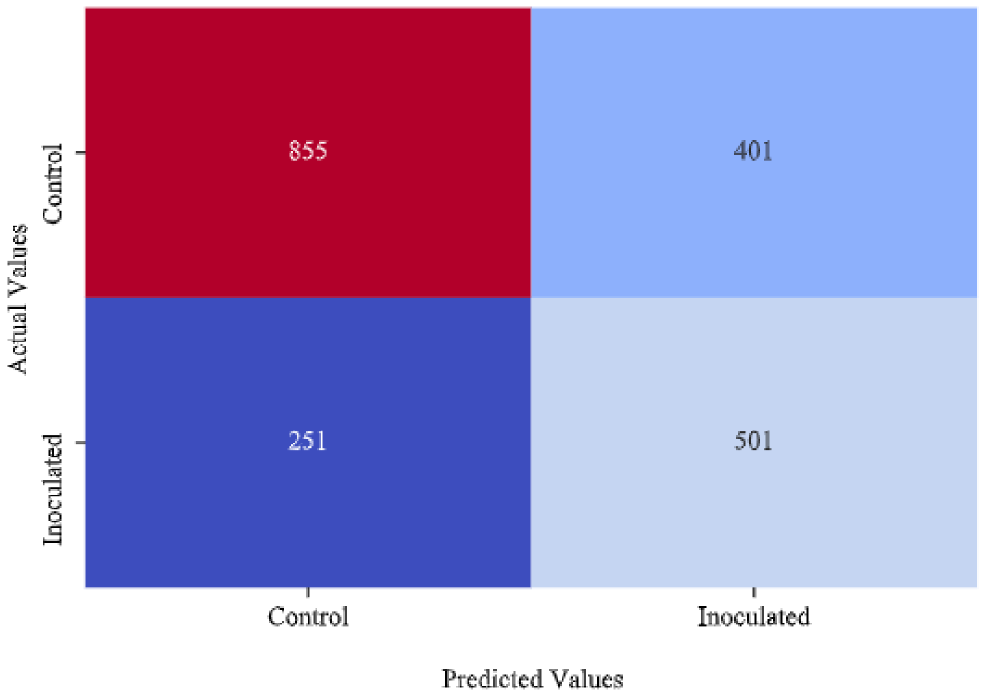
Data split

cultivar	Total	Health status		HAI			
		Control	Inoculated	0h	18h	36h	96h
Aracy	191	119	72	48	48	48	47
Bintje	271	172	99	68	68	67	68
Desiree	288	180	108	72	72	72	72
King Edwards	281	175	106	70	70	70	71
Kuras	271	169	102	68	68	67	68
Magnum Bonum	361	222	139	92	93	92	84
Matilda	345	219	126	86	86	86	87
Total	2008	1256	756	504	505	502	497

Limits of PLSDA :



(a) Cumulative classification error (%) per class for PLSDA



(b) Confusion matrix for 7 LV

Limits of PLSDA : detail per cultivar and HAI

Cultivar	Control	Inoculated
Aracy	50	15
Bintje	29	38
Desiree	54	15
King Edwards	27	26
Kuras	23	30
Magnum Bonum	23	41
Matilda	24	54

Classification error (%) of PLSDA detailed by cultivar

Classification error (%) of PLSDA detailed by HAI

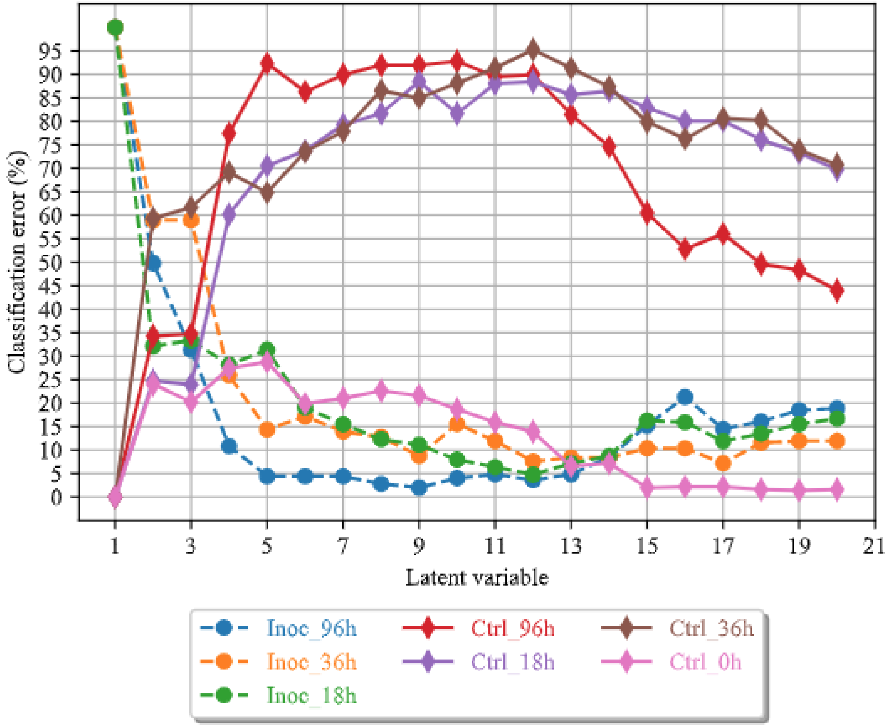
HAI time-points	0h	18h	36h	96h
Control	4	81	48	**20**
Inoculated		64	20	**15**

Problem statement

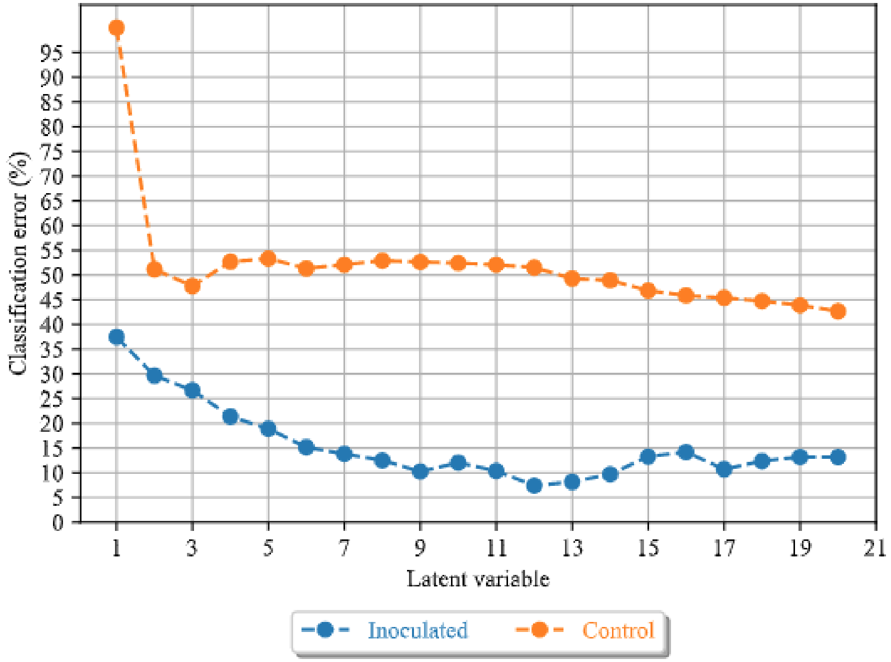
- Variability caused by HAI → sub-classes problem?

- Time induced effect?
  - Measurment
  - Environment

PLSDA2: predicting HAI-time points and health status altogether



(a) mixed classes classification cumulated CE



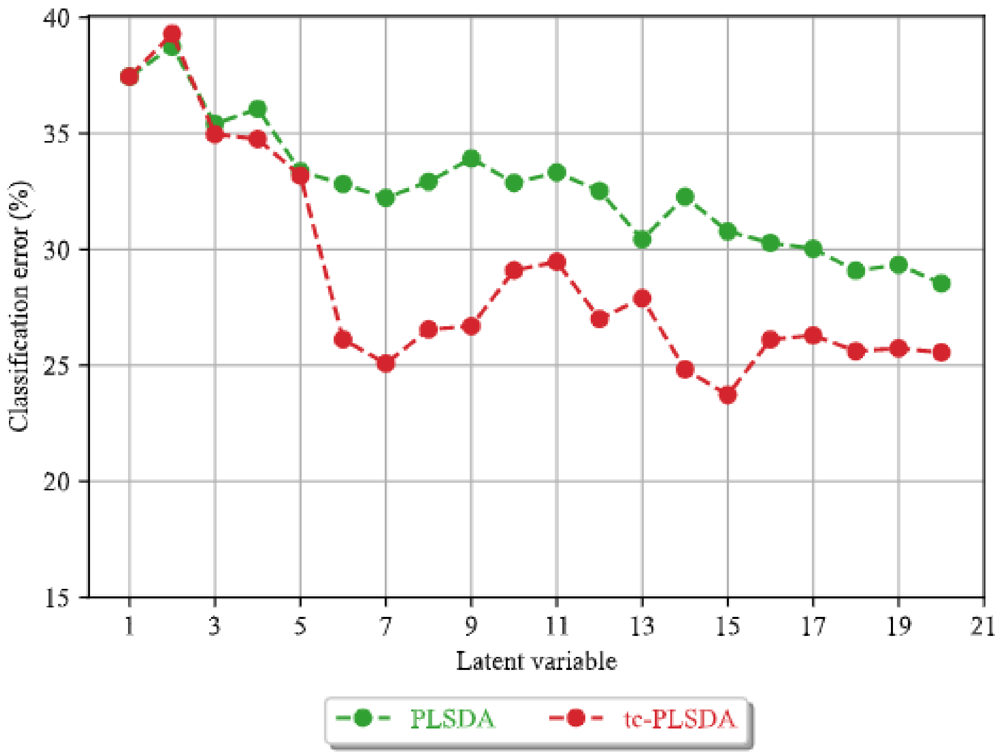
(b) Summarised binary classification cummulated CE

Time-constrained PLSDA

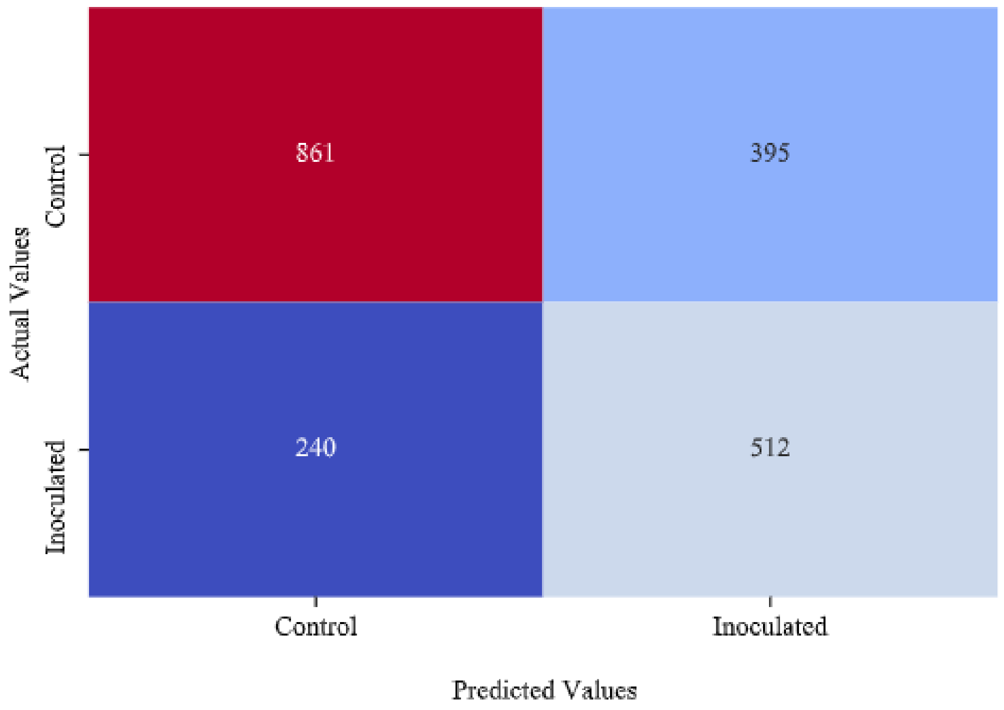
- Calibrating with Y health status and HAI time-points
- Predicting **only health status** (with one column of B-coefs)
- Florian Wülfert, Wim Th. Kok, Onno E. de Noord, Age K. Smilde, Linear techniques to correct for temperature-induced spectral variation in multivariate calibration

~ An implicit orthogonalisation within the PLS model

Comparison of tc- PLSDA and PLSDA



(a) Comparison of mean classification error between between PLSDA and tc-PLSDA



(b) confusion matrix of tc-plsda

Comparison of tc- PLSDA and PLSDA

cultivar	PLSDA		tc-PLSDA	
	Control	Inoculated	Control	Inoculated
Aracy	50	15	29	37
Bintje	29	38	23	17
Desiree	54	15	30	12
King Edwards	27	26	24	9
Kuras	23	30	25	21
Magnum Bonum	23	41	31	30
Matilda	24	54	27	29

C.E. (%) of tc-PLSDA detailed by cultivar

C.E. (%) of tc-PLSDA detailed by HAI

Classification error (%)	tc-PLSDA			
HAI	0h	18h	36h	96h
Control	12	54	33	25
Inoculated		22	31	13

HAI time-points	0h	18h	36h	96h
Control	4	81	48	**20**
Inoculated		64	20	**15**

C.E. (%) of PLSDA detailed by HAI

EPO-PLSDA

1: Compute the difference matrix of the "external factors"

$$D[k=4, p=2151] = \begin{pmatrix} \bar{X}-\bar{X}_{t=0} \\ \bar{X}-\bar{X}_{t=18} \\ \bar{X}-\bar{X}_{t=36} \\ \bar{X}-\bar{X}_{t=96} \end{pmatrix},$$

where $\bar{X}$ denotes the column-wise mean of X, *i.e.* the mean spectrum.

2: Perform singular value decomposition (SVD) of the covariance matrix of D

$[U, s, V] = SVD(D^{\mathrm{T}}.D)$,

where "T" denotes the transposition operator and . denotes the matrix multiplication operator.

3: Compute the "detrimental space" G

$G = D^{\mathrm{T}}.V$

4: Determine G_{opt} the optimal subset of G

$\hat{G}_{opt} \subset G$

5: Compute the vector orthonormal basis $Proj_{\perp[p,p]}$ to project X orthogonally on G_{opt}

$Proj_{\perp[p,p]} = Id_{[p,p]} - (G_{opt}.(G_{opt}{}^{\mathrm{T}}.G_{opt})^{-1}.G_{opt}{}^{\mathrm{T}})$,

6: where $Id_{[p,p]}$ denotes the identity matrix of dimension $Proj_{\perp} = Id - [\hat{G}_{opt}.(\hat{G}_{opt})^{\mathrm{T}}.(\hat{G}_{opt}^{-\frac{1}{2}\mathrm{T}})].\hat{G}_{opt}.[\hat{G}_{opt}^{\mathrm{T}}.\hat{G}_{opt})^{-\frac{1}{2}}]$

7: Project X orthogonally to G_{opt}

$X_{orth} = X.Proj_{\perp}$

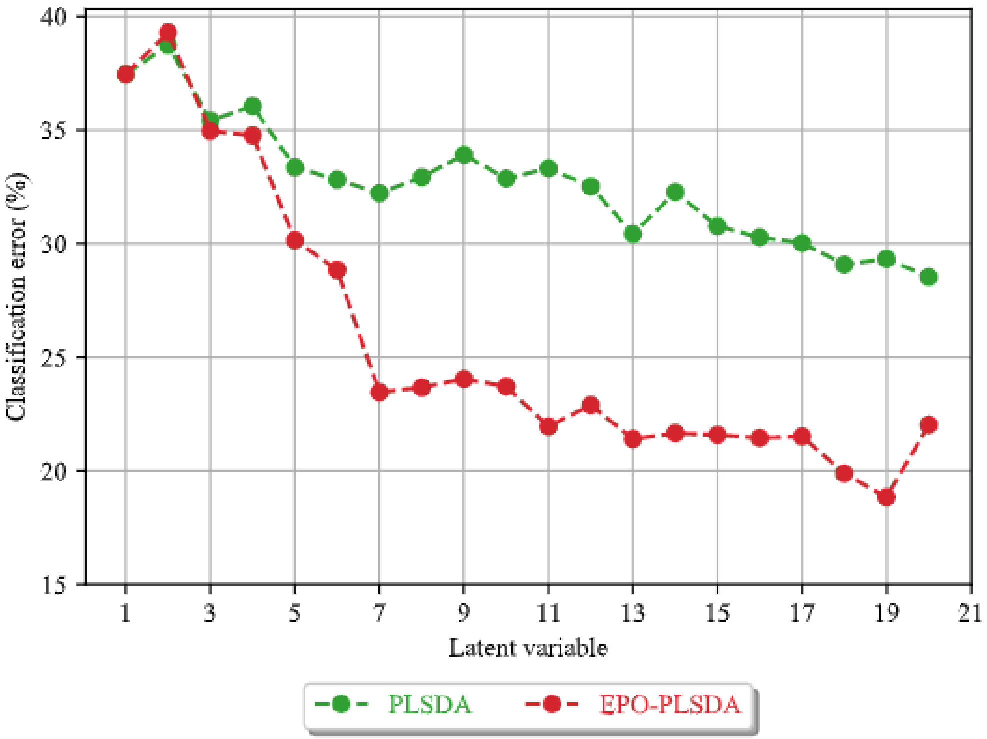
8: Calibrate a robust PLS model invariant to the factors in D, with X_{orth}

dv = PLSDA.calibration(X_{orth}), so that:

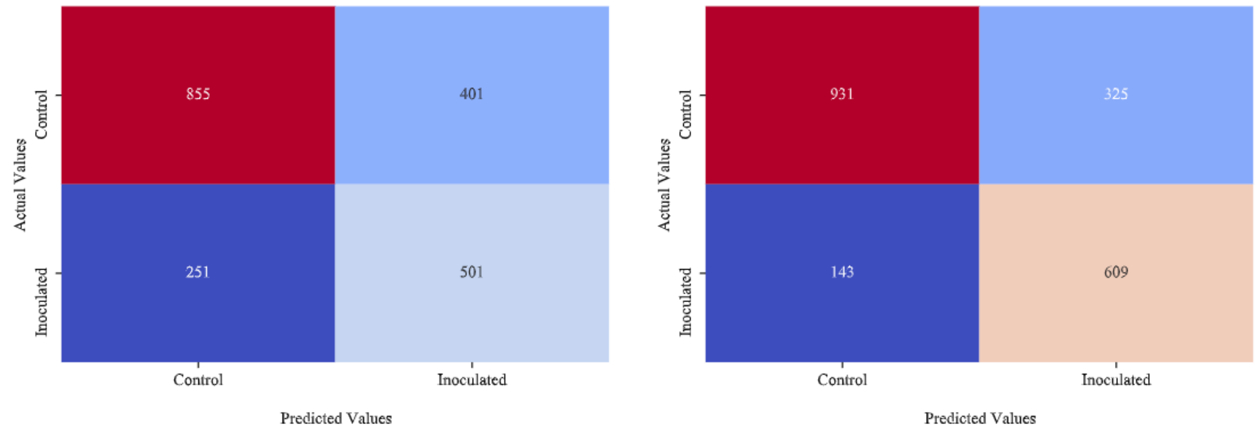
for any given measured X_{new}, regardless the HAI, $\hat{Y} = dv.X_{new}$,

$\hat{Y}$ denotes the estimation of the sample class, from the data X_{new} and dv denotes the discriminant vector of the PLSDA model.

Comparison between PLSDA / EPO-PLSDA



Comparison of MCE between PLSDA and EPO-PLSDA



(a) PLSDA

(b) EPO-PLSDA

Test confusion matrices

EPO-PLSDA: detail per cultivar HAI

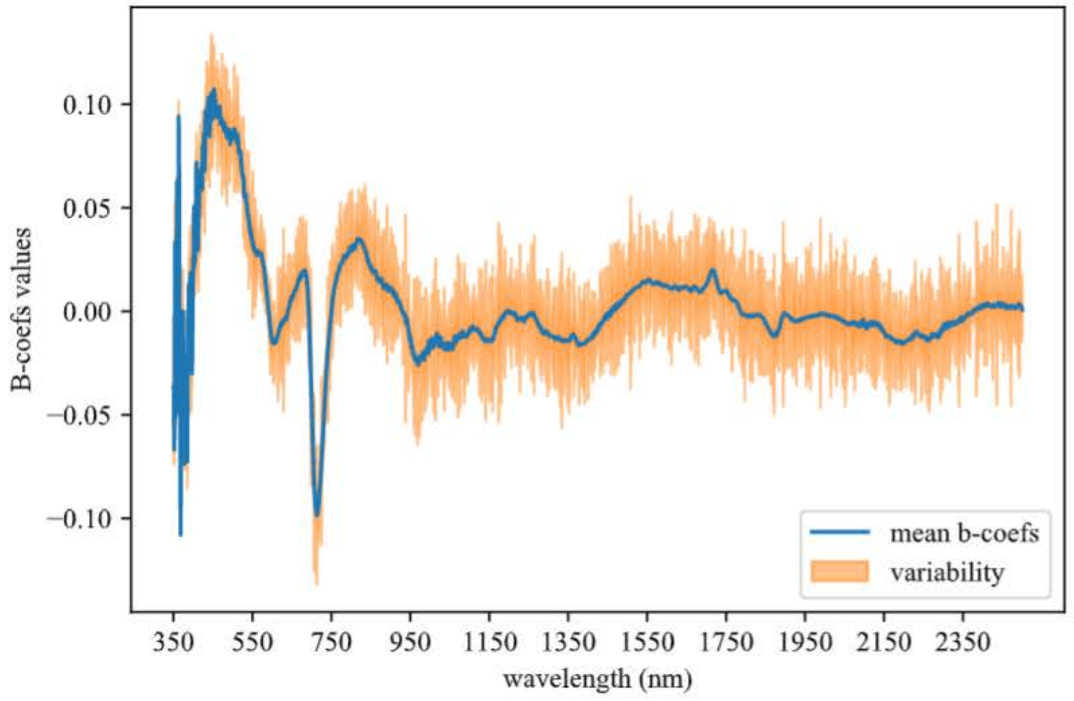
cultivar	PLSDA		EPO-PLSDA	
	Control	Inoculated	Control	Inoculated
Aracy	50	15	25	38
Bintje	29	38	21	11
Desiree	54	15	31	7
King Edwards	27	26	24	12
Kuras	23	30	23	19
Magnum Bonum	23	41	25	25
Matilda	24	54	25	23

C.E. error (%) of EPO-PLSDA detailed by cultivar

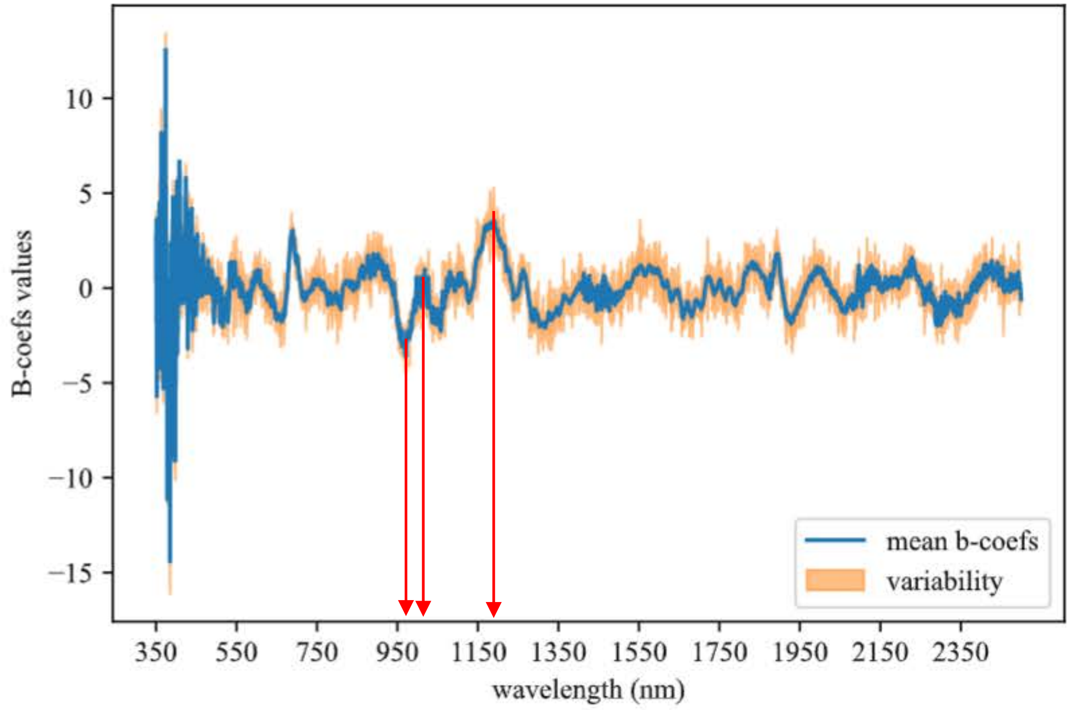
Method	PLSDA				EPO-PLSDA			
HAI	0h	18h	36h	96h	0h	18h	36h	96h
Control	4	81	48	**20**	8	63	**30**	**20**
Inoculated		64	20	**15**		27	**10**	**19**

C.E. error (%) of EPO-PLSDA detailed by HAI

Stability of the model and new contributions



(a) PLS



(b) EPO-PLS

Conclusion

- Discrimination problems in agriculture (phenotyping pathology)

  = diffcult even in controled conditions

- Significant external variability sources, especially time induced measurement effect > Studied effect

- Correction with EPO-PLS : other applications for transfer and correction method

Merci

Questions ?