Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

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

Electrophysiological assessment of plant outside a Faraday cage using supervised machine learning

Machine Learning Workshop

C. Camps
GreenHouse Crops

12.11.2019

Content – General Info

The group GreenHouse Crops

Team - 3.8 ETP + 1 ETP + apprentices and students
Facilities - 2000m^2 greenhouses (soilless & soil) - 3x phytotrons
Project (BO, OFAG, INNOSUISSE, EU-project, direct company, etc.)

General Topics of the group

Energy management in greenhouse
Cropping systems (conv. Vs. Organic, soil vs. Soilless)
Pest & diseases in greenhouse
Non-destructive analyses of fruit quality - chemometric
Digitized solutions for Small and Mid-size production

Agroscope

Content

1 - Machine learning last 10 Years –

Application to *Non-destructive analyses of fruit quality*

2 - Electrophysiological assessment of plant a Faraday cage using supervised machine learning -

Application to *tomato growth in greenhouse*

1 - Machine learning last 10 Years –

Application to *Non-destructive analyses of fruit quality*

«Classic» machine learning

Methods - PLS, PCR, StepWiseR, etc. models (Matlab®)

Training a model

Calibration + Cross-validation = Calibration

Validate a model

Test set = validation

External validation set = validation and improve the model with time

Maintain a model

Data sets from production years, sites over the years of the model exploitation …

3 short Ex. ---

molecules MDPI

Article

Optimized Prediction of Reducing Sugars and Dry Matter of Potato Frying by FT-NIR Spectroscopy on Peeled Tubers

Cédric Camps * and Zo-Norosoa Camps

Institute for Plant Production Sciences IPS, Agroscope, CH-1964 Conthey, Switzerland; zonorosoa@gmail.com

* Correspondence: cedric.camps@agroscope.admin.ch

Contents lists available at ScienceDirect

LWT - Food Science and Technology

journal homepage: www.elsevier.com/locate/lwt

Prediction of raspberries puree quality traits by Fourier transform infrared spectroscopy

Zo-Norosoa Andrianjaka-Camps [a,*], Daniel Baumgartner [b], Cedric Camps [a], Elena Guyer [b], Eva Arrigoni [b], Christoph Carlen [a]

[a] Agroscope, Institute for Plant Production Sciences IPS, CH-1964 Conthey, Switzerland

[b] Agroscope, Institute for Food Sciences, CH-8820 Wädenswil, Switzerland

J Sci Food Agric 2014; **94**: 1397–1402

Prediction of essential oil content of oregano by hand-held and Fourier transform NIR spectroscopy

Cédric Camps,[a*] Marianne Gérard,[a,b] Mélanie Quennoz,[b] Cécile Brabant,[c] Carine Oberson[c] and Xavier Simonnet[b]

molecules

Article

Optimized Prediction of Reducing Sugars and Dry Matter of Potato Frying by FT-NIR Spectroscopy on Peeled Tubers

Cédric Camps * and Zo-Norosoa Camps

Institute for Plant Production Sciences IPS, Agroscope, CH-1964 Conthey, Switzerland; zonorosoa@gmail.com
* Correspondence: cedric.camps@agroscope.admin.ch

Table 1. Partial least square values of dry matter content prediction. PDTE: entire and unpeeled potatoes, PDTP: entire and peeled potatoes; PDTC: potatoes cut transversally.

Model Fitting

Expression of actual vs. Predicted values.

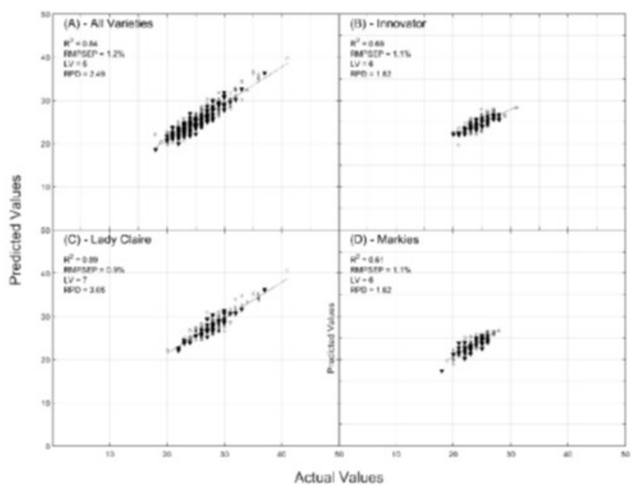


Figure 1. Actual vs. predicted values of dry matter content (DMC) (g of dry weight/100 g fresh weight). Calibration (○), validation (▲). **(A)** Entire and peeled tubers of the three tested varieties, **(B)** Entire and peeled tubers of Innovator, **(C)** Entire and peeled tubers of Lady Claire, **(D)** Entire and peeled tubers of Markies.

Model Beta-Coefficients

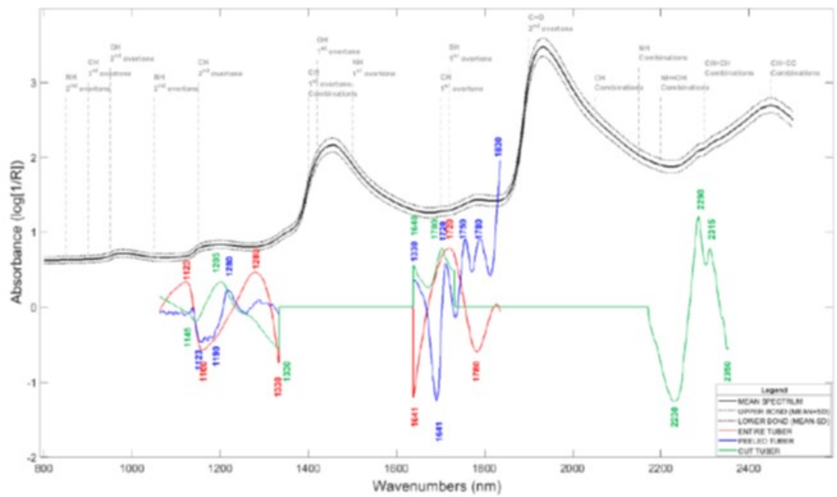


Figure 2. Mean spectra and beta-coefficients of the first PLS models latent variable to predict the DMC based on spectral data acquired on entire tubers (red line), peeled tubers (blue line), and cut tubers (green line).

Contents lists available at ScienceDirect
LWT - Food Science and Technology
journal homepage: www.elsevier.com/locate/lwt

Prediction of raspberries puree quality traits by Fourier transform infrared spectroscopy

Zo-Norosoa Andrianjaka-Camps [a,*], Daniel Baumgartner [b], Cedric Camps [a], Elena Guyer [b], Eva Arrigoni [b], Christoph Carlen [a]

[a] Agroscope, Institute for Plant Production Sciences IPS, CH-1964 Conthey, Switzerland
[b] Agroscope, Institute for Food Sciences, CH-8820 Wädenswil, Switzerland

Table 2
Statistical summary of calibration and cross-validation of the FTIR-PLS prediction of quality traits of raspberries.

	Quality traits	LV	Spectral range	R^2	RMSEC	RPD	Slope	Bias
Calibration	pH	9	[2461–1952],[1435–926]	0.98	0.02	7.73	0.98	$-4.5\ 10^{-12}$
	TA	8	[5011–4505],[1435–926]	0.99	0.02	12.93	0.99	$2.6\ 10^{-15}$
	SSC	3	[3992–3487],[1435–926]	0.99	0.12	11.51	0.98	$-1\ 10^{-14}$
	Glucose	4	[3992–3487],[1435–926]	0.98	0.07	8.67	0.99	$7.6\ 10^{-15}$
	Fructose	2	[4501–3996],[1435–926]	0.98	0.08	7.41	0.98	$2.1\ 10^{-15}$
	Sucrose	5	[2461–1952],[1435–926]	0.97	0.12	6.12	0.97	$9.7\ 10^{-16}$
	Vitamin C	14	[2974–2465],[1435–926]	0.89	1.5	3.14	0.89	$-5.4\ 10^{-13}$
	Phenolics	12	[2974–2465],[1435–926]	0.82	18	2.4	0.86	$7.7\ 10^{-12}$
	Anthocyanins	9	[4501–3996],[2461–1952]	0.68	10.7	1.76	0.72	$-2.4\ 10^{-11}$
Cross-validation	pH	9	[2461–1952],[1435–926]	0.96	0.03	4.94	0.96	$-2.9\ 10^{-3}$
	TA	8	[5011–4505],[1435–926]	0.99	0.02	9.97	0.98	$2.0\ 10^{-3}$
	SSC	3	[3992–3487],[1435–926]	0.99	0.14	9.72	0.96	$-4.0\ 10^{-2}$
	Glucose	4	[3992–3487],[1435–926]	0.98	0.08	7.24	0.91	$-1.9\ 10^{-2}$
	Fructose	2	[4501–3996],[1435–926]	0.98	0.09	7.04	1.01	$-7.0\ 10^{-3}$
	Sucrose	5	[2461–1952],[1435–926]	0.96	0.14	5.22	0.99	$-4.6\ 10^{-3}$
	Vitamin C	14	[2974–2465],[1435–926]	0.69	2.7	1.78	0.63	0.73
	Phenolics	12	[2974–2465],[1435–926]	0.65	26.01	1.64	0.56	−4.8
	Anthocyanins	9	[4501–3996],[2461–1952]	0.33	15.7	1.16	0.39	−2.6

SSC: soluble solids content; TA: total acid; LV: the number of latent variables introduced in PLS models; R^2: the coefficient of determination; RMSEC/CV: the root mean square error of calibration/cross-validation; RPD: the ratio of performance to deviation. SSC is expressed in °Brix; TA in g malic acid equivalent/100 g fw; glucose, fructose and sucrose in g/100 g fw; vitamin C in mg/100 g of fw; phenolics in mg gallic acid equivalent/g of 100f fw and anthocyanins in mg cyanidin-3-glucoside equivalent/100 g fw. In column Val (validation) 1st and 2nd indicate the first and second validations performed on PLS models.

Agroscope

J Sci Food Agric 2014; **94**: 1397–1402

Prediction of essential oil content of oregano by hand-held and Fourier transform NIR spectroscopy

Cédric Camps,[a*] Marianne Gérard,[a,b] Mélanie Quennoz,[b] Cécile Brabant,[c] Carine Oberson[c] and Xavier Simonnet[b]

Table 1. PLS data of EOC determination (hand-held NIR and FT-NIR)

		PHAZIR 1018			FT-NIR		
PLS data	Unit	Calibration	Cross-validation	Validation	Calibration	Cross-validation	Validation
N	—	74	74	27	74	74	27
EOC range	mL per 100 g	0.23–10.1	0.23–10.1	0.4–8.1	0.23–10.1	0.23–10.1	0.4–8.1
EOC mean value	mL per 100 g	4.8	4.8	4.89	4.8	4.8	4.89
EOC SD	mL per 100 g	2.61	2.61	2.34	2.61	2.61	2.34
λ range	nm	939–1797	939–1797	939–1797	1000–2500	1000–2500	1000–2500
LV	—	3	3	3	6	6	6
R^2	—	0.92	0.92	0.58	0.93	0.94	0.91
SE(C/CV/P)	mL per 100 g	0.75	0.77	2.20	0.7	0.68	0.69
Bias	mL per 100 g	1.40×10^{-2}	1.55×10^{-2}	−2.04	4.5×10^{-7}	1.08×10^{-2}	8×10^{-2}
$SE(C/CV/P)_c$	mL per 100 g	0.75	0.77	0.81	0.7	0.68	0.68
$RSE(C/CV/P)_c$	Relative %	15	15	18	15	15	14
CV_c	Relative %	15	15	18	15	15	14
RPD_c	—	3.54	3.44	3.51 (1.30)[a]	3.7	3.82	3.24
$RPIQ_c$	—	6.19	6.01	5.03 (1.87)[a]	6.6	6.8	4.55
RER_c	—	13.17	12.78	9.45 (3.50)[a]	14.04	14.51	11.31
Spectral treatment			Golay second derivative (step 3)			SNV + detrending	

N, number of samples; EOC, essential oil content; SD, standard deviation; λ range, wavelength range of PLS model; LV, number of latent variables; R^2, determination coefficient; SE, standard error; RSE, relative standard error of prediction; CV, coefficient of variation; RPD, ratio of performance to deviation; RPIQ, ratio of performance to interquartile; RER, ratio of error to range; subscript 'c' (SE_c, RSE_c, CV_c, RPD_c, $RPIQ_c$ and RER_c), parameters calculated after bias correction; SNV, standard normal variate.
[a] Values in parentheses are RPD, RPIQ and RER before correction for bias.

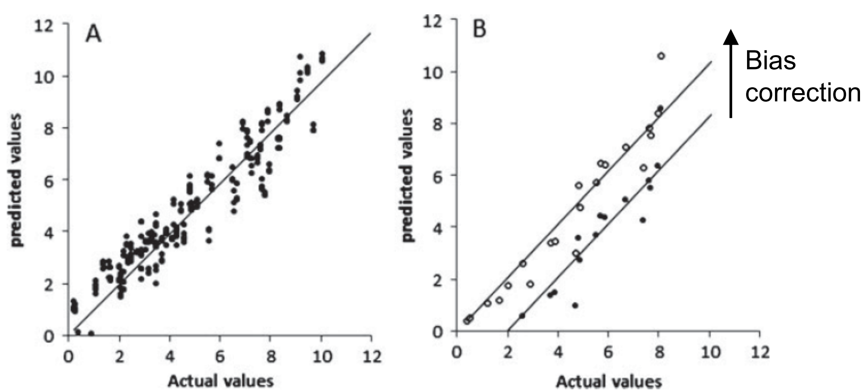


Figure 2. Actual *versus* predicted values of EOC using hand-held NIR device PHAZIR 1018: A, calibration; B, validation; ●, data without bias correction; ○, data corrected for bias value.

$$R^2\,(C/CV/P) = 1 - (PRESS/TSS)$$

$$SE\,(C/CV/P) = \left[\sum_{i=1}^{n} (y_i - \hat{y}_i)^2/n\right]^{1/2}$$

$$\text{bias} = \sum_{i=1}^{n} (\hat{y}_i/n) - \sum_{i=1}^{n} (y_i/n) = \bar{\hat{y}} - \bar{y}$$

$$SE\,(C/CV/P)_c = \left[\sum_{i=1}^{n} (\hat{y}_i - \text{bias} - y_i)^2/n\right]^{1/2}$$

$$RSE\,(C/CV/P)_c\,(\%) = (100/\bar{y})\left[\sum_{i=1}^{n} (\hat{y}_i - \text{bias} - y_i)^2/n\right]^{1/2}$$

where $\hat{y}_i$ is the predicted value, $\bar{y}_i$ the mean value and y_i the actual value of EOC in the PLS model, n is the number of samples in the PLS model, PRESS is the prediction residual error of the sum of squares, TSS is the total sum of squares and the subscript 'c' indicates that the parameters (SE(C/CV/P) and RSE(C/CV/P)) have been corrected for bias.

The accuracy and robustness of the PLS models are discussed according to the following parameters, all corrected for bias value:

$$\text{coefficient of variation, } CV_c\,(\%) = SEP_c/\text{mean}$$

$$\text{ratio of performance to deviation, } RPD_c = SD/SEP_c$$

where SD is the standard deviation;

$$\text{ratio of } SEP_c \text{ to reference data range, } RER_c = (y_{max} - y_{min})/SEP_c$$

where y_{max} and y_{min} are the maximum and minimum reference values of EOC respectively;

$$\text{ratio of } SEP_c \text{ to interquartile,}^{21}\ RPIQ_c = (Q_3 - Q_1)/SEP_c$$

where Q_3 and Q_1 are the values of the third and first quartiles of reference data respectively.

2 - Electrophysiological assessment of plant a Faraday cage using supervised machine learning -

Application to tomato growth in greenhouse

PISA-Project - …………………………………………………………. *InnoSuisse (endding)*

From development to implementation of a ML-based model

Yield-Project - …………………………………………………………. *OFAG (start 2020)*

Implementation of an autonomous system using AI

Big Data :

→ 24h recording is about $4.3*10^{7}$ data points /plant

→ 1 Month recording is $1.3*10^{9}$ data points / plant

→ 1 Month recording on 8 plants (8 channels Device) is $1.1*10^{10}$ data points

Agroscope

PISA-Project

Digitalization approaches

- Plant level
  - Electrophysiology sensors development
  - Data modelling
    - Whole data – features extraction
    - Spectrogram image analyses

Electrophysiology sensors development

Device

- **M**ono-channel prototype
- **M**ulti-channels prototype

A recorded signal on plant?

Software

Data

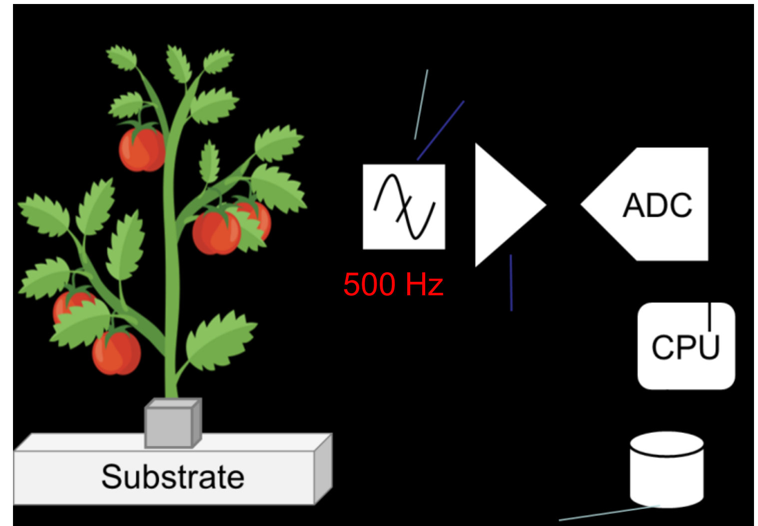
- **W**hole data – features extraction (2D-line vectors)
- **S**pectrogram image analyses (3D-images)

Indexing data by experts

Modelling machine learning approaches

Calibration/cross-validation

test sets on independant plants



PhytlSign Device

Schematic representation of the PhytlSigns composed of an amplifier-voltmeter. Digitized data are logged on a Raspberry Pi.

Enabling electrophysiological recordings in greenhouse

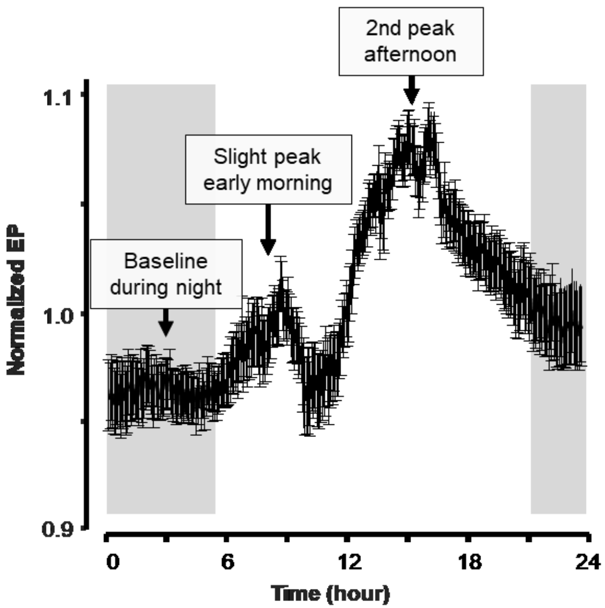
Experiments are performed on hydroponic tomatoes grown in greenhouse. The PhytlSigns device allows monitoring of electric signal in 'real' environment without Faraday cage. Electrode is inserted in the tomato petiole at the top of the plant (*bottom*).

Electrophysiology sensors development

A recorded signal on plant?

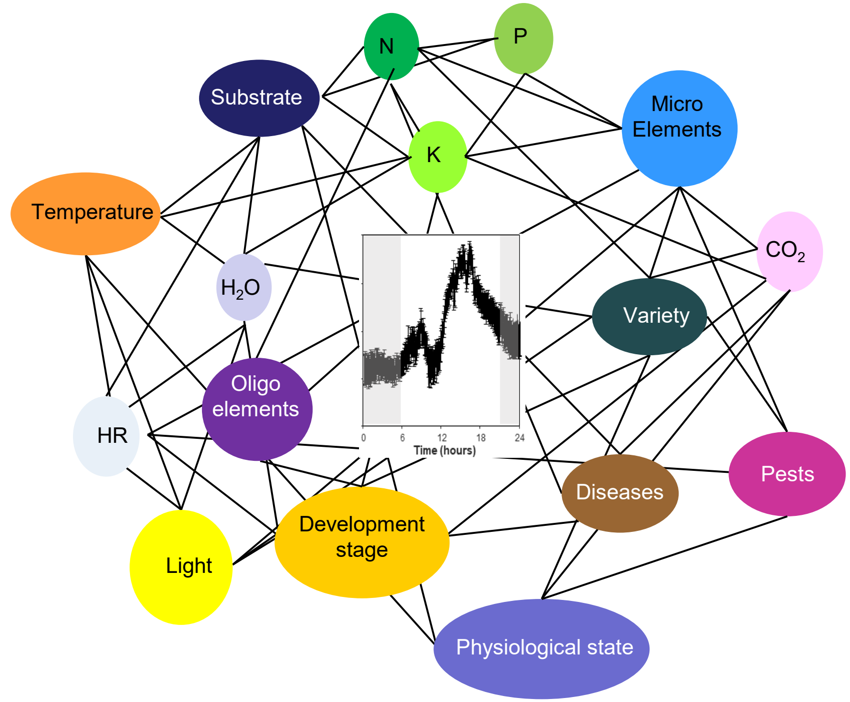
Whole data – features extraction (2D-line vectors)

Indexing data by experts



Electrical signal

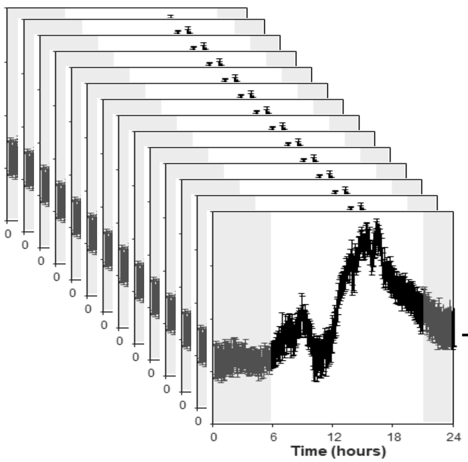
Daily electrical potential (EP) variations in a tomato plant growing in soilless conditions.



Electrophysiology sensors development

Whole data – features extraction (2D-line vectors)

Indexing data by experts



Library

Collect data to create a library of electrical signal records. A given 24h long signal is composed of about 4.3×10^{7} data points.

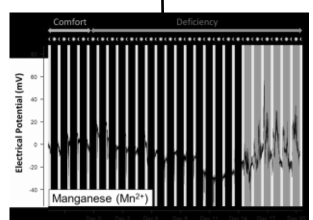
Matrix [*n* x *p*]

n = time duration of the records

p = number of plant

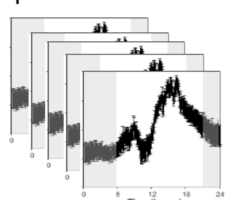
Extract features from signals

- Different features (*p* = *238*)
- Different modalities for extractions



Increase library

By adding new records from new experiments and so on.



Annotations of each recorded signal

- Days vs. Night
- Confort vs. Water stressed plants
- Confort vs. nutritional deficiency plants
- Confort vs. Spider mites related stress

Machine learning

Machine learning applied on features.
Calibration/cross-validation.

Model prediction

Machine learning applied on features.
Test with external plant records.

Water stress

The **G**eneralized **B**oosted **T**ree method allowed to classify **95%** of plants according to their water status (Comfort or stressed)

Daniel Tran, Fabien Dutoit, Elena Najdenovska, Nigel Wallbridge, Carrol Plummer, Marco Mazza, Laura Elena Raileanu and Cédric Camps. Electrophysiological assessment of plant status outside a Faraday cage using supervised machine learning. ***Scientific Reports***, 2019.

Models	LR	DL	DT	RF	GBT
Accuracy (%)	73.2	83.5	62.0	61.4	94.6
Precision (%)	75.9	87.4	61.4	61.0	95.4
Recall (%)	81.2	84.8	99.6	99.8	95.6

Agroscope

Light / Night cycles

The **G**eneralized **B**oosted **T**ree method allowed to classify **98%** of plants according to their reaction to the presence or absence of light at a level of 100µmol.m^{-2}.s^{-1}.

Elena Najdenovska, Fabien Dutoit, Daniel Tran, Carrol Plummer, Nigel Wallbridge, Marco Mazza, Cédric Camps and Laura Elena Raileanu. Insights of plant electrophysiology – Using signal processing techniques and machine learning algorithms to associate tomatoes reaction to external stimuli. ***31st Conference of the International Biometric Society of the Austro-Swiss Region*** *https://wp.unil.ch/ibs-roes2019/, Lausanne, 9th-12th of September 2019.*

Stimuli	Accuracy	STD	Precision	Recall
Light (100µmol.m^{-2}.s^{-1})	97.3	0.0007	98.0	97.0
Water Deficit	97.4	0.0032	97.0	98.0

Agroscope

Spider Mites Detection

Features subset	# features	Model accuracy (algorhitm: **GBT**)	Prediction rate on unseen data		
			Control (B ch1)	Infested (F ch8)	Average
All features	238	59+/-0.8%	96%	79%	88%
Feature Selection (FS)	158	56+/-0.8%	99%	71%	86%
Non-Correlated (<0.95)	130	72+/-1.6%	96%	80%	88%
FS (Non-Correlated)	70	70+/-0.8%	97%	77%	87%
Non-Correlated + Target-correlated (>0.01)	128	72+/-1.9%	94%	82%	88%
FS (Non-Correlated + Target-correlated)	97	69+/-1.1%	90%	73%	81%

External validation with the electrical signal of <u>non-infested</u> tomato plant.

External validation with the electrical signal of <u>infested</u> tomato plant.

Agroscope

Next Modelling Approach

Live Data

Recorded Data

Preprocessing
Filters, windows, amplify features, remove noise

Convert to Image
STFT
Develop algorithm / code or …

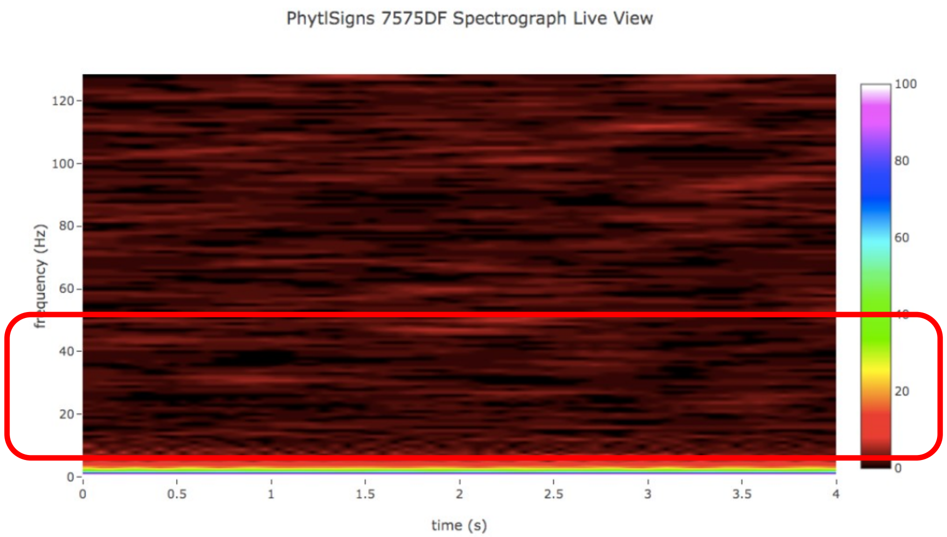
Create Model
STFT
Inputs, layers, nodes, outputs, loss function

Train Model
Raw Data
Labels
Training metric

Run Model to detect Stimuli
Prediction of probability that stimuli is present

Convert to image (STFT)

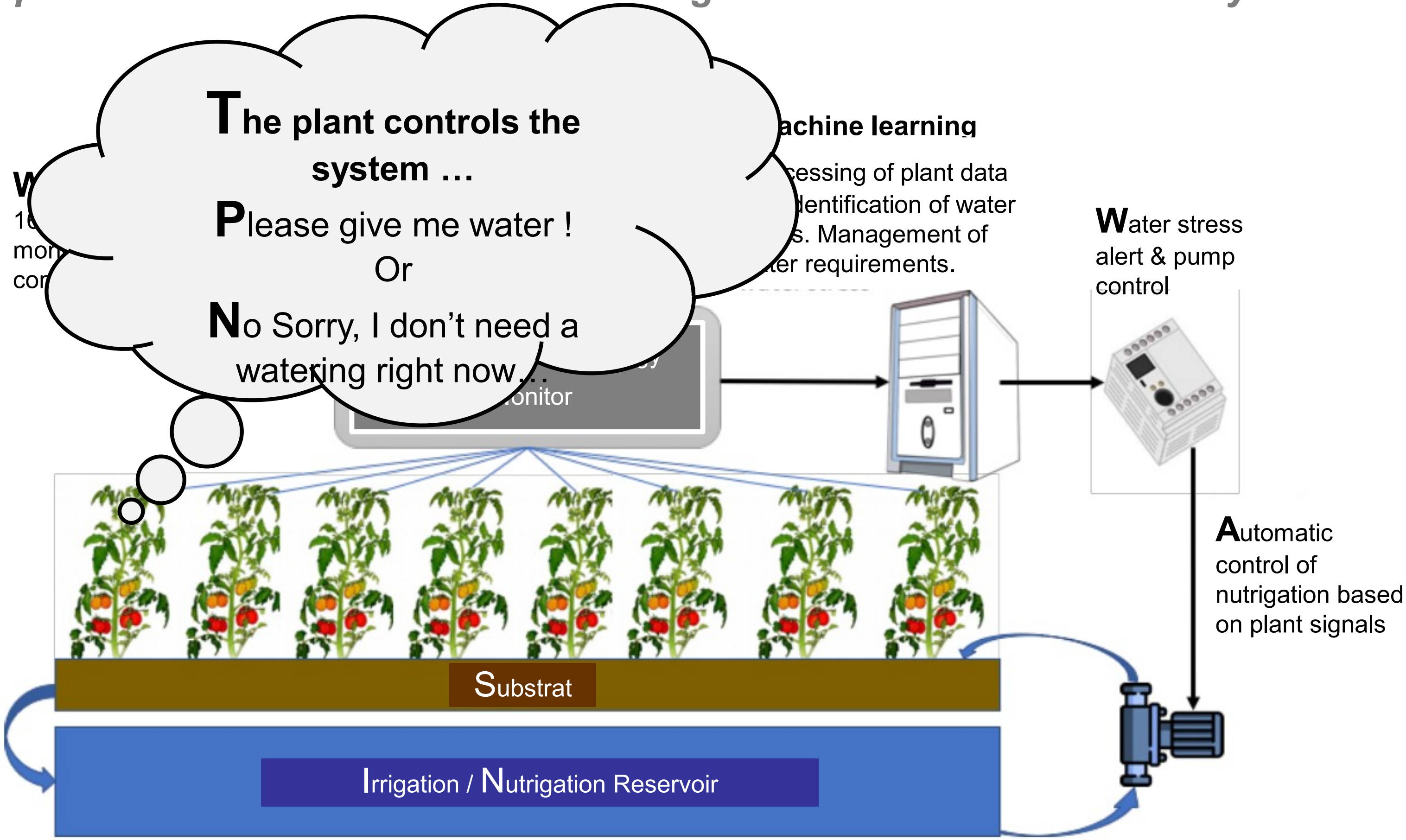
- Short time Fourier transform
- Power vs. frequency graph
- Colours show spectral power (amplitude of the signal at specific frequencies).



YIELD – Project

Yield Improvement using Electrophysiology Device

Implementation of the Machine learning model in an autonomous system



Agroscope







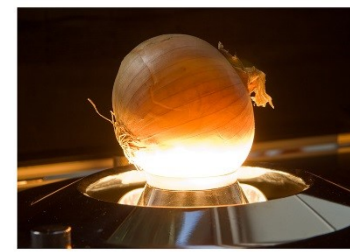





Merci pour votre attention

Prénom Nom

prenom.nom@agroscope.admin.ch

Agroscope une bonne alimentation, un environnement sain

www.agroscope.admin.ch