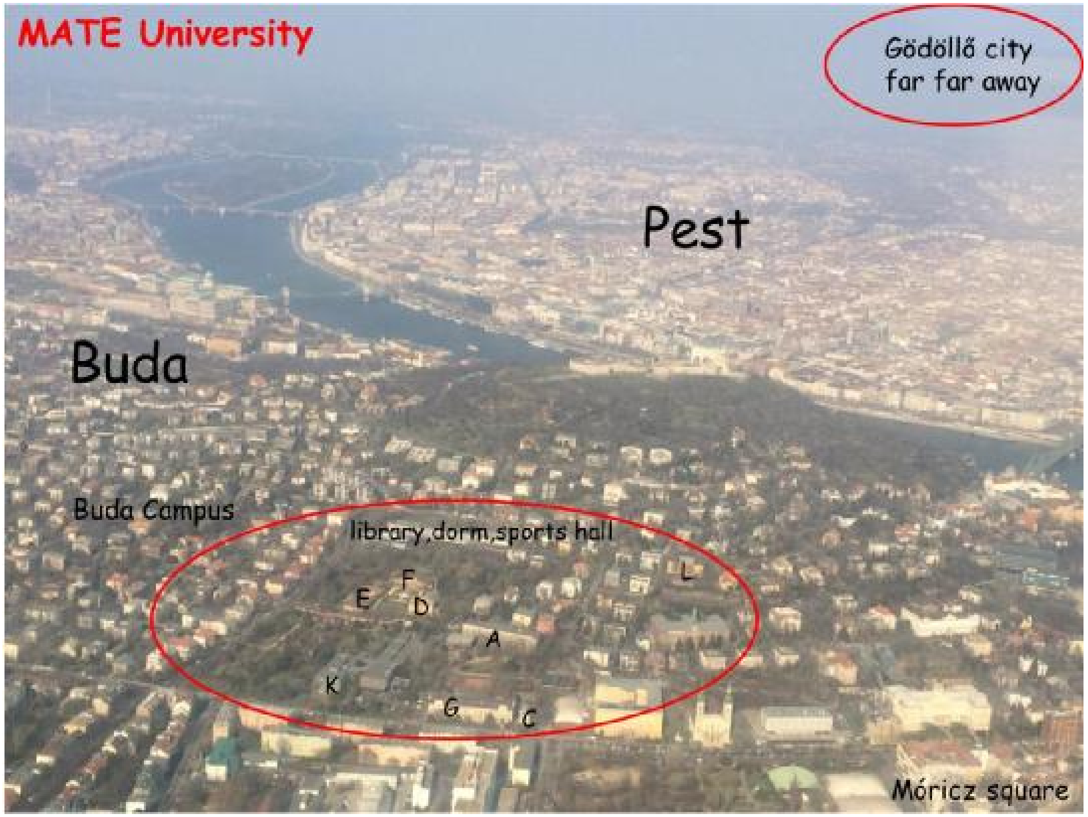
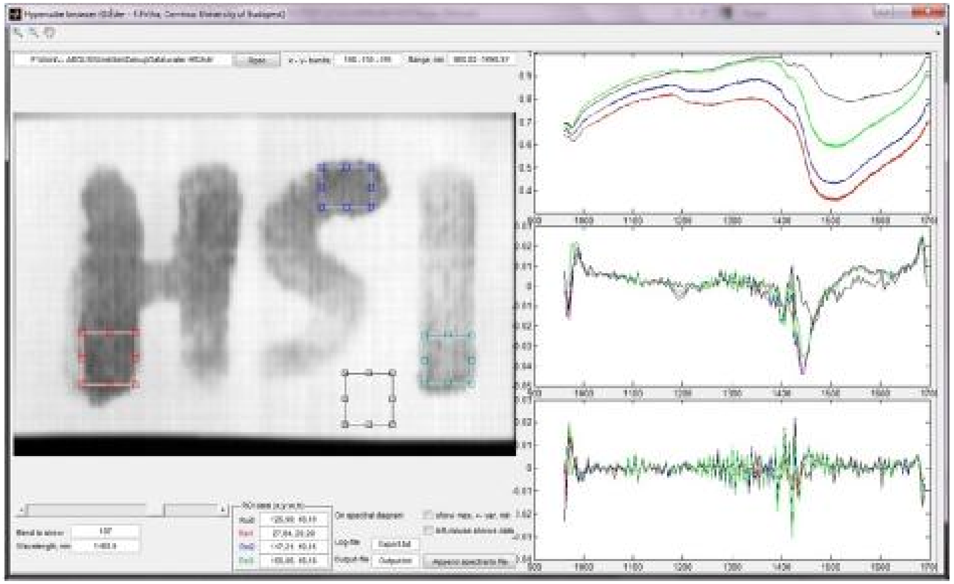

MATE University
Gödöllő city
far far away
Pest
Buda
Buda Campus
library,dorm,sports hall
F
L
E
D
A
K
G
C
Móricz square

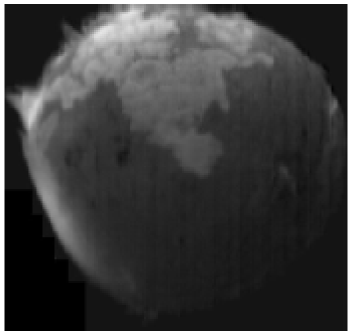
# Early detection of fungal infection by hyperspectral imaging based on secondary effects

MATE, Buda Campus, Dept. of Measurement and Control in Food Procesing
Ferenc Firtha PhD



**What is spectral imaging? Seeing the invisible:**
water fingerwriting becomes visible on 1450 nm image

RGB image



1456nm image

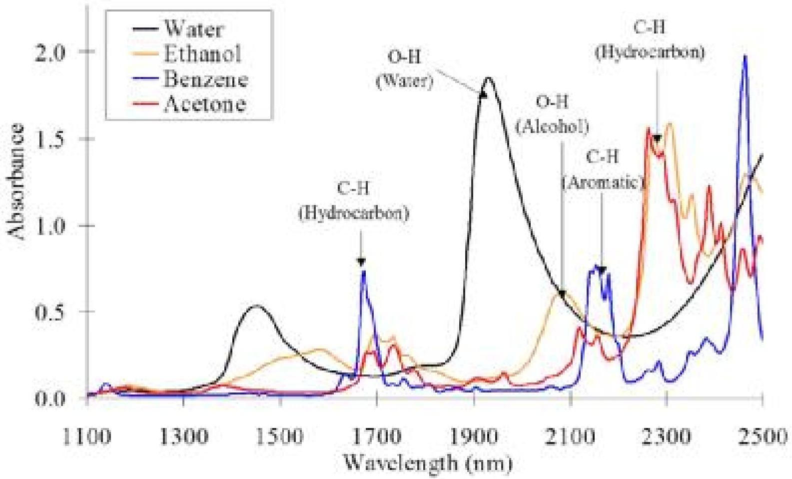
## Colorimetry: what like?

quick, but contact :
-> average RGB/Lab/Lch

## Image processing: where? shape and pattern?


remote sensing + data reduction:
-> position (segmentation), colour, shape, pattern

## Spectroscopy: what from?



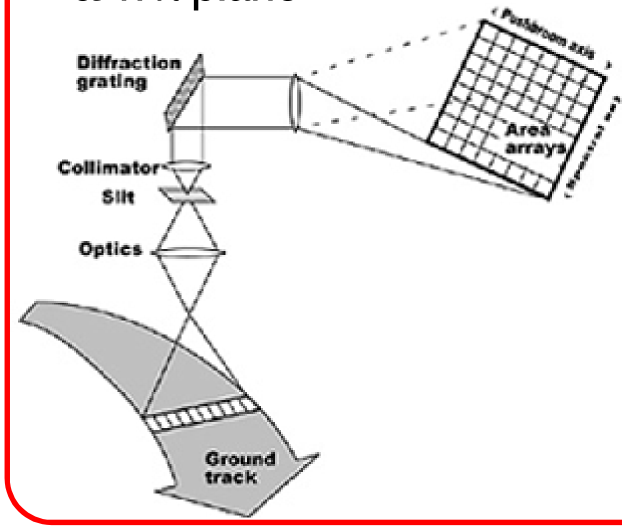

Contact method + statistical analysis:
NIR -> water, fat, oil, protein,...

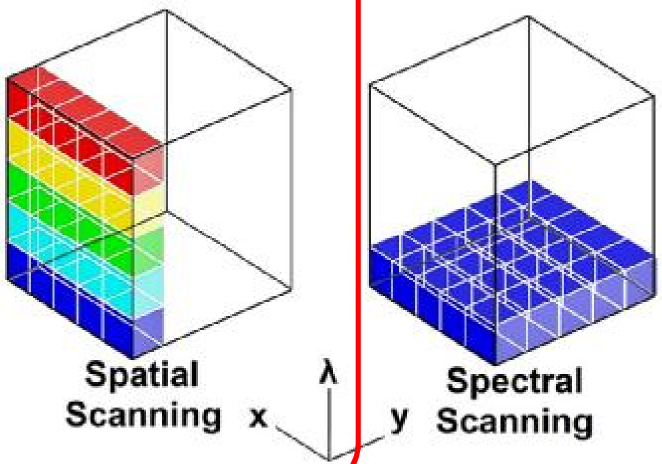
## Spectral imaging: where and what from?

remote + stat. analysis + image processing
-> position, distribution of compounds

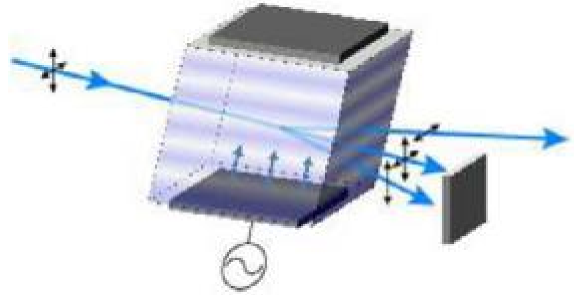
# Hypercube: spectra of each pixels ß several data acquisition methods

1. push-broom (előretoló módszer)
   à x-λ plane



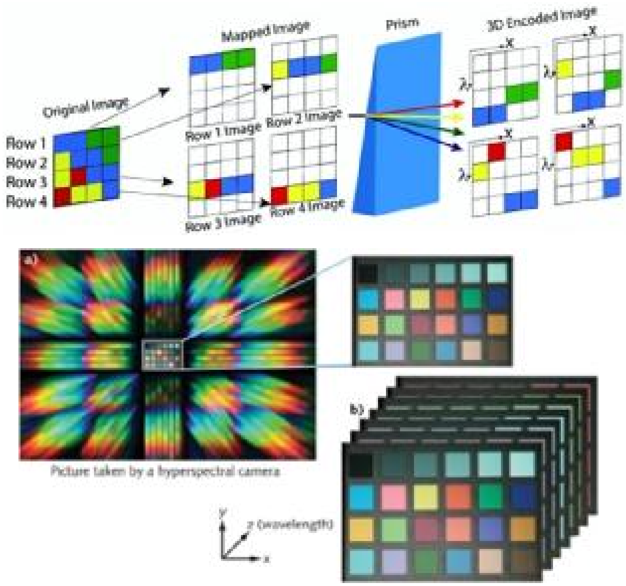



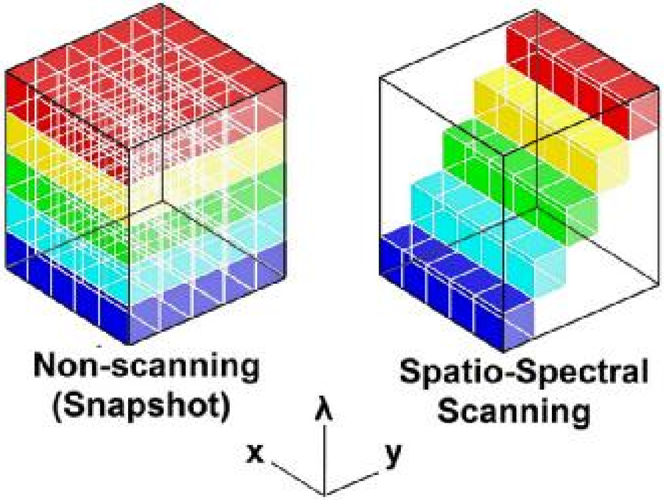

2. tunable filter (hangolható szűrő)
   à x-y plane



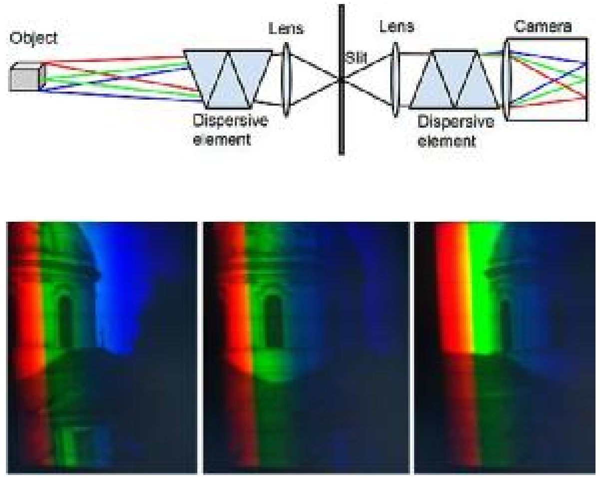
*3. Snapshot:*

*h*w* sensor -> *x*y*λ* hypercube resolution




4. Spatio-spectral scanning



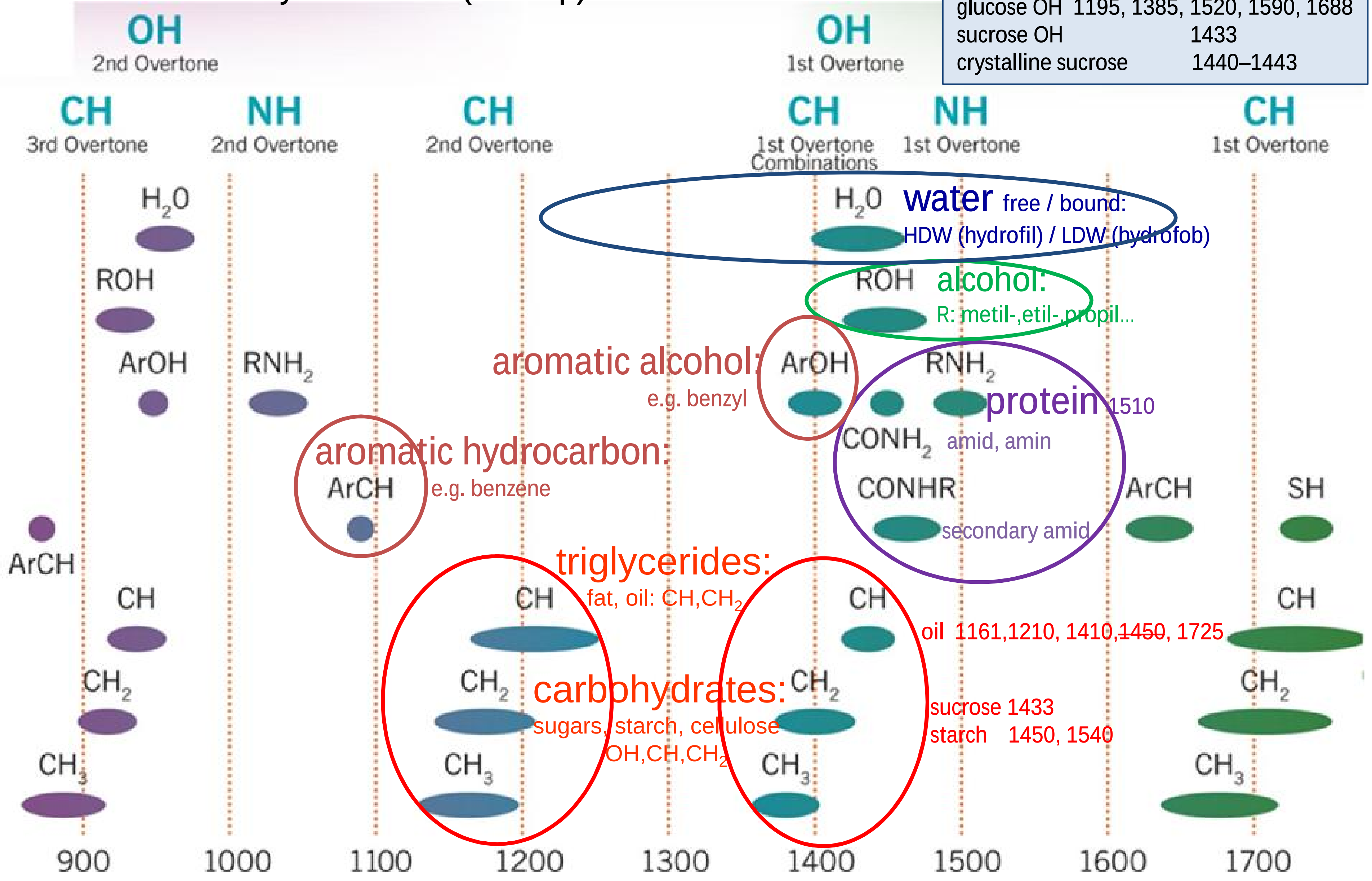
# What to measure in NIR 900–1700?

*broad* absorption peaks of vibrations of
*O-H, C-H, N-H* chemical bonds
affected also by the *matrix* (overlap)

| | |
|---|---|
| OH bond: | 970, 1450, 1980 |

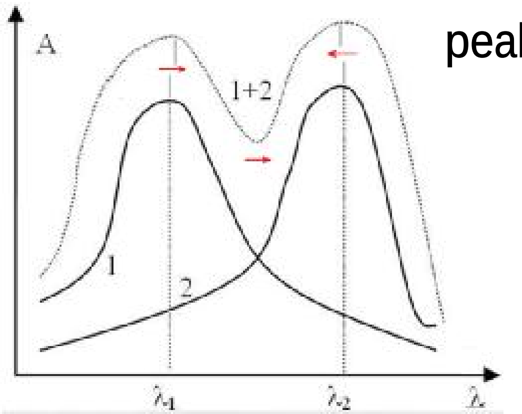
| | |
|---|---|
| Fiber: | 1100, 1300, 1350, 1403, 1483, 1500, 1534 |
| Cellulose (green plants): | 1490 |
| Lignin (wood): | 1170, 1410, 1417, 1420, 1440 |

| | |
|---|---|
| glucose OH | 1195, 1385, 1520, 1590, 1688 |
| sucrose OH | 1433 |
| crystalline sucrose | 1440–1443 |



## Problem of assignation (peak identification):

broad absorption peaks cause overlapping




peaks are shifted

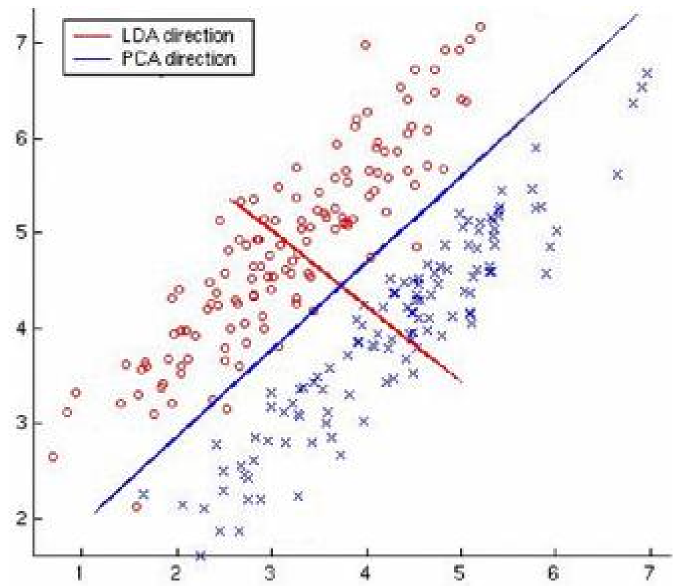
component interacts with matrix

## Evaluation by chemometry:

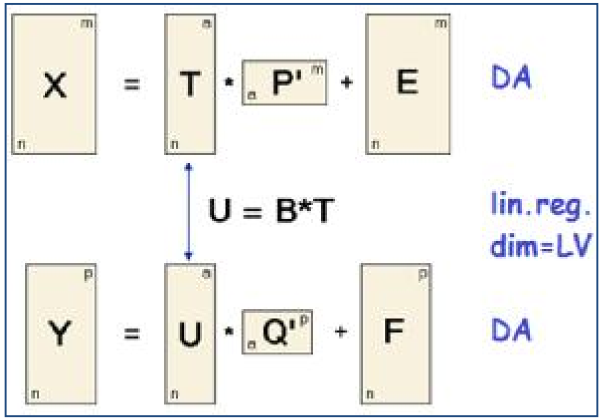
sample, X independent, Y dependent vars

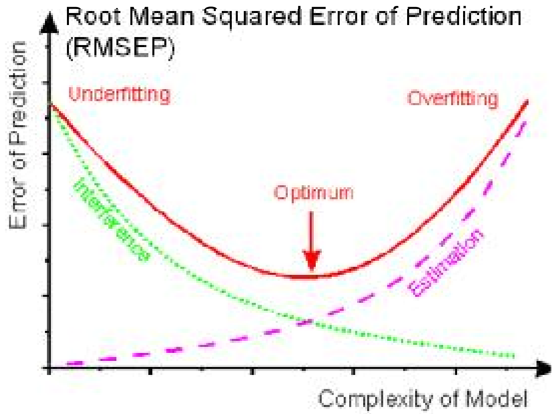
| | A | B | C | D | E | EZ | FA | FB | FC |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Sample | Independant: Absorbance on 155 spectral bands | | | | | | | Dependant |
| 2 | # | 953,95 | 958,78 | 963,61 | | 1692,66 | 1697,48 | 1702,31 | Fat measured |
| 3 | 1 | 0,003637 | 0,001893 | 0,001109 | ... | 0,020039 | 0,022956 | 0,027763 | 0,02 |
| 4 | 2 | 0,260903 | 0,259142 | 0,252833 | ... | 0,492475 | 0,521888 | 0,542394 | 0,23 |
| 5 | 3 | 0,040375 | 0,033535 | 0,029435 | ... | 0,143669 | 0,135637 | 0,117872 | 0,78 |
| 6 | 4 | 0,011148 | 0,005165 | 0,001917 | ... | 0,013151 | 0,017277 | 0,019343 | 0,32 |
| 7 | 5 | 0,303865 | 0,305647 | 0,299407 | ... | 0,461204 | 0,48644 | 0,499642 | 0,45 |
| 8 | 6 | 0,050206 | 0,043487 | 0,038126 | ... | 0,145994 | 0,148761 | 0,150271 | 0,76 |
| 9 | 7 | 0,139376 | 0,132027 | 0,123925 | ... | 0,155064 | 0,150772 | 0,141379 | 0,12 |
| 10 | 8 | 0,026476 | 0,019536 | 0,014589 | ... | 0,061249 | 0,063689 | 0,066535 | 0,87 |
| 11 | 9 | 0,205776 | 0,209365 | 0,202549 | ... | 0,381126 | 0,400397 | 0,413213 | 0,02 |
| 12 | 10 | 0,143362 | 0,141262 | 0,133336 | ... | 0,199697 | 0,201683 | 0,202408 | 0,23 |
| 13 | 11 | 0,099449 | 0,093974 | 0,085127 | ... | 0,009464 | 0,00876 | 0,008829 | 0,78 |
| 14 | 12 | 0,064664 | 0,057447 | 0,050311 | ... | 0,091355 | 0,09299 | 0,091857 | 0,32 |
| 15 | 13 | 0,039362 | 0,036098 | 0,032012 | ... | 0,151284 | 0,150967 | 0,148637 | 0,45 |
| 16 | 14 | 0,015409 | 0,010616 | 0,006445 | ... | 0,061 | 0,059842 | 0,059355 | 0,76 |
| 17 | 15 | 0,143221 | 0,141403 | 0,134046 | ... | 0,249926 | 0,264618 | 0,272407 | 0,12 |
| 18 | 16 | 0,112132 | 0,109324 | 0,102869 | ... | 0,164316 | 0,175072 | 0,183038 | 0,87 |

categorical Y: DA for classification




continuous Y: PLS for prediction

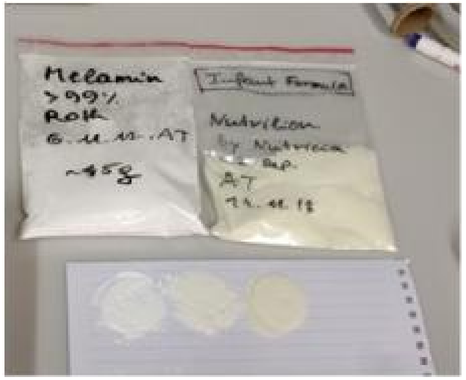


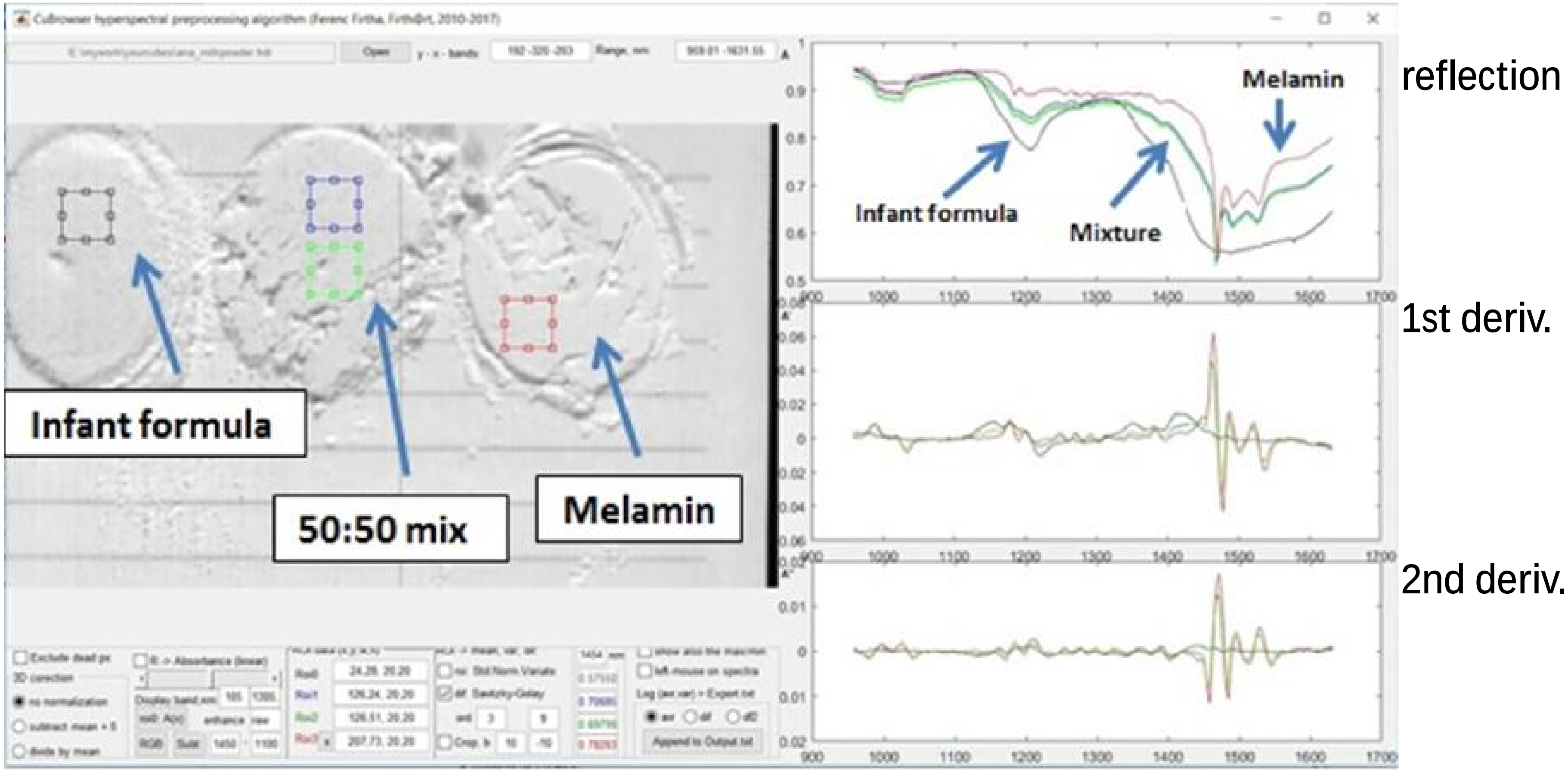

number of latent variable

## Applications1: simple NIR usage

Detection of melamine in infant formula (chineese milk scandal, 2008)
*experiment in Boku, (Firtha, 2017)*




Samples:
1. Infant formula (Nutrilon, Pronutra)
2. Melamine (Roth, p.a. 99%)
3. 50:50 mixture of Infant formula and Melamine

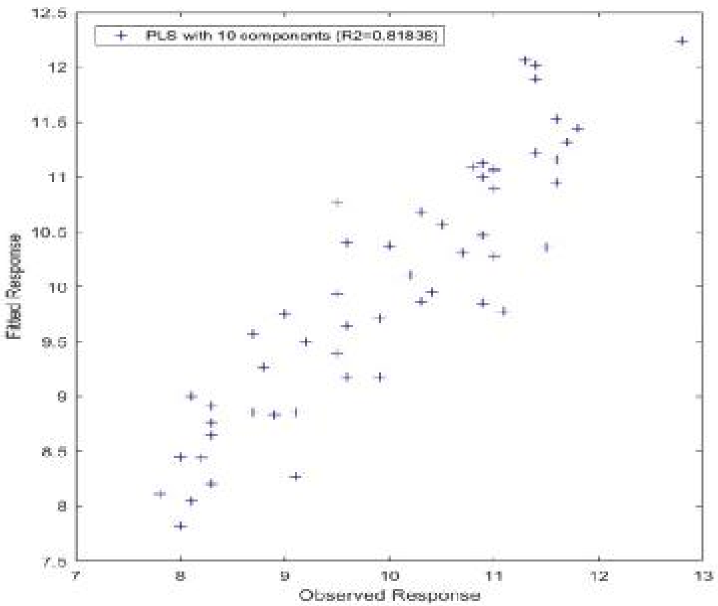



Advantage: Remote sensing

# Applications2: non-homogeneous distribution of inspected property

Brix and pH in watermelon *(Firtha, Kaszab, 2013)*

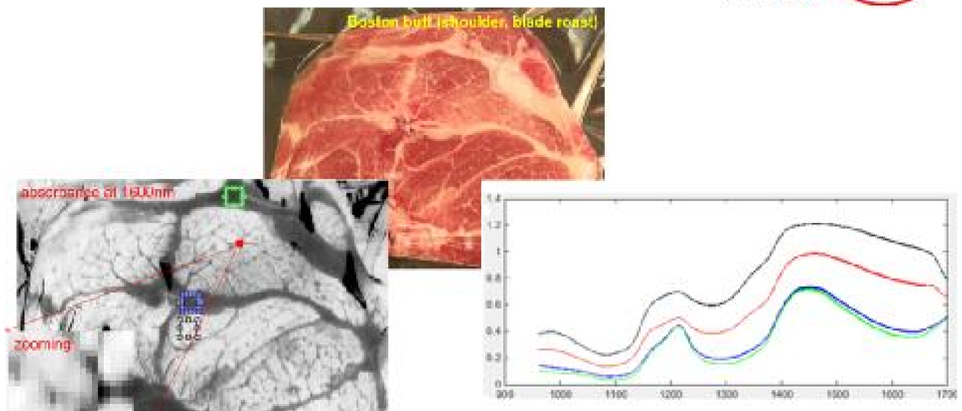
PLS for Brix: $R^2$=0.82




Meat marbling *(Firtha, Friedrich, Romvári 2013)*

How to make difference between

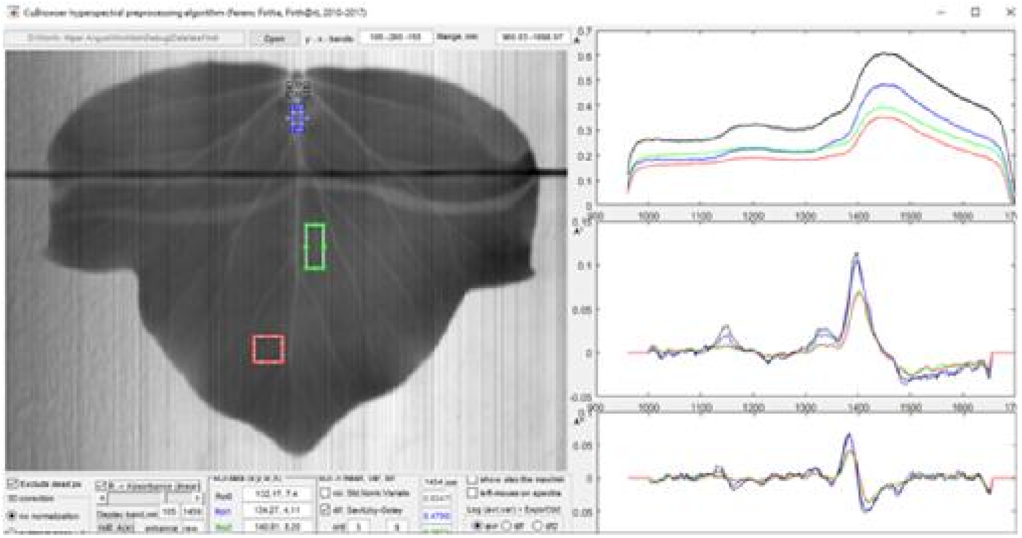
- lean meat,
- intermuscular fat deposite and
- intramuscular connective tissue (IMCT)?


spectra of lean - IMCT - fat

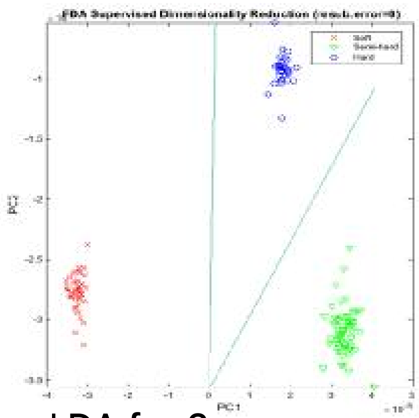
Additional applications:

Water in the vein of a leaf (*Firtha*)

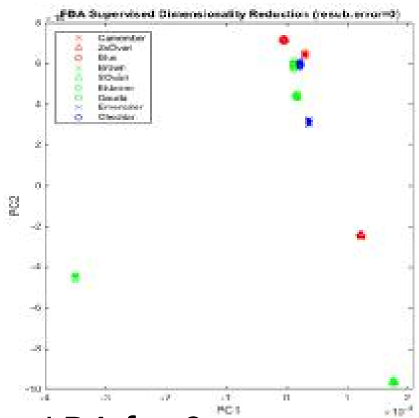


Classification of different cheeses. Checking effect of storage temperature *(Králik, Firtha)*

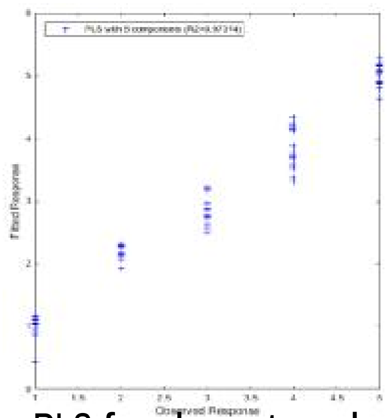
Fat in cheese *(Darnay, Firtha)*



LDA for 3 groups



LDA for 9 types



PLS for days stored

Spectral Imaging is a novel, remote sensing method for estimation of components

But it usually cannot detect trace elements

# Material and Method

## Sample experiment: Early detection of fungal infection of white button mushroom (*Viktória Parrag, 2013*)

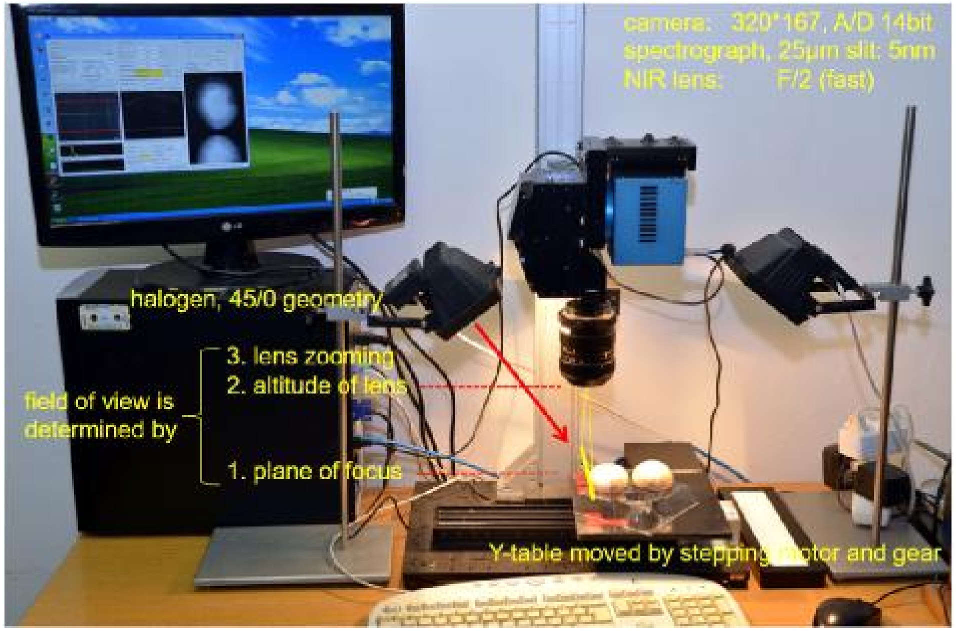
days: 1st 4th 7th 11th

Two sets of champignon (*Agaricus bisporus*) were inspected:

- Infected by cobweb (*Dactylium dendroides*) parasite fungi
- Control group

Hardware: Headwall push-broom HSI system:

- Xeneth InGaAs senzor: 900-1700nm range
  14bit quantization (16’384 level)
  320*256 (x*band) resolution
  à 5nm spectral resolution
- Specim spectrograph (splits light into comp)
- Canon NIR Lens, F/2.0 (fast), FL25mm
- Y-table: moves object to inspect it line by line

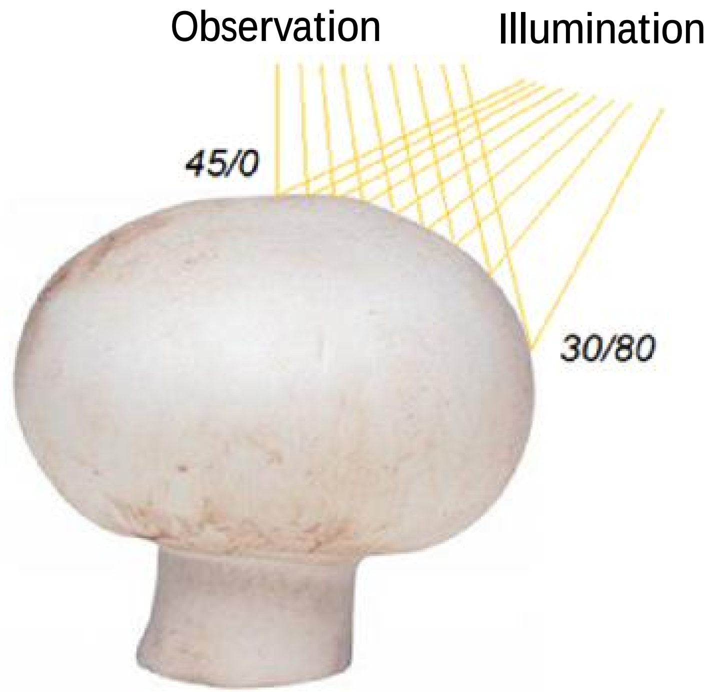



Software:

- Argus: for controling measurement *(Firtha)*
- CuBrowser: for preprocessing
- PLS and SVM (Support Vector Machine): for classification

## Problem 1: Non-homogeneous illumination geometry

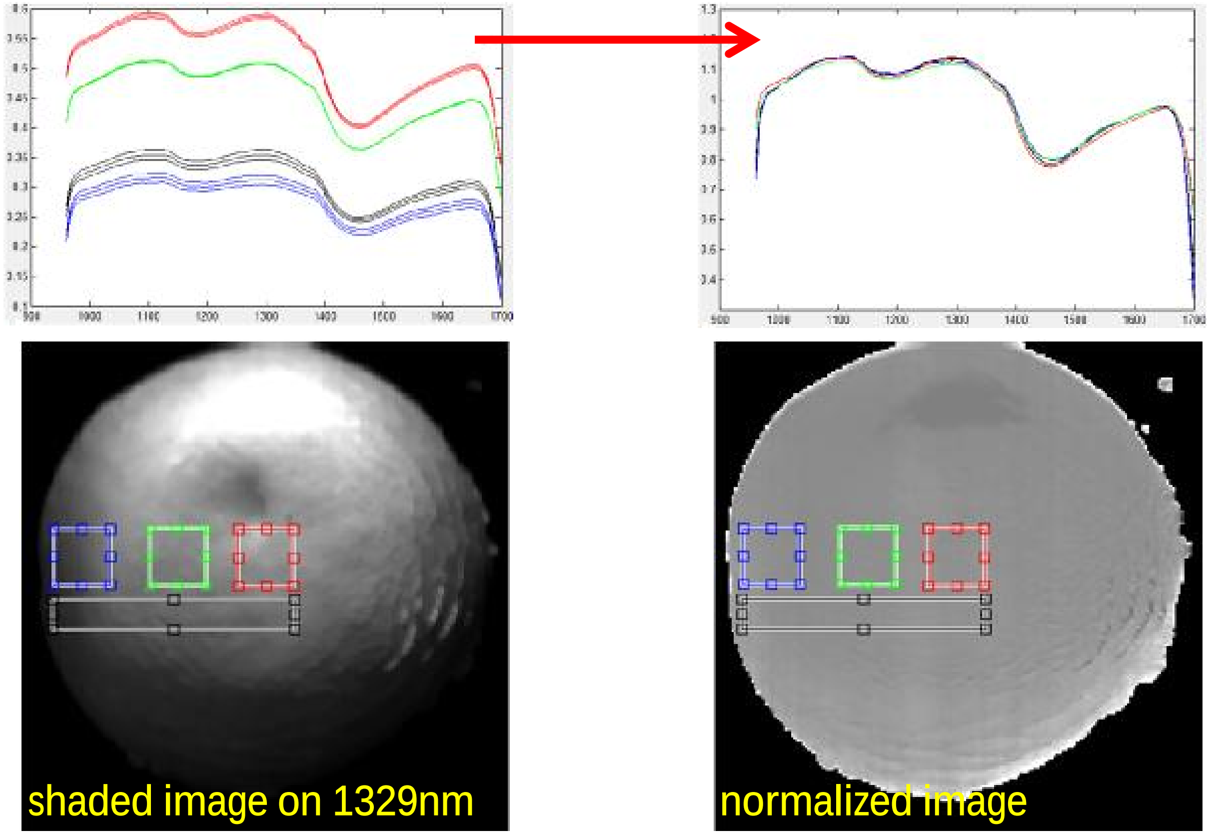
Geometry: Illumination- / Observation angle
Usual: 45/0




## Solution: normalization of ROI's spectra

- Shifting to same level by subtraction s-mean
- Stretching by dividing by mean s/mean
- SNV (Standard Normal Variates): center & scale (s-mean)/stdev

Optimal method depends on the object type



Non-homogeneous illumination, Normalization:

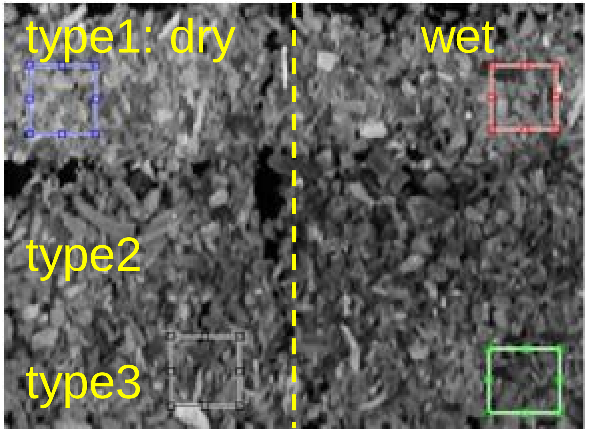
From a disadvantage to an advantage

| NIR spectrometer | Spectral imaging |
|---|---|
| Isolated from outer noises ✓ | Remote: signal is loaded with noises |
| But cannot handle uneven surface | Geometry is handled by normalization ✓ |

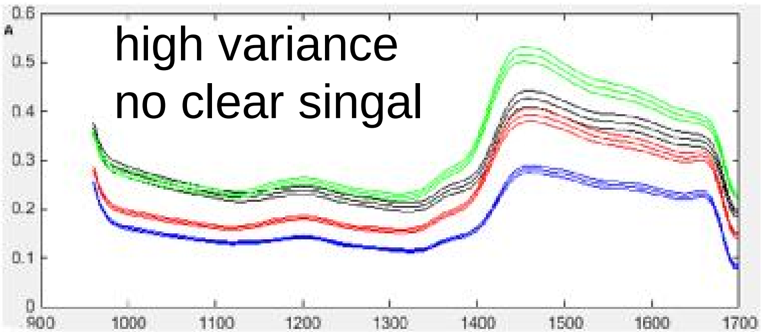
Sample application: Moisture in tea leaves *(Firtha, 2013)*

Right half were in humid air for 1 hour
(only 1 drop water in the chamber)

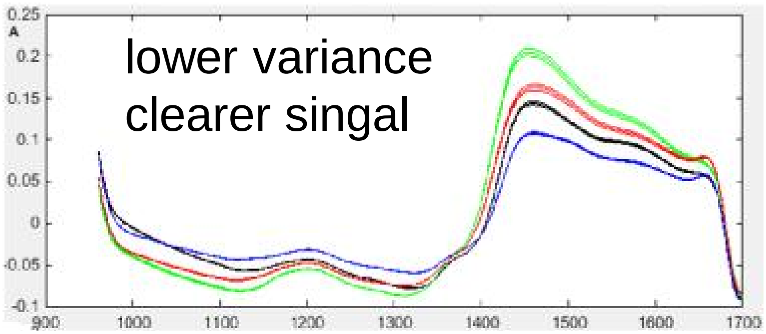
ROIs:
Dry: Black, Blue
Wet: Green, Red




Without normalization:



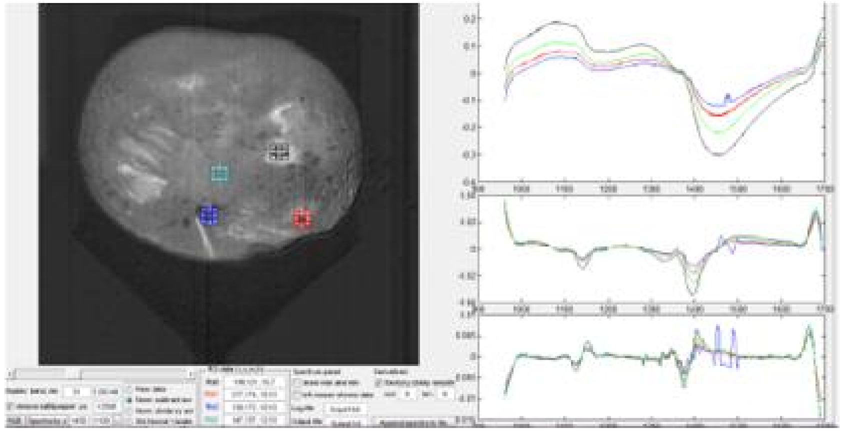

With normalization:



Back to „Early detection of fungal infection” experiment

ROIs were selected on infected & control samples every three days using preprocessing methods

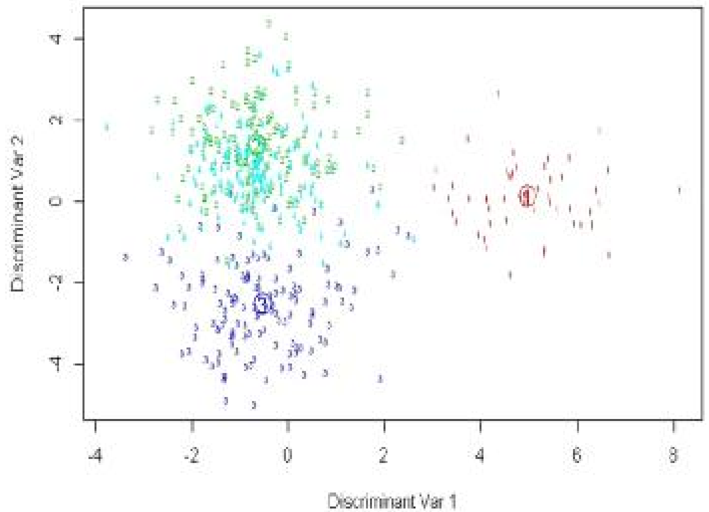
- Normalization
- Exclude extreem pixels (kind of median alg.)
- Savitzky-Golay: smooth, 1st and 2nd derivatives
- Crop noisy edge of 900-1700nm range
- Reflectance – Absorbance conversion



CuBrowser *(Firtha)*

DA and SVM models were built for classification by

Pretreatment




1: untreated (79,7%), 2: natamycin, 3: prochloraz-Mn,
4: Bacillus subtilis -treated

Infection

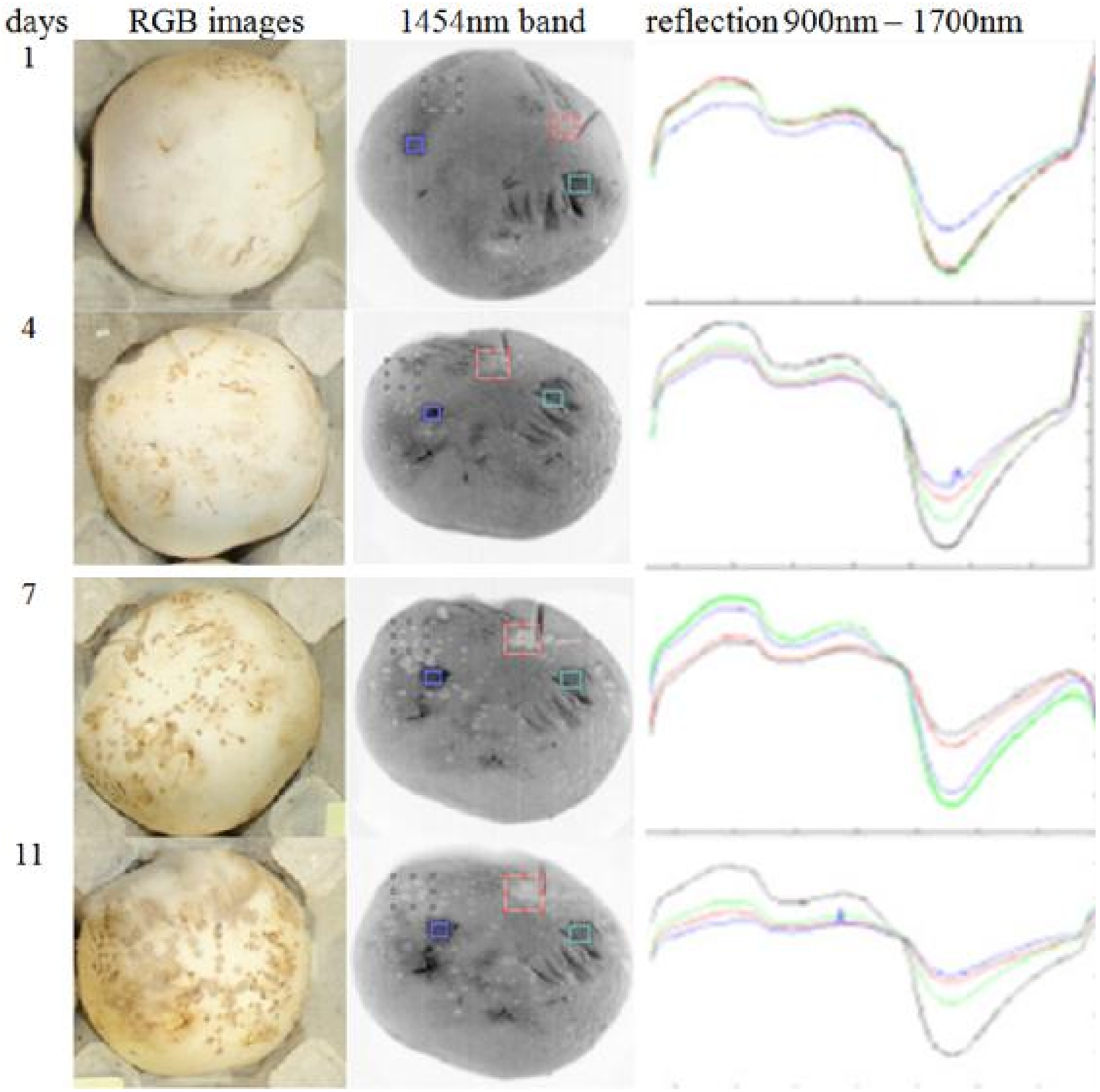
Results of SVM classification

| Not treated | | | Natamycin | |
|---|---|---|---|---|
| | true | | true | |
| predicted | dactylium | control | dactylium | control |
| dactyilum | 100.00% | 0.00% | 60.13% | 5.60% |
| control | 0.00% | 100.00% | 39.87% | 94.40% |
| Correct classification | 100% | | Correct classification | 80,78% |
| **Prochloraz-Mn** | | | ***Bacillus subtilis*** | |
| | true | | true | |
| predicted | dactylium | control | dactylium | control |
| dactyilum | 95.42% | 16.96% | 88.89% | 4.24% |
| control | 4.58% | 83.04% | 11.11% | 95.76% |
| Correct classification | 89,71% | | Correct classification | 93,16% |

groups are well separated.
a beautiful result, like a fairy tale

Problem 2:

- The spectra of later measurement days were also used for calibration.
  The statistical model is probably not enough for „Early detection"
- These models can be easily overfitted, because of the huge number of independent variables (spectral bands). They might not work on unknown sample set.




In next data evaluation method:

3 types of ROI were selected on infected samples

1. infected (not visible in early stage): RED
2. not infected
3. mechanically injured

The reflection spectra at 1450nm shows, that the infected areas started to dry at the very first day.
The reflectance at 1450nm was growing there.

Not the absolute value of the reflectance delivers information, but its distribution on the surface.

Further investigation proved that the difference between two wavelength images highlights best the infected islands: 1450 nm – 1080 nm

Dactilium infected spots at the very first day on 1450nm – 1080nm

No major change in chemical components, but the secondary effect of disease is detectable.

Human eyes cannot see, but measuring at only the significant, mochromatic wavelengthes, an image processing algorithm can detect the rounded areas on the surface.

1. significant wavelength should be assigned first
2. areas segmented by image processing method
3. spot shape and spectral differences together will identify the infection

# Result, more examples

This result showed the fundamental difference in NIR & HSI methods, for case when only a secondary effect is detectable in the spatial distribution of a spectral feature

NIR measurement in laboratory (X, Y)

Building statistical model (calibration)

NIR measurement in industry (X)

Using statistical model

Classification / estimation of a property

HSI measurement in laboratory

Segmentation of ROIs

Building statistical model (calibration)

Assignation of significant wavelengths

Multispectral measurement in industry

Image processing

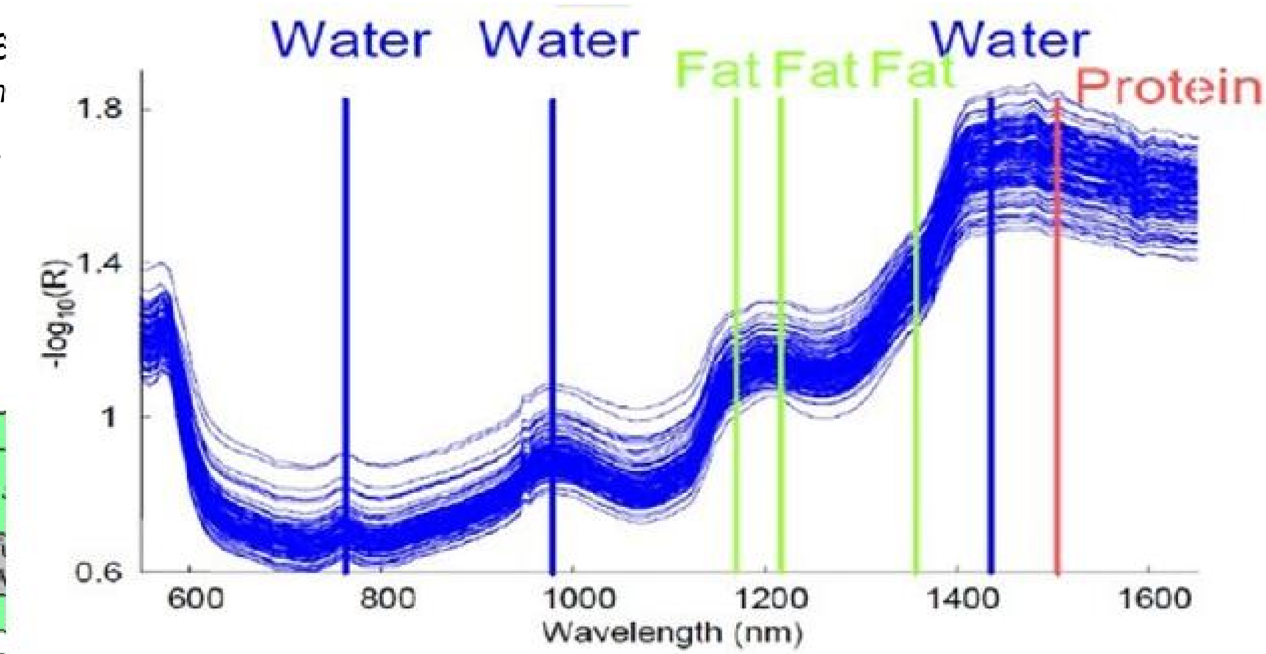
Description of property

Additional examples of secondary effect:
Whether the moisture content can be predicted in VIS range?

Spectra of beef show very weak signal in VIS (980nm)

*Spectral and Hyperspectral Inspection of Beef Ageing State (Firtha, Jasper Friedrich, 2011)*

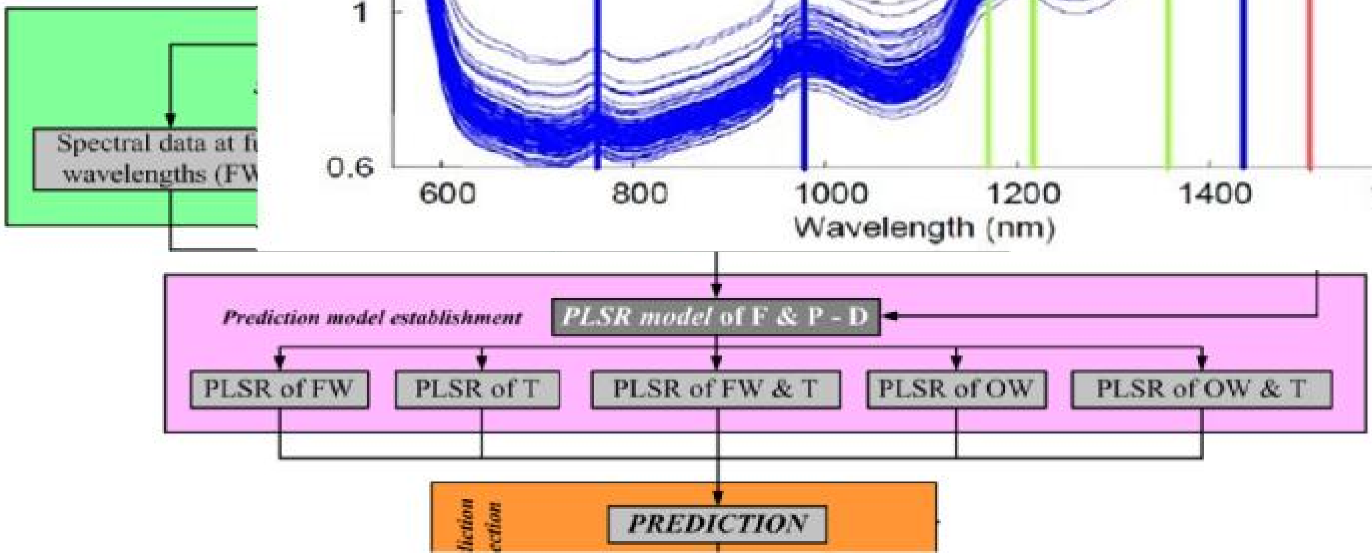
Fig.1: General spectra of beef

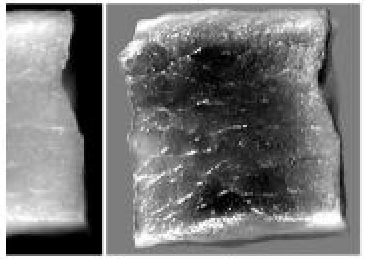



Spectral data ... ties

*Prediction of MC in ... Wen, Hongbin, 2015 for rev)*

Fig.1: Flowchart ... two PC images



Grayscale props:

- Entropy
- Energy
- Contrast
- Homogeneity
- Dissimilarity

Sophisticated, but this method mostly detects a secondary effect:
The texture depends on moisture content & probably many other unknown factors (*Firtha*)
This Da-Wen Sun article was published in 2017, then several similar articles were born referencing each other.

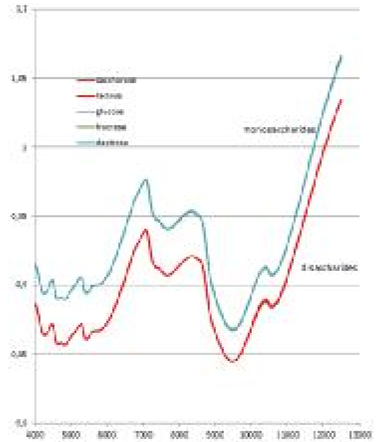
# Example 3: Whether fructose can measured in marcipane?

(*Szabina Németh – Katalin Kerti – Firtha, 2012*)

In marzipan invertase converts sucrose into glucose and fructose.
The mixture of sucrose, glucose and fructose has a lower viscosity and shows less tendency to crystallize than sucrose alone. The product stays softer.

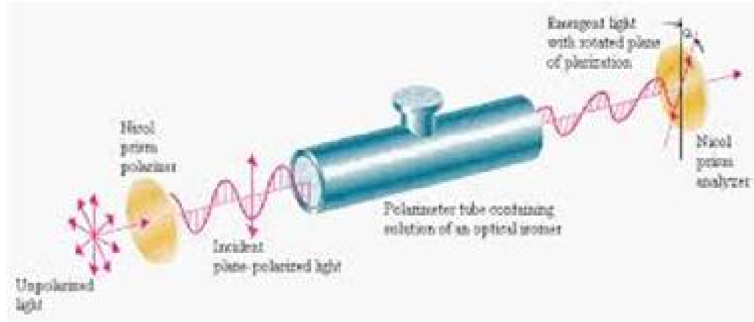
How to measure fructose content?

a.) no difference in NIR



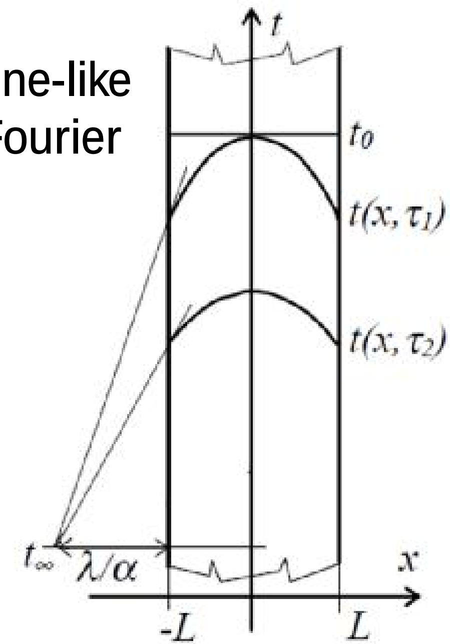
b.) polarimetry handles only 1 comp.
each has its $\alpha_D$ specific rotation

$$a = a_D \cdot c \cdot l$$



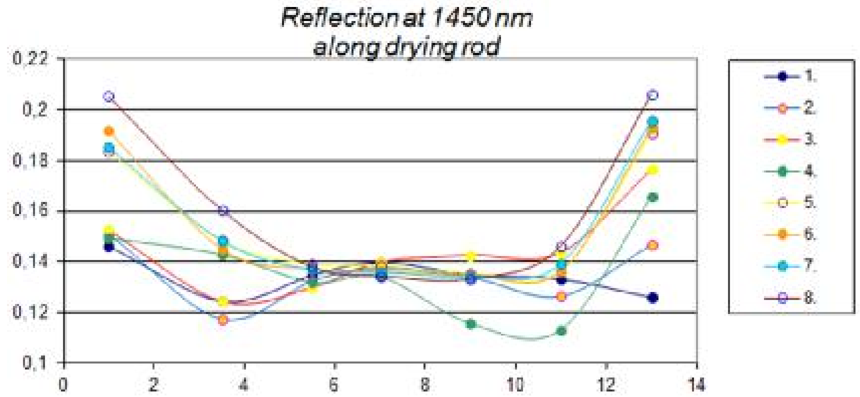

c.) by moisture distribution of drying rod

For normal material: cosine-like like cooling model by Fourier




In case of hygroscopic material, like fructose: constant in the middle and changing on the edges only

Measureable by: penetrometer, impedance, SI




Spatial distribution of MC is measured instead of fructose

# Conclusion

In spectral imaging sometimes we cannot directly measure a chemical component,

instead the secondary effects of comp. can be detected on images at characteristic wavelengths,

but in this case we cannot say, that we measured the component,

because other unknown factors might also effect this secondary effect.

On the other hand, this method can still be used for industrial use

## Finally some simple multispectral application: Inspection of grape leaves using 1 channel

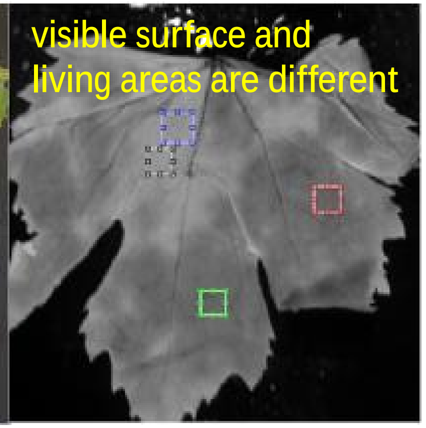
*(Nyitrai, Báló, Firtha, 2018)*



On 1450nm reflection images
the dark areas are living:

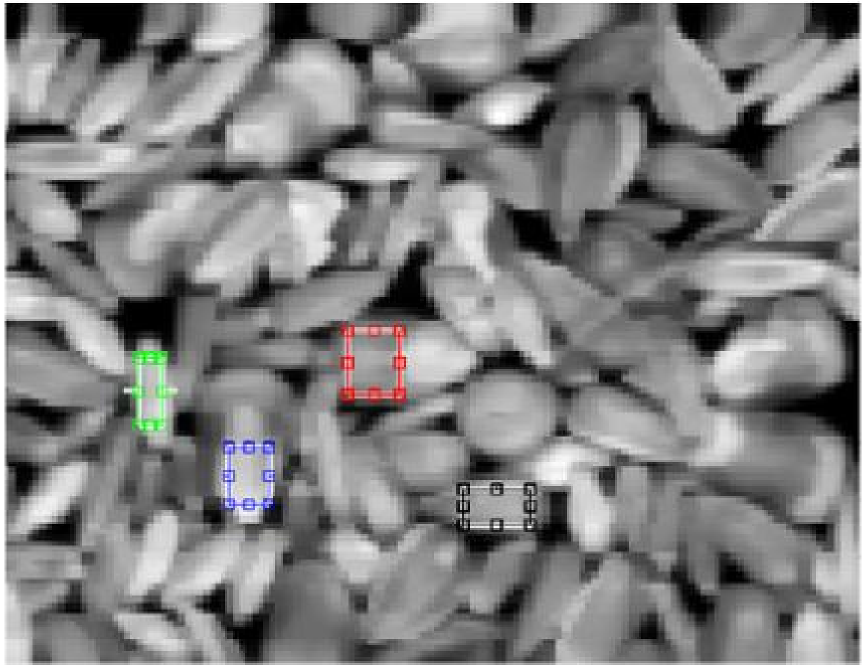
less reflectance means
higher absorbance
higher moisture content

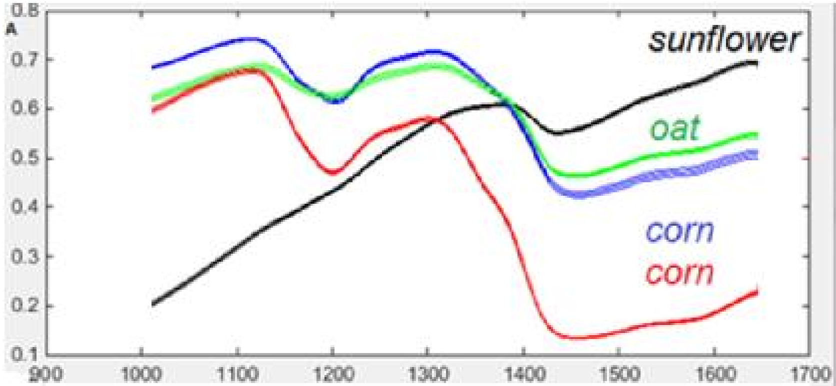


Example 2: Selection of sunflower seeds using 2 channels

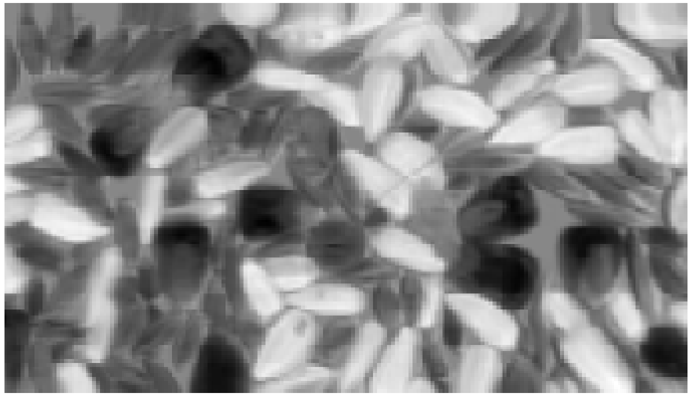
Manual selection of ROI-s on 1329 nm cross-section of hypercube






RGB image of parrot food

subtraction of 1600nm and 1100nm bands highlights seeds



Job is done

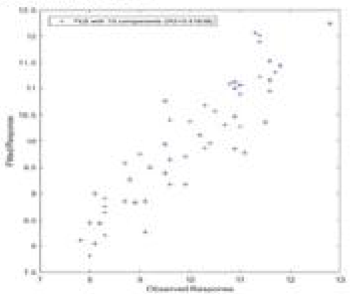
Cinderella can go to dance

Content:

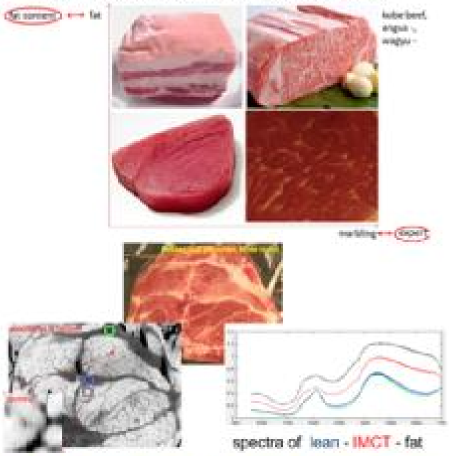


**Brix and pH in watermelon** *(Firtha, Kaszab, 2013)*

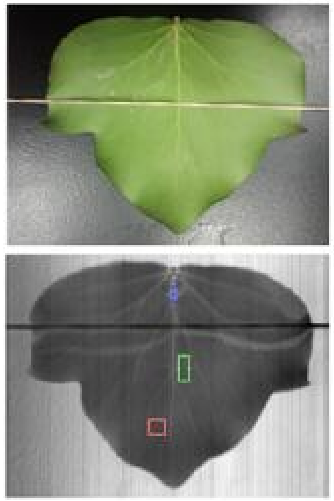
PLS for Brix: $R^2$=0.82



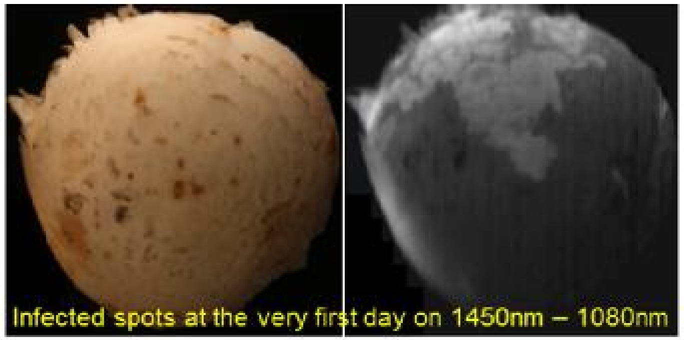
**Meat marbling** *(Firtha, Friedrich, Romvári 2013)*
How to make difference between
- lean meat,
- intermuscular **fat** deposite and
- intramuscular connective tissue **(IMCT)**?

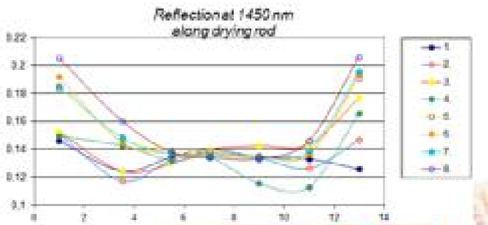


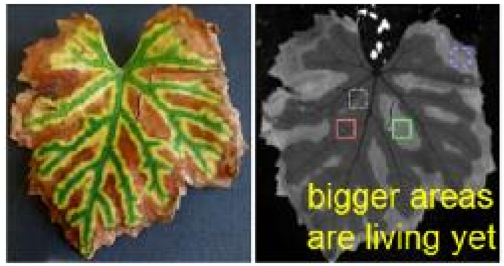

Water in the vein of a leaf *(Firtha)*



## Early detection of fungal infection by hyperspectral imaging based on secondary effects



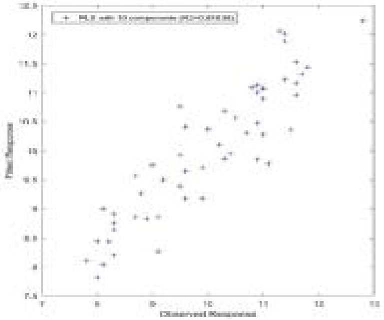




## Thank you for your attention

**Brix and pH in watermelon** *(Firtha, Kaszab, 2013)*

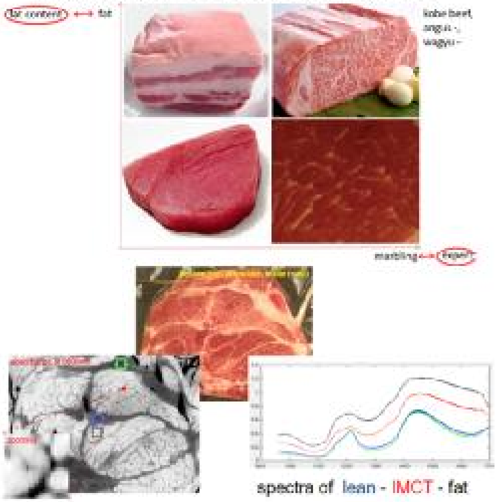
PLS for Brix: $R^2$=0.82



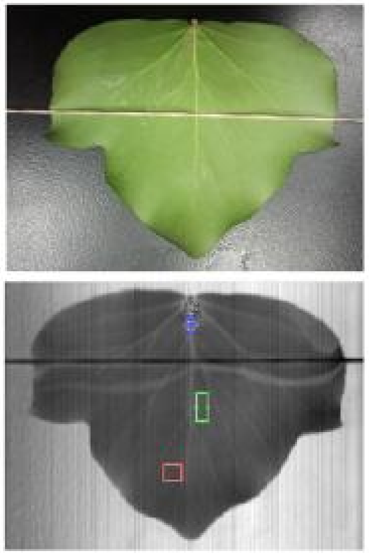
**Meat marbling** *(Firtha, Friedrich, Romvári 2013)*

How to make difference between

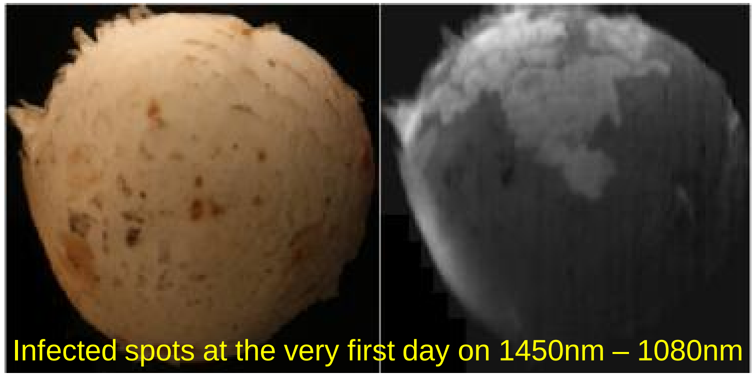
- lean meat,
- intermuscular **fat** deposite and
- intramuscular connective tissue (**IMCT**)?

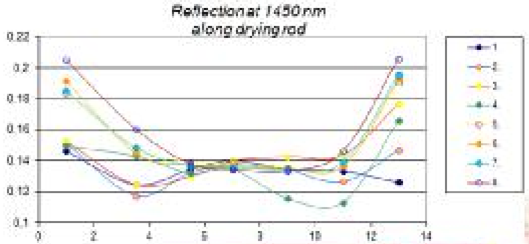


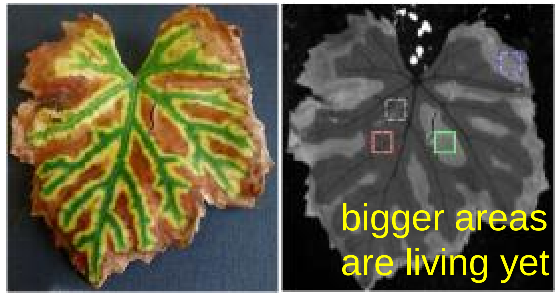

Water in the vein of a leaf (*Firtha*)



# Early detection of fungal infection by hyperspectral imaging based on secondary effects




# Thank you for your attention

Angol nyelvű tudományos előadás

30 perc előadás + vita (max 45 perc)
20-24p előadás ok, 30p a max

*Option 1:*

Title: Early detection of fungal infection by hyperspectral imaging based on secondary effects

Topic:
Hyperspectral image processing (HSI) is used to estimate the major components of a substance based on its reflection spectrum. Good examples of the applicability are sugar content in melon, melanin in baby food, meat marbling. But trace elements of fungal infections (mushrooms, grape leaves) are difficult to detect at an early stage.

I examine the advantages of the HSI method, data pre-processing possibilities, and the applicable chemometric models. The constructed DA models appear to be effective for classificating infection on a known sample set, but do not work for an unknown sample group. It is more effective to detect the consequence of infection, the drying of the surface spots. In the monochromatic image recorded at the water-specific wavelength (1454nm), the rounded islands can be recognized by image processing.

Publications often promise to estimate chemical composition by spectra (e.g., estimating the moisture content of meat in the visible range) while, in fact, they only measure secondary effects (change of pattern). Finally other examples of measuring secondary effects on only a few channels will be shown, like fructose in marzipan, segmentation of sunflower seeds.

Abbrevations:

| | |
|---|---|
| AOTF | (acousto-optic tunable filter) |
| CCD | (Charge Coupled Device) |
| CMOS | (complementary metal-oxide-semiconductor) |
| FLD | (Fisher's linear discriminant) |
| FOV | (field of view) |
| GLCM | (Gray Level Co-occurrence Matrix) |
| HSI | (Hyperspectral Imaging) |
| Light Emitting Diodes | (LED) |
| LL | (Lifelong learning) |
| LS-SVM | (Least-squares support-vector machines) |
| MCR | (multivariate curve resolution) |
| MCR-ALS | (multivariate curve resolution-alternating least squares) |
| MLR | (Multiple linear regression) |
| NIR | (Near-Infrared Spectroscopy) |
| PLS-DA | (Partial least squares discriminant analysis) |
| PLS-R | (Partial Least. Squares Regression) |
| QC | (quality control) |
| R2 | (correlation coefficient) |
| R2c | (calibration correlation coefficient) |
| R2p | (prediction correlation coefficient) |
| RMSE | (root mean square error) |
| RMSEC | (root mean square error of calibration) |
| RMSEP | (root mean square error of prediction) |
| RPD | (ratio of performance to deviation) |
| SEP | (standard error of prediction) |
| SPA-PLSR | (Successive projections algorithm- Partial Least. Squares Regression) |
| SVM | (support vector machine) |