#NIR4 - HASH NIRRRR

How to Analyse, Structure & Handle NIR Results in R Resiliently

Zoltan Kovacs[1], Flora Vitalis[1], Mercy Mukite[1], John-Lewis Zinia Zaukuu[2], Bernhard Pollner[3]

[1]Department of Food Measurement and Process Control, Institute of Food Science and Technology, Hungarian University of Agriculture and Life Sciences (**MATE**), H-1118 Budapest, Hungary,
zoltan.kovacs.food@uni-mate.hu

[2]Department of Food Science and Technology, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

[3]Dept. f. Hygiene and Medical Microbiology, Medical University of Innsbruck, Schöpfstraße 41, 6020 Innsbruck, Austria

https://github.com/zoltankovacs/NIR2025-HASH_NIR4

SUNDAY JUNE 8, 2025

10:00-12:00 **Sala dell'Editoria**

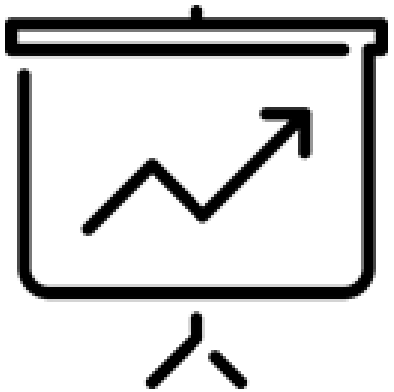
Precourse 2

#NIR4 - How to Analyse, Structure & Handle NIR Results in R Resiliently

How to Analyse, Structure & Handle NIR Results in R Resiliently

Outline

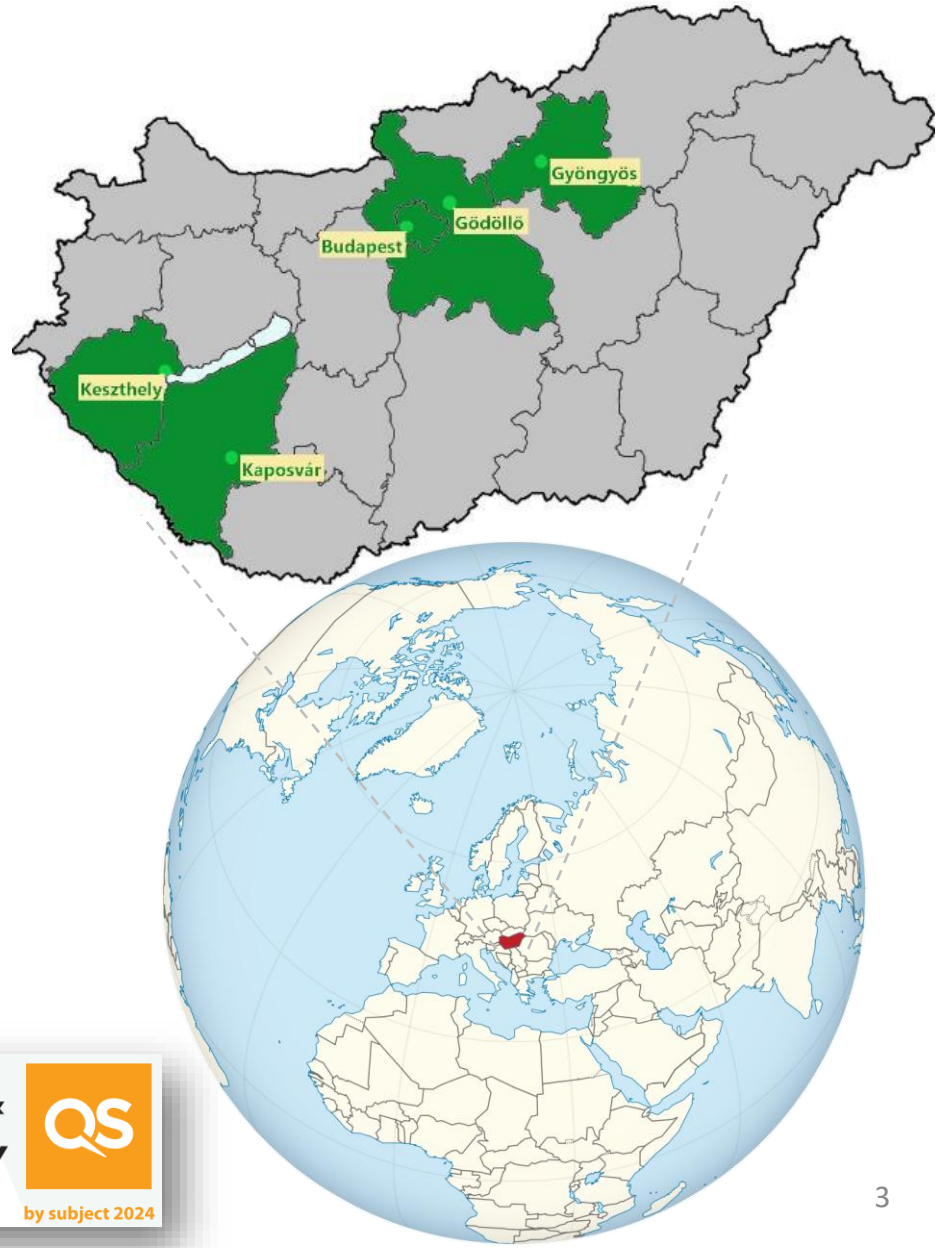
- Brief self introduction
- Importance of NIRS in recent agriculture and food production
- Objective of this precourse
- An overview of aquap2
- Case study: adding data to a sample NIR – R database
- Aquap2 practice



MATE
The Hungarian University of Agriculture and Life Sciences

MATE HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

- **Founded on 1 February 2021** (successor to former Szent Istvan University, Kaposvar University and National Agricultural Research and Innovation Centre)
- **One of the largest agricultural-focused, multi-disciplinary higher education institutions in CEE**
- **provides world-class education in agriculture-related fields**
- **No. of institutes: 21**
- **No. of students: 13 621**
- **No. of international students: 2076** (100+ countries) **15.24%**
- **Number of PhD schools: 12**
- **Number of PhD students: 864** (international PhD students 382) **44.2%**
- **Number of academic staff: 1060**
- **Number of staff: 1412**
- **Languages of instruction: Hu, Eng**

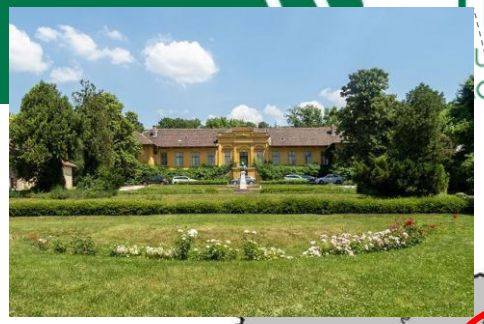


The Best Agriculture & Forestry University in Hungary

QS by subject 2024

MATE Campuses and Buda Campus Institutes

MATE
HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES



Hungarian University of Agriculture and Life Sciences:

- **Szent Istvan Campus, Gödöllő – Head Office**
- **<u>Buda Campus</u>**
  - **Institute of Food Science and Technology**
  - Institute of Horticultural Sciences
  - Institute of Landscape Architecture, Urban Planning and Garden Art
  - Institute of Plant Protection
  - Institute for Viticulture and Oenology
- **Kaposvár Campus**
- **Gyöngyös Campus**
- **Keszthely Campus**

HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

Institute of Food Science and Technology
Research topics in focus I.

Across the entire Food Chain:

production of food industry raw materials – food processing – food trading

Food processing technologies

- Animal product technologies (meat, milk, fish, egg):
  - Modern meat preservation and processing techniques etc.
  - Modern technologies, products, special diet, functional food etc.
- Fermentation technologies:
  - Beer, spirits, fermented functional food, pro/pre/synbiotics etc.
- Cereal science:
  - Modern bakery products, smart products etc.
- Fruit and vegetable processing technologies:
  - Drying technologies, aseptic technologies, PEF technology, etc.
- Modern postharvest and preservation technologies
  - Packaging, cold chain, minimal processing technologies (HHP, ultrasound)
  - Smart fresh technologies, etc.

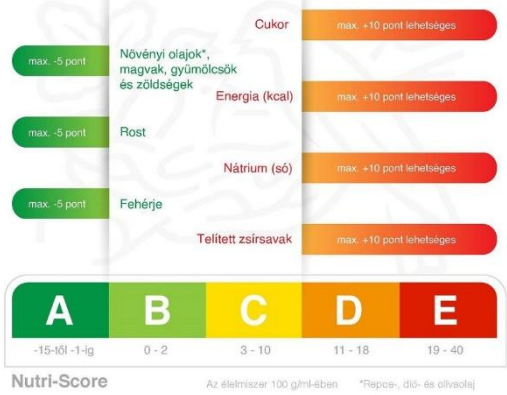
Institute of Food Science and Technology
Research topics in focus II.

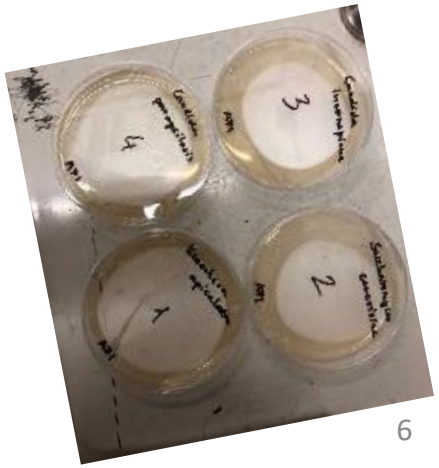
Across the entire Food Chain:

production of food industry raw materials – food processing – food trading

Food safety, quality, nutrition and trading

- Food safety:
  - Food microbiology, food analytics and food chemistry, microplastics, toxins, pesticide residues etc.
- Nutrition
  - Nutritional value analysis
  - Biological value enhancement technologies
- Food trading:
  - Consumer acceptance and preferences, logistics, marketing and trading
- *State of the art* research methodologies and instruments:
  - Analytical laboratories, sensory laboratory etc.





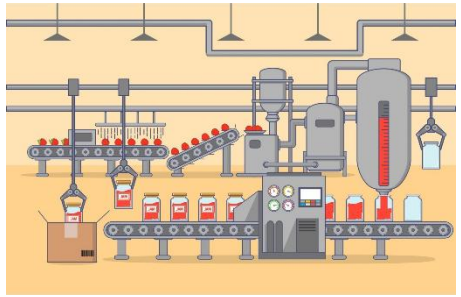
HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

Institute of Food Science and Technology
Research topics in focus III.

Across the entire Food Chain:

production of food industry raw materials – food processing – food trading

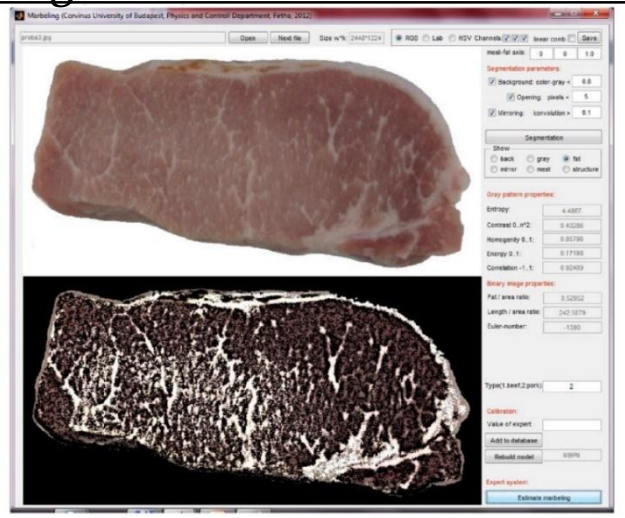
Measurement techniques, automatization, digitalization



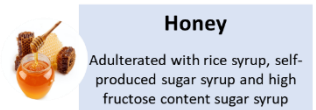
- Robotization
- Data processing and evaluation, bigdata, chemometrics
- Digitalization, industry 4.0, IoT, M2M
- Modern measurement methods:
  - Non-destructive methods, contactless measurement technologies
  - Instrumental taste and odor sensing, NIR, hyperspectral, machine vision, etc.

Meat marbling measurement– machine vision

3 classes:
- lean
- fat
- IMCT



Quality assessment of food – NIRS and aquaphotomics



Honey
Adulterated with rice syrup, self-produced sugar syrup and high fructose content sugar syrup

Coffee drinks
Classification and quantification of different origin Arabica and Robusta coffee

Meat mixture extracts
Detection and quantification of pork in beef

Yogurt quality
With different lactobacillus strains grown with different water

Mung bean juice
Monitoring of germination and prediction of ascorbic acid

Tokaj wine
Adulterated with grape must concentrate

Monilia (brown rot) on plums
Early phase detection of *Monilia fructigena* infection on plums

Tomato concentrate
Adulterated with starch, paprika seed powder, NaCl and sucrose

Paprika powder dilution
Adulterated with corn flour at different concentration levels

Mineral water
With different mineral content mixed and analyzed

Importance of the non-destructive analytical techniques

– Precision farming and food production

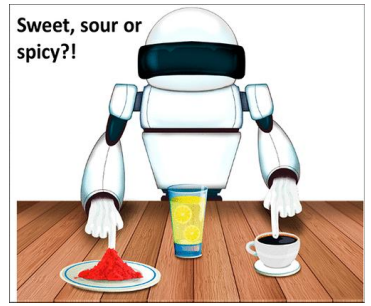
– Increasing quantitative needs ←→ the standard quality food ←→ more economically

– Data driven decision making

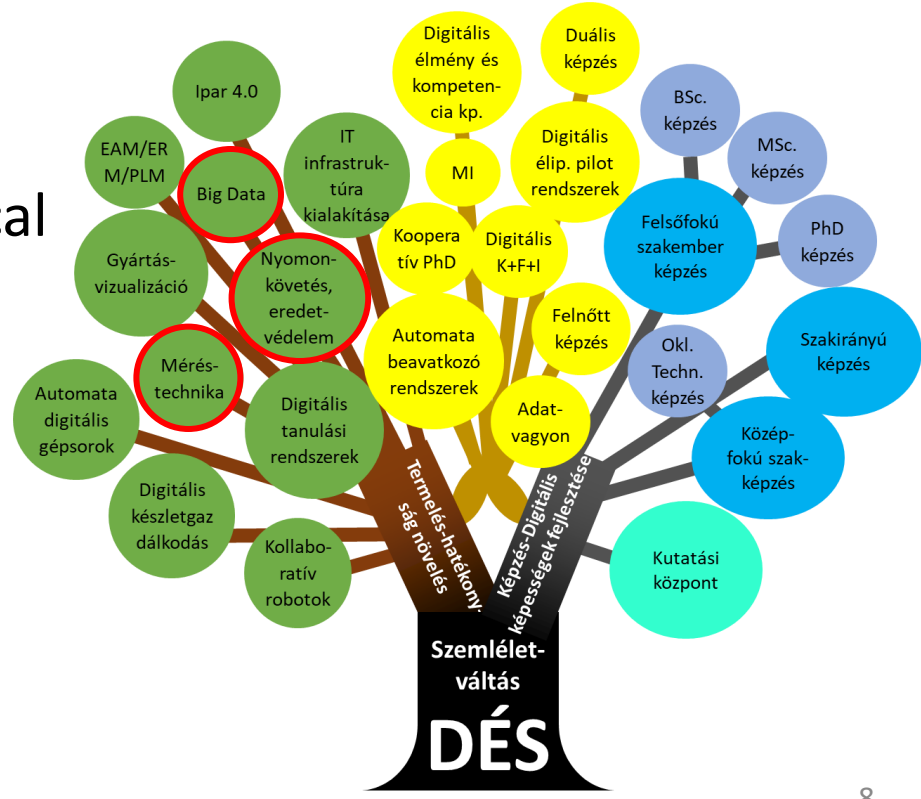
– Rapid and precise quality determination of raw materials and food products

– Immense need for advanced and non-destructive analytical techniques in the agriculture and food industry:

  - ultrasonic measurement,
  - bioelectric properties, impedance spectroscopy,
  - acoustic measurement,
  - electronic tongue and electronic nose
  - machine vision,
  - near-infrared spectroscopy and aquaphotomics etc.



ACS Sens. 2018, 3, 11, 2375–2384

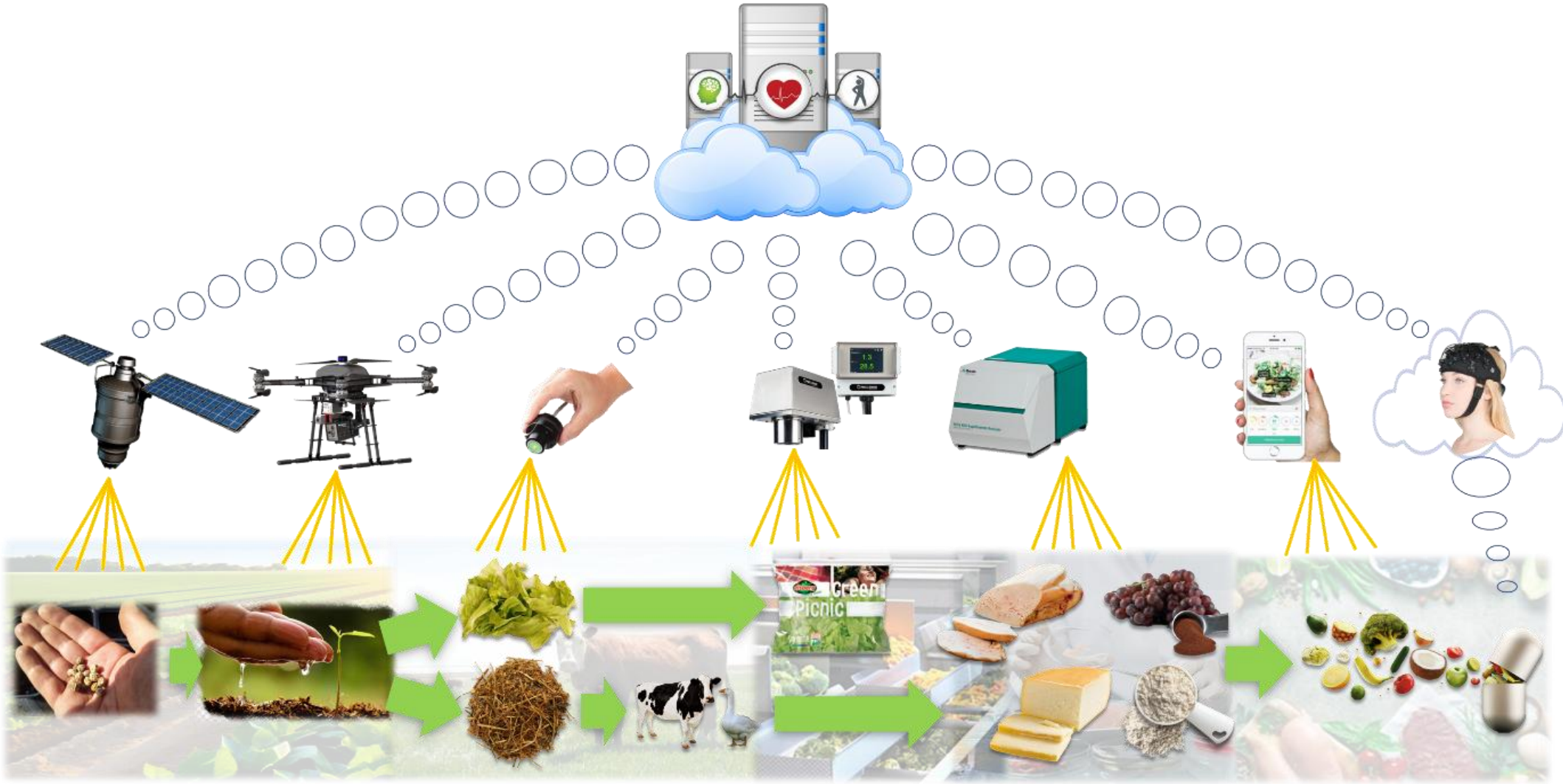


Hungary's digital food industry strategy (2022)

Objective of this presentation

HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

to provide insights about recent results of NIRS and aquaphotomics supporting the monitoring of food supply chain “from farm to fork” – and beyond.



Food quality assessment with NIR and aquaphotomics

Lettuce quality

Monitoring of lettuce during storage in argon gas packaging

Monilia (brown rot) on plums

Early phase detection of *Monilia fructigena* infection on plums

Coffee drinks

Classification and quantification of different origin Arabica and Robusta coffee

Goose Liver quality

Determination of the blood content in fattened goose liver

Meat mixture extracts

Detection and quantification of pork in beef

Tokaji wine

Adulterated with grape must concentrate

Honey authentication

Detection of heat treatment and syrup adulteration of honeys

Mung bean juice

Monitoring of germination and prediction of ascorbic acid

Mineral water

With different mineral content mixed and analyzed

Fruit Juice Fortification

Understanding water structure changes in fruit juices induced by plant extracts using aquaphotomics

Enriched cheese quality

Impact of Treatment and Aging on PUFA-Enriched Cheese: an Aquaphotomics focused approach

Enriched cheese quality

Comparison of regression models for the prediction of caffeine content in aqueous solutions + ANN

Objectives of this course

https://www.inno-spectra.com

to introduce some possible applications of R project and the aquap2 package

- to design experiments,
- to structure, curate and archive spectral data effectively

All in all, to present hands-on practical examples – from experimental design to saving and structuring the collected data as part of a larger NIR database.

An overview of aquap2

Multivariate analysis tools for aquaphotomics

Data analysis with the “aquap2” R-package: a free multivariate data analysis framework

https://github.com/bpollner/aquap2

https://github.com/zoltankovacs/NIR2025-HASH_NIR4

Package "aquap2"

Main Advantages

- drastically speeds up analysis time
- highly repetitive tasks get completely scriptable (i.e., automated)

Key Features

Experiment Design

- truly randomize samples
- facilitates time resolved experiments

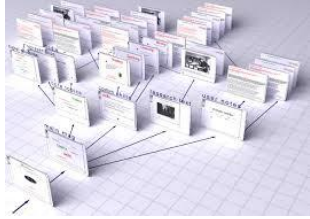
Data Import & Organization

- import of metadata: class.- and numerical variables
- import of spectral data
- consistent coloring
- align environmental data (T, rH) to timestamp on spectra
- add custom, device-specific noise to dataset

Data Analysis & Visualization

- User-defined splitting of dataset
- Bootstrapped Aquagram: custom temperature calibration data
- User-defined splitting of dataset by wavelength or any class-variable
- Highly flexible input and control system

Workflow of the “aquap2“ package



Acquire and import data | Analyze data

一
Build the structure of folders

Copy/paste an earlier one
OR
in R call function genFolderStr()
after aquap2 package was loaded

二, a
Make experimental design

二, b
Produce the sample list

Set up experimental parameters in the file metadata.r and run the function esl()

三
Perform the experiment

- Apply the correct sample order
- Write and/or record RH, Temp…
- Complete the sample list with reference data

四
Copy spectral data (.da), TRH & sample list into the resp. folders

- spectral data:
  to „rawdata” folder
  name: e.g. „ExpName.da”*
- TRH logger data:
  to „rawdata” folder
  name: e.g. „TRHlog.txt”*
- sample list:
  to „sampleLists/sl_in ”
  folder
  name: e.g. „ExpName-in.xlsx”*

*various file formats are available

五
Import and save raw data

call function gfd() specifying:
- file types**
- trhLog type**
- multiplyRows**

**if not default

六
Modify the analysis procedure list

call function gdmm()

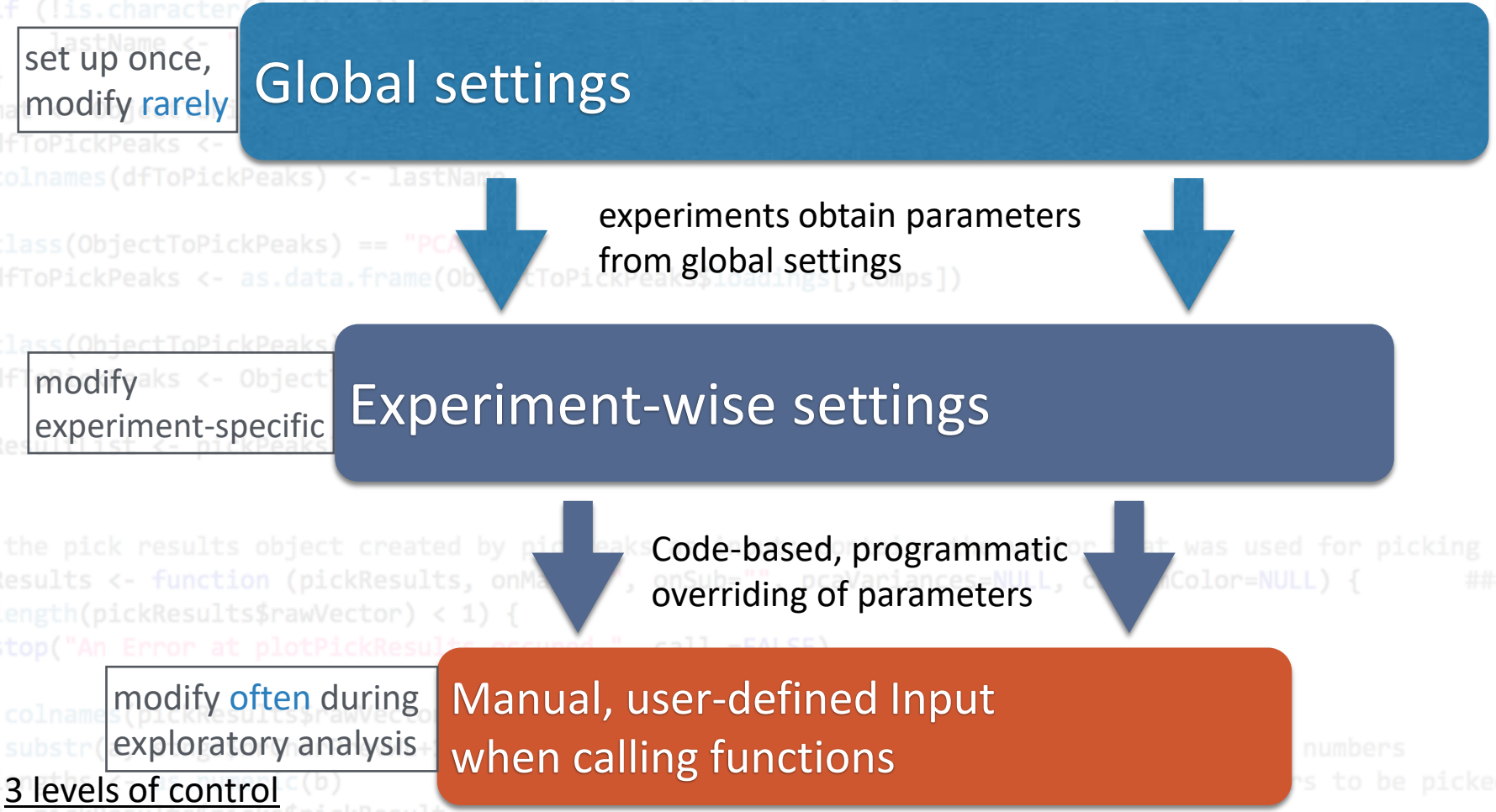
七
Plot the results

call function plot()

In R-project

NOT in R-project

Input & Control



- 3 levels of control
- flexible, highly customizable
- global settings: change input globally in a single, central file („settings.r“)
- experiment-wise settings: the specific metadata and analysis procedure file for each experiment.
- code-based overriding of parameters: Enables programmatic, i.e. scriptable modification of input-parameters.

Measurement preparation

- ✓Generate the most important folders and files
- ✓Create and export samples list based on „metadata.R”

```
17  TimePoints <- FALSE            ## how many points in time does the experiment cover? Provide a label for eac
18  nrConScans <- 3              ## how many consecutive scans for every sample?
19  spacing <- FALSE               ## the space between environmental control samples (set to FALSE for not inse
20  ###
21  columnNamesL1 <- c("C_Water", "Y_waterTemp")      ## please note that there is a special prefix for class-
22  columnNamesL2 <- c("C_DELETE", "C_DELETE")      ## "DELETE" is the default character for those columns that
23  L1  <- list(list("MQ", "TAP", "DW", "AIR"), list(seq(20, 74, 2)))          ## Please look at the vignette
24  L2  <- L1
25  ###
26  Repls <- 1                  ## how many replications for each sample?
27  Group <- c("no")              ## additional groups to split the above generated classes into, like e.g. exper
28
```

Metadata

Measurement preparation

✓Generate the mo

✓Create and expo

```
17  TimePoints <- FALSE
18  nrConScans <- 3                ##
19  spacing <- FALSE
20  ###
21  columnNamesL1 <- c("C_Water",
22  columnNamesL2 <- c("C_DELETE",
23  L1  <- list(list("MQ", "TAP",
24  L2  <- L1
25  ###
26  Repls <- 1                   ## h
27  Group <- c("no")               ##
28
```

Metadata

Console

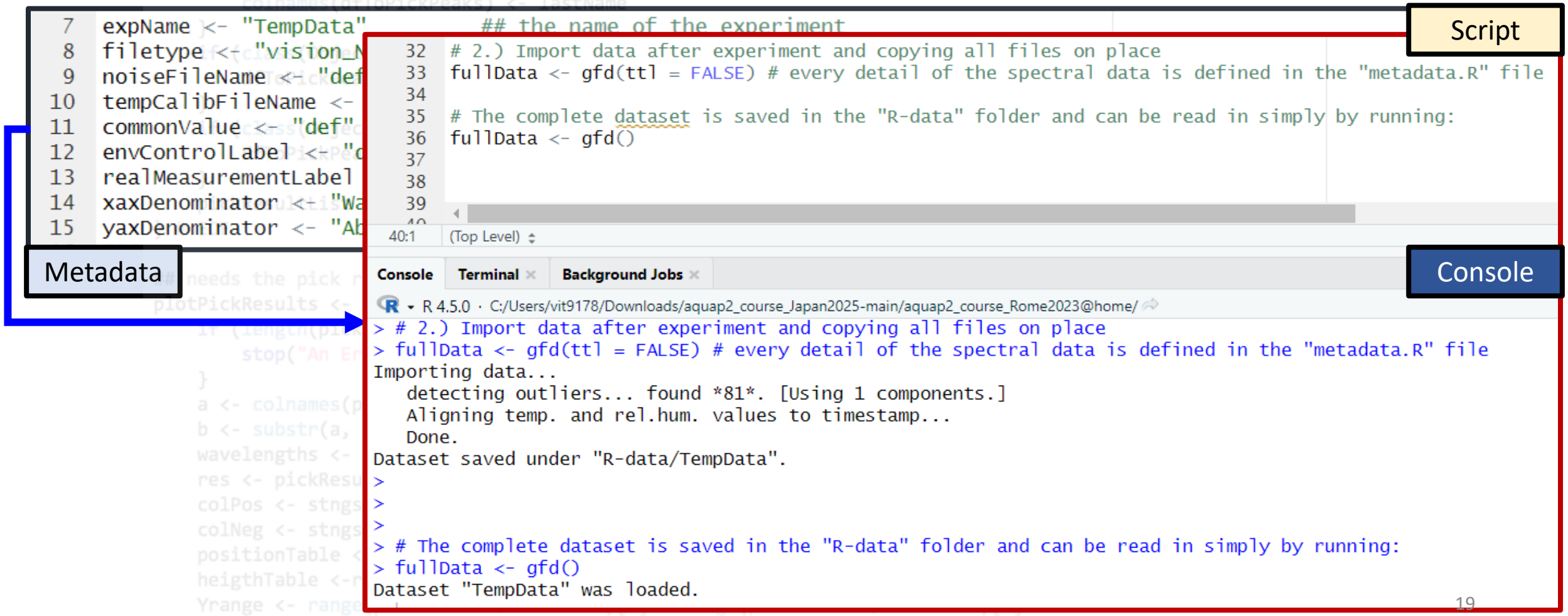
```
> # 1.) Creat sample list based on metadata file
> esl(form = "xls", rnd = FALSE, showFirstRows = TRUE, timeEstimate = TRUE)
Creating sample list...
Writing sample list to file...
A sample list in .xlsx format with 112 rows has been saved to "sampleLists/sl_out".
Minimum working time: 9.5 hours.
Have fun!

   Y_SampleNr C_Time C_ECRM C_Water Y_waterTemp C_Repl C_Group Y_Temp Y_RelHum
1           1      NT     RM      MQ          20     R1      no
2           2      NT     RM     TAP          20     R1      no
3           3      NT     RM      DW          20     R1      no
4           4      NT     RM     AIR          20     R1      no
5           5      NT     RM      MQ          22     R1      no
6           6      NT     RM     TAP          22     R1      no
7           7      NT     RM      DW          22     R1      no
8           8      NT     RM     AIR          22     R1      no
9           9      NT     RM      MQ          24     R1      no
10         10      NT     RM     TAP          24     R1      no
11         11      NT     RM      DW          24     R1      no
12         12      NT     RM     AIR          24     R1      no
13         13      NT     RM      MQ          26     R1      no
14         14      NT     RM     TAP          26     R1      no
15         15      NT     RM      DW          26     R1      no
16         16      NT     RM     AIR          26     R1      no
17         17      NT     RM      MQ          28     R1      no
18         18      NT     RM     TAP          28     R1      no
19         19      NT     RM      DW          28     R1      no
20         20      NT     RM     AIR          28     R1      no
21         21      NT     RM      MQ          30     R1      no
```

Data analysis preparation

✓Specify the raw data details in „metadata.R”

✓ and import



MVA Methods – *close look at the „anproc.R" file*

Already Implemented

- Data pre-treatment

  smoothing, 2nd derivative, SNV, MSC, EMSC, deTrend, gap derivative, …
- PCA; SIMCA; PLSR
- DA
- Aquagram (classic & extended)

In development

- PLS-DA; ANN; SVM; ICA; KNN

Key spectral prepocessing

Data splitting

- By grouping variables,
- By wavelength ranges…

Data treatment

- Specific spectral processing,
- Averaging,
- Outlier elimination…

Anproc

```
10  spl.var <- NULL              ## by which variables should the dataset be split?
11  spl.wl <- NULL               ## which wavelengths to use? leave empty c() or set
12  #######
13  dpt.pre <- NULL              ## Character vector, which of the available modules
14  #######
15  spl.do.csAvg <- FALSE          ## if all the consecutive scans of a single sampl
16  spl.csAvg.raw <- TRUE          ## if, should the consecutive scans of a single s
17  #
18  spl.do.noise <- FALSE          ## if artificial noise should be added to the dat
19  spl.noise.raw <- TRUE          ## if, should the noise-test be performed, the ra
20  #
21  spl.do.exOut <- FALSE          ## if exclusion of outliers should be performed
22  spl.exOut.raw <- FALSE           ## if, should exclusion of outliers be performe
23  spl.exOut.var <- c("C_FooBar")       ## A vector with valid class variable names
24  #######
25  dpt.post <- NULL             ## Character vector, which of the available modules
```

User-defined splitting of dataset

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345

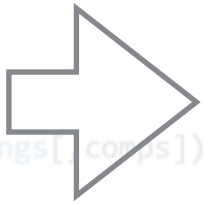
Key Features II: User-defined splitting of dataset

Split by „C_Time“

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345

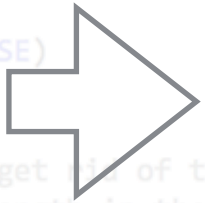
Key Features II: User-defined splitting of dataset

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362



Separate Analysis and Visualization

T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345



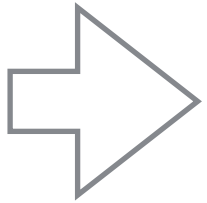
Key Features II: User-defined splitting of dataset

Split by „C_Group“

C_Time	C_Group	w1300	w1300.5	w1301	w...	w1600
T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345

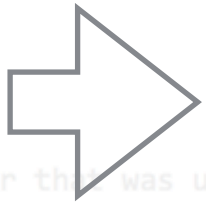
Key Features II: User-defined splitting of dataset

T0	Treatment_AAA	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_AAA	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_AAA	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_AAA	0.2345	0.2362	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.4352	0.2345	0.2362	0.4352	0.2362
T1	Treatment_AAA	0.2362	0.4352	0.2345	0.2362	0.2345



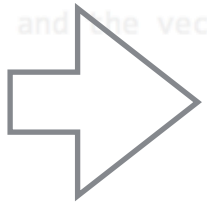
Separate Analysis and Visualization

T0	Treatment_BBB	0.2345	0.4352	0.2362	0.2345	0.2345
T0	Treatment_BBB	0.4352	0.2362	0.2345	0.4352	0.4352
T0	Treatment_BBB	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_BBB	0.2345	0.2362	0.4352	0.2362	0.4352
T1	Treatment_BBB	0.4352	0.2345	0.2362	0.2345	0.2362
T1	Treatment_BBB	0.2362	0.4352	0.2345	0.4352	0.2362



Separate Analysis and Visualization

T0	Treatment_CCC	0.2345	0.4352	0.2362	0.4352	0.2345
T0	Treatment_CCC	0.4352	0.2362	0.2345	0.2362	0.4352
T0	Treatment_CCC	0.2362	0.4352	0.4352	0.2345	0.2362
T1	Treatment_CCC	0.4352	0.2362	0.4352	0.2362	0.2345
T1	Treatment_CCC	0.2362	0.2345	0.4352	0.2362	0.2362
T1	Treatment_CCC	0.2345	0.4352	0.2362	0.2345	0.2345



Spectrum-based calculations – *raw spectra*

Script

```
37
38
39
40  # 3.) Plot raw spectra from imported dataset
41  plot(fullData, colorBy = "C_waterTemp")
42  plot_spectra(fullData, colorBy = "C_waterTemp", pg.where = "")
43
44
45
46
```

60:1 (Top Level)

Console

Console | Terminal | Background Jobs

R 4.5.0 · C:/Users/vit9178/Downloads/aquap2_course_Japan2025-main/aquap2_course_Rome2023@home/

```
> # 3.) Plot raw spectra from imported dataset
> plot(fullData, colorBy = "C_waterTemp")
Plotting raw spectra ... ok
> plot_spectra(fullData, colorBy = "C_waterTemp", pg.where = "")
>
```

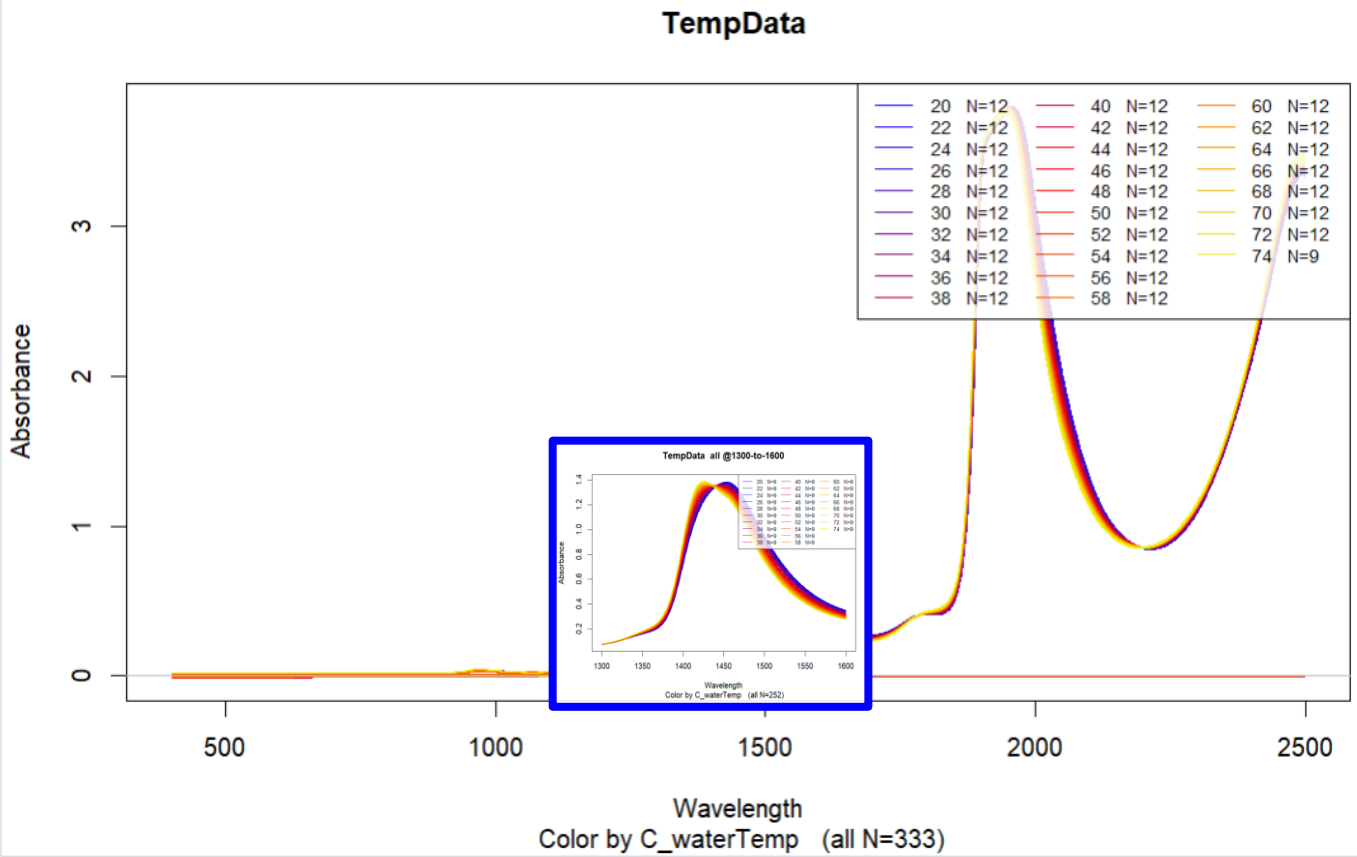
R | Global Environment

Data

fullData	333 obs. of 3 variables

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export | Publish

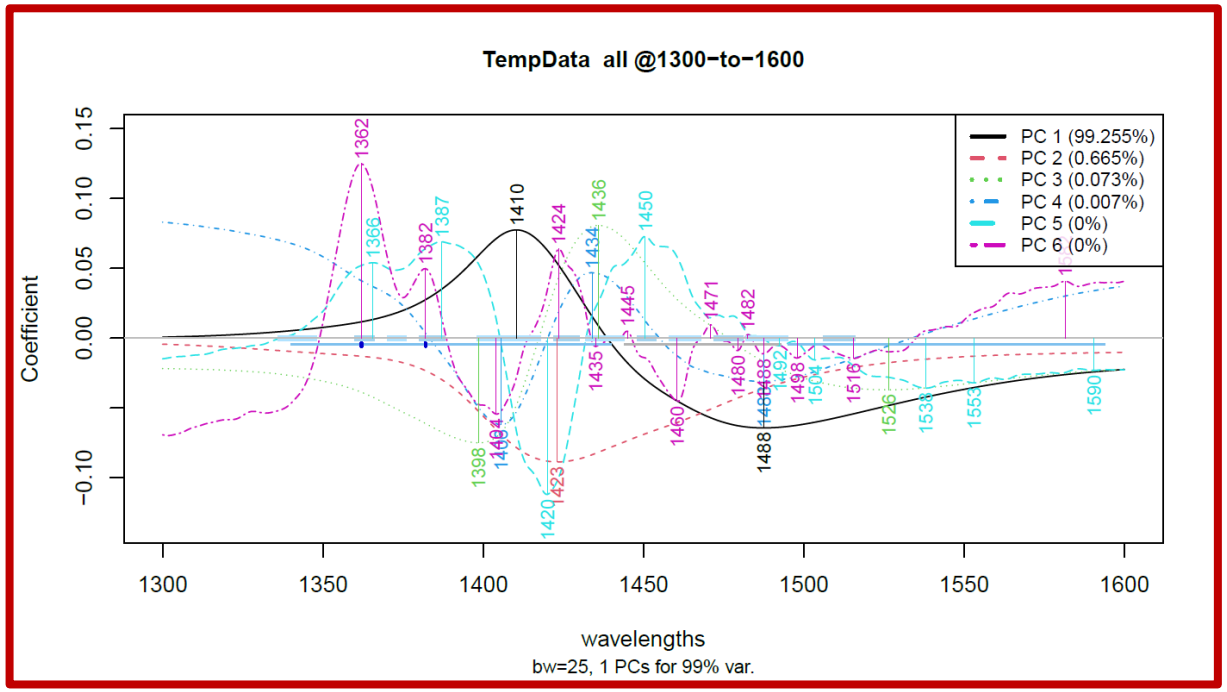


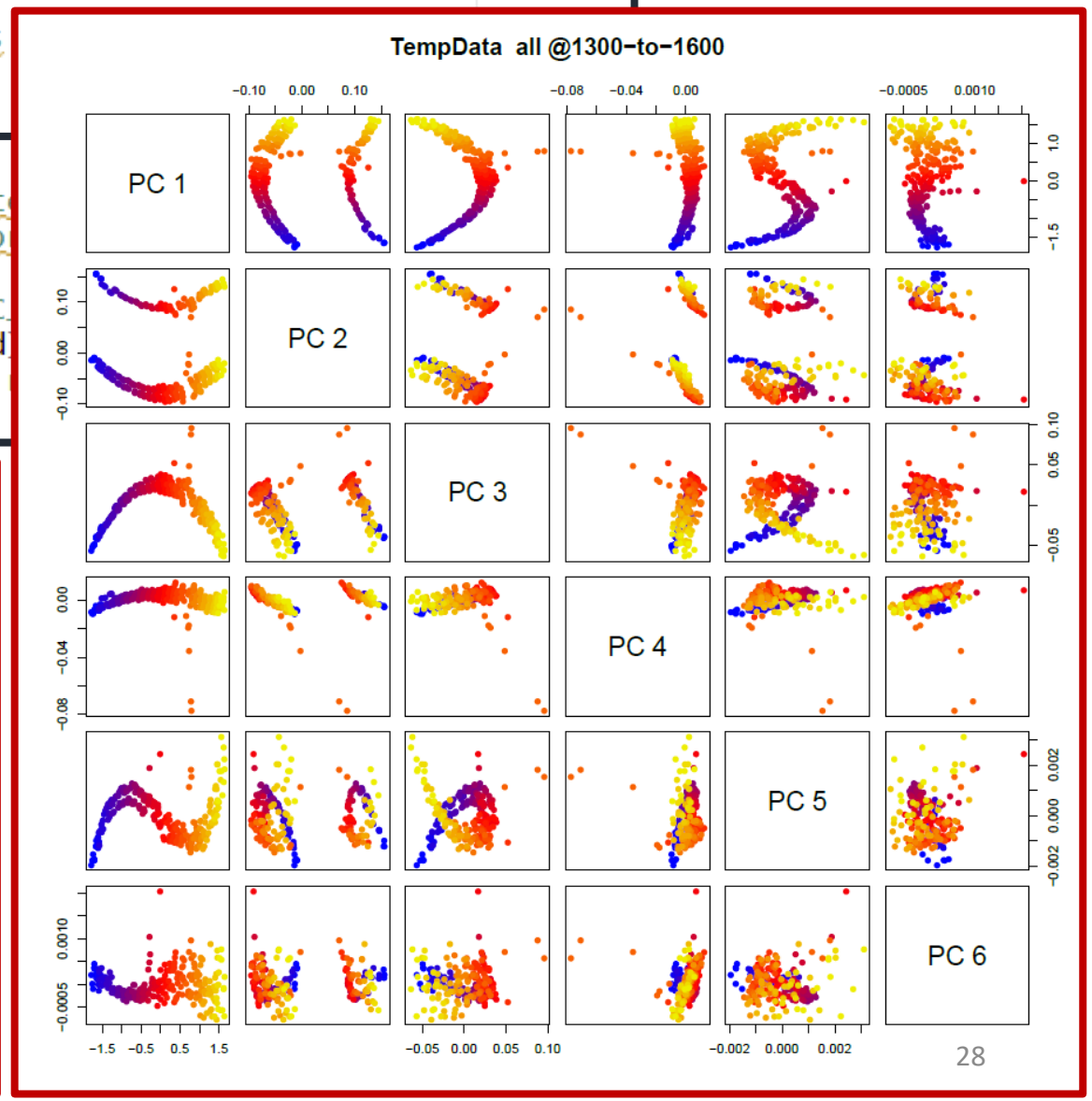
Spectrum-based calculations – *PCA*

Anproc

```
35  ##########################
36  ### PCA
37  do.pca <- T               ## if PCA of the given datasets
38
39  # plotting
40  pca.colorBy <- NULL
41  pca.elci <- "def"
42  pca.elcolorBy <- NULL
43  pca.what <- "both"
44  pca.sc <- c(1, 2)
45  pca.sc.pairs <- 1:6
46  pca.lo <- 1:6
47  ##########################
```

```
59
60  # 6.) Do PCA on the dat
61  ## set in "matedata/anp
62  ### do.pca <- TRUE
63  ### pca.colorBy <- c("C
64  cu2 <- gdmm(dataReduced
65  plot(cu2)    # plot the
66
```





Spectrum-based calculations – *PLSR*

Anproc

```
61  ##########################
62  ### PLSR
63  do.pls <- FALSE              ## if PLSR models of the given datasets should be calculated
64  pls.ncomp <- NULL            ## number of components, leave at NULL for automatic detection
65  pls.regOn <- NULL            ## which variables should be used to predict on? Leave empty c
66  pls.valid <- "def"
67  pls.exOut <- "def"
68
69  # plotting
70  pls.colorBy <- NULL
71  pls.what <- "both"
72  pls.rdp <- FALSE
73  ####################
```

Script

```
107
108 # 10.) Do regression on water temperature with PLSR
109 ## set in "metadata/anproc.r":
110 ### do.pca <- FALSE
111 ### do.pls <- TRUE
112 ### pls.ncomp <- 2 # set the number of latent variables (LV) to be used in the modelling
113 ### pls.regOn <- c("Y_waterTemp")  # variable to be predicted
114 ### pls.valid <- "C_SampleNr"  # default: 10-fold cross-validation but it can be changed to act
115 ### pls.colorBy <- "C_waterTemp" # colouring of the points
116 cu6 <- gdmm(dataReduced) # calculate PLSR
117 plot(cu6, pg.where = "_byWater_sgol+msc") # plot the results of PLSR
118
119
```

108:1 (Top Level)

Console

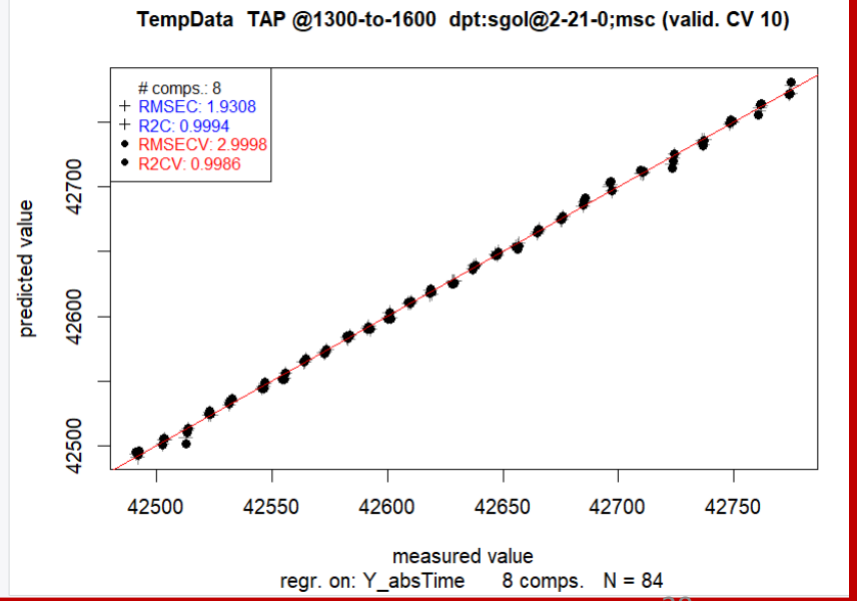
```
R 4.5.0 · C:/Users/vit9178/Downloads/aquap2_course_Japan2025-main/aquap2_course_Rome2023@home/
      Data treatment: sgol@2-21-0
      Data treatment: msc
   Working on #3 of 3
      Data treatment: sgol@2-21-0
      Data treatment: msc
Done.

Calculating models...
   Working on dataset #1 of 3 (DW @1300-to-1600)
      calc. PLSR, 7 models (Y_SampleNr, Y_conSNr, Y_waterTemp, Y_Temp, Y_RelHum, Y_absTime, Y_chron)
in parallel...  ok
   Working on dataset #2 of 3 (MQ @1300-to-1600)
      calc. PLSR, 7 models (Y_SampleNr, Y_conSNr, Y_waterTemp, Y_Temp, Y_RelHum, Y_absTime, Y_chron)
in parallel...  ok
   Working on dataset #3 of 3 (TAP @1300-to-1600)
      calc. PLSR, 7 models (Y_SampleNr, Y_conSNr, Y_waterTemp, Y_Temp, Y_RelHum, Y_absTime, Y_chron)
in parallel...  ok
Done.
> plot(cu6, pg.where = "_byWater_sgol+msc") # plot the results of PLSR
*** PCA model not available or not selected for plotting
*** SIMCA model not available or not selected for plotting
*** Aquagram not available or not selected for plotting
*** DA model not available or not selected for plotting
*** Random forest model not available or not selected for plotting
*** SVM model not available or not selected for plotting
*** NNET model not available or not selected for plotting
>
```

Console

Data	
cu	Large aquap_cube (3.3 MB)
cu2	Large aquap_cube (3.3 MB)
cu6	Large aquap_cube (3 elements, 37.9 MB)
dataReduced	252 obs. of 3 variables
fullData	333 obs. of 3 variables

Files | Plots | Packages | Help | Viewer | Presentation



Spectrum-based calculations – *aquagrams*

Anproc

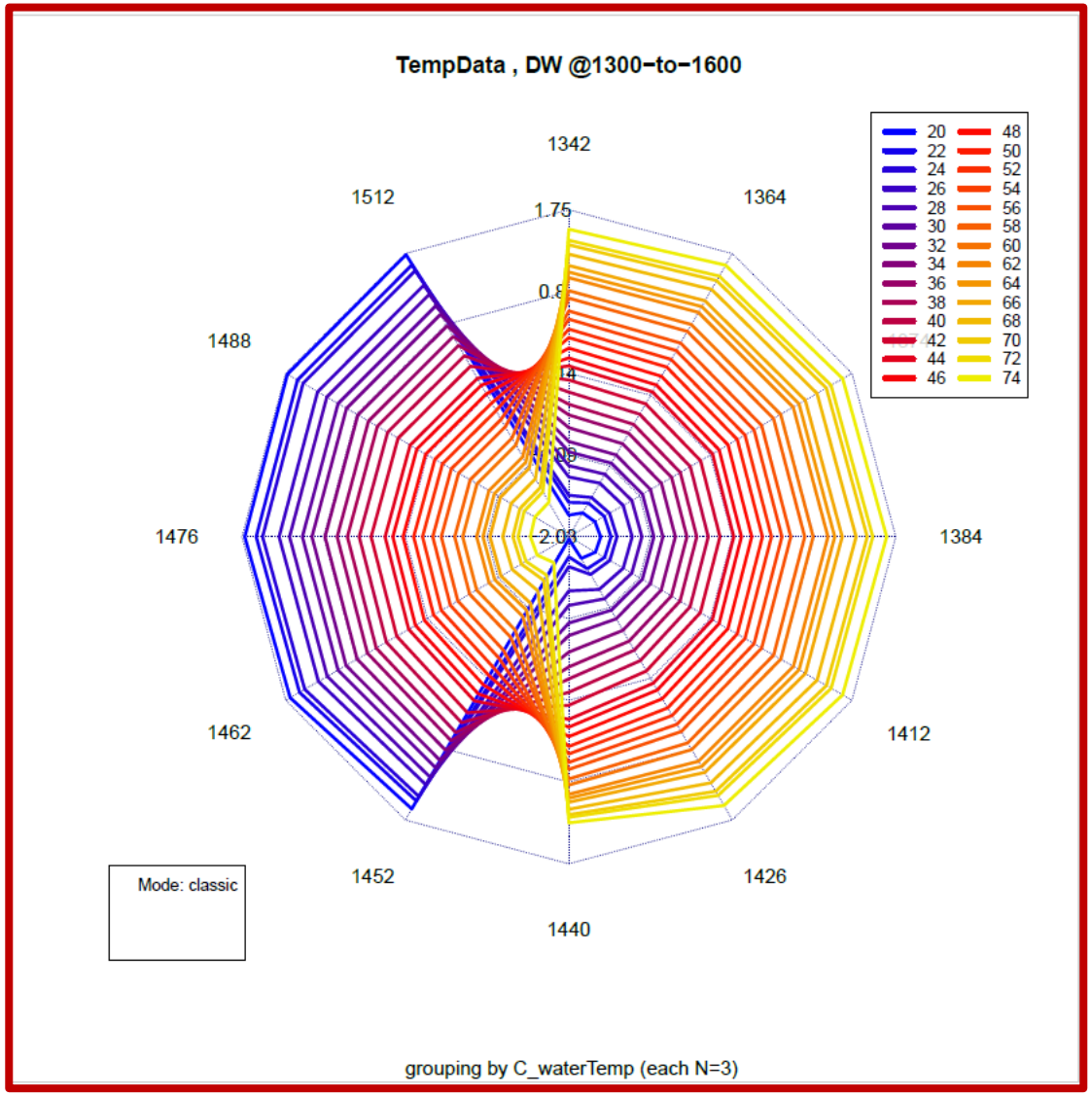
```
77  ### Aquagram
78  do.aqg <- FALSE
79  aqg.vars <- NULL
80  aqg.nrCorr <- "def"
81  aqg.spectra <- FALSE
82  aqg.minus <- NULL
83  aqg.mod <- "def"
84  aqg.TCalib <- "def"
85  aqg.Texp <- "def"
86  aqg.bootCI <- "def"
87  aqg.R <- "def"
88  aqg.smoothN <- 21
89  aqg.selWls <- "def"
90  aqg.msc <- TRUE
91  aqg.reference <- NULL
```

Script

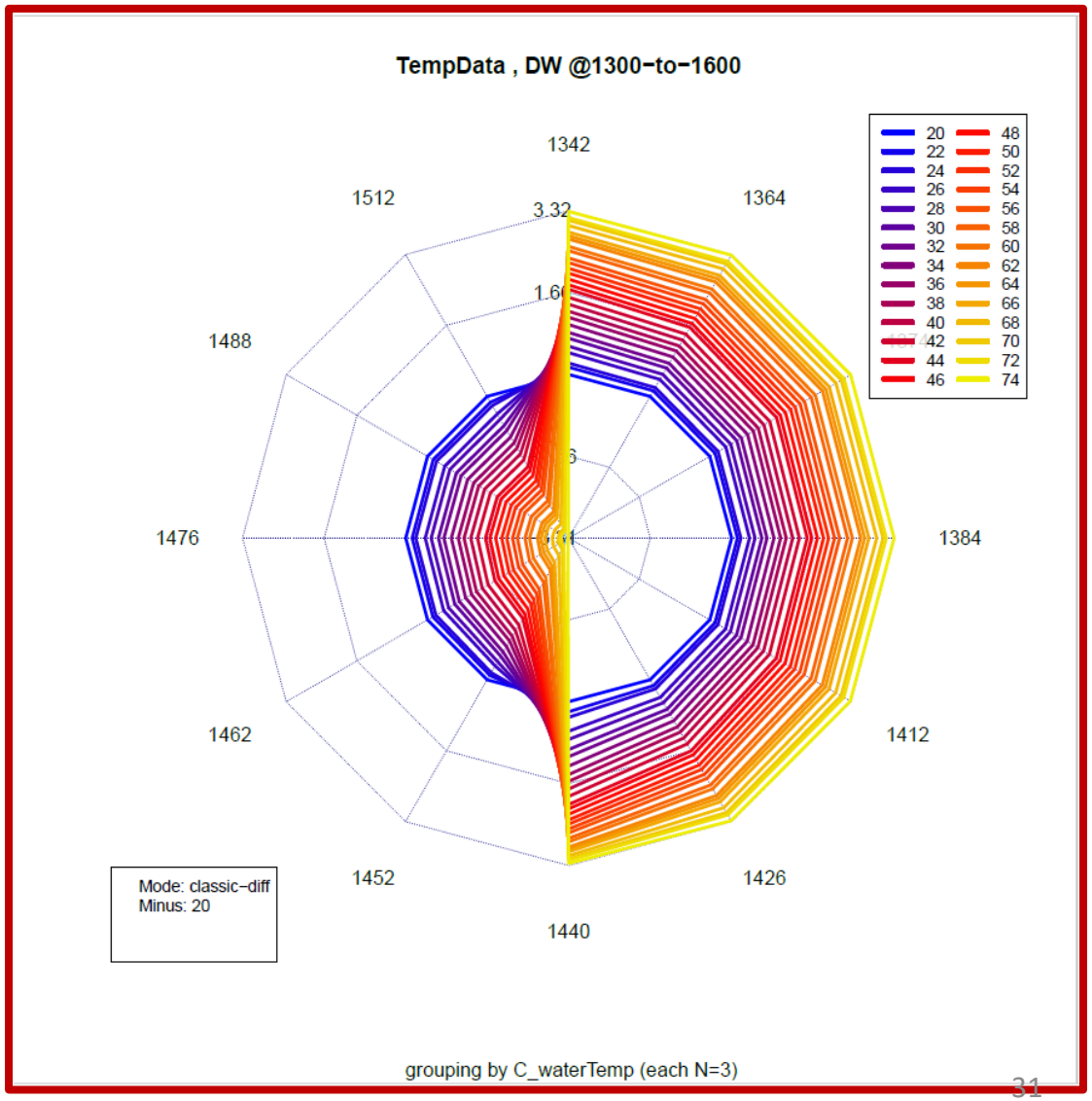
```
119
120 # 11.) Create classic aquagrams
121 ## set in "metadata/anproc.r":
122 ### do.pls <- FALSE
123 ### do.aqg <- TRUE
124 ### aqg.vars <- "C_waterTemp"
125
126
127 # 11a.) Calculate classic aquagram
128 ### aqg.mod <- "classic"
129 cu7 <- gdmm(dataReduced)
130 plot(cu7, pg.fns = "_byWater") # plot "classic" aquagram
131
132
133 # 11b.) set anproc to do classic aquagram by subtracting lowest temperature
134 ## set in "metadata/anproc.r":
135 ### aqg.spectra <- "all"
136 ### aqg.minus <- "20"
137 ### aqg.mod <- "classic-diff"
138 cu8 <- gdmm(dataReduced)
139 plot(cu8, pg.fns = "_byWater")  # plot "classic-difference" aquagram
140
```

Spectrum-based calculations – *aquagrams*

Classic

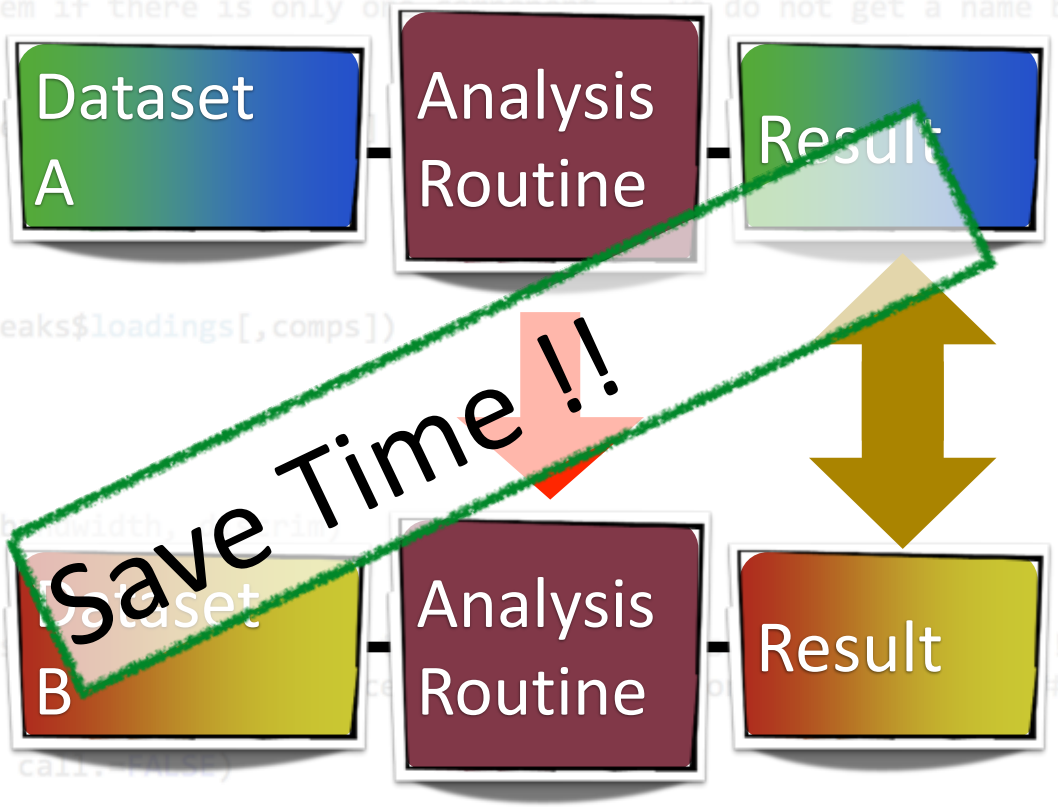


Classic-diff



Summary - Main Advantages

- fully scriptable analysis routines
  - consisting of (user´s) custom code, and
  - easy-to-use analysis modules
- same analysis of an exact repetition of a previous experiment can so be done in a few seconds - just by plugging new data into the old (copied) analysis routine



Think - Copy&Paste - Enjoy

Case study:

adding data to a sample

NIR – R database

Case study

- Individual experiments
  - Presentation of sample experiments
    - Paprika – adulterated with corn flour
    - Powdered sugar – adulterated with wheat flour
    - Grape seed extract – adulterated with plant extracts
  - Quick on spot experiment
    - Paprika – adulterated with corn flour
- Importing the data of individual experiments into a database
- Effective data evaluation

Aquap2 practice

https://github.com/bpollner/aquap2

https://github.com/zoltankovacs/NIR2025-HASH_NIR4