



Demonstrations of new tools for spectral data analysis &
NIRS of Waters
(treated with Magnetic Fields)

Bernhard Pollner, Zoltan Kovacs

Aquaphotomics: Understanding Water in the Biological World
at the 5th Kobe University Brussels European Center Symposium
14th October 2014, Brussels, Belgium

Outline & Resume

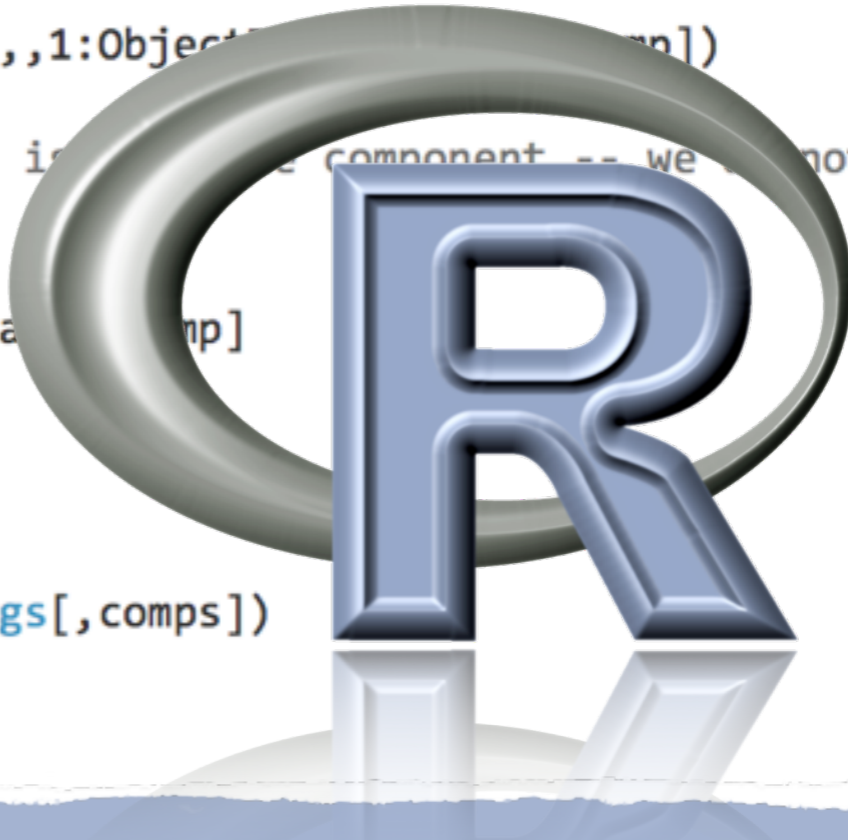
R-Package „Aquaphotomics“

- Experiment-design, Data-Import, Data-Analysis
- Prelim. Results:
 - distinguish between different water species
 - show effect of different magnetic treatments of water

Online „Aquagram-Maker“

- upload data to make Aquagram
- easy and fast

```
the universal input, checking for the class of the input-object
pickPeaks <- function(ObjectToPickPeaks, bandwidth=25, comps=1:4, discrim=FALSE) {    ### universal peak picker
  if (class(ObjectToPickPeaks) == "mvr") {
    allColNames <- colnames(ObjectToPickPeaks$coefficients[, , 1:ObjectToPickPeaks$comps])
    lastName <- allColNames[length(allColNames)]
    if (!is.character(lastName)) {    ## problem if there is no component -- we do not get a name back
      lastName <- "1 comps"
    }
    mat <- ObjectToPickPeaks$coefficients[, , ObjectToPickPeaks$comps]
    dfToPickPeaks <- data.frame(X=mat)
    colnames(dfToPickPeaks) <- lastName
  }
  if (class(ObjectToPickPeaks) == "PCA") {
    dfToPickPeaks <- as.data.frame(ObjectToPickPeaks$loadings[, , comps])
  }
  if (class(ObjectToPickPeaks) == "data.frame") {
    dfToPickPeaks <- ObjectToPickPeaks
  }
  pickResultsList <- pickPeaksInner(dfToPickPeaks, bandwidth, discrim)
  EOF
}
```



R-Package „Aquaphotomics“

a very brief introduction

```
needs the pick results object created by pickPeaks as input; contains the vector that was used for picking !
pickResults <- function (pickResults, onMain="", onSub="", pcaVariances=NULL, customColor=NULL) {    #####
  if (length(pickResults$rawVector) < 1) {
    stop("An Error at plotPickResults: no data")
  }
  a <- colnames(pickResults$rawVector)
  b <- substr(a, stngs$nrCharPrevWL+1, nchar(a))    ## to get rid of the "w" in front of the numbers
  waveLengths <- as.numeric(b)    # so we have the wavelength in the column, and the vectors to be picked f
  res <- pickResults$picks$pickResults
  colPos <- stngs$colPosPeaks
  colNeg <- stngs$colNegPeaks
  positionTable <- res[1: (nrow(res)/2) , ]
  heightTable <- res[((nrow(res)/2)+1):nrow(res) , ]
  Yrange <- range(pickResults$rawVector)[2] - range(pickResults$rawVector)[1]
  onSub <- paste(onSub, ", bw=", pickResults$picks$pickWindow, sep="")
}
```



The R Project for Statistical Computing

About R

- [What is R?](#)
- [Contributors](#)
- [Screenshots](#)
- [What's new?](#)

Download, Packages

[CRAN](#)

R Project

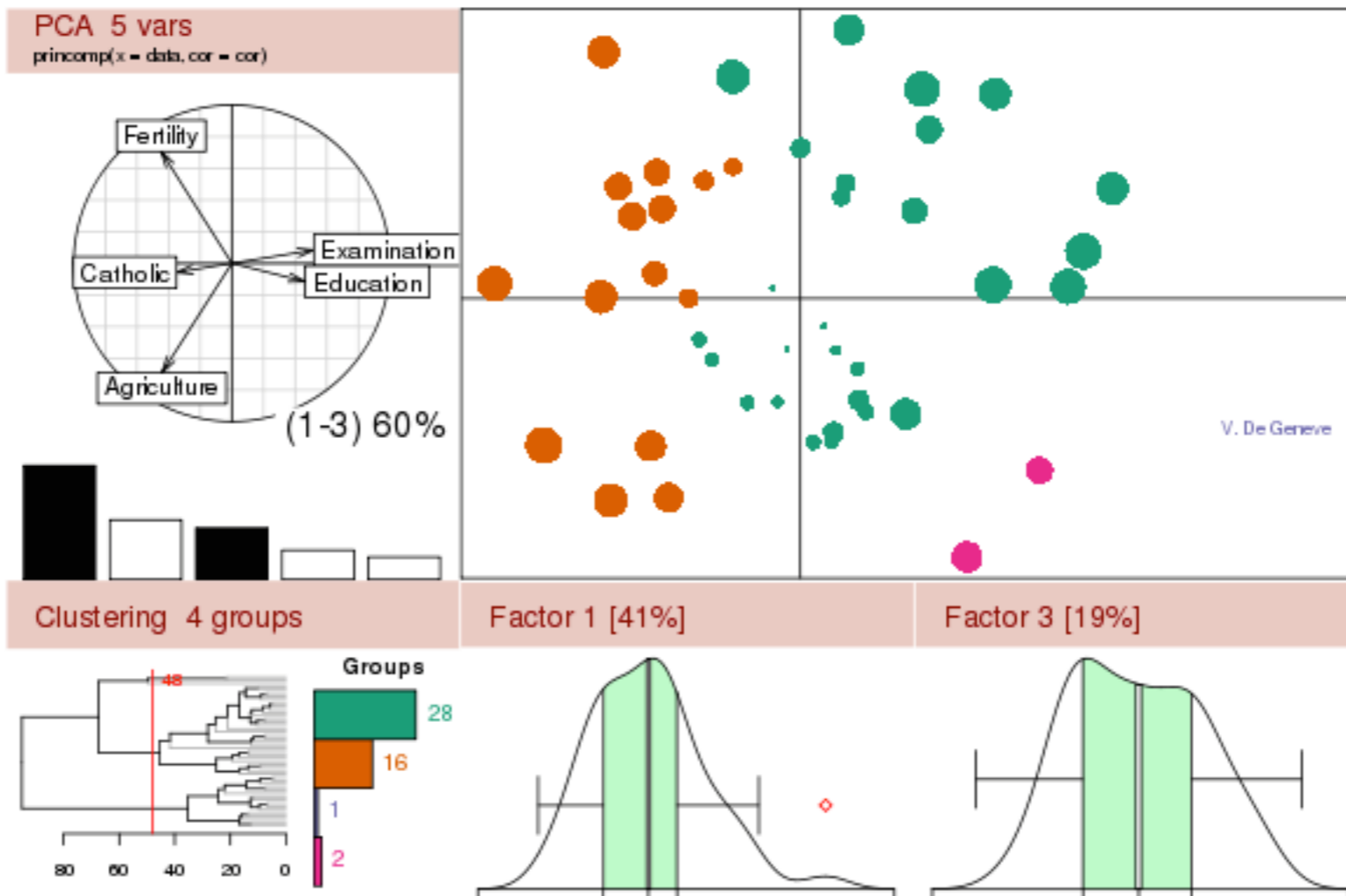
- [Foundation](#)
- [Members & Donors](#)
- [Mailing Lists](#)
- [Bug Tracking](#)
- [Developer Page](#)
- [Conferences](#)
- [Search](#)

Documentation

- [Manuals](#)
- [FAQs](#)
- [The R Journal](#)
- [Wiki](#)
- [Books](#)
- [Certification](#)
- [Other](#)

Misc

- [Bioconductor](#)
- [Related Projects](#)
- [User Groups](#)
- [Links](#)



Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News :

- **R version 3.1.1** (Sock it to Me) has been released on 2014-07-10.
- **R version 3.0.3** (Warm Puppy) has been released on 2014-03-06.
- [The R Journal Vol.5/2](#) is available.
- [useR! 2014](#), took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.
- [useR! 2015](#), will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.



Why R-Project ?

- high level programming language
- open source - freeware !
- cross-platform
- powerful - extendable
- possible: something really new
- widely used in academia

[About R](#)
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)
[Download Packages](#)
[CRAN](#)

[R Project](#)
[Foundation](#)
[Members / Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Development](#)
[Conferences](#)
[Search](#)

[Documentation](#)
[Manual](#)
[FAQ](#)
[The R Journal](#)
[Wiki](#)
[Book](#)
[Certification](#)
[Other](#)

[Science](#)
[Bioconductor](#)
[Related Projects](#)
[User Groups](#)
[Links](#)

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN](#) [mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

- R version 3.1.1 (Sock it to Me) has been released on 2014-07-10.
- R version 3.0.3 (Warm Puppy) has been released on 2014-03-06.
- [The R Journal Vol.5/2](#) is available.
- [useR! 2014](#), took place at the University of California, Los Angeles, USA June 30 - July 3, 2014.
- [useR! 2015](#), will take place at the University of Aalborg, Denmark, June 30 - July 3, 2015.

R-Package „Aquaphotomics“

Experiment Design

- randomize samples
- facilitates time resolved experiments

Data Import

- e.g. straight from binary (.da) file
- controlled source and structure for class- and numerical variables

Data Analysis

- easy splitting and grouping of data
- parallelization of different data pre-treatments
- tools to compare those parallel results

Data Import

```
data.frame: 3676 obs. of 22 variables:
 $ C_ExpName   : Factor w/ 12 levels "DZ1hSeq","DZ12h",...: 1 1 1 1 1
 $ C_ExpClass  : Factor w/ 3 levels "ImmCuv1","ImmCuv2",...: 3 3
 $ Y_SampleNr  : num  1 1 1 1 1 2 2 2 2 2 ...
 $ Y_ConSNr    : int   1 2 3 4 5 1 2 3 4 5 ...
 $ C_Time      : Factor w/ 17 levels "Seq1","Seq2",...: 1 1 1 1 1 1
 $ C_ECRM      : Factor w/ 2 levels "MQ","RM": 1 1 1 1 1 2 2 2 2 2
 $ C_water     : Factor w/ 2 levels "MQ","water": 1 1 1 1 1 2 2 2
 $ C_waterNames: Factor w/ 6 levels "MQ","StU","Ob",...: 1 1 1 1 1
 $ C_Repl      : Factor w/ 13 levels "MQ","R1","R2",...: 1 1 1 1 1
 $ C_Group     : Factor w/ 4 levels "Cont","GDum",...: 4 4 4 4 4
 $ C_all       : Factor w/ 1 level "all": 1 1 1 1 1 1 1 1 1 1
 $ Y_Temp      : num  24.2 24.2 24.2 24.2 24.2 23.4 23.4 23.4 23.4
 $ Y_RelHum    : num  53 53 53 53 53 56 56 56 56 56 ...
 $ C_SampleNr  : Factor w/ 626 levels "1","10","1001",...: 1 1 1
 $ C_ConSNr    : Factor w/ 7 levels "1","2","3","4",...: 1 2 3 4
 $ C_Temp      : Factor w/ 8 levels "22","22.5","23",...: 5 5 5 5 5
 $ C_RelHum    : Factor w/ 9 levels "54","57","60",...: 1 1 1 1 1 2
 $ Time        : POSIXct, format: "2014-07-09 09:14:51" "2014-07-0
 $ absTime     : num  55695 55695 55696 55696 55697 ...
 $ chron       : int   1 2 3 4 5 6 7 8 9 10 ...
 $ numRep      : 'data.frame': 3676 obs. of 13 variables:
 ..$ C_ExpName   : chr  "blue4" "blue4" "blue4" "blue4" ...
 ..$ C_ExpClass  : num  3 3 3 3 3 3 3 3 3 3 ...
 ..$ C_Time      : chr  "blue4" "blue4" "blue4" "blue4" ...
 ..$ C_ECRM      : num  1 1 1 1 1 2 2 2 2 2 ...
 ..$ C_water     : num  1 1 1 1 1 2 2 2 2 2 ...
 ..$ C_waterNames: num  1 1 1 1 1 2 2 2 2 2 ...
 ..$ C_Repl      : chr  "blue4" "blue4" "blue4" "blue4" ...
 ..$ C_Group     : num  4 4 4 4 4 2 2 2 2 2 ...
 ..$ C_all       : num  1 1 1 1 1 1 1 1 1 1 ...
 ..$ C_SampleNr  : chr  "blue4" "blue4" "blue4" "blue4" ...
 ..$ C_ConSNr    : num  1 2 3 4 5 1 2 3 4 5 ...
 ..$ C_Temp      : chr  "#FC2100" "#FC2100" "#FC2100" "#FC2100"
 ..$ C_RelHum    : chr  "#0000FF" "#0000FF" "#0000FF" "#0000FF"
 $ NIR         : num [1:3676, 1:4200] 0.0199 0.0201 0.0203 0.0204
 ..- attr(*, "dimnames")=List of 2
 .. ..$ : chr  "1-Seq1-MQ0" "1-Seq1-MQ0.1" "1-Seq1-MQ0.2" "1-Seq1
 .. ..$ : chr  "w400" "w400.5" "w401" "w401.5"
```

- standardized data structure

- encapsulated and „transportable“ color-coding

- nice to look at: view of NIR-data is minimized (but of course still possible)

- (regress on absTime)

Data Analysis

- splitting / slicing / grouping of data
- recyclable, separate data-analysis procedure file
- strict separation of calculations and plotting
- customizable parallel processing (if applicable; on UNIX)
- various methods implemented
- and much more already on the „list“

sdrc_calc_plot_PLSR_LDA	Calculate and / or plot a pls-lda.
sdrc_groupAvgSubtract	Subtractions in raw-spectra
sdrc_plotAquagram	Plot an Aquagram
sdrc_plotCompScores	Plot comparison scores; advanced grouping
sdrc_plotCompScores_simple	Plot comparison scores; simple grouping
sdrc_plotGroupAvgSubtract	Plot averaged subtracted spectra
sdrc_plotPCAScores	Plot PCA scores
sdrc_plotRaw	Plot Spectra
sdrc_plotRMSEx	Error / calibration / validation plots of PLSR models
sdrc_plotSimcaClassDist	Plot SIMCA interclass distances (sdrc)
sdrc_plotVector	Plot Vector
sdrc_plotVectorCollection	Plot Vector Collection
sdrc_printSvmTable	Classify using Support Vector Machines
sdrc_showGroups	Show SDRC set by grouping
sdrc_selectWLS	Select Wavelengths
sdrc_showAllClassRepresentations	Show the character representation of a factor
sdrc_showSingleClassRepresentation	Show the character representation of a factor
sdrc_nca_mmp	Plot SIMCA interclass distances (dataset)
sdrc_splitSequence	Manually split within sequence
sdrc_ssc	Select rows according to class-membership

Data Analysis - Implemented Methods

Data Pre-Treatment

- smoothing & derivatives, SNV, MSC, EMSC
- special Aquaphotomics data pretreatments (parallelized)

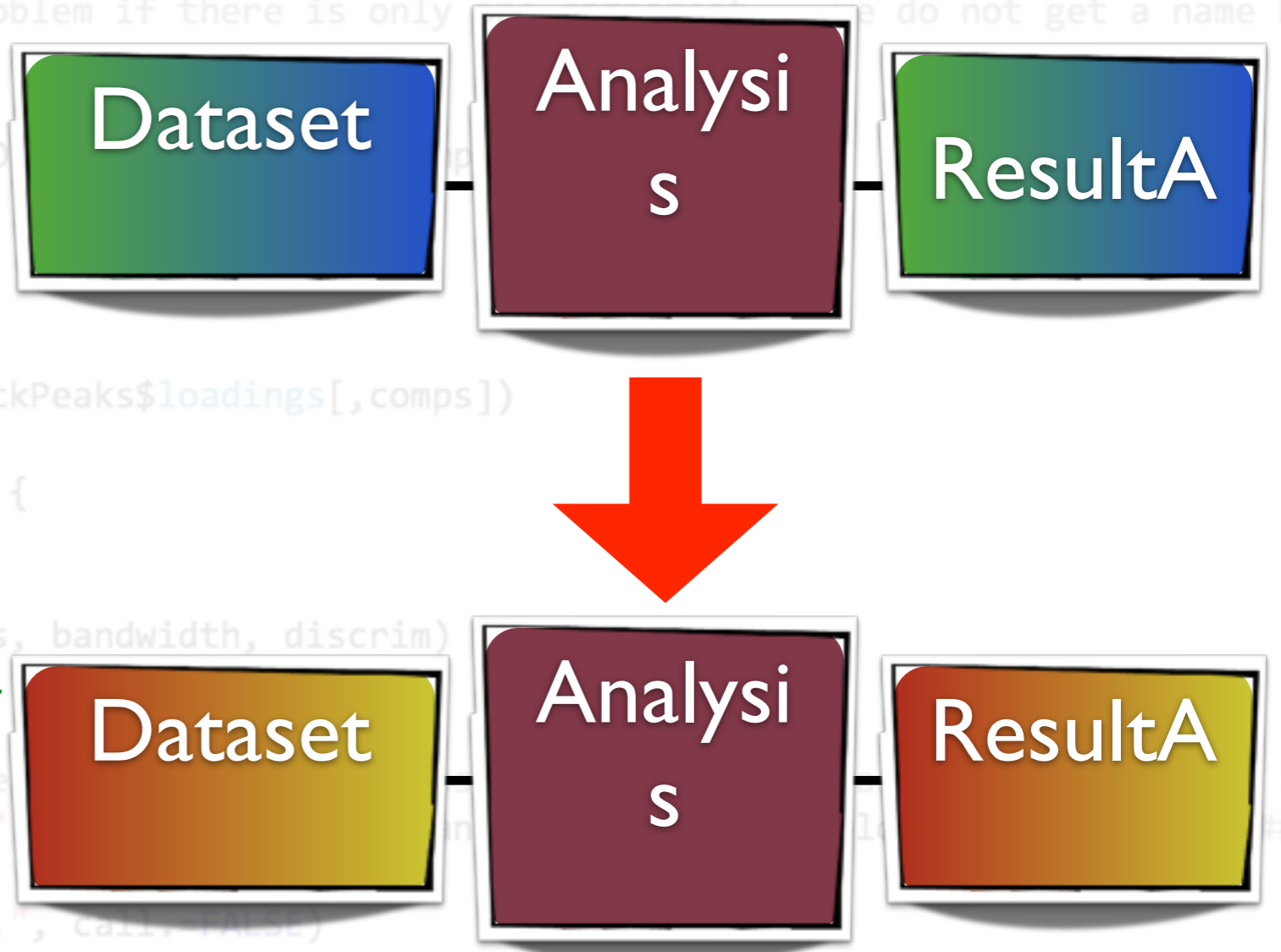
Evaluation / Analysis:

- subtractions in raw-spectra
- PCA, PLSR
- SIMCA
- (SVM, bootstrap, ...)
- Aquagram !


sdrc_calc_plot_LDA	Calculate and / or plot a pls lda
sdrc_groupAvgSubtract	Subtractions in raw-spectra
sdrc_plotAquagram	Plot an Aquagram
sdrc_plotCompScores	Plot comparison scores; advanced grouping
sdrc_plotCompScores_simple	Plot comparison scores; simple grouping
sdrc_plotGroupAvgSubtract	Plot averaged subtracted spectra
sdrc_plotPCAScores	Plot PCA scores
sdrc_plotRaw	Plot Spectra
sdrc_plotRMSEx	Error / calibration / validation plots of PLSR models
sdrc_plotSimcaClassDist	Plot SIMCA interclass distances (sdrc)
sdrc_plotVector	Plot Vector
sdrc_plotVectorCollection	Plot Vector Collection
sdrc_printSvmTable	Classify using Support Vector Machines
sdrc_showGroups	Show SDRC set by grouping
selectWLS	Select Wavelengths
showAllClassRepresentations	Show the character representation of a factor
showSingleClassRepresentation	Show the character representation of a factor
simca_mmp	Plot SIMCA interclass distances (dataset)
splitSequence	Manually split within sequence
select	Select observations (rows) according to class membership

R-Package „Aquaphotomics“

- **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



Preliminary Results

Different Water Species

Spectrometer:

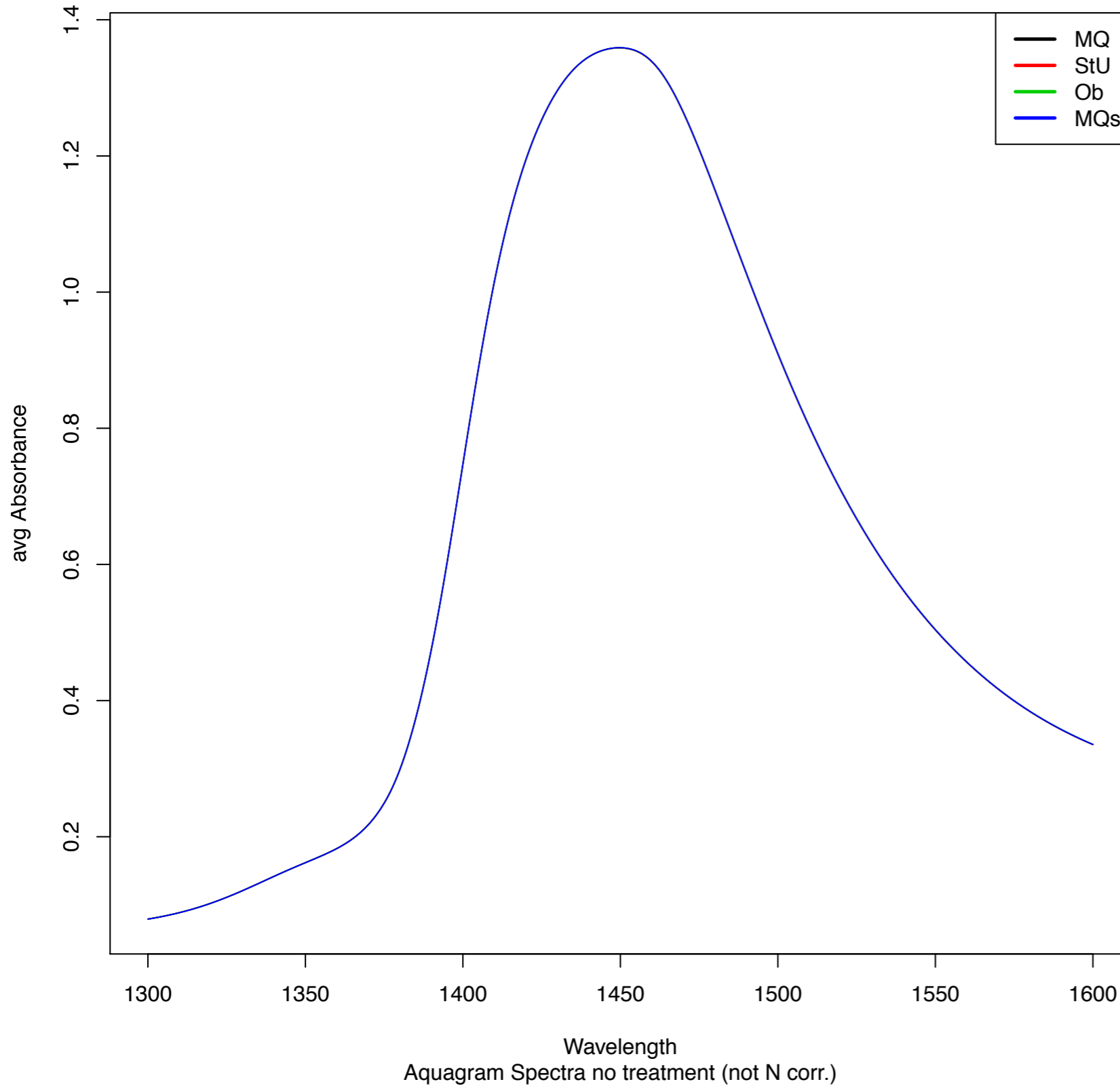
- FOSS XDS
- transmittance, pathlength = 1mm
- 400nm - 2500nm; $\Delta=0.5\text{nm}$
- $T = 28.6^\circ\text{C}$

- up to 5 different water species
- MilliQ as standard
- Abbreviations showing origin of water
 - StU: St. Ulrich
 - Ob: Oberndorf
 - MQ: MilliQ

- Analysis in 1st overtone of OH/water (1300-1600nm)

3(4) Different Water Species - no Treatment

Fusion-AllWaters no Treatment, ProxCuv@1300-to-1600

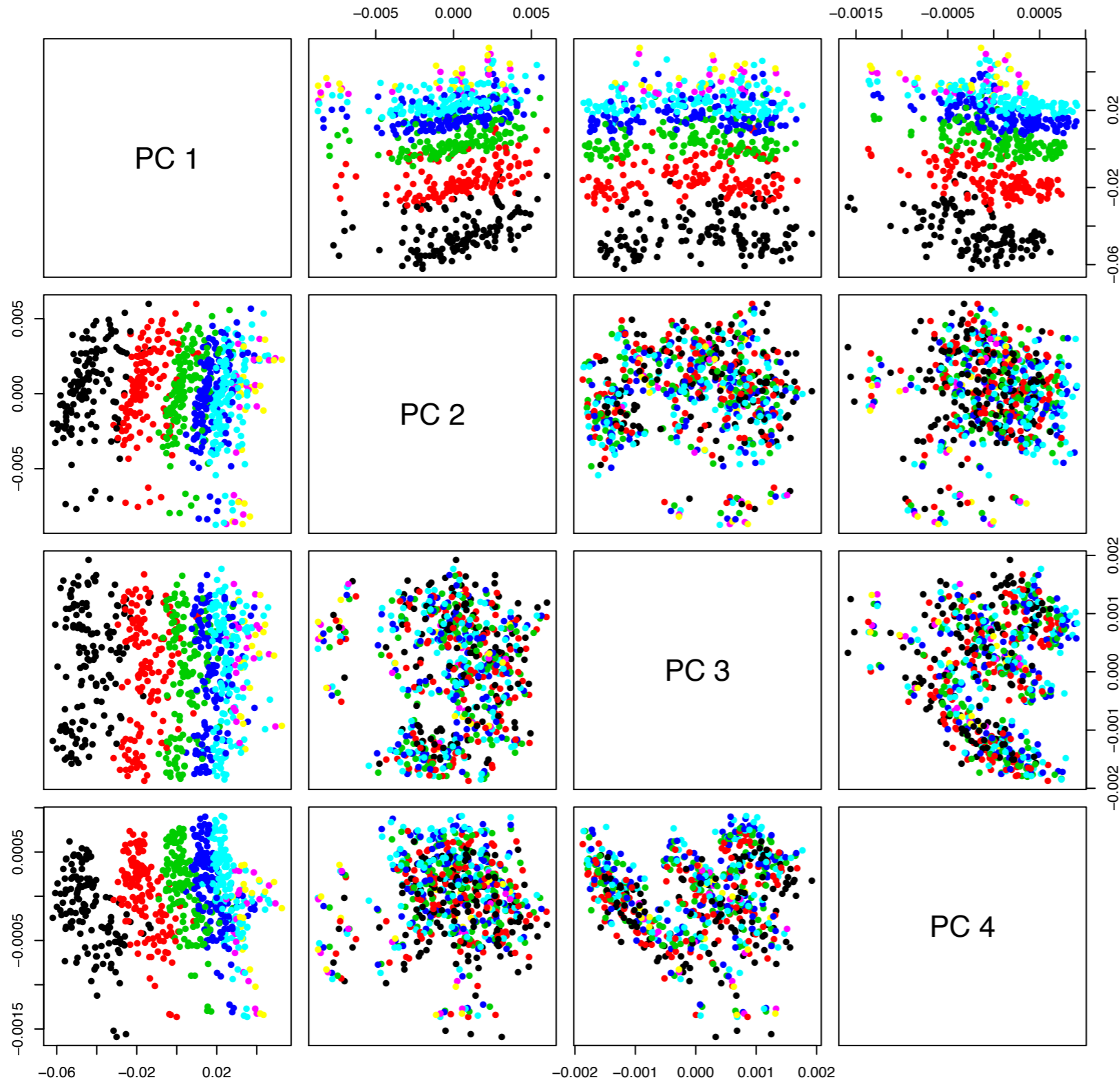


N= 216, 254, 190, 40

● raw spectra

3(4) Different Water Species - no Treatment

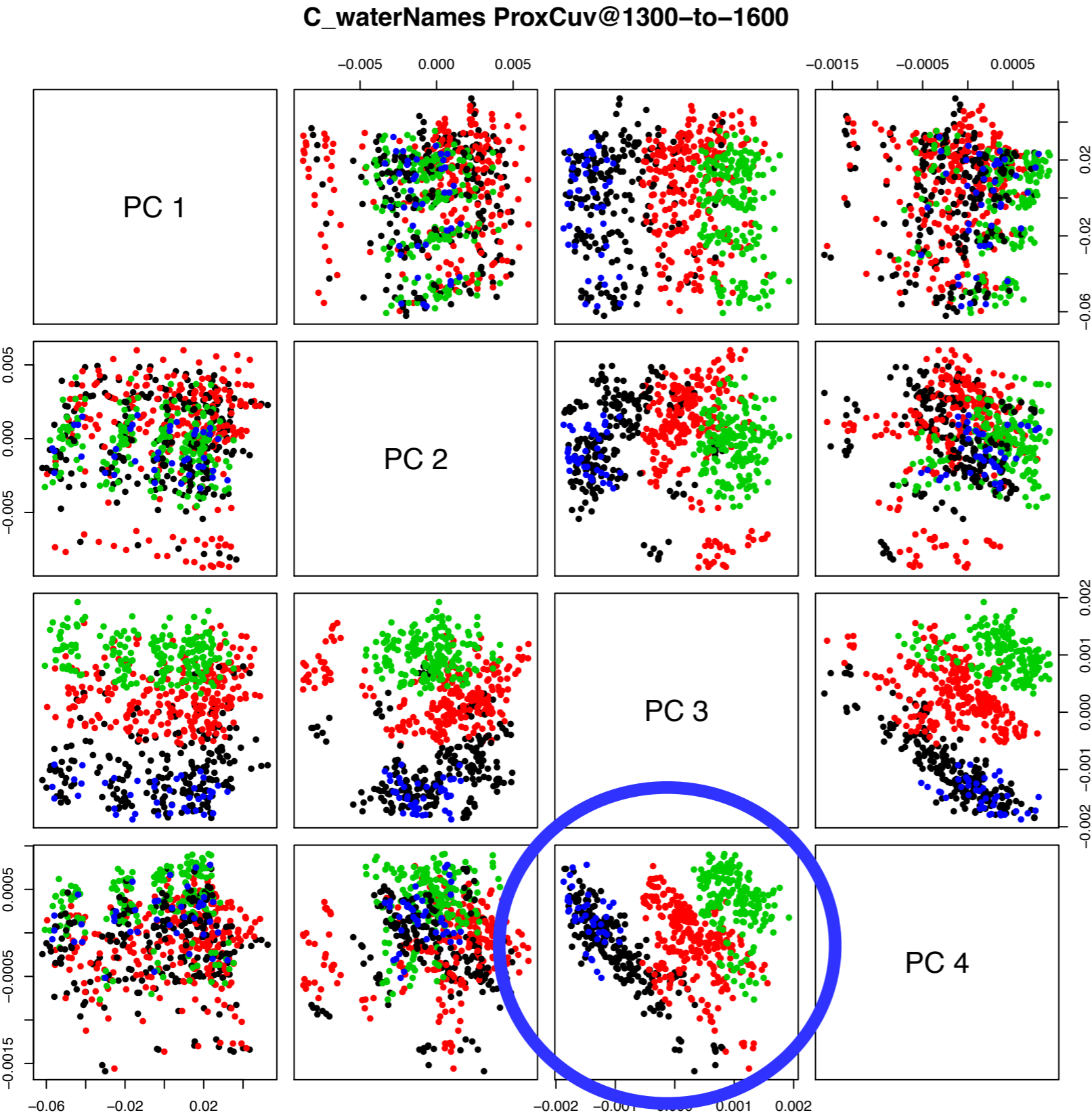
C_ConSNr ProxCuv@1300-to-1600



PCA:

- biggest effect: consecutive scans (98.7%)

3(4) Different Water Species - no Treatment

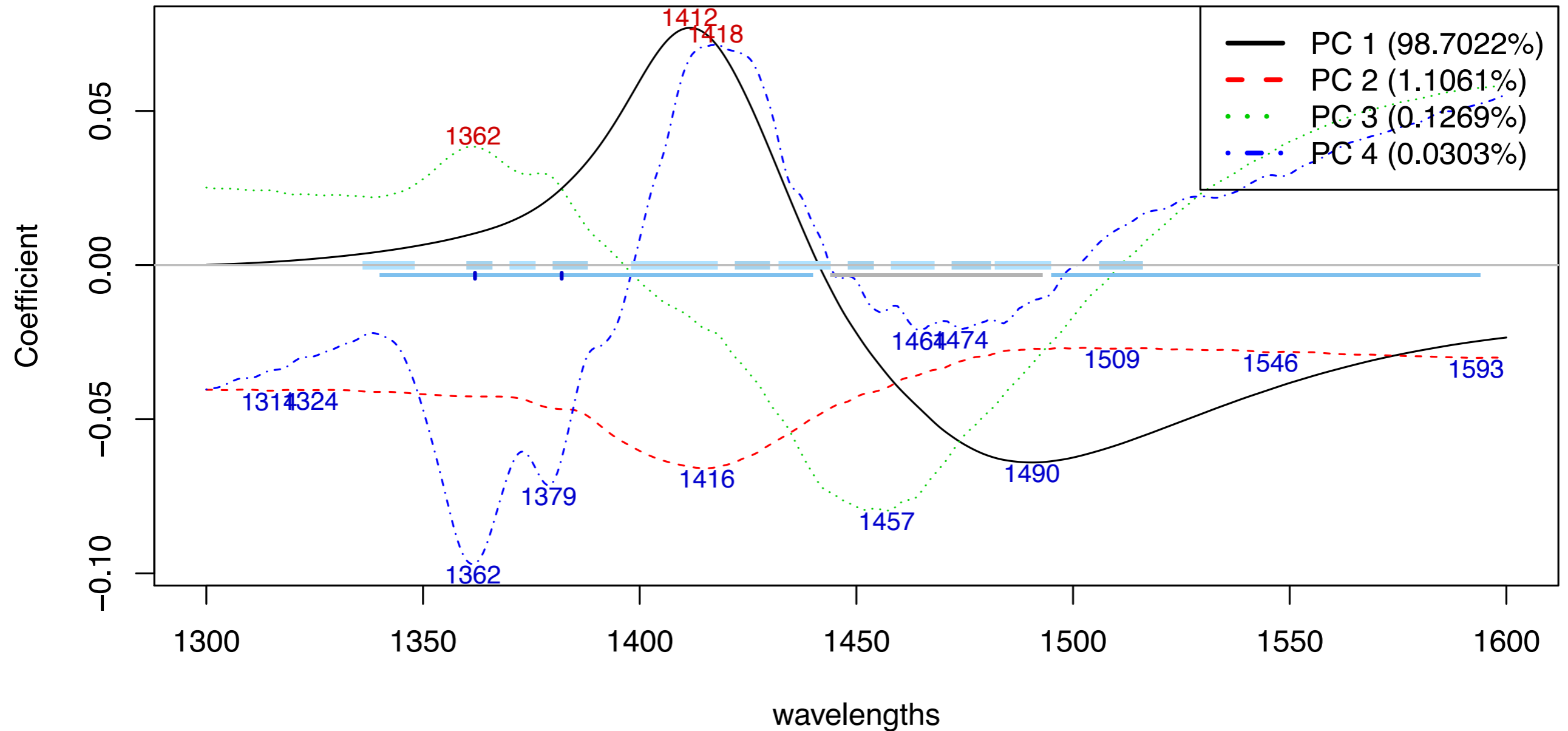


PCA:

- single water species
- PC3: 0.127%
- PC4: 0.030%

Loadings

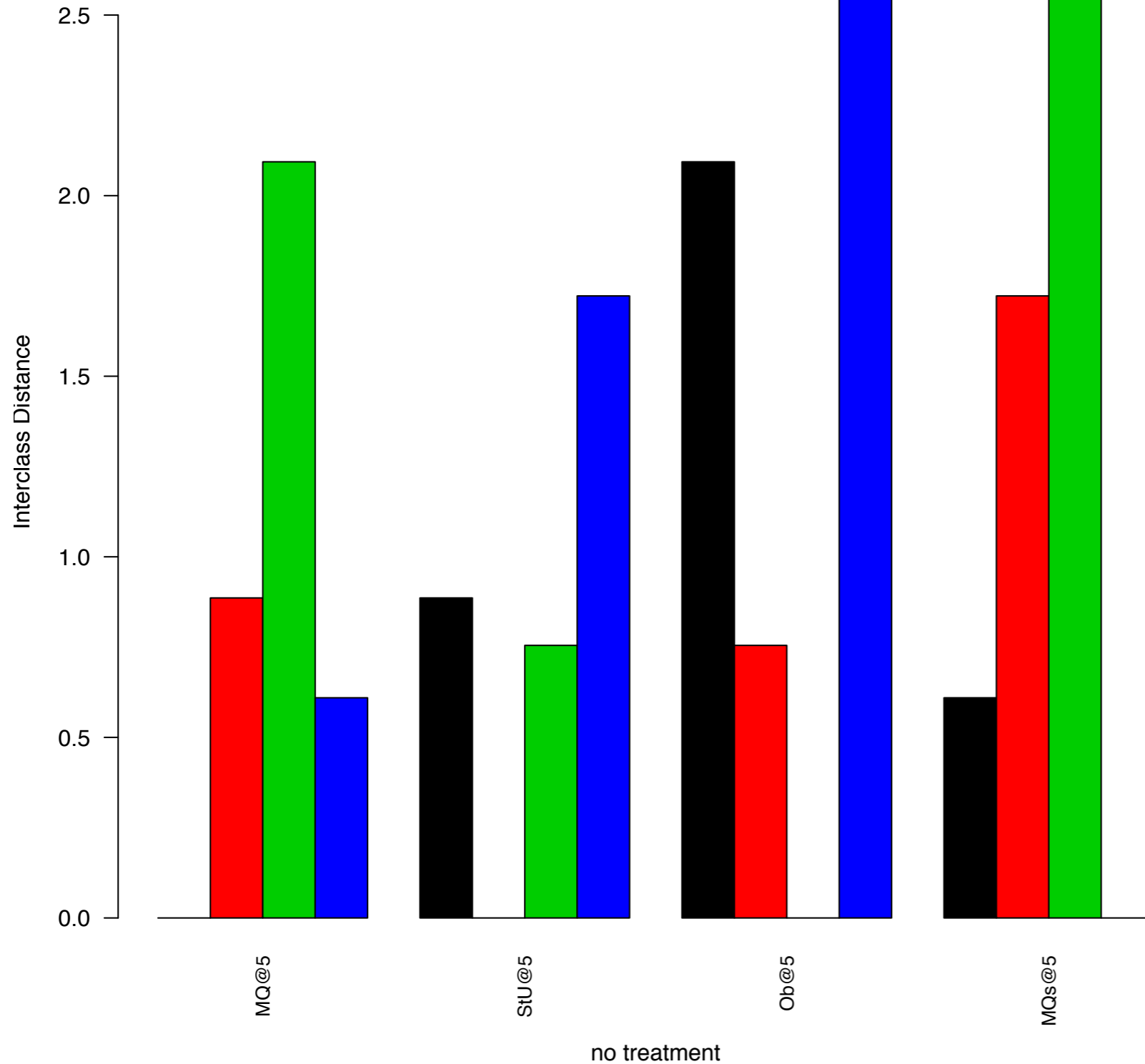
ProxCuv@1300-to-1600



no treatment PCA , bw=25 , 2 PCs for 99% var.

SIMCA - Interclass Distances

MQ StU Ob MQs
no Treatment ProxCuv@1300-to-1600

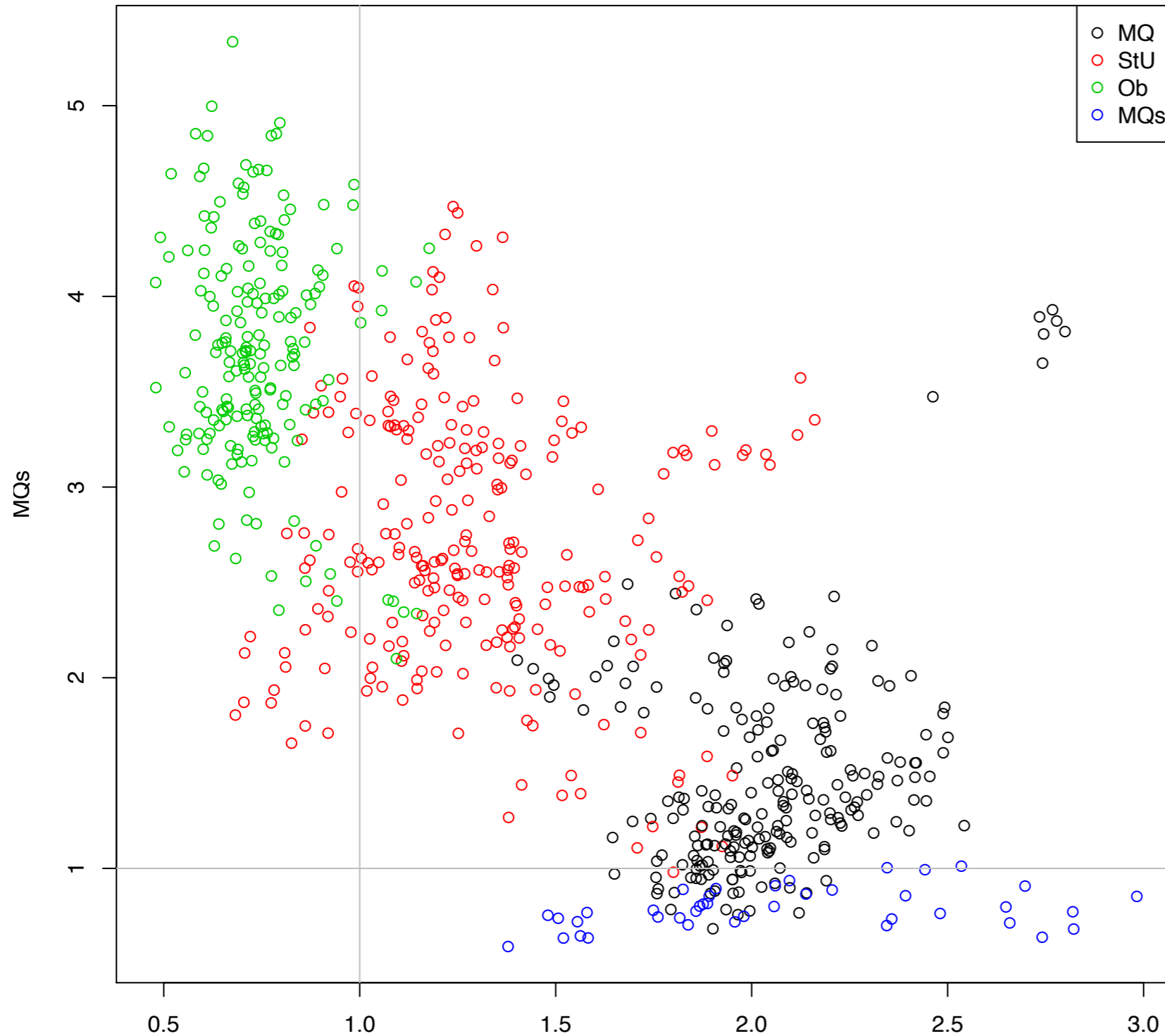


observe:

- MQs and MQ - same water in different containers

SIMCA - Interclass Distances

no Treatment ProxCuv@1300-to-1600



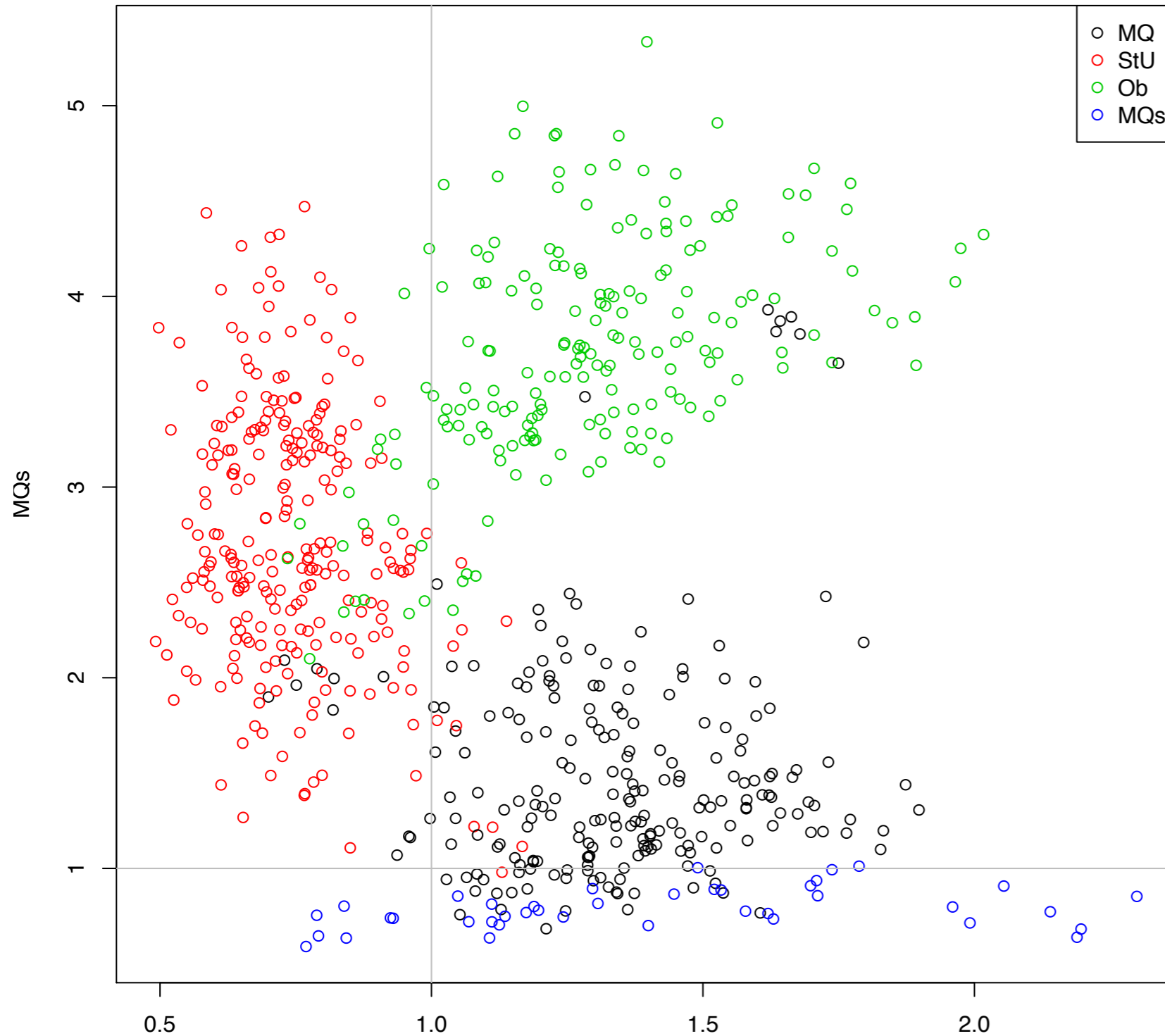
Ob / MQs:

- nice separation

Ob@5
(classic) no treatment class distance (Ob, MQs) = 2.72

SIMCA - Interclass Distances

no Treatment ProxCuv@1300-to-1600



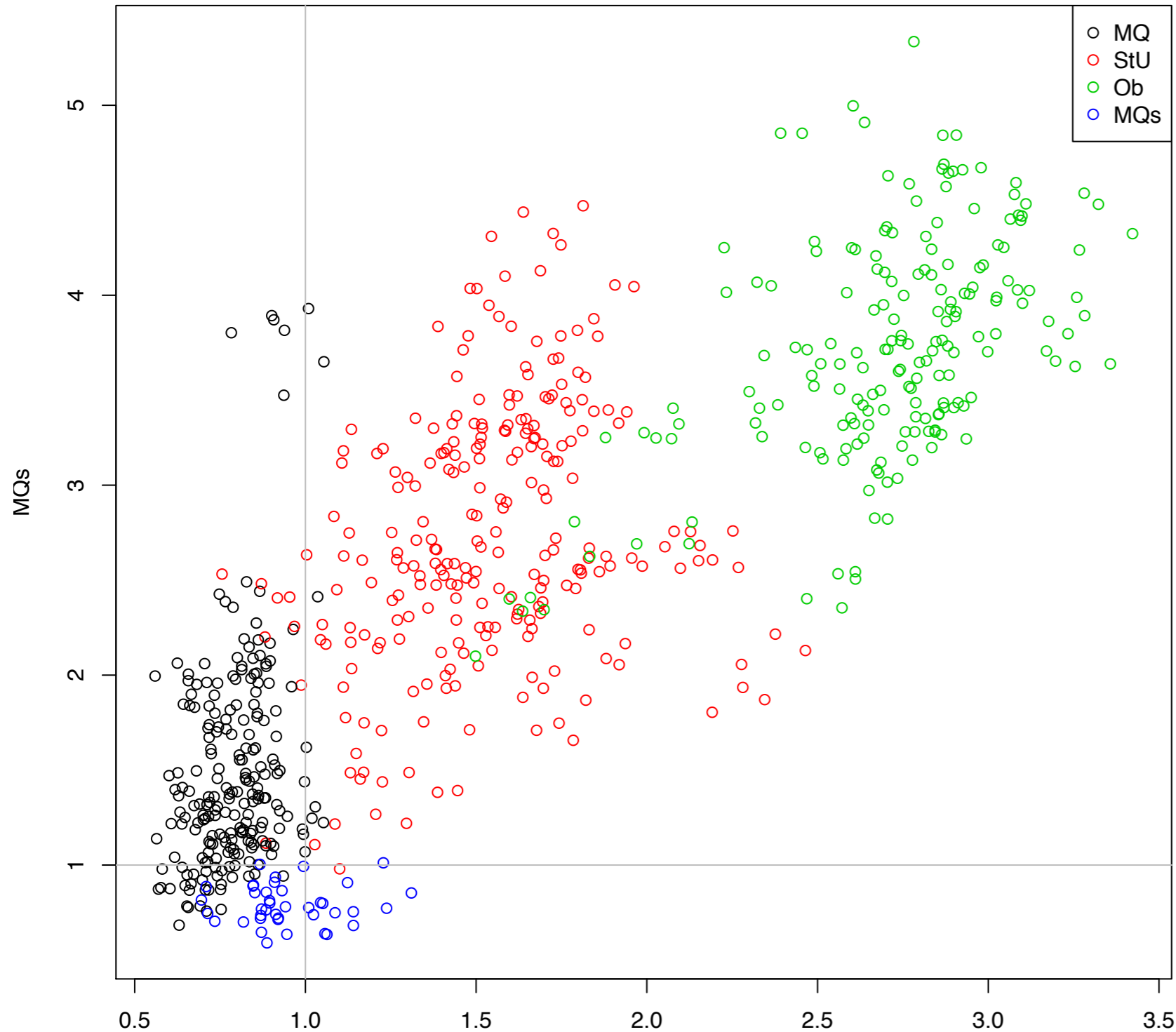
StU / MQs:

- nice separation

StU@5
(classic) no treatment class distance (StU, MQs) = 1.72

SIMCA - Interclass Distances

no Treatment ProxCuv@1300-to-1600



MQ / MQs:

- same water - different container

MQ@5
(classic) no treatment class distance (MQ, MQs) = 0.61

SIMCA - Classification

Apparent error rate 0.0371

	Predicted			
Actual	MQ	StU	Ob	MQs
MQ	208	7	0	1
StU	1	253	0	0
Ob	0	5	185	0
MQs	12	0	0	28

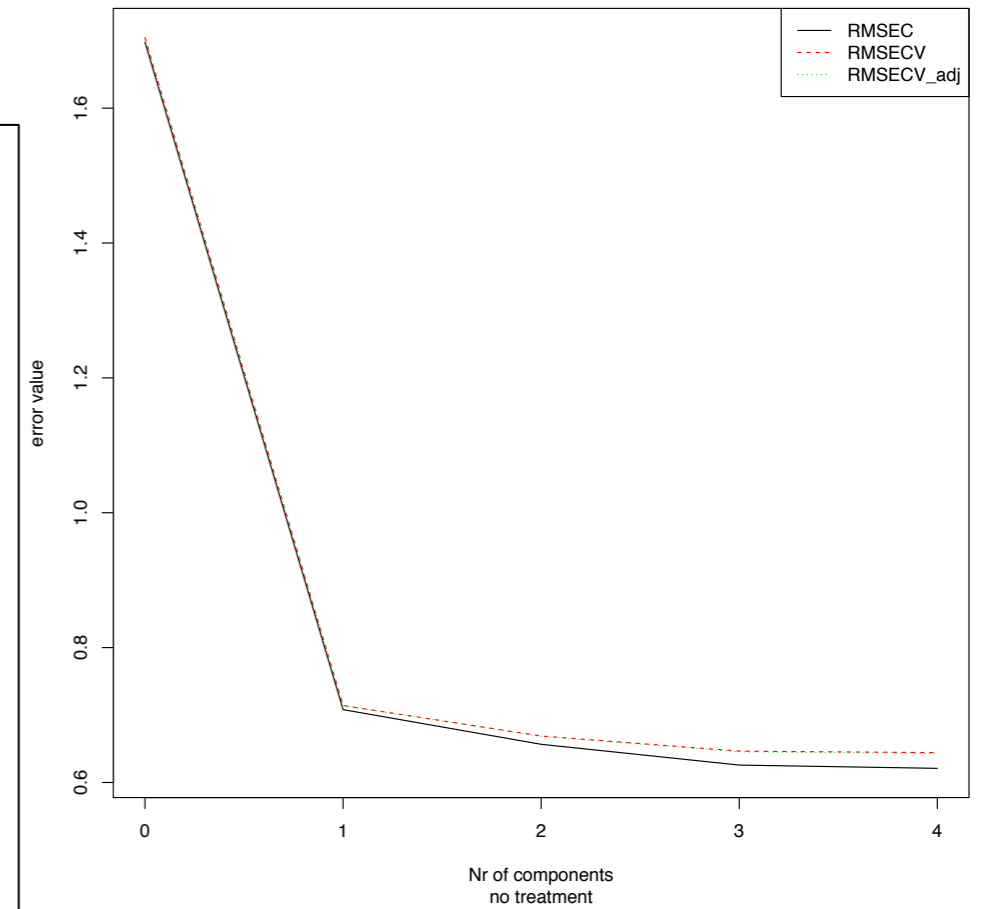
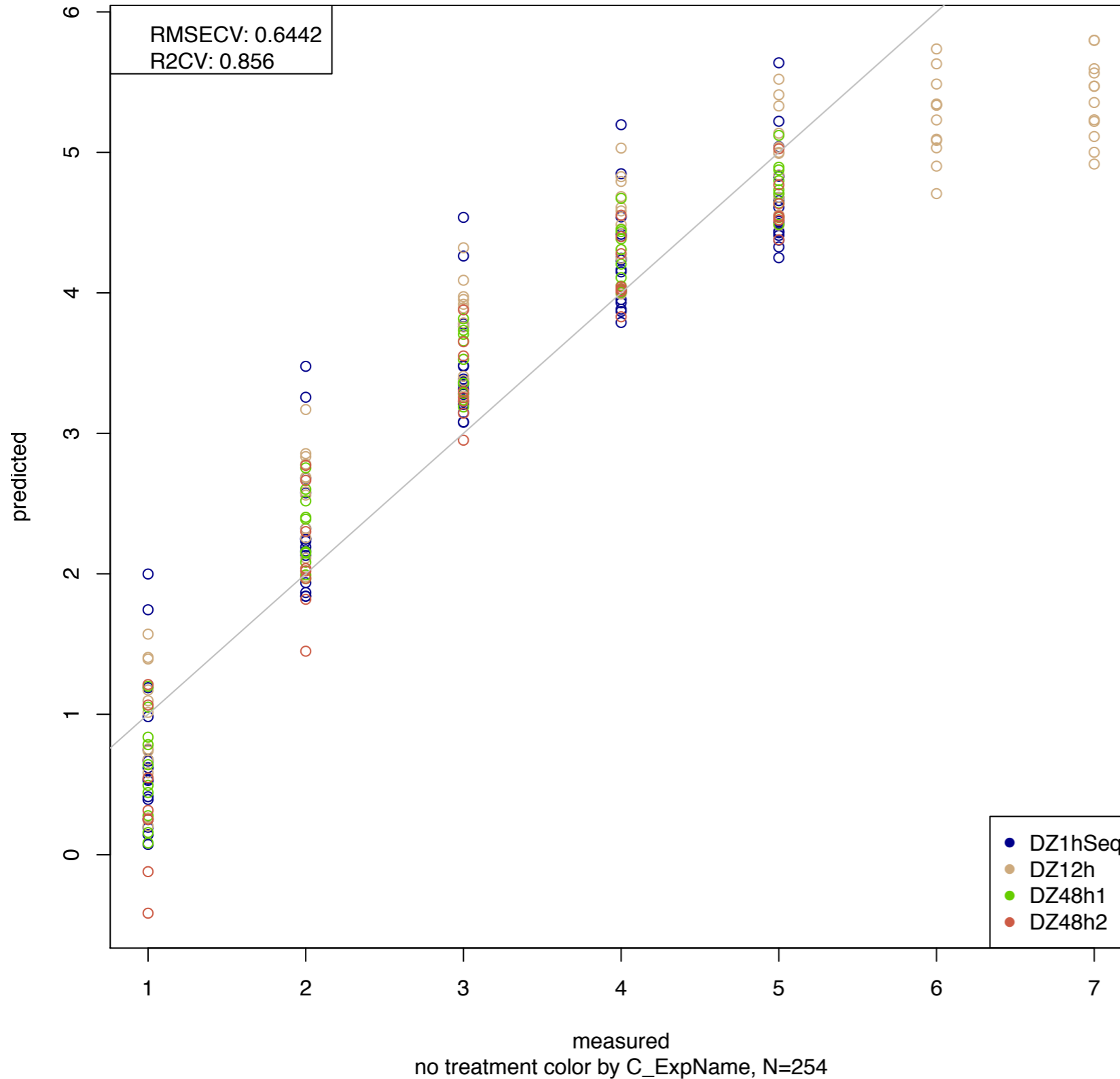
50% new data - Apparent error rate 0.0657

	Predicted			
Actual	MQ	StU	Ob	MQs
MQ	107	3	0	0
StU	1	127	0	0
Ob	0	3	89	0
MQs	15	1	0	4

PLSR (number of consec. scan)

StU@1300-to-1600

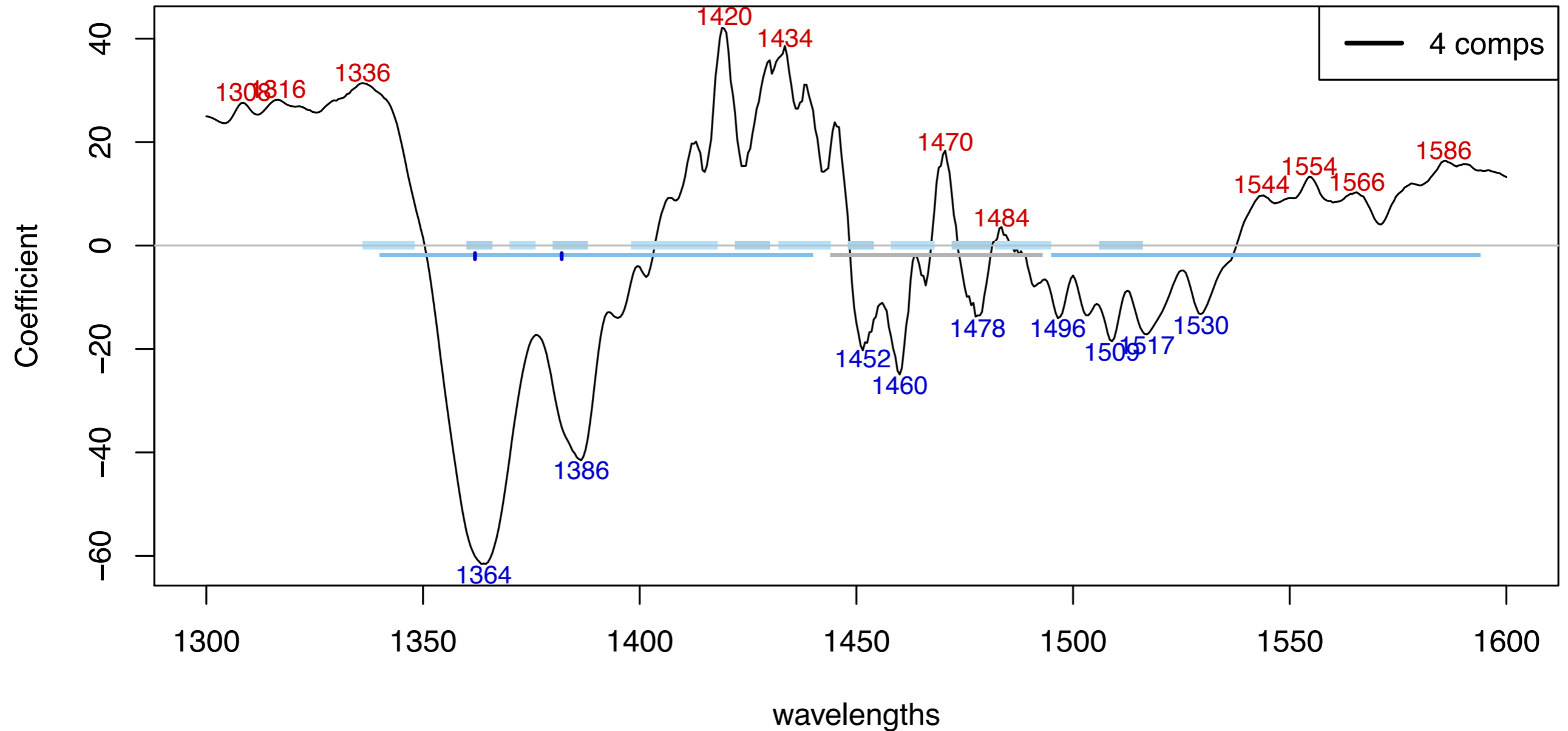
StU@1300-to-1600 - Validation



St. Ulrich Water

PLSR (number of consec. scan)

StU@1300-to-1600

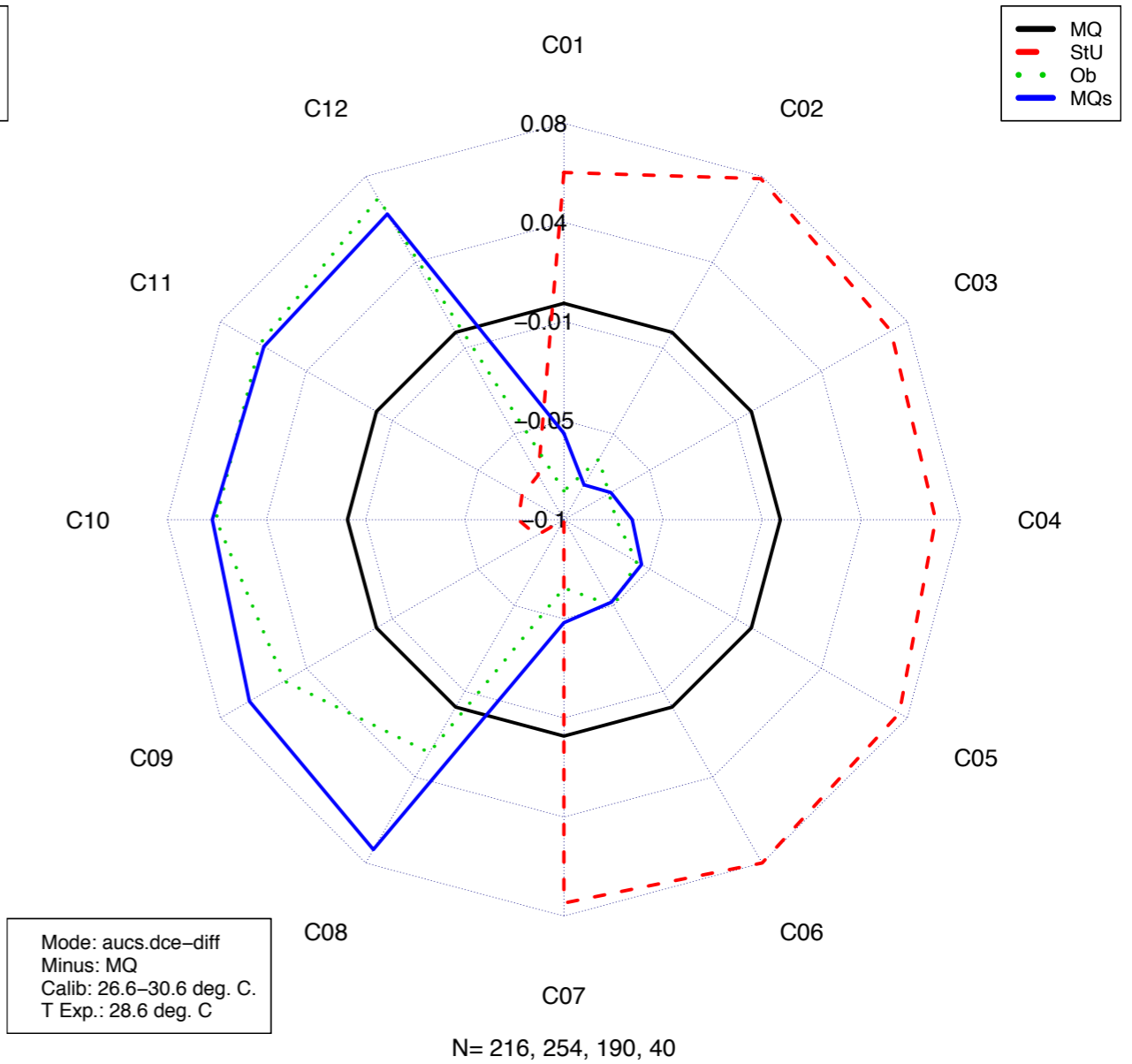
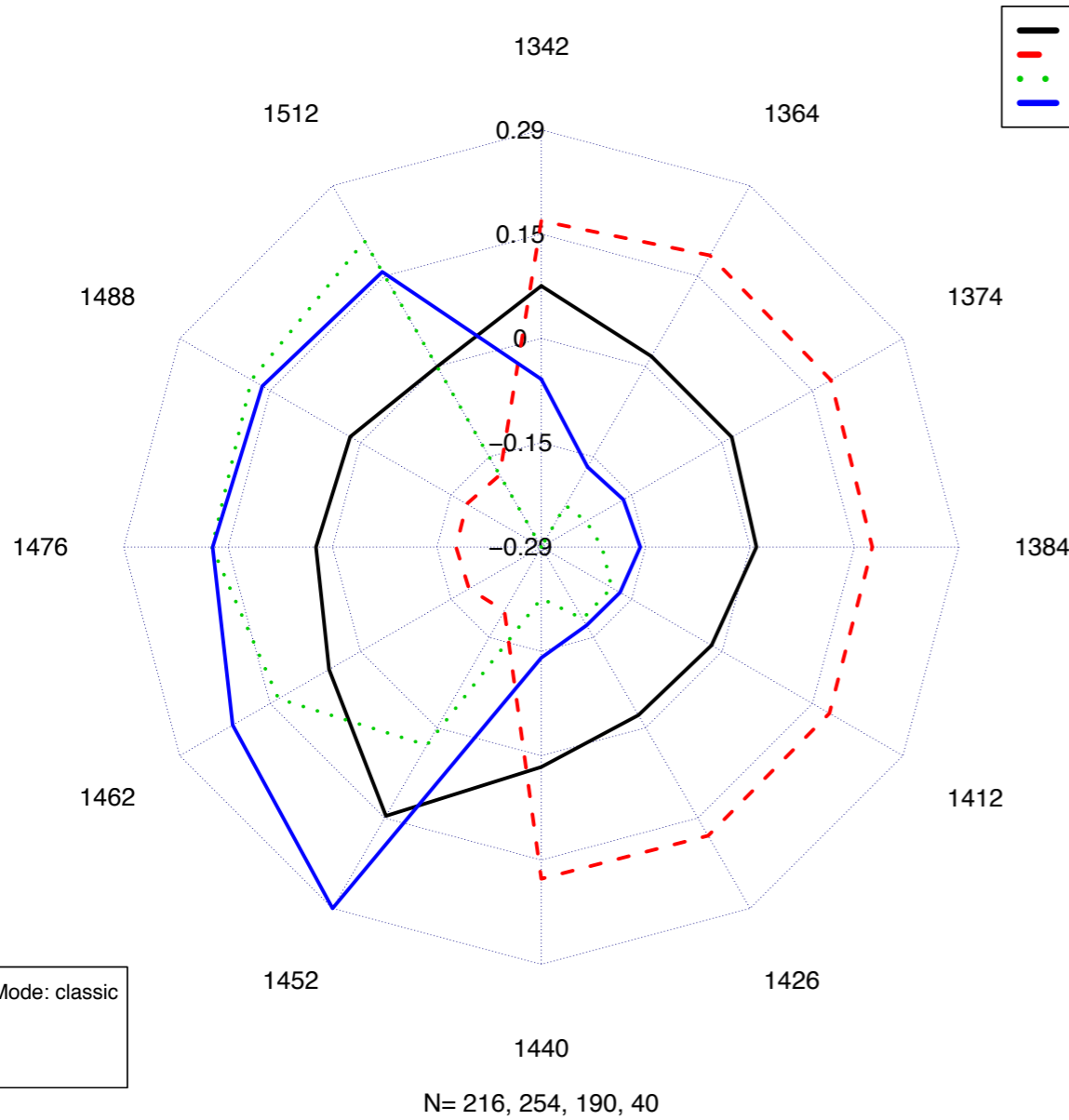


no treatment mvr regressed on Y_ConSNr , RMSECV= 0.6442 , R2CV= 0.856 , bw=25

Aquagram

Fusion-AllWaters no Treatment, ProxCuv@1300-to-1600

Fusion-AllWaters no Treatment, ProxCuv@1300-to-1600



Classic

aucs.dce-diff

Aquagram

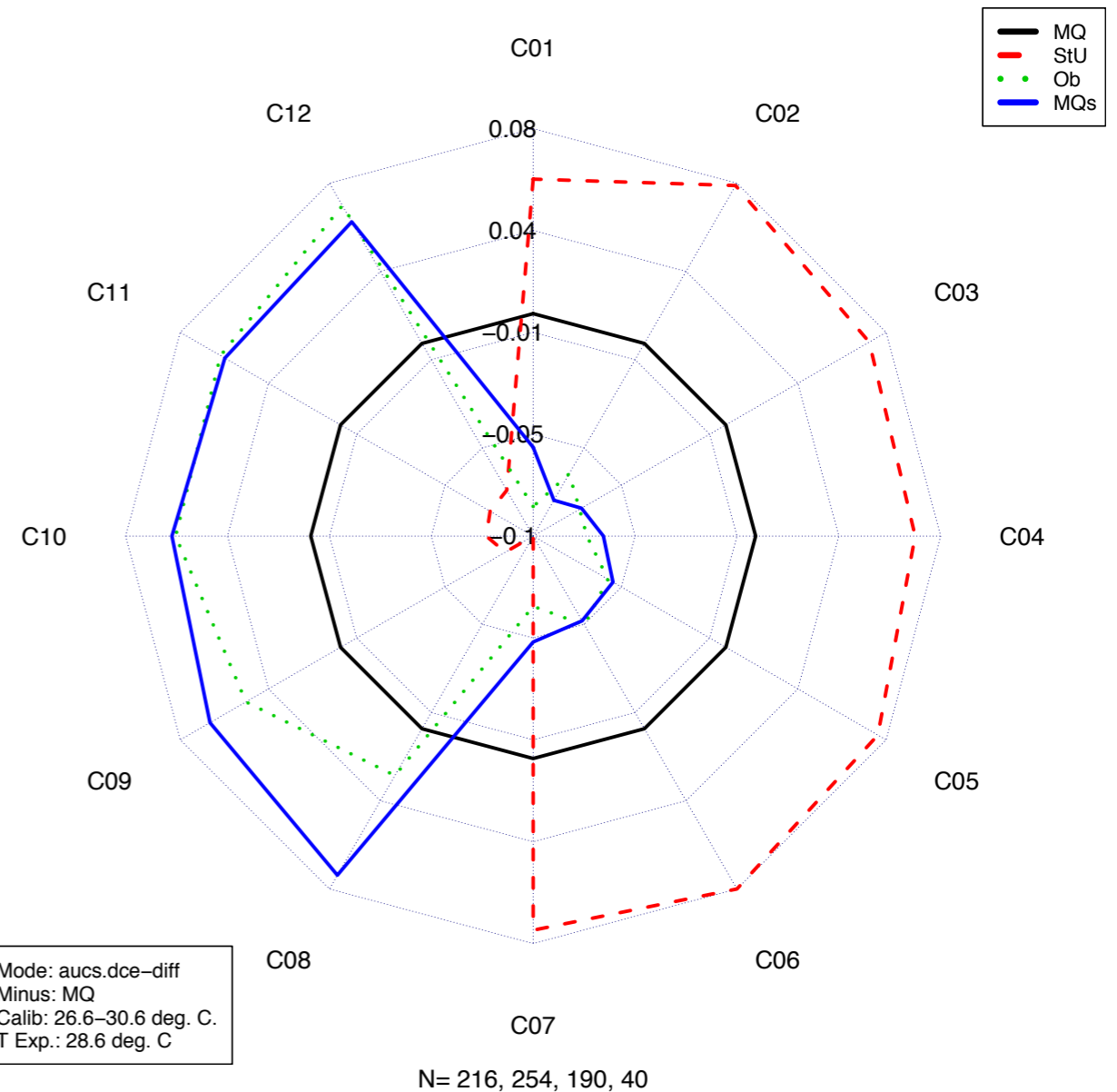
Fusion-AllWaters no Treatment, ProxCuv@1300-to-1600

New Aquagram Algorythm

- completely different mathematics,
- but showing high similarity with results of „classic“ code
- independent, stable scale
- unit: °C

now possible:

- use fix-scale aquagrams in time-resolved experiments



no treatment grouping by C_waterNames (not N corr.)

aucs.dce-diff

Effect of Magnetic Fields on Water

Spectrometer:

- FOSS XDS
- transmittance, pathlength = 1mm
- 400nm - 2500nm; $\Delta=0.5\text{nm}$
- $T = 28.6^\circ\text{C}$

- 2 different kinds of magnetic treatment
- MilliQ as standard
- Analysis in 1st overtone of OH/water (1300-1600nm)
- Preliminary Results: as this was an industry-financed experiment, I can not give out hardcopies of even the preliminary results before we published them — I humbly ask for your understanding. Thank you.

water-temperature

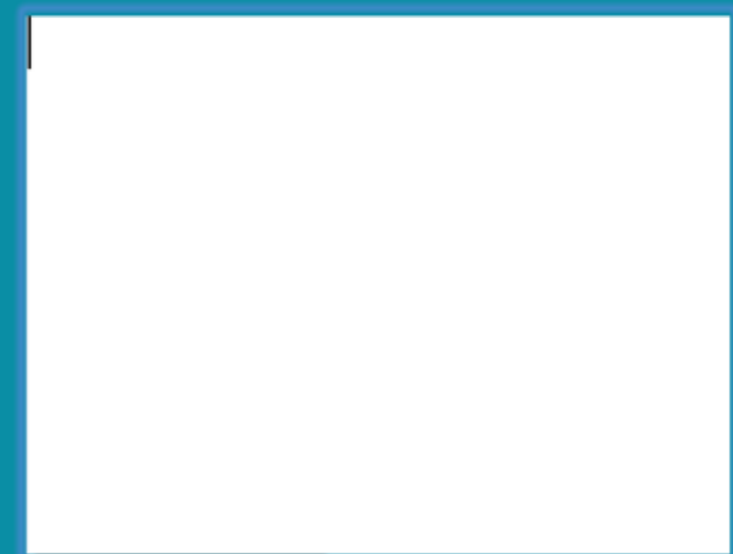
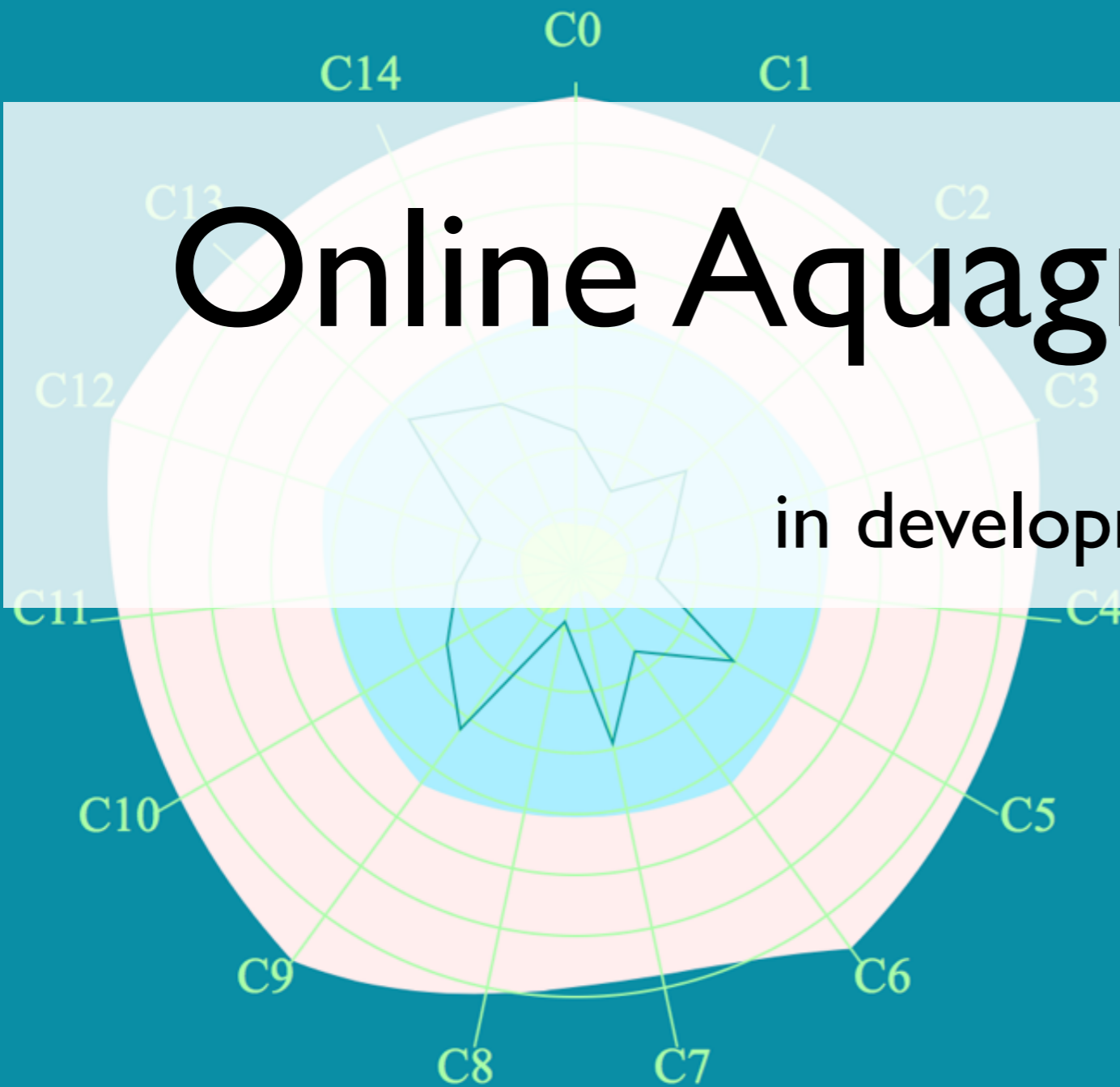
L1

©.Tsenkova: 1st overtone free OH stretch (OH-(H₂O)₄)

Additional information should be provided by the site editors.

Online Aquagram Maker

in development ...

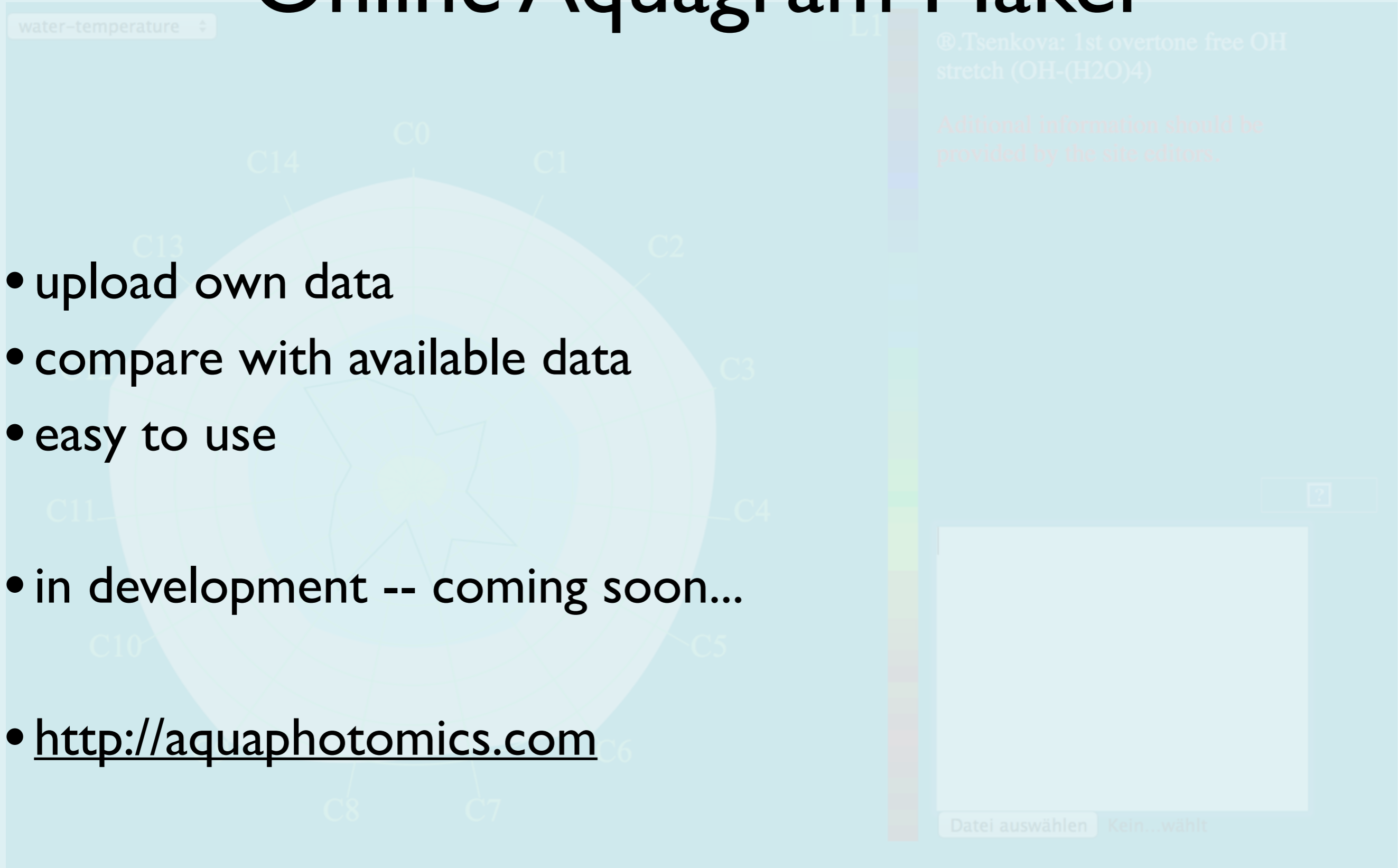


Datei auswählen

Kein...wählt

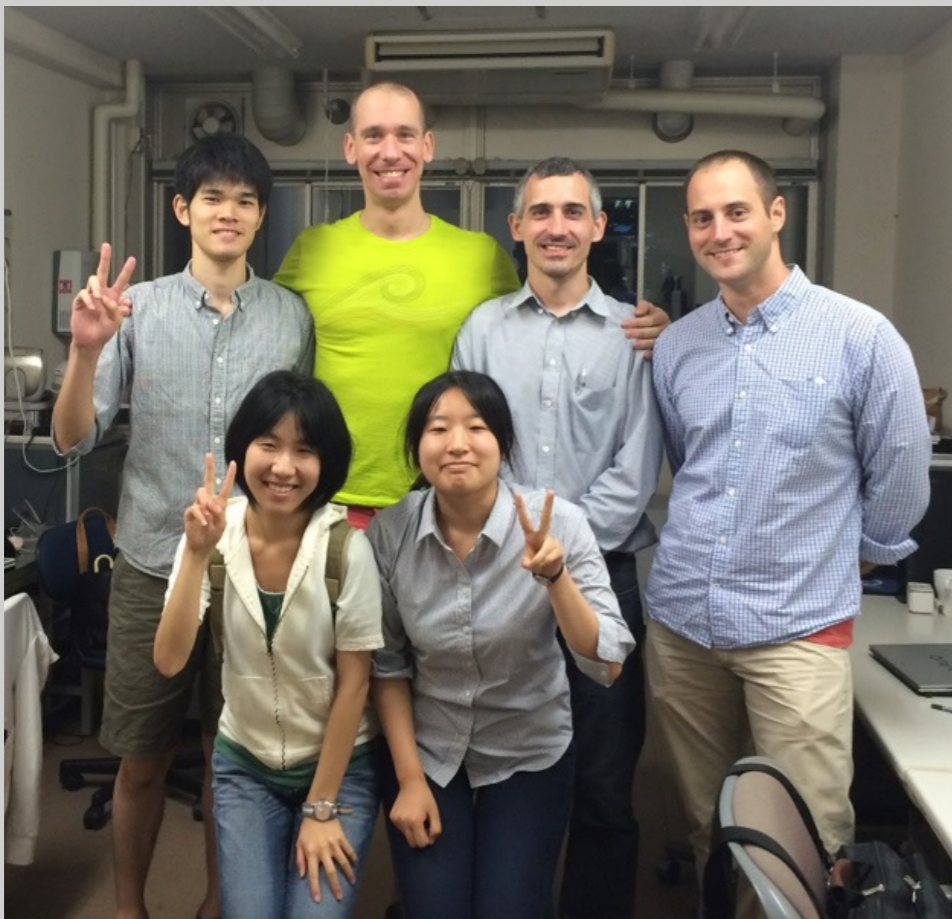
Online Aquagram Maker

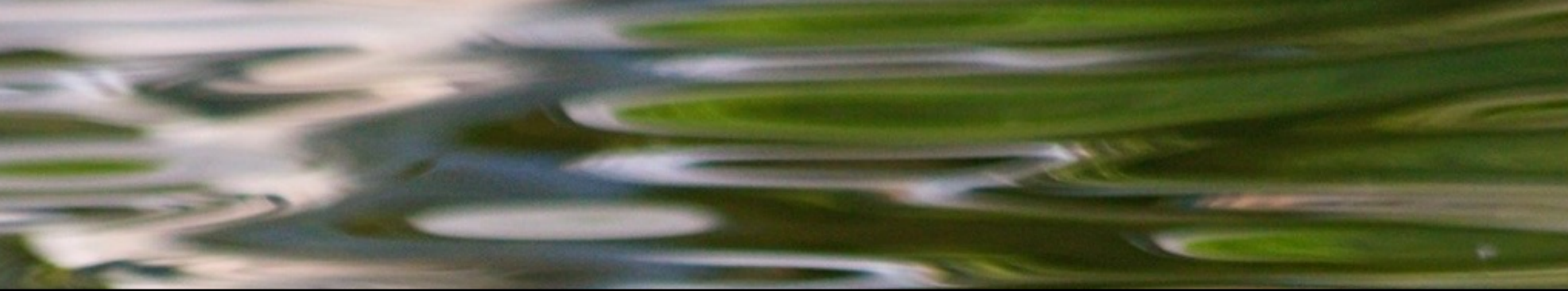
- upload own data
- compare with available data
- easy to use
- in development -- coming soon...
- <http://aquaphotomics.com>



Acknowledgements

- Prof. Roumiana Tsenkova & Zoltan Kovacs
- Zoltan Kovacs
- George Bázár & Alexandar Slavchev
- everybody at the Biomeasurement Technology Laboratory





Thank You !