Calibration Practices Beyond 50 Samples

David Honigs, Ph.D.

Perten Instruments, Springfield, IL

Perten INSTRUMENTS a PerkinElmer company

Illustrating Data Set
Sugar Beet Brei

Brei – cut/chopped sugar beets

12,424 samples

(4 years, 4 instruments, 2 locations)

Two Parameters

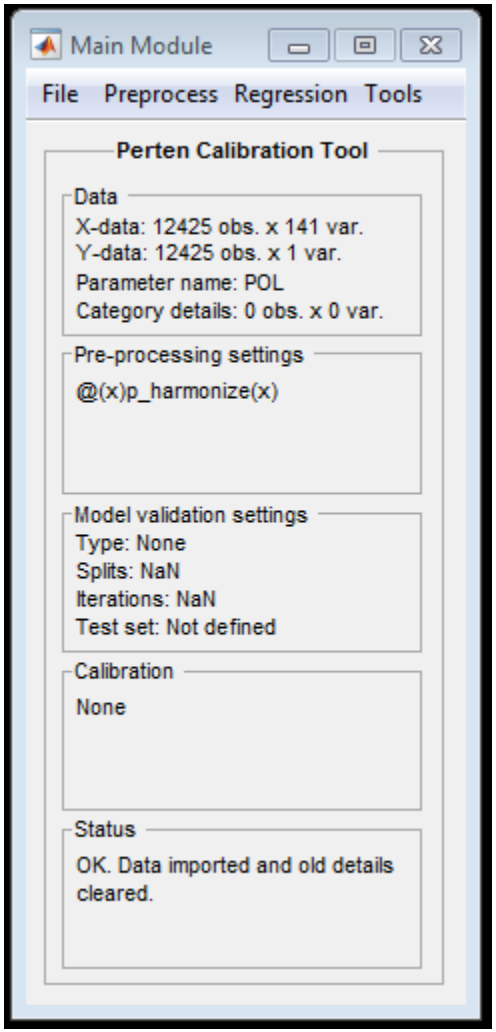
POL (Polarity, sugar content)

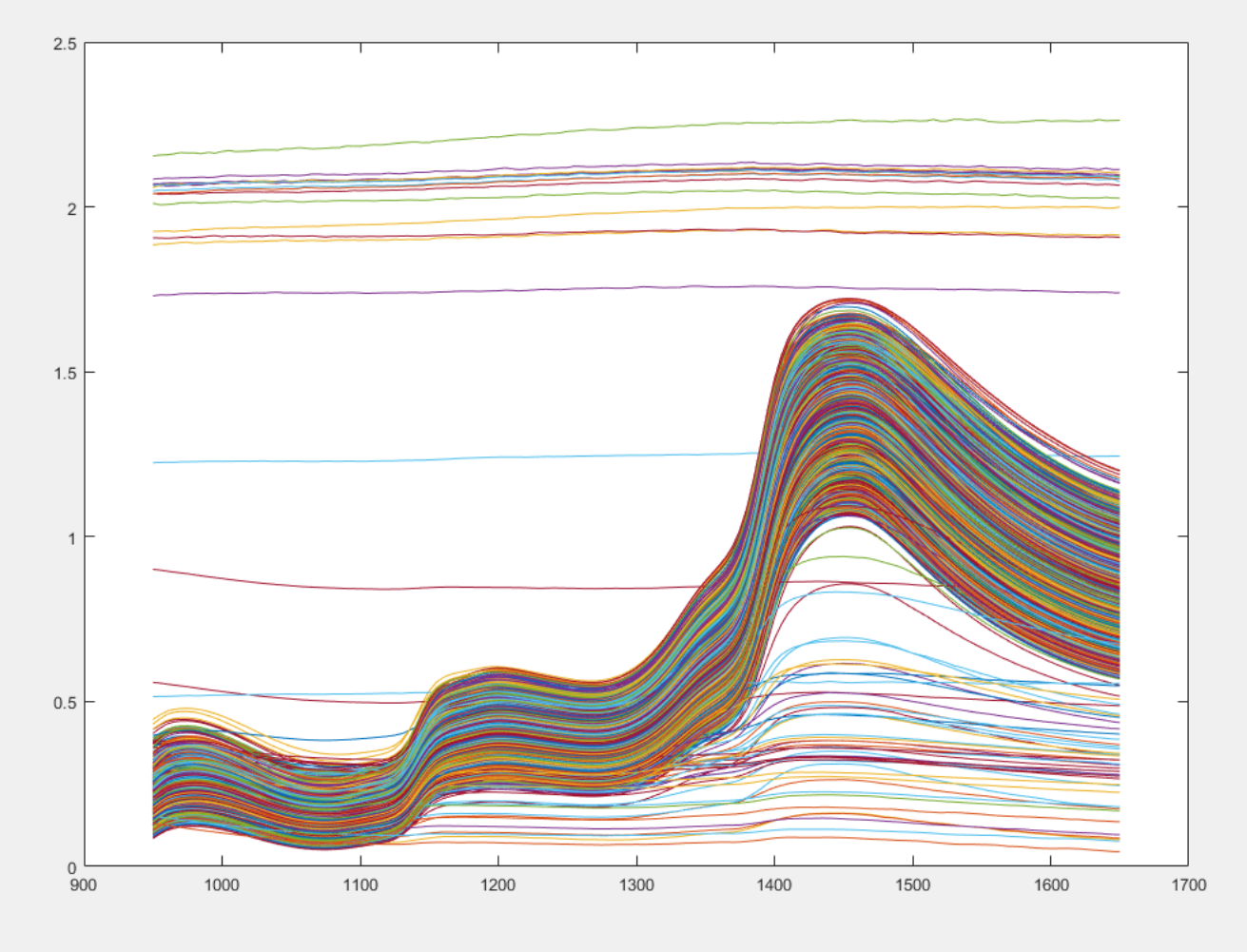
Conductivity – salt, protein content

Pretreatment

Detrending, 2nd order

SNV



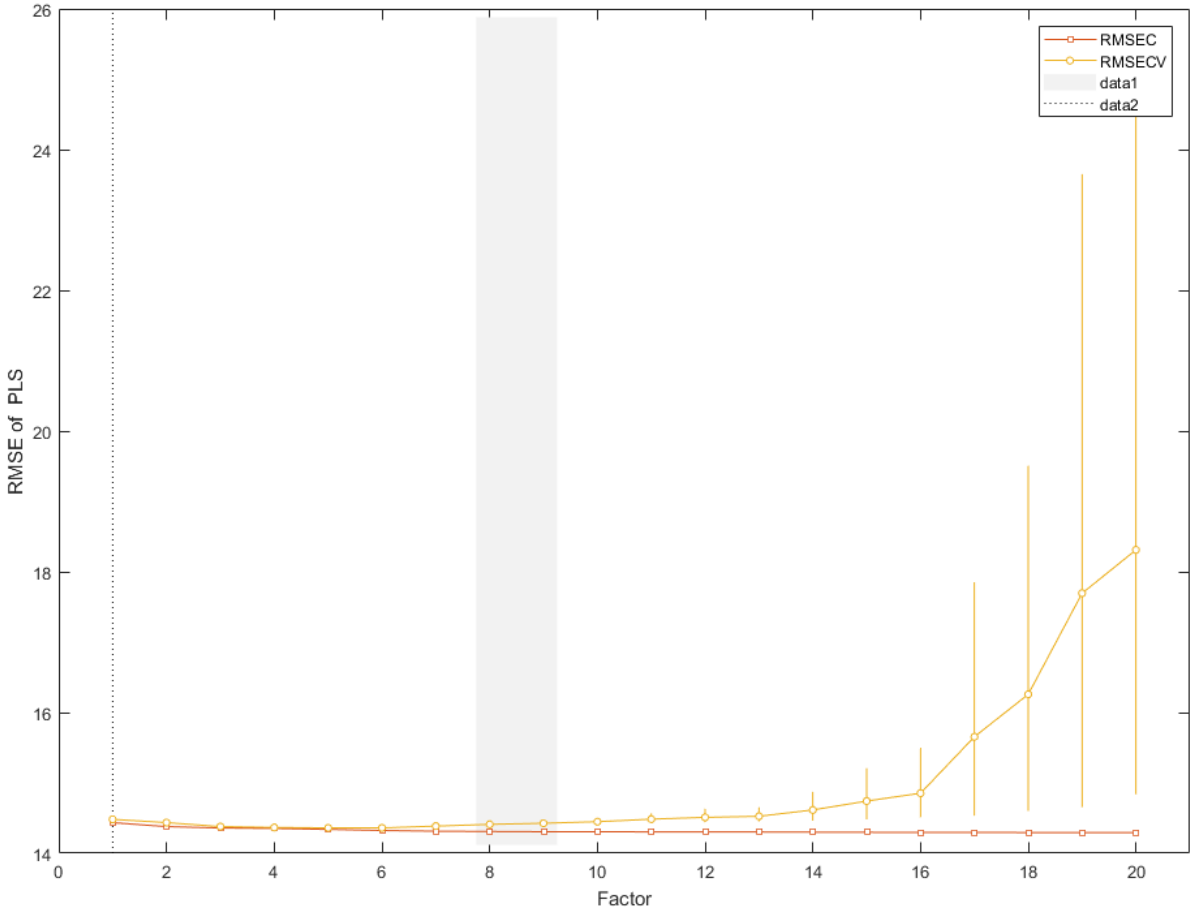


- Step 1: Plot the spectra even if it takes a while. Out of 10K, some mistakes will be made.

Perten INSTRUMENTS
a PerkinElmer company

How Do You…

- Test Set and Validation Set
  - What is the probability of overfitting 10K?
  - Keep half out for testing or validation?
  - What should this be, a percentage or a fixed number or unimportant?



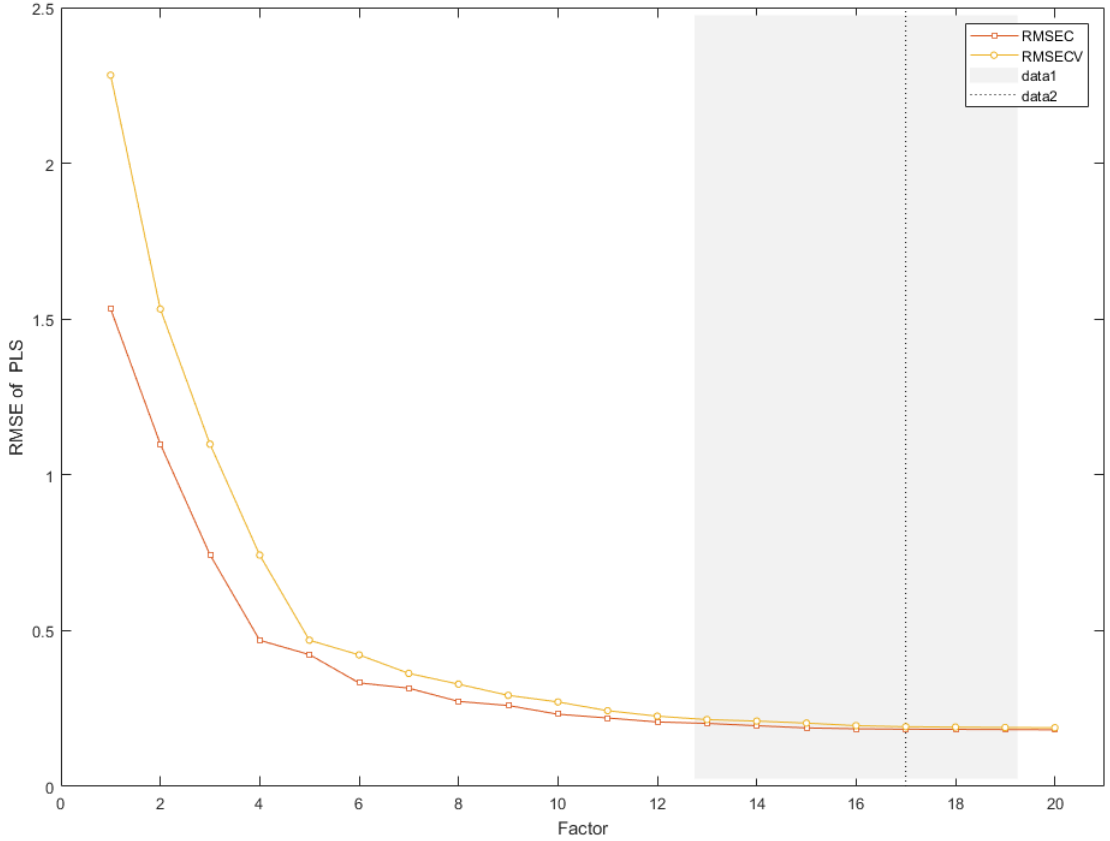
PLS with “Auto Tune”

- Auto Tune: 146 possible outliers, factors to 17. That’s 1.175%. Just about 1 in 100.

Perten INSTRUMENTS
a PerkinElmer company

How Good is 1% Outliers?

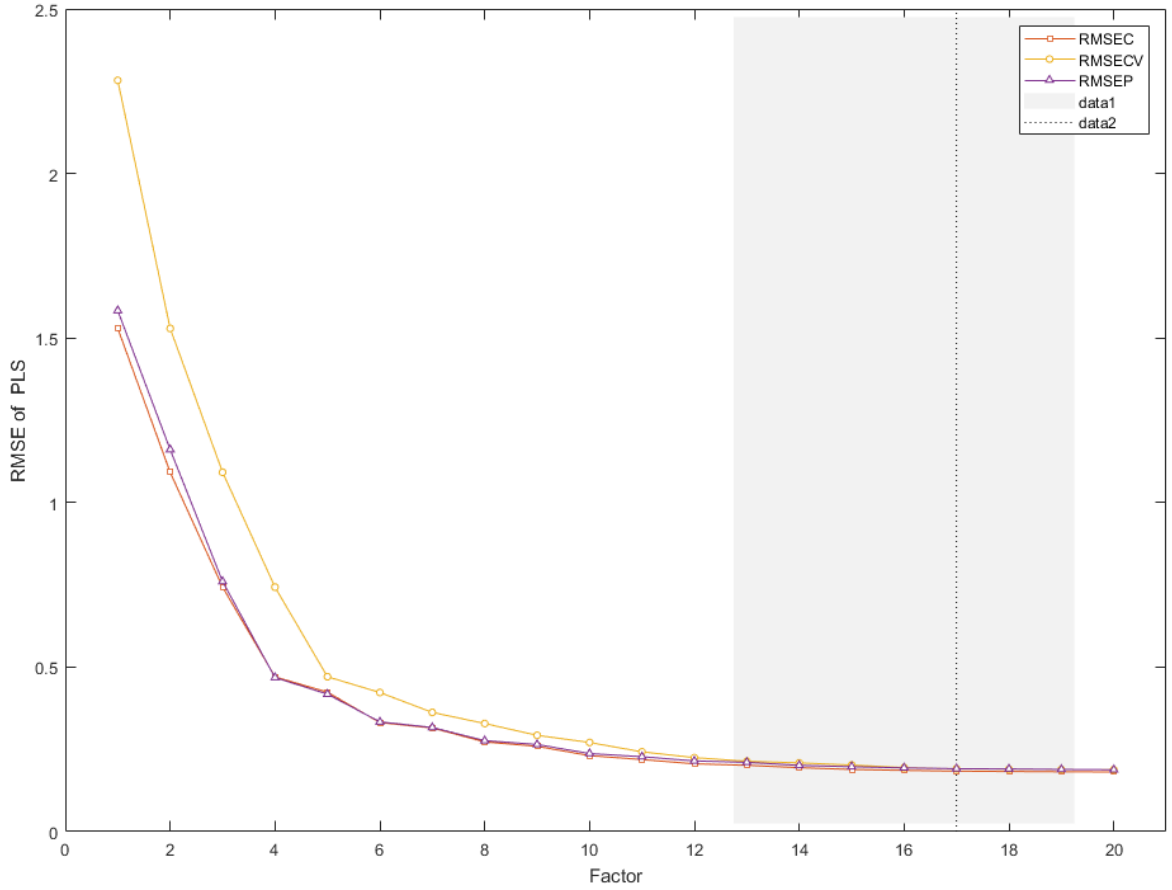
- How many times in a 100 sample calibration do you only drop 1 sample?
- Myth: PLS fails when there are many samples.
- Is 17 factors a lot? Why not more?



SEC and SECV

- Values converge at this number of samples

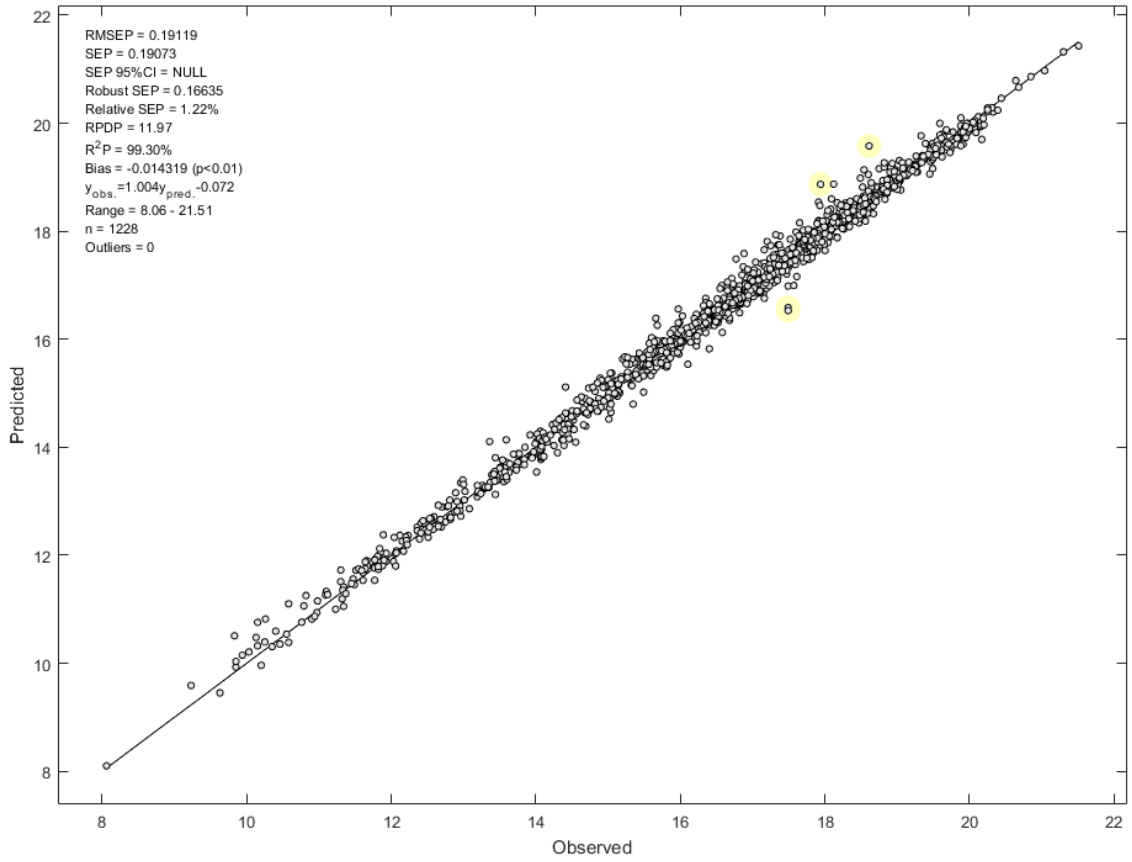
Perten INSTRUMENTS
a PerkinElmer company



SEC / SECV / SEP

- With enough samples there is no essential difference between them at high enough factors

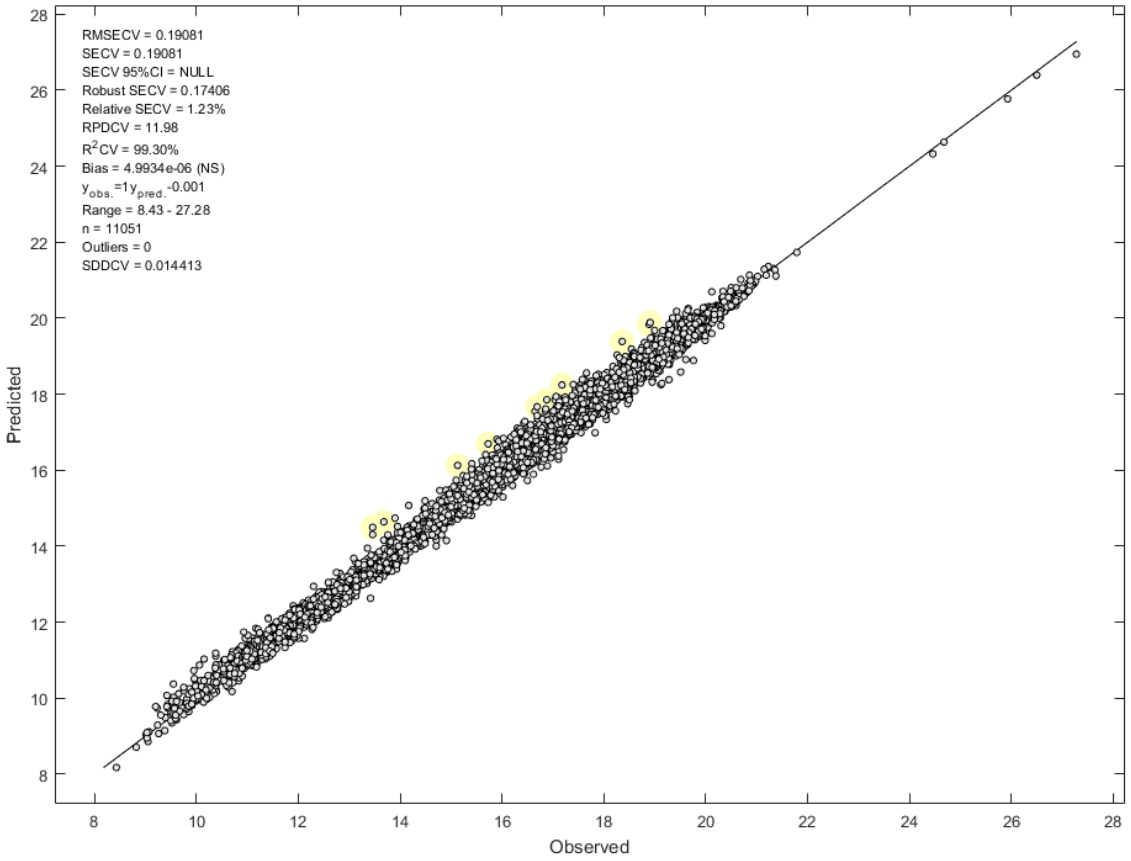
Perten INSTRUMENTS
a PerkinElmer company



SEP Value

- Can compute a “Robust SEP”. 0.19 to 0.16 polarity
- Are 4 outliers really outliers?

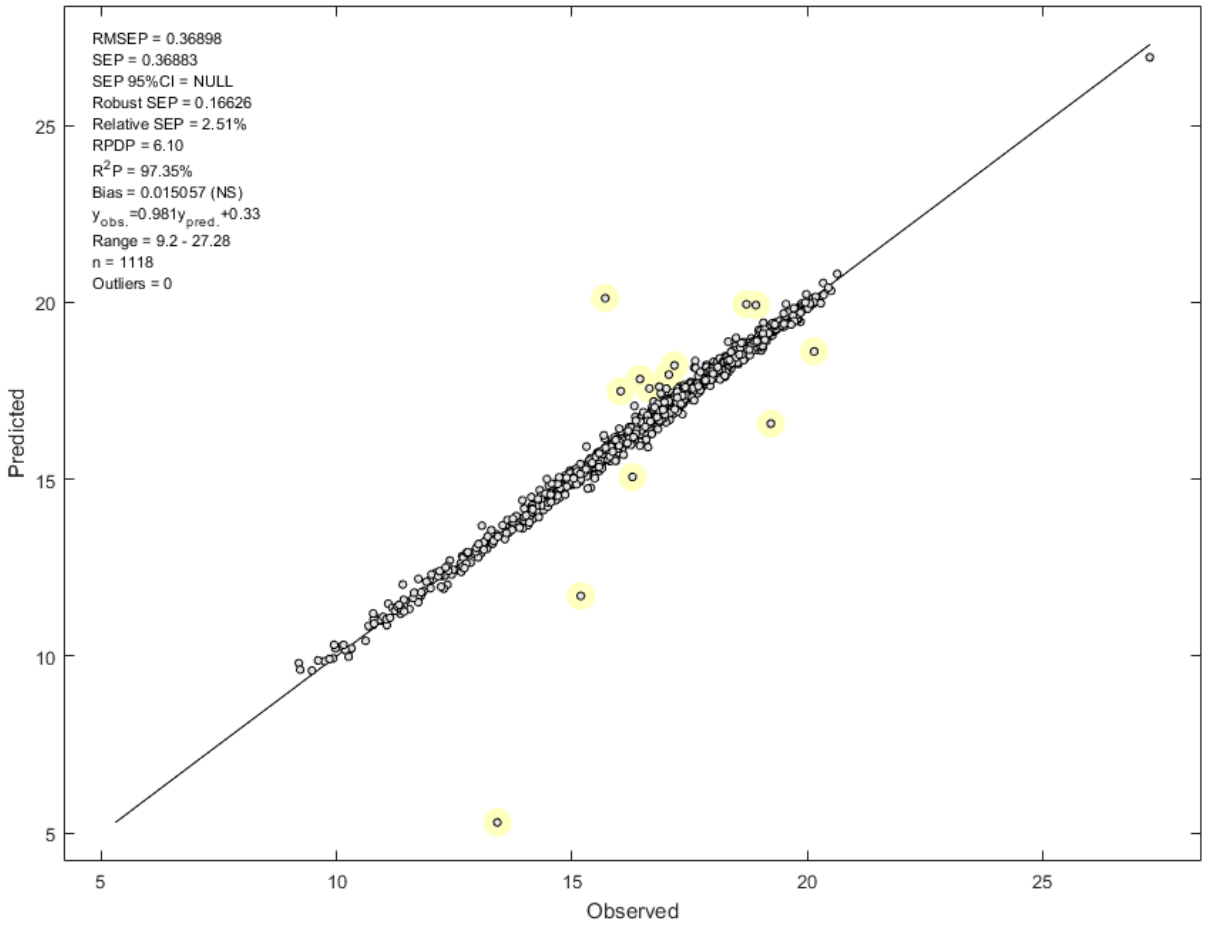
Perten INSTRUMENTS
a PerkinElmer company



SECV Value With 10% Test Set

- 0.19 to 0.175 polarity

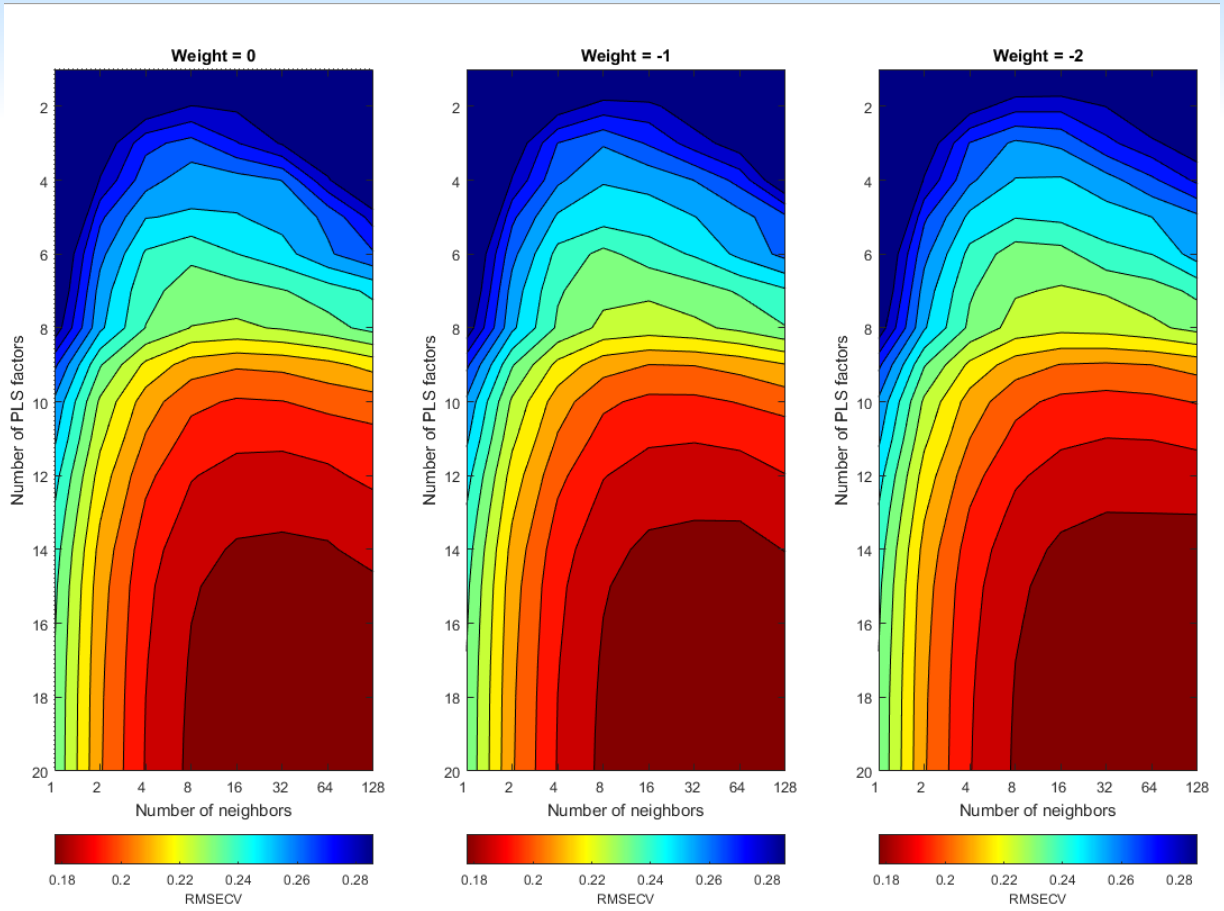
Perten INSTRUMENTS a PerkinElmer company



Random Test Sets

- Are not a good idea. Everything needs to be cleaned first.

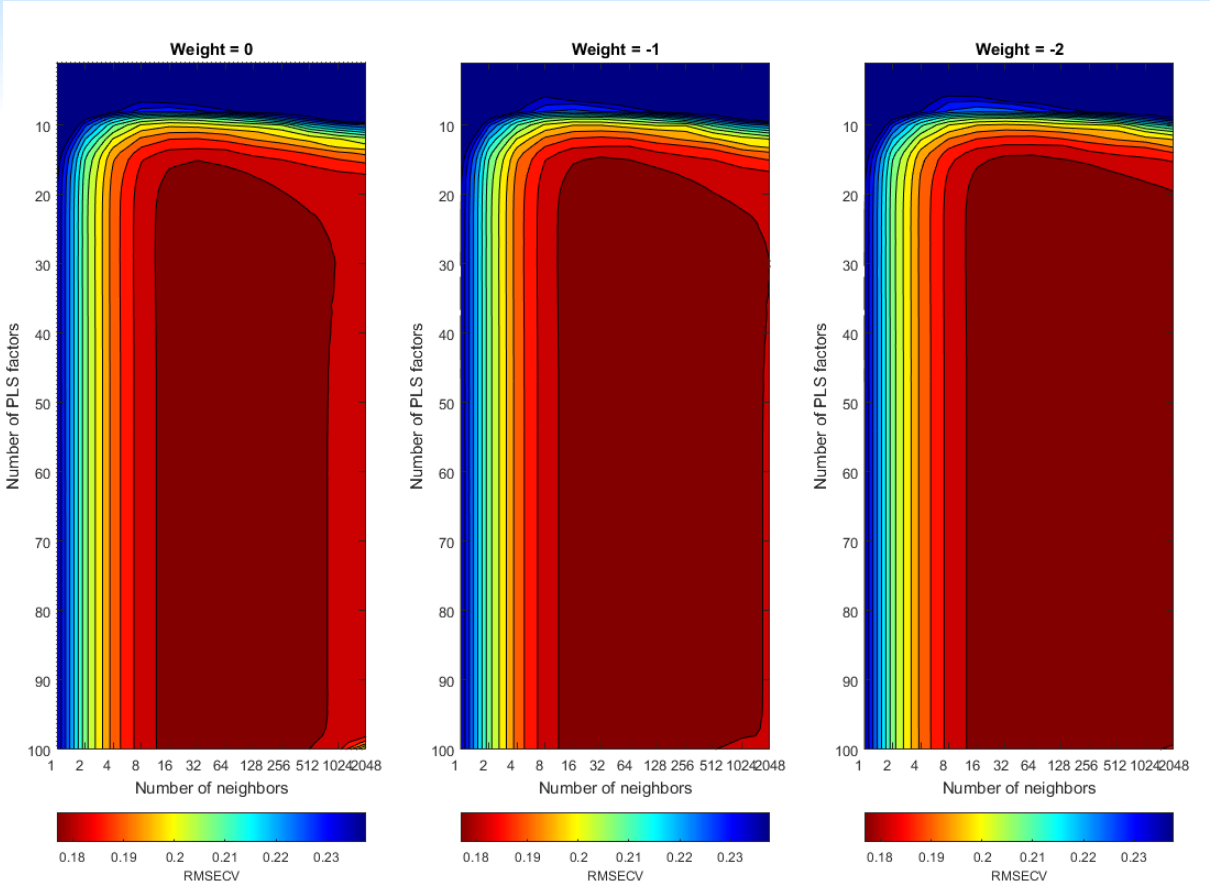
Perten INSTRUMENTS
a PerkinElmer company



HR Calibration Results

- Nicely behaved. Maximum area is broad and extends off the chart.

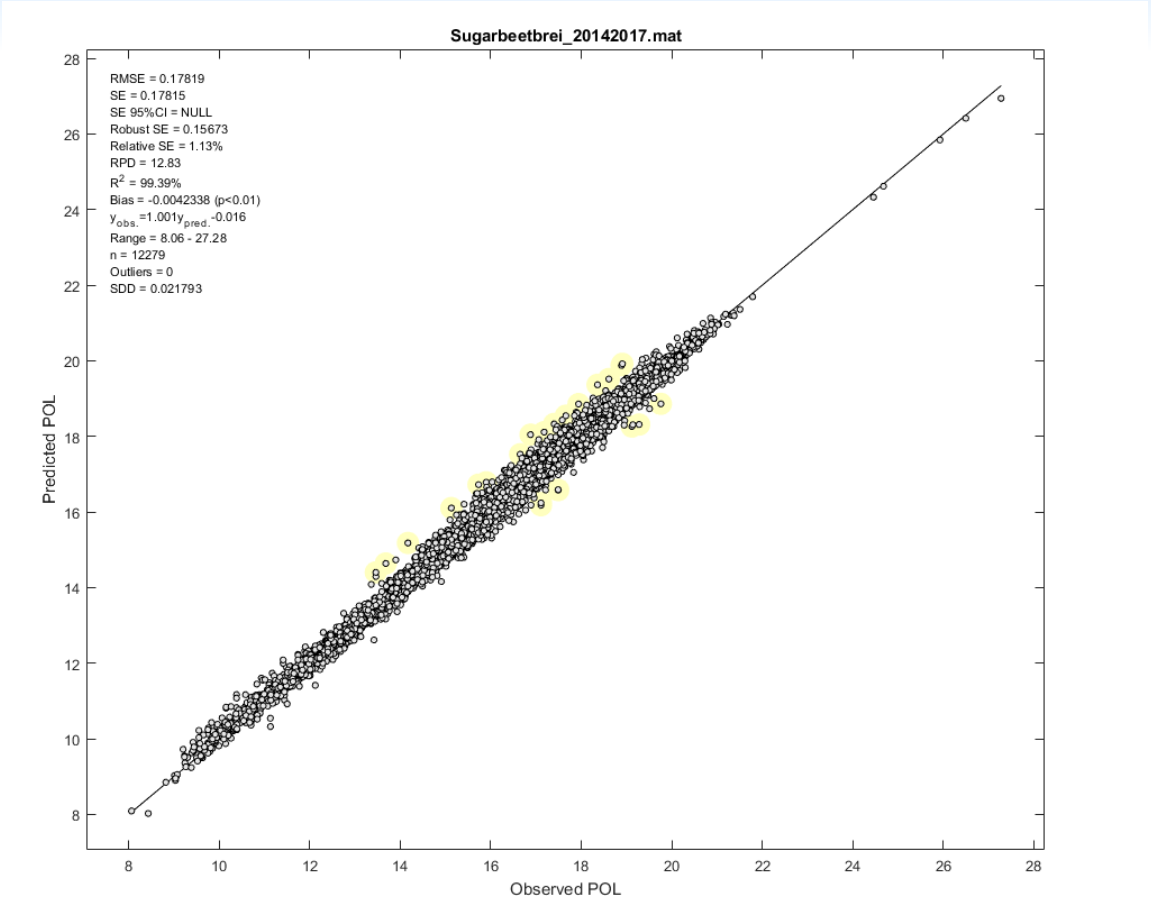
Perten INSTRUMENTS
a PerkinElmer company



HR – Just How Far Off The Chart?

- Around 1000 samples in Weight 0.
- Around 100 Factors. If we keep the ratio; 100 samples per factor.

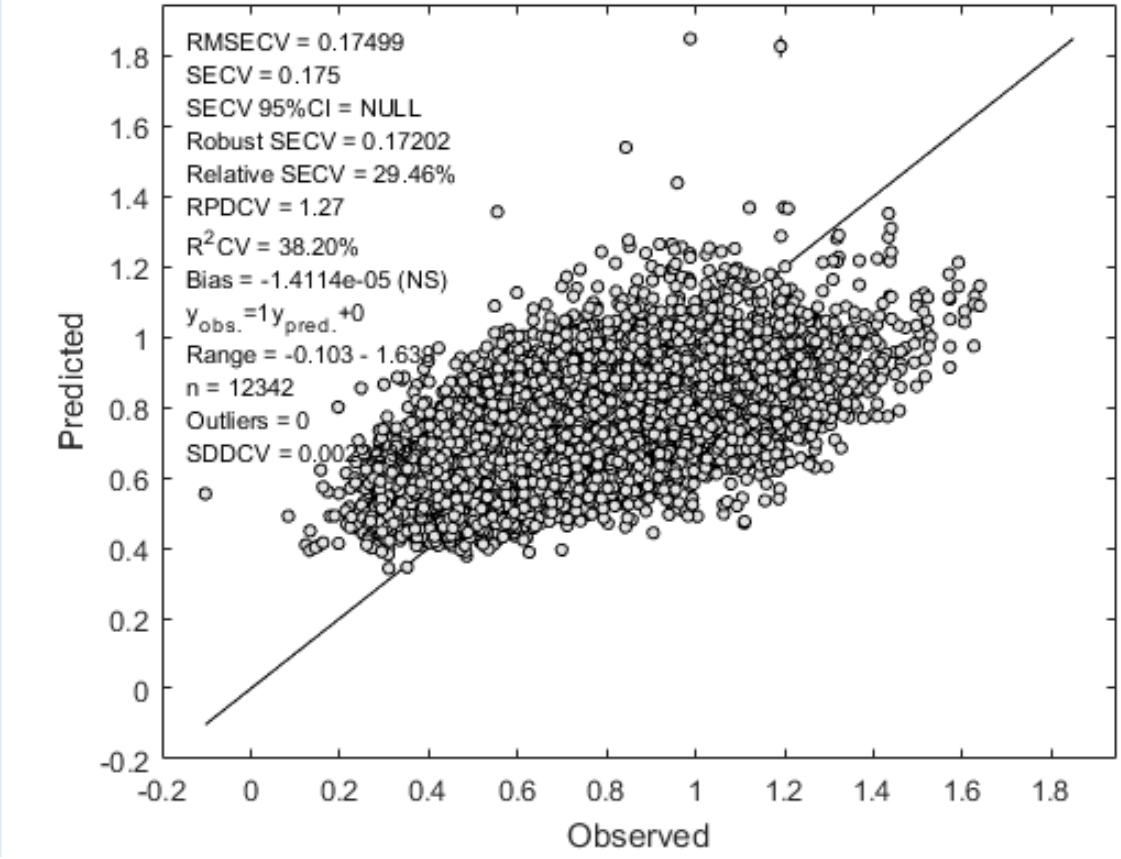
Perten INSTRUMENTS
a PerkinElmer company



HR Results

- RMSE 0.178. PLS was 0.191
- Maybe a 5% relative improvement with the non-linear technique.

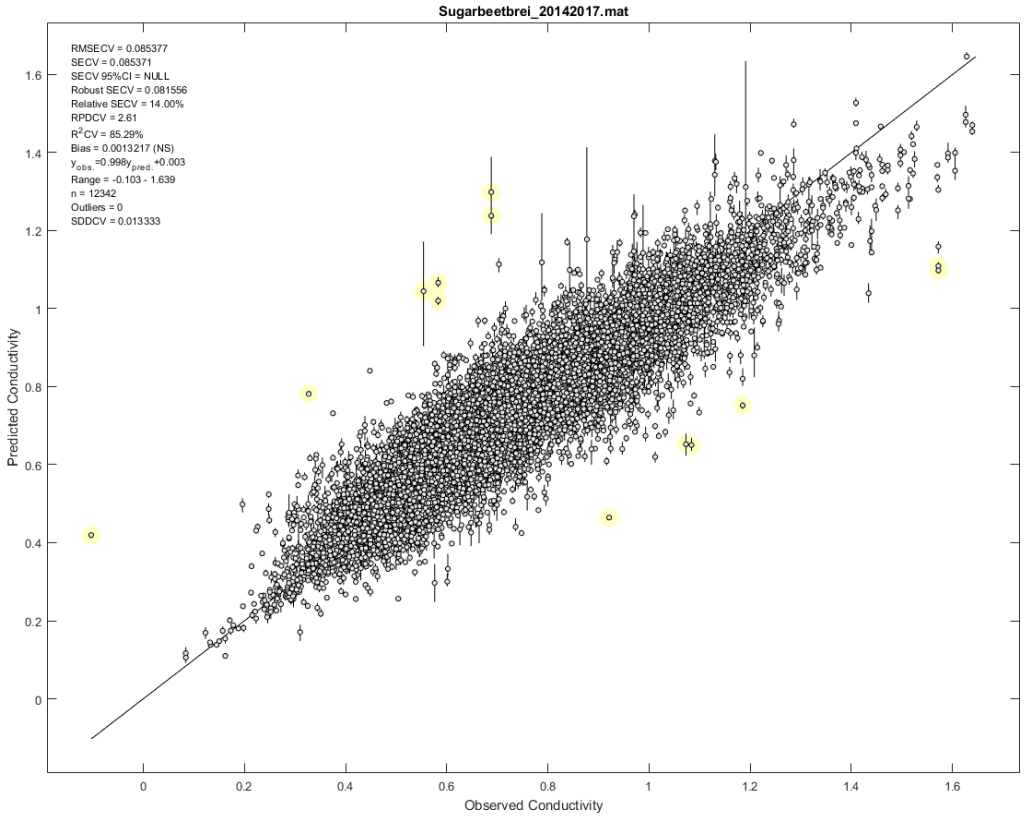
Perten INSTRUMENTS a PerkinElmer company



PLS of Conductivity after Auto Tune

RPD of 1.27. SECV of 0.175.

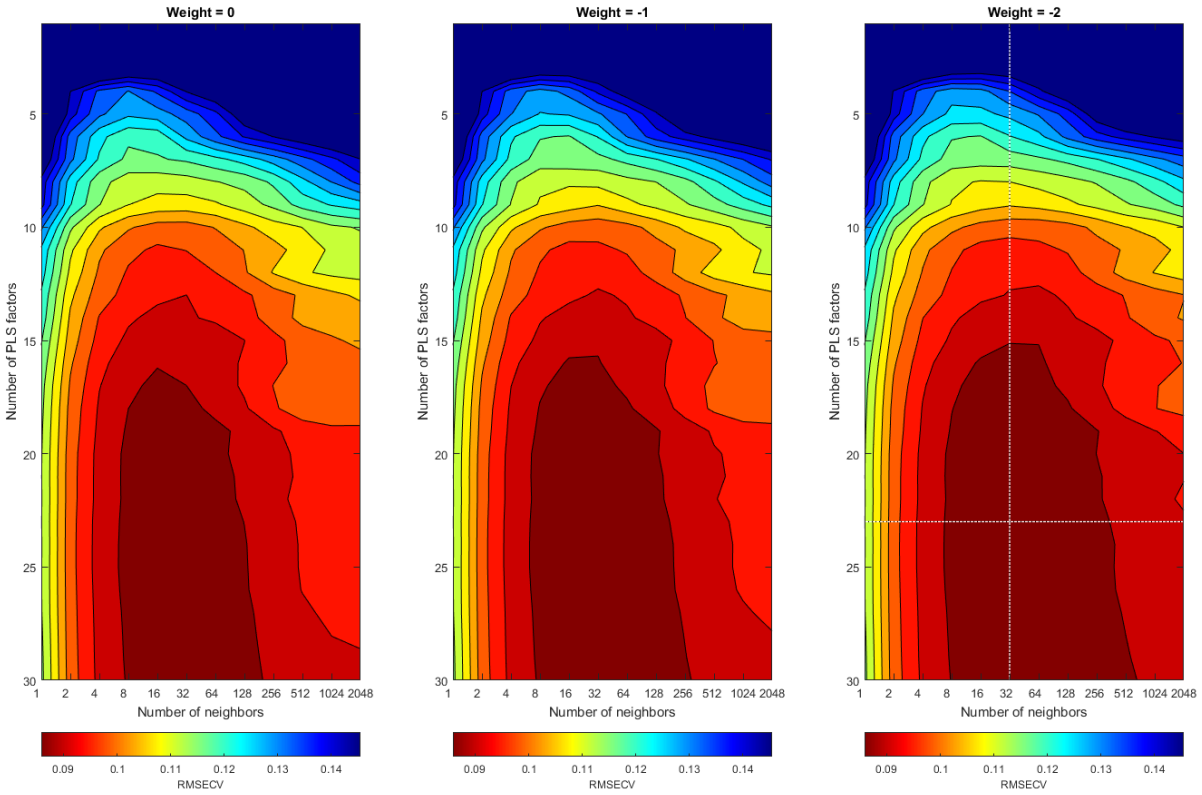
The data suggest you can/can't measure salt by NIR.



HR Calibration for Conductivity

RMSECV 0.085, RPDCV 2.6, R Sq 0.853

You can measure salt with NIR, but it is non-linear.



HR Color Map

PLS begins to fail with too many samples. This happens in a non-linear situation. Modest changes are linear. The full range of changes are not.

Conclusions

- At some point random Test Sets become unimportant.
- PLS has no mathematical issue with number of samples.
- Non-linear systems become more pronounced at higher number of samples.
- Given enough samples, time, instruments, even very linear calibrations are slightly non-linear.