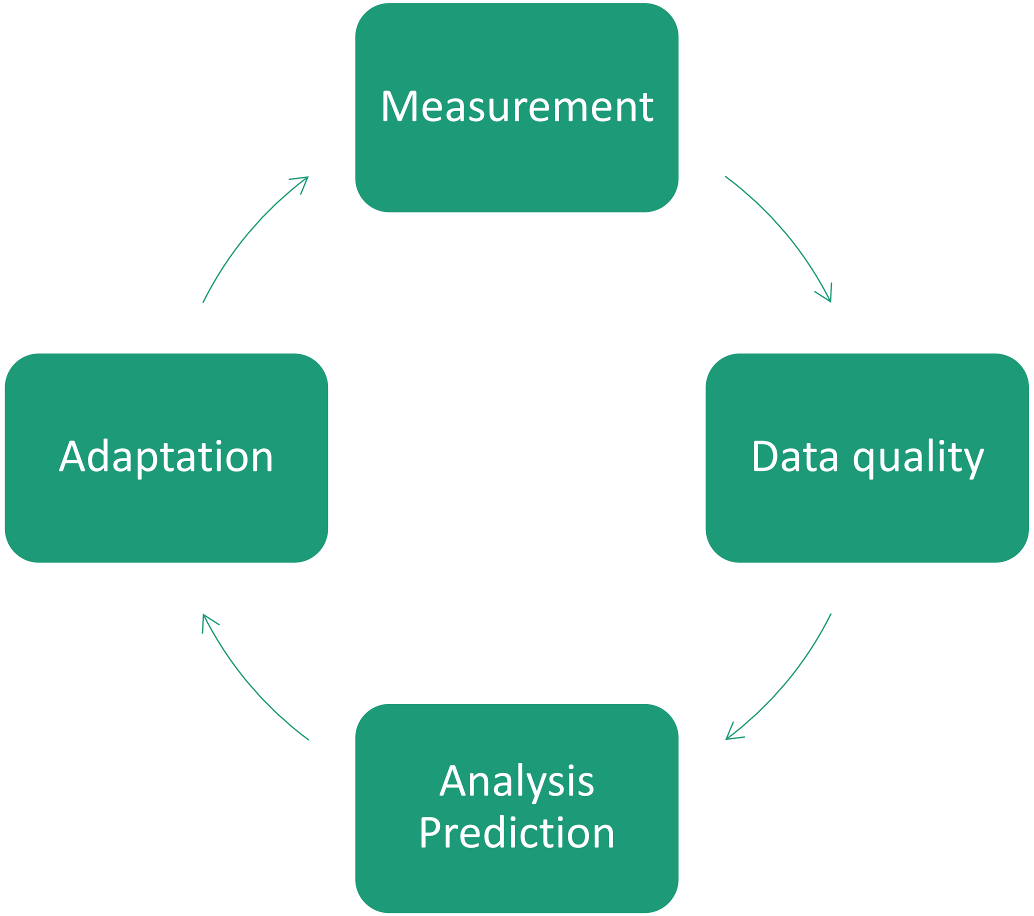
Statistic tools for the on-line interpretation of DPF on-board control signals

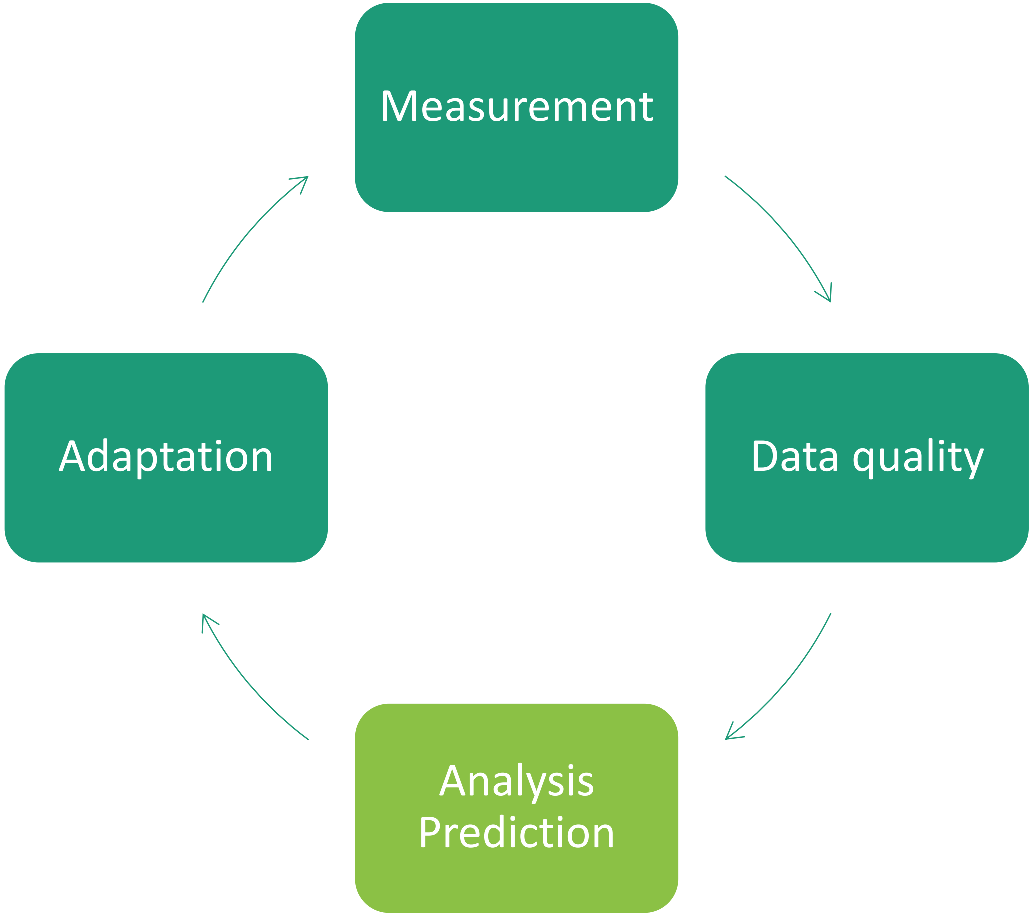
Lukas Fabrykowski (lukas@fabrykowski.com)

Friedrich Legerer

Measurement-cycle

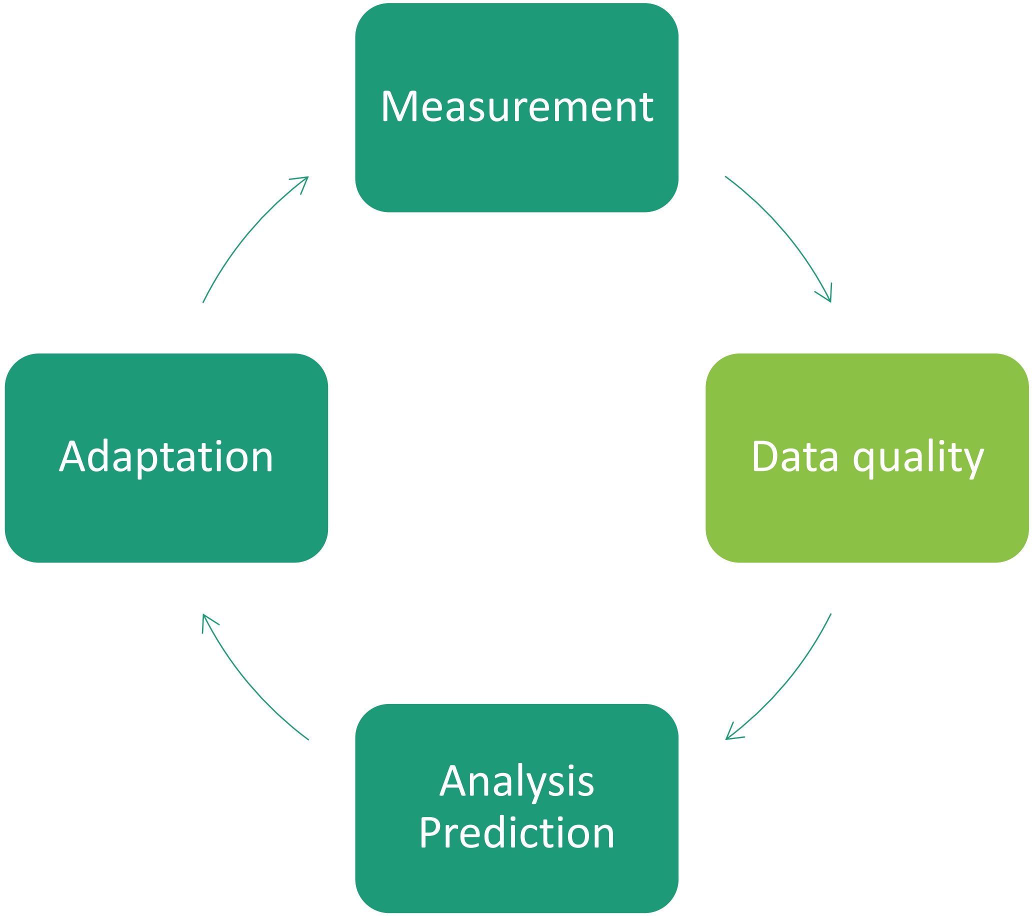


Targets



Analysis and prediction of key values to support fleet management and engineering

Targets



Automated analysis and detection of data faults and preparation of data for further analysis and prediction

IF YOU TORTURE THE DATA LONG ENOUGH, IT WILL CONFESS.

Prof. Ronald Coase

British economist, laureate in Nobel Memorial Prize in Economic Sciences 1991

Unexpected data

Measurement

- Sensor
- On-site chip
- Data trasmission
- Off-site chip
- etc.

Filter

- Coking
- Partial regeneration
- leakage
- etc.

Increasing complexity of analysis

Data stream

1st order

2nd order

Data stream

- No data manipulation
- On-line

- Multiple time stamps
- Hard upper or lower bounds

1^{st} order

- Grouping, splitting
- Immediate history

- Too short measurements
- Operation cycle
  Block of continuous measurements
- Upper or lower bounds
  Ratio of cylce

2^{nd} order

- Statistical analysis, calculus
- Entire history

- 1^{st} / 2^{nd} derivative
- Variance
- Entropy / efficiency

Entropy

- 19th century: Rudolf Clausius (classical thermodynamics)
- 2nd half of 19th century: Ludwig Boltzman (statistical mechanics)
- 1948: Claude Shannon (information theory)

$$H(X) = -\sum_{i=1}^{n} P(x_i)\ln(P(x_i))$$

- Shannon-Index

Efficiency

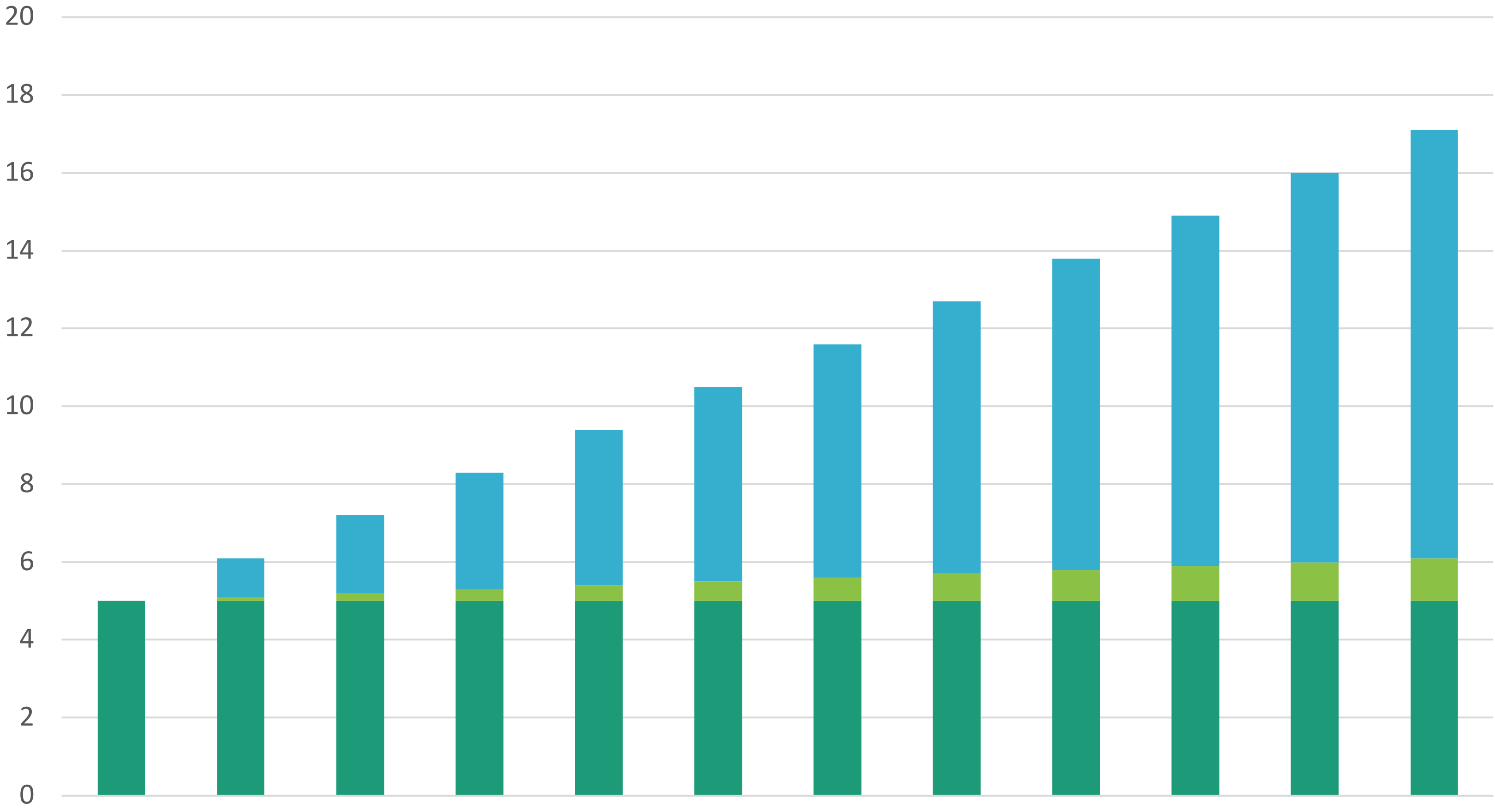
- Maximal Entropy at uniform distribution 15th

$$H(X) = -\sum_{i=1}^{n} \frac{1}{n} \cdot \ln\left({}^{1}/_{n}\right) = \ln(n)$$

- Normalised entropy

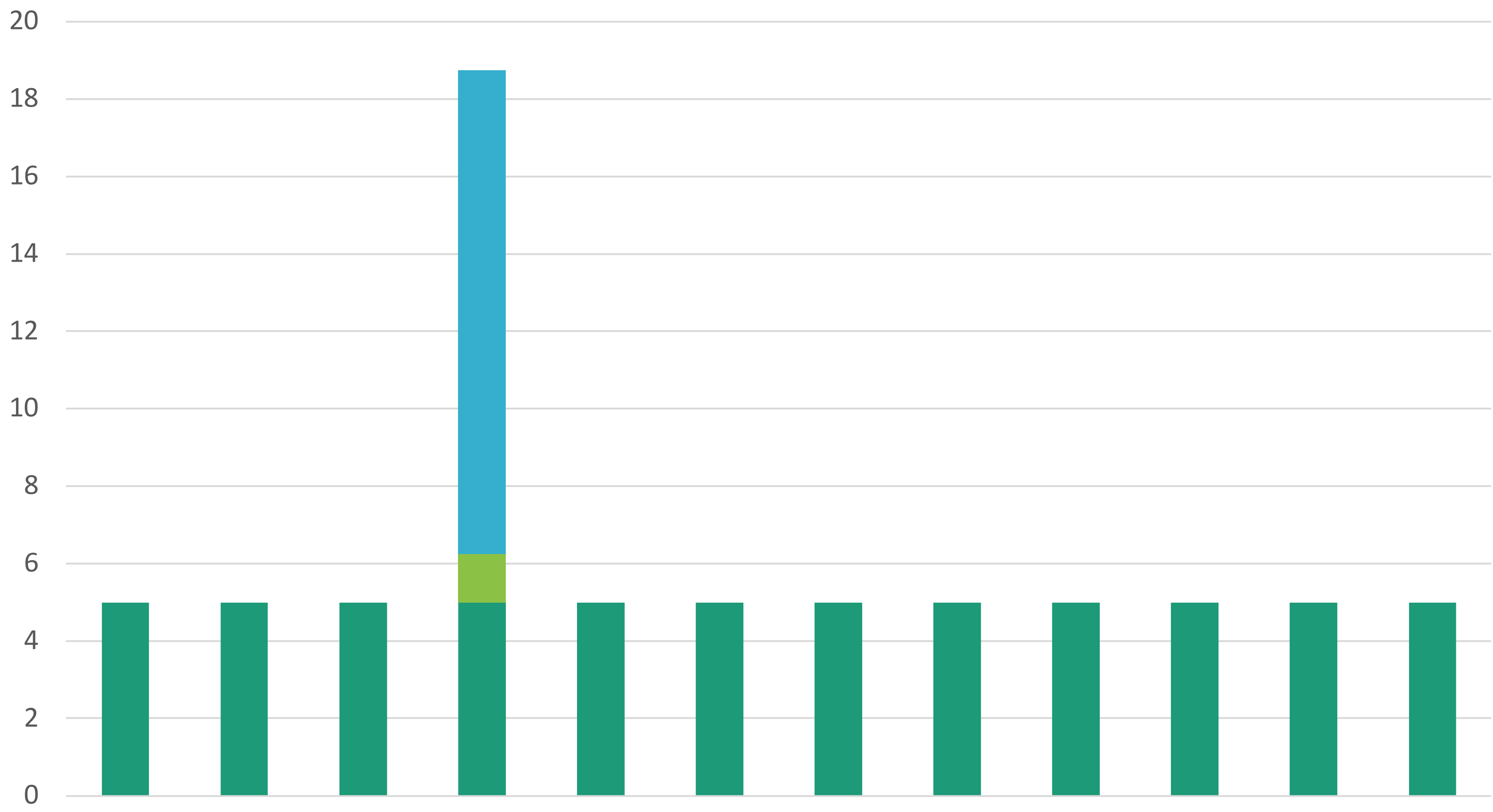
$$\eta(X) = {}^{H(X)}/_{\ln(n)}$$

Relation entropy / variance



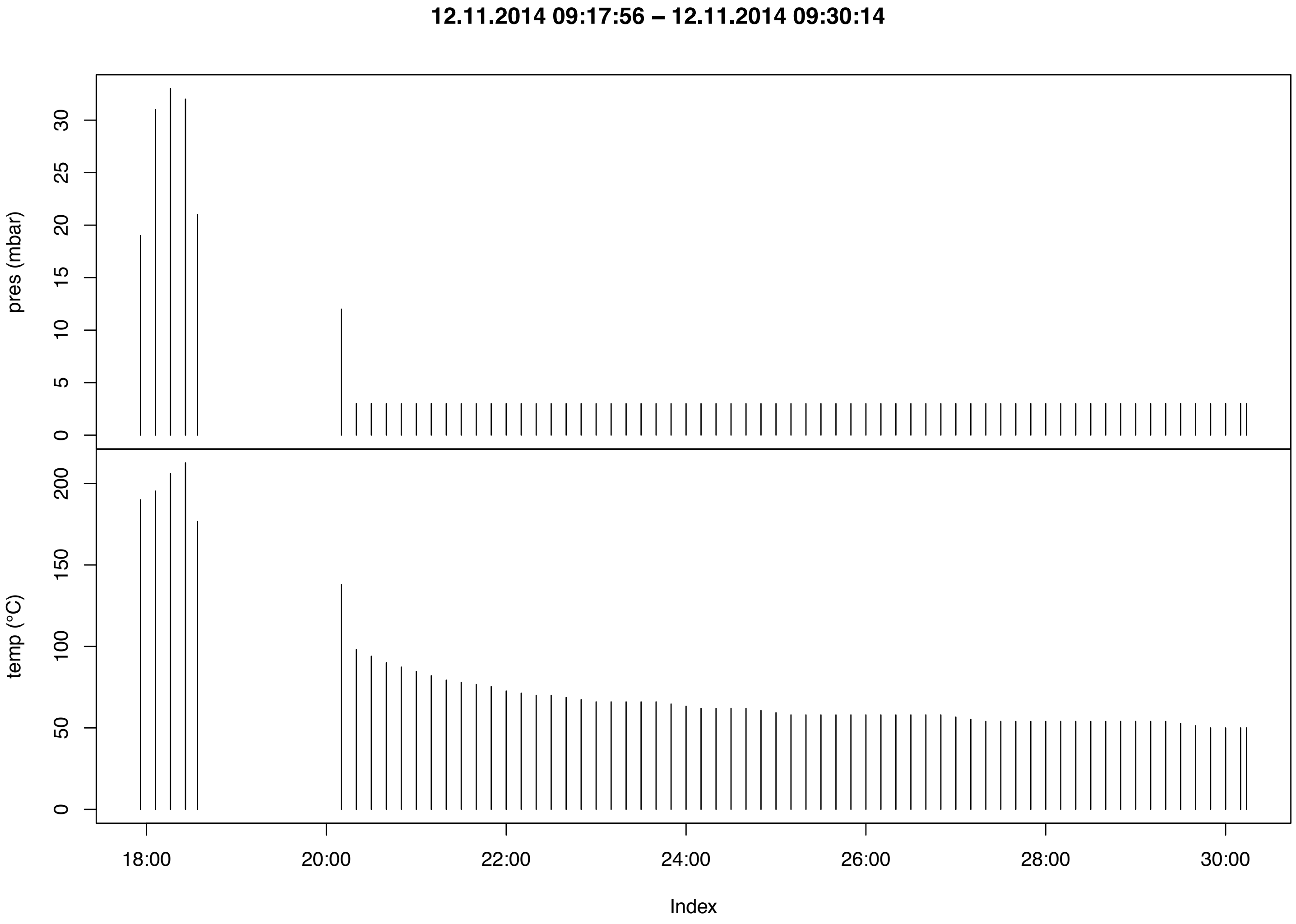
Var(X)	H(X)
0.00	0.00
0.13	2.48
15.73	2.48

Relation entropy / variance



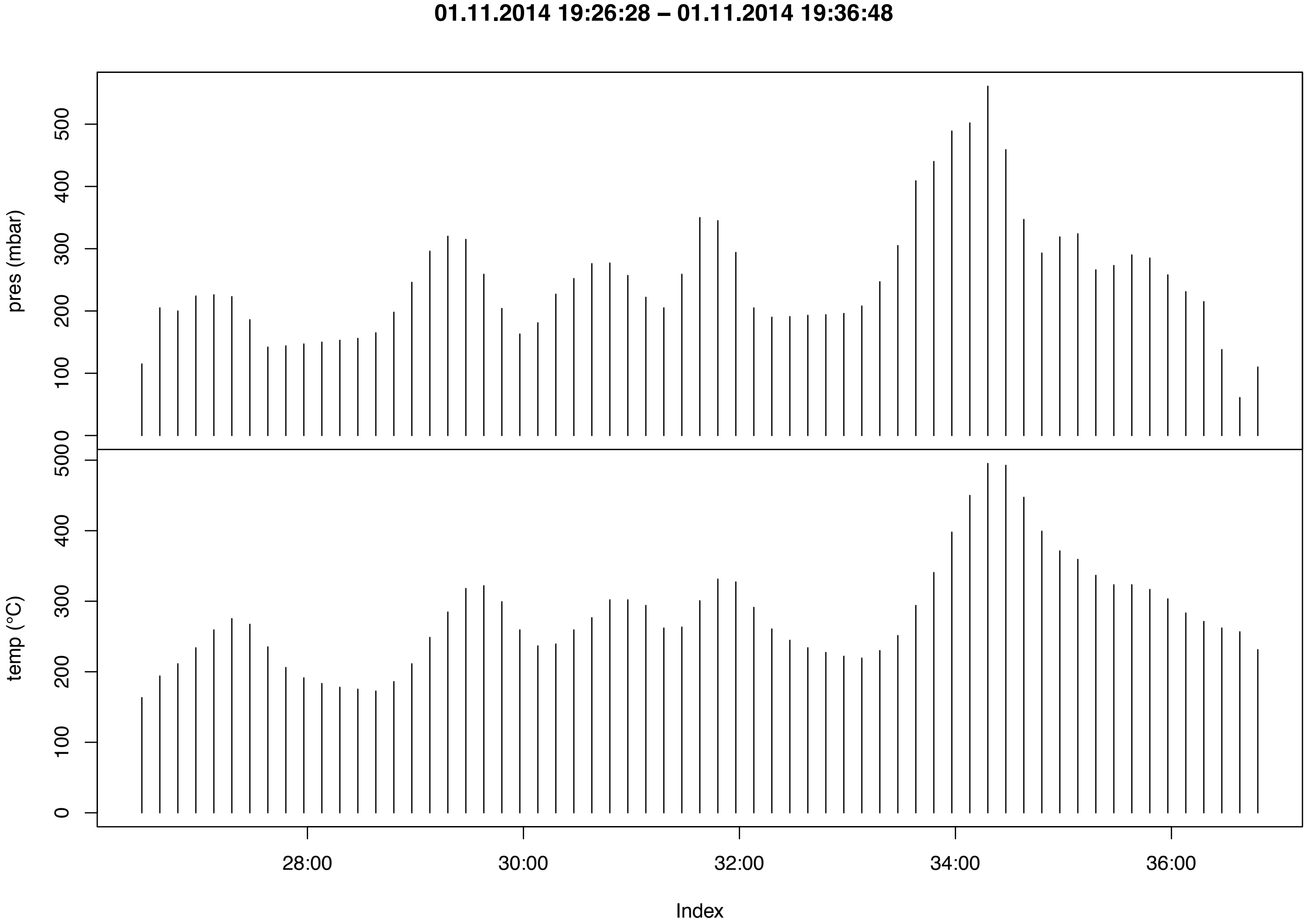
Var(X)	H(X)
0.00	0.00
0.13	0.29
15.73	0.29

Real-world measurements



$\eta(X) = 0.1098669$

Real-world measurements



$\eta(X) = 0.982062$

Current state and outlook

- 20 data series
- > 19 mil. data points

- Further working and faulty series as reference data

- Which predictors?
- Which characteristics?