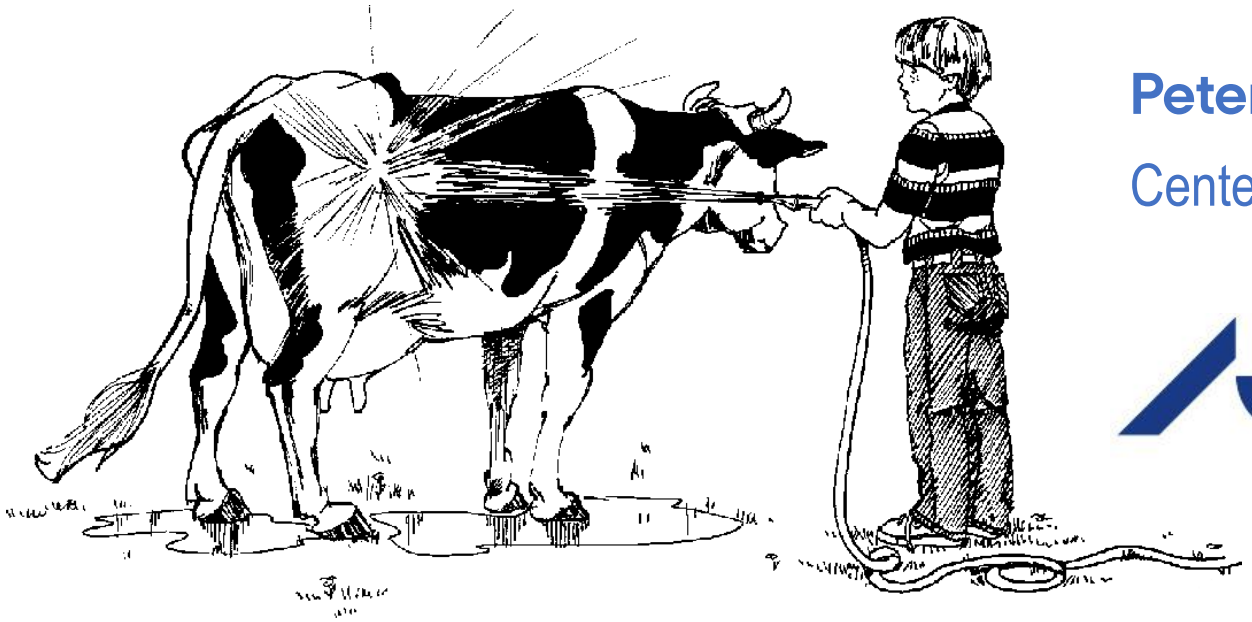
BIG DATA

.. may be useful to dairy herd managers – after washing and drying



Peter Løvendahl

Center for Quantitative Genetics and Genomics -QGG

AARHUS
UNIVERSITY

DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

Acknowledgement

- Data from DeLaval OCC units in private Danish herds and from DCRC
- Funding by the Danish Research Council
- Colleagues at AU-Foulum: Lars Peter Sørensen and Martin Bjerring

The issue … partly …

Many organizations face a data quality problem and this adds another dimension to the question of whether there's too much data and what data to keep.

The problem is partly one of missing or inaccurate data, but often one of redundant data which is why solid governance policies are so important.

https://enterprisersproject.com/article/2015/11/get-value-out-many-dimensions-big-data

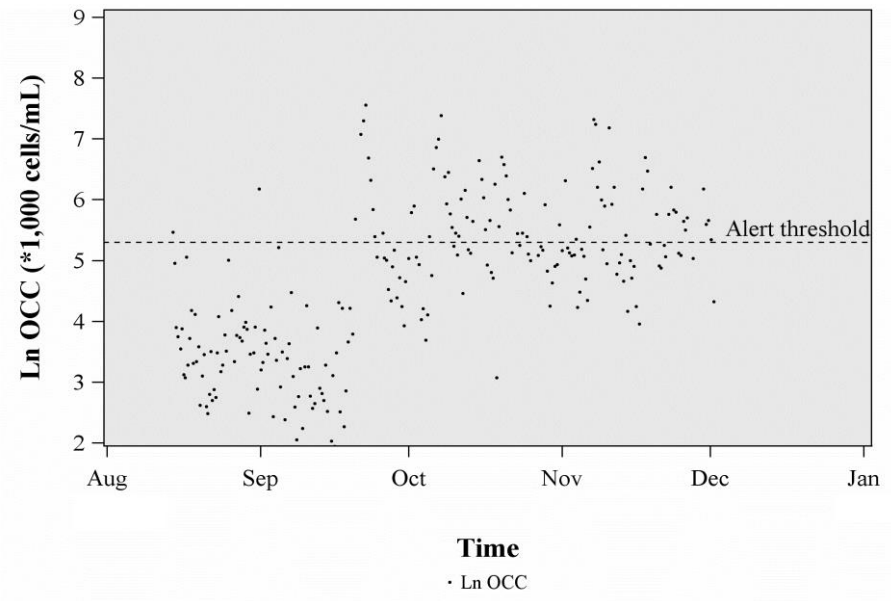
An experience with sensor data …….

An in-line sensor for detection and monitoring of udder health in dairy cows:

OCC = Online Cell Counter (DeLaval)

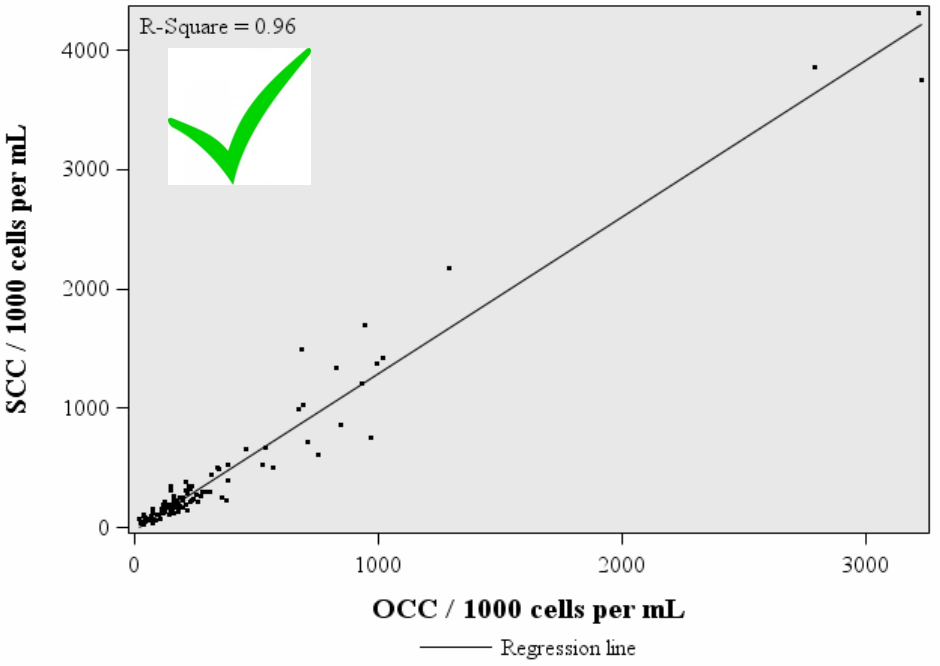
Measures OCC every day, or every milking in AMS herd

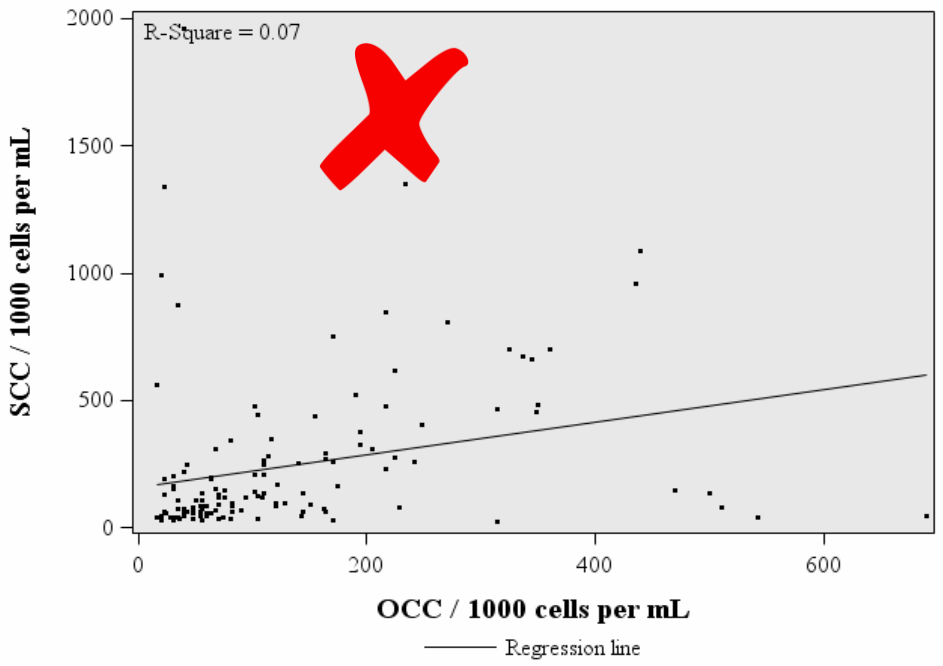
Typical raw data – **is that cow healthy or not**?



Are measurements correct, precise, accurate?

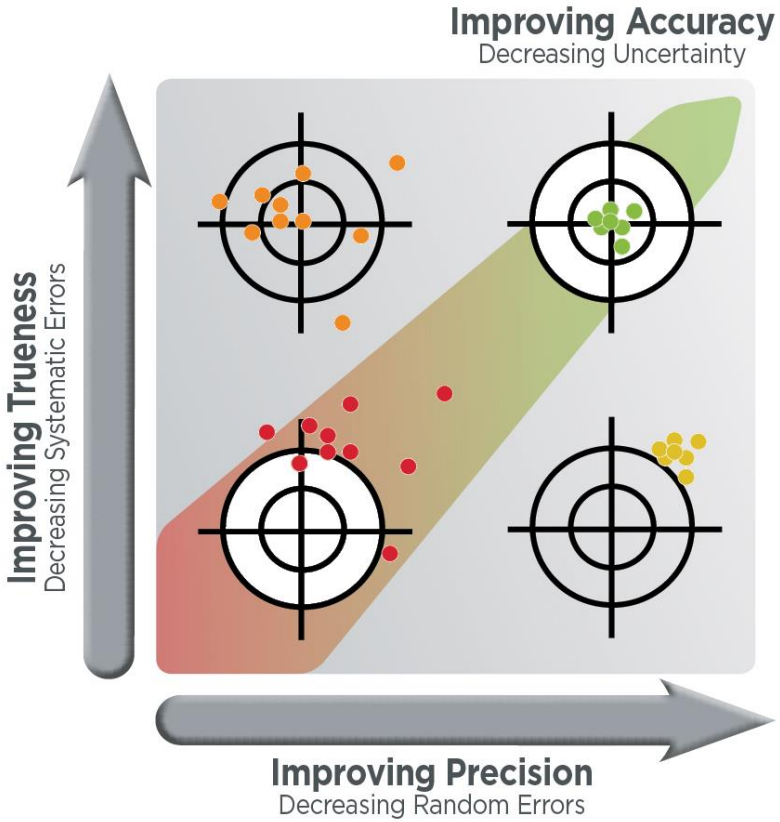
- Are the instruments **actually** working?
- Checkpoints – warning lights – dashboard?
- Performance indicators R^2 ….





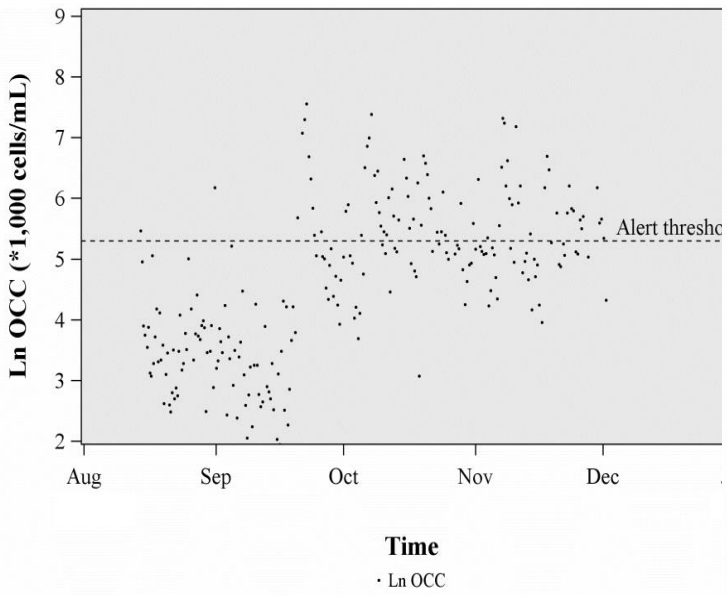
Checkpoints and adjustments using internal and external refence data

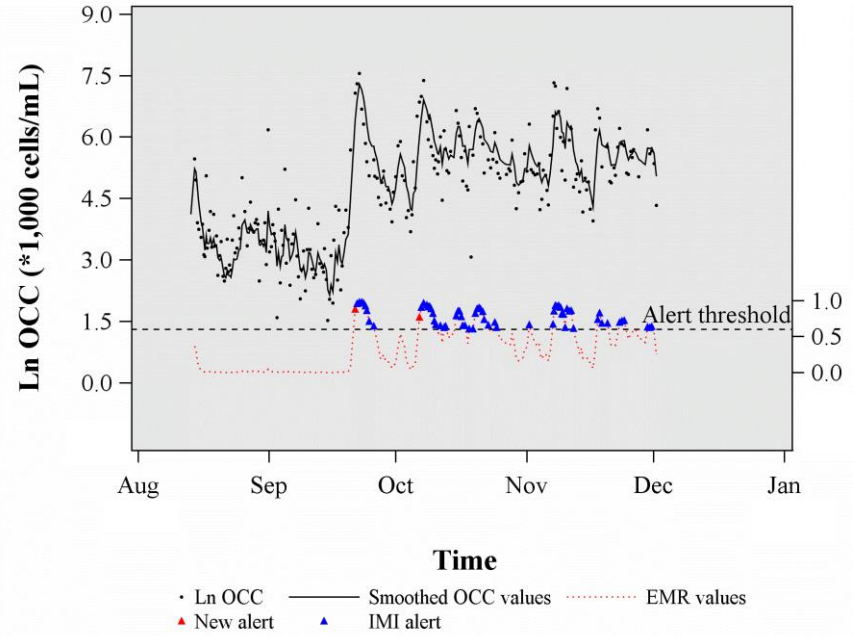
- Drift in instruments – daily mean value comparisons
- Aligning different instruments – using readings of the same animals on two or more instruments
- Align to external standard- supply with ”calibrator fluid” or re-test samples with gold standard equipment

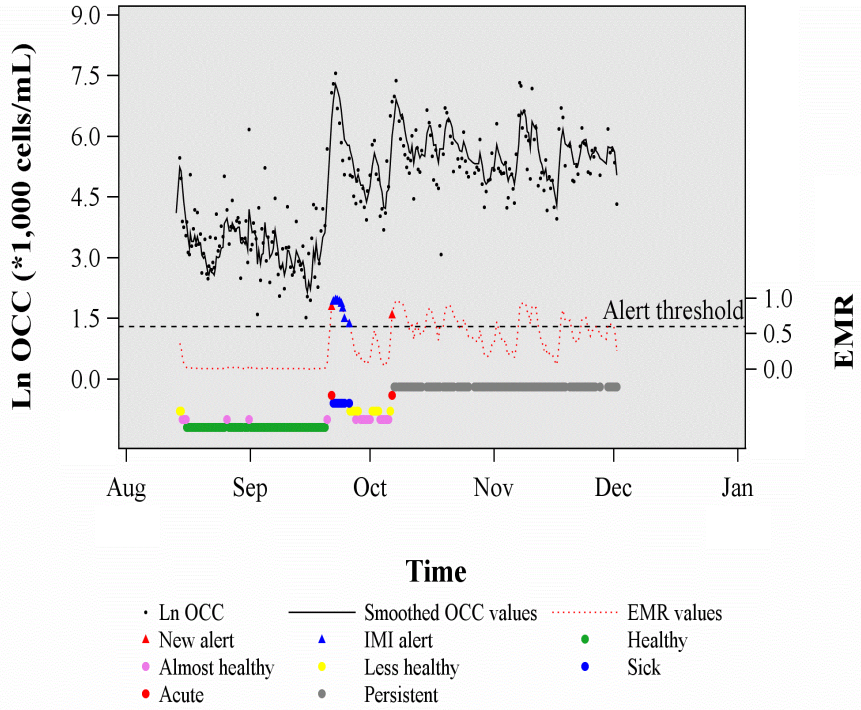


Sensor data become time series

Smoothed values – and assignment to health categories?







Exponential smoothing was used here – simple and effective

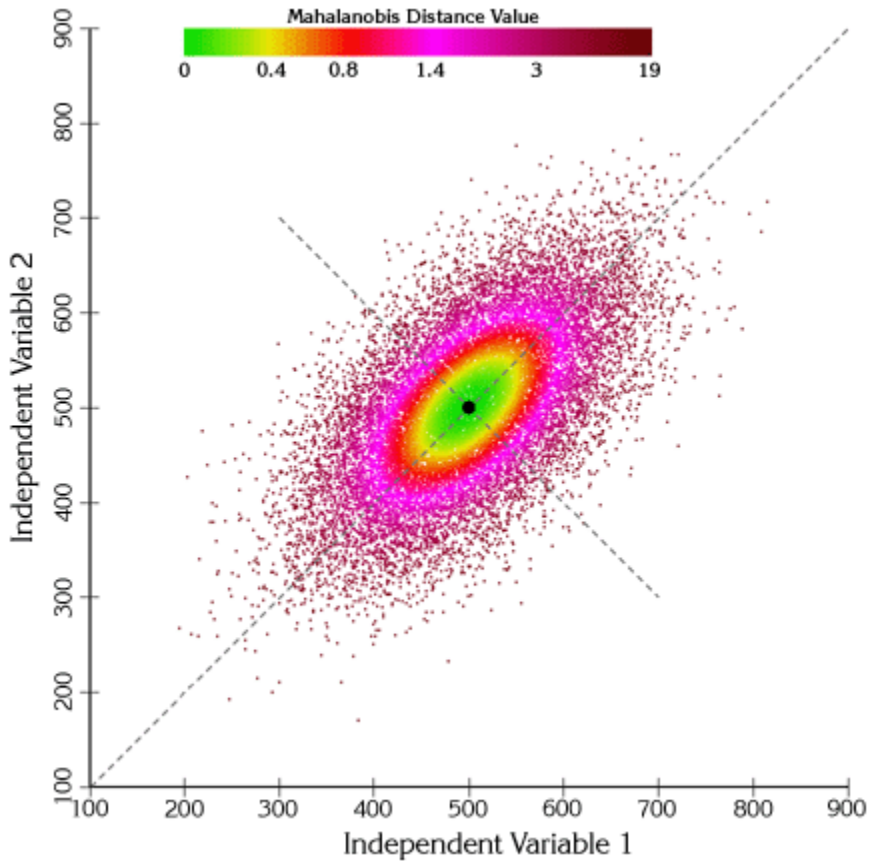
Combining sensor alarms -

- Signals in two or more dimensions
- Factor analysis or :

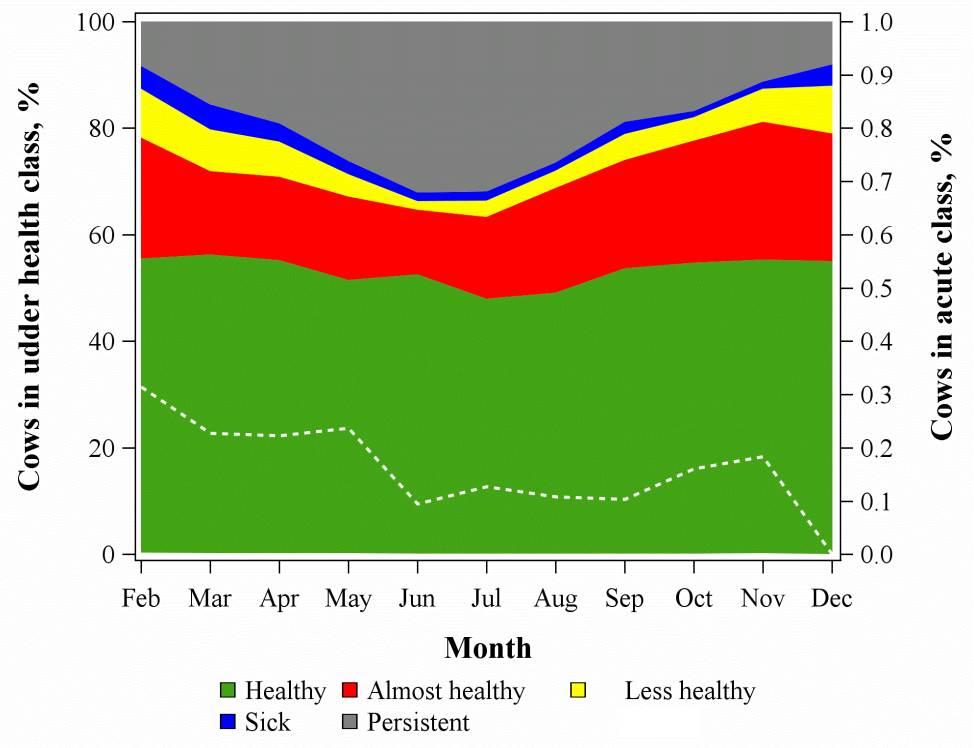
Mahalanobis distance:

Dist = SQRT ($a^2 + b^2 + c^2 + d^2$ ….)

where a, b, c … are **standardized** deviations from the mean – *MD* gives the length of the ”deviation vector”

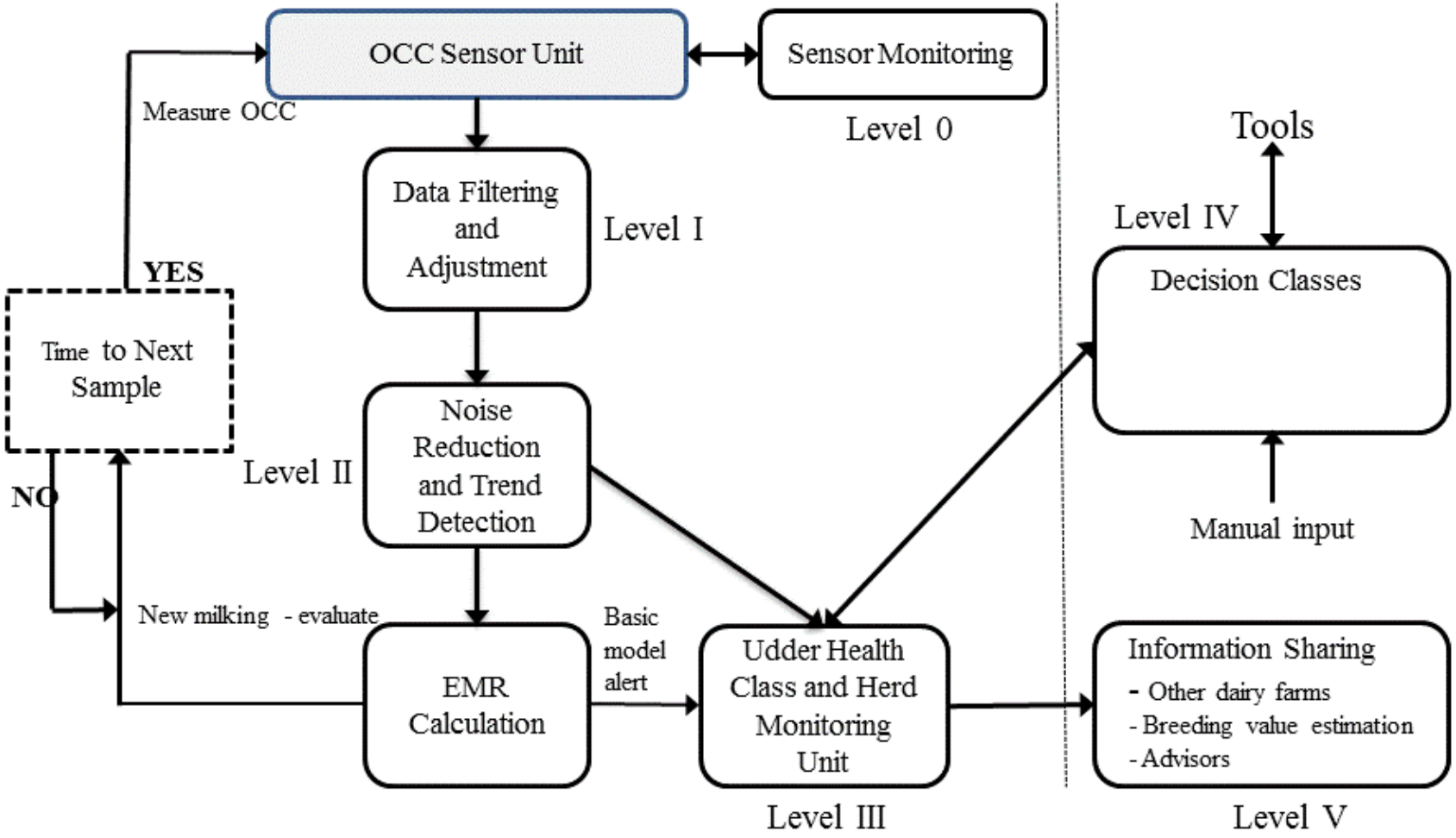


Categories for management
Suggested operating procedures – SOP's
Validation procedures – what is the gold standard?



- Classical validation, sensitivity, specificity
- Need test in commercial conditions
- What's the reference – is it a true gold standard?

Feed-back system and exchange of data



Modules **must** include fix-up tools to mend holes in data!

Wash and dry

- **Wash** away bad, wrong or noisy data when the instrument (sensor) is not OK
- Re-calibrate sensors using reference data – internal or external

- Fit ***time series*** to stabilize data, and use deviations as ”**dry**ed signal”

Thanks …

Løvendahl P & LP Sørensen, 2016. Frequently recorded sensor data may provide health status of cows if data are handled carefully and errors filtered away. Biotechnol. Agron. Soc. Environ. 20(1). OA.