ISESS 2015, Melbourne Australia
March 26, 2015

Quality Control of Environmental Measurement Data with Quality Flagging

Mauno Rönkkö, Okko Kauhanen, Markus Stocker, Harri Hytönen, Ville Kotovirta, Esko Juuso, and Mikko Kolehmainen

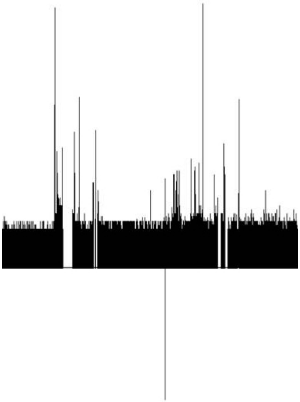
@envinf

UNIVERSITY OF EASTERN FINLAND

Introduction

- Increased interest in environmental monitoring
  - For process understanding, assess human impact, ...
  - Increasingly automated and large scale
  - Example projects include SMEAR, ICOS, NEON, GLEON
- Measurement, the process, prone to disruptions
- Resulting data often of low quality
- Standard data *representation* models exist, e.g. OGC
  - Attribute `resultQuality` to represent quality (value)
- However,
  - How to *assess* quality?
  - How to assess quality in real-time?

Household water consumption



- Quantity of water consumed between consecutive time points
- Positive values possible
  - Peaks plausible
- Zero values possible
  - Most frequent?
- Negative values make no sense!

Aims

- Discuss quality control of measurement data
  - Using quality flagging
- Implement quality flagging
  - Using an ESB-based software architecture

Quality control of measurement data

- Applications may want quality of individual data point
  - Meaning quality at *dataset* level is insufficient
- In such cases, utilize quality flagging
- With a flagging scheme, such as that of
  - Nordic meteorological institutes[1]
  - Four quality control levels: QC0, QC1, QC2, HQC
  - Ten quality flag values: $0 \ldots 9$
  - Formula to compute overall quality flag
  - Interpretation specific for weather measurements
  - We propose a generic interpretation

Quality control levels

Level	Performed by	Mode
QC0	Device or station	Real-time
QC1	Data acquisition system	Real-time
QC2	Data management system	Batch
HQC	Human operator	Batch

Quality flag values and interpretations

Value	Original interpretation	Generic interpretation
0	No check performed	Value not checked
1	Observation is ok	Approved value
2	Suspected small difference	Suspicious value
3	Suspected big difference	Anomalous value
4	Calculated value	Corrected value
5	Interpolated value	Imputed value
6	(Not defined originally)	Erroneous value
7	(Not defined originally)	Frozen value
8	Missing value	Missing value
9	Deleted value	Deleted value

Quality flag and example data

The quality flag C is computed as follows:

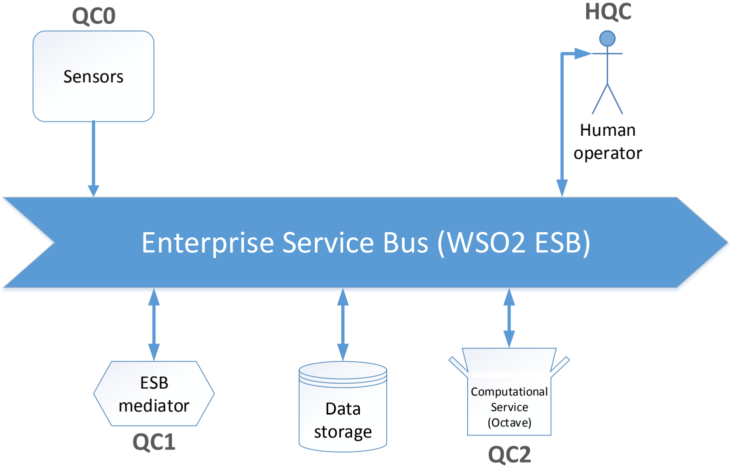
$$C = E_{QC0} + 10 \times E_{QC1} + 100 \times E_{QC2} + 1000 \times E_{HQC}$$

where $E_{QC0}, E_{QC1}, E_{QC2}, E_{HQC}$ are quality flag values for the corresponding quality control level.

time	room temperature	quality flag
2015-03-10T09:30	3.1	9330
2015-03-10T09:30	21.8	4000

9330 = Value not checked by device; anomalous value by data acquisition and management systems; deleted value by human operator. 4000 = Corrected value by human operator.

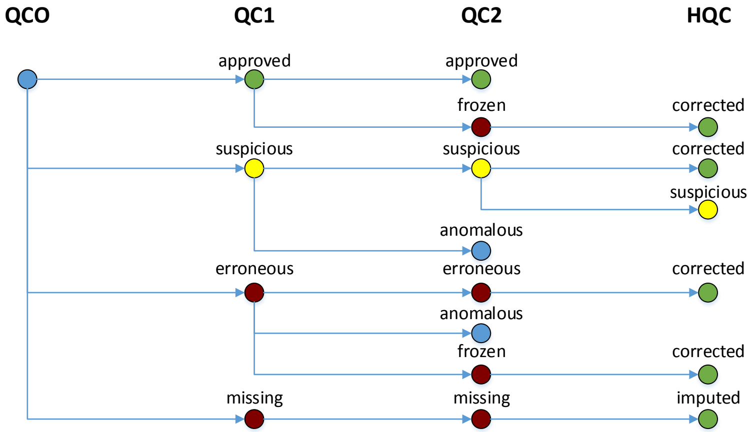
Implement quality flagging



Case study

- Monitoring of residential buildings
- Tested for room temperature and water consumption
- Implemented ESB architecture
- Tested various statistical methods
- Low-cost sensors do not perform QC0
- Instead, QC1 also performs QC0 checks

Room temperature data quality control



Related work

- UncertML[2]
  - Utilized in measurement data quality control[3,4]
  - Proposed also as extension to OGC standards
  - Interoperable representation of probabilistic uncertainties
  - However, uncertainty primarily at dataset level
  - Also, uncertainty $\neq$ quality
- Quality flagging done at device level, e.g. Vaisala

Take aways

- Quality flagging for measurement data quality control
- Requires some flagging scheme
- Advantages of
  - Flagging: individual data point, querying, diagnostics
  - ESB: reconfiguration, data format support, scalability
- Disadvantages
  - Quality flag interpretation is implicit
  - Flags may be too coarse

References

[1] F. Vejen (ed), C. Jacobsson, U. Fredriksson, M. Moe, L. Andresen, E. Hellsten, P. Rissanen, T. Palsdottir, and T. Arason. Quality Control of Meteorological Observations. Automatic Methods Used in the Nordic Countries. Climate Report 8/2002, Norwegian Meteorological Institute, 2002.

[2] M. Williams, D. Cornford, L. Bastin, and E. Pebesma. Uncertainty markup language (UncertML), OpenGIS Discussion Paper 08-122r2, Open Geospatial Consortium Inc., 2009.

[3] M. Williams, D. Cornford, L. Bastin, R. Jones, and S. Parker. Automatic processing, quality assurance and serving of real-time weather data, Computers & Geosciences, 37, pp. 351-362, 2011.

[4] M. Williams, D. Cornford, B. Ingram, L. Bastin, T. Beaumont, E. Pebesma, G. Dubois. Supporting interoperable interpolation: the INTAMAP approach. In Proceedings of the In-ternational Symposium on Environmental Software Systems, Prague, 2007.