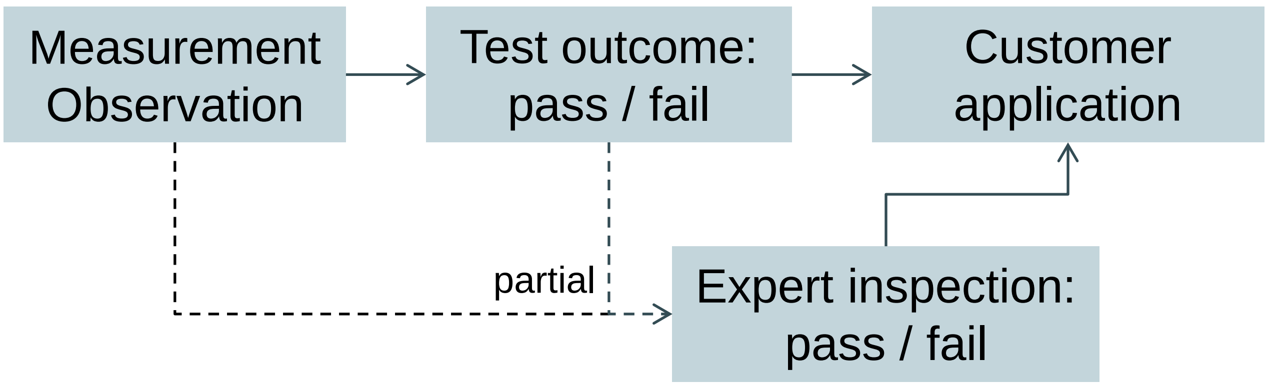
Probabilistic Plausibility of Surface Data

EUMETNET Data Management Workshop 2017

Christian Sigg – christian.sigg@meteoswiss.ch
Markus Abbt, Deborah van Geijtenbeek and Marc Musa

In a Nutshell: Legacy Data Model



Categorical plausibility information:

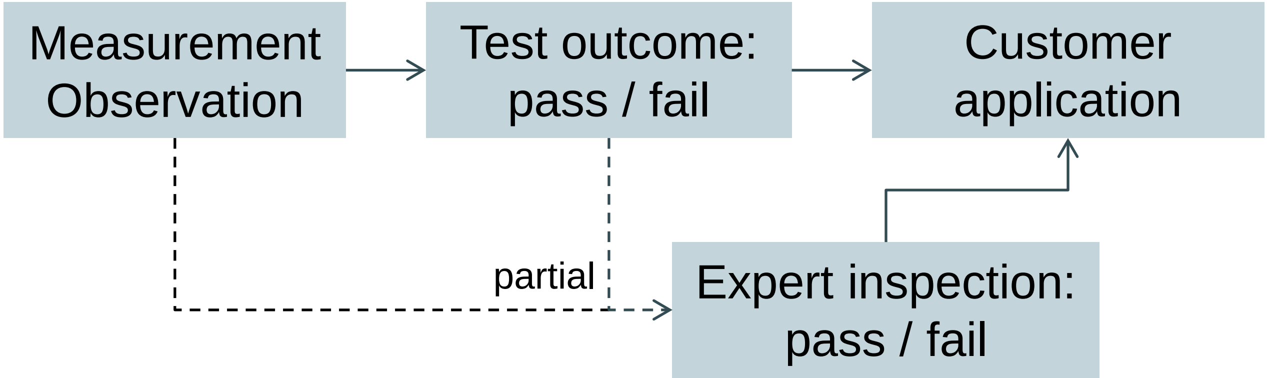
- Define categories of test failures (physical impossibility, climatological limits, …)
- Record failure category with each measurement

\+ Straightforward representation

– Test outcomes have different evidence-strength

– Hard to integrate in customer application

In a Nutshell: Proposed Data Model



Continuous plausibility information:

1. Store individual test outcomes (both “pass” and “fail”)
2. Compute *probabilistic plausibility:* chance that measurement would pass expert inspection, given all test outcomes

+ Test outcomes contribute according to their evidence
+ Customer sets plausibility threshold 0 - 100 % for their application
– Needs computation to obtain probabilistic plausibility

The Bigger Picture

An overview of the modernization of our complete data processing chain:

“Next Generation of Quality Management Tools at MeteoSwiss”, presented by Marc Musa in Session 2 on Thursday

Overview

1. Introduction
2. QC at MeteoSwiss
3. Probabilistic plausibility:
   - Prior
   - Test likelihood
   - Posterior
   - Combining multiple test outcomes
   - Integrating expert inspections
4. Toy examples
5. Discussion

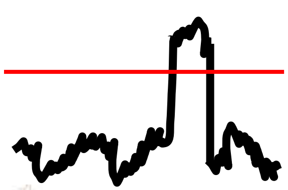
Quality Control at MeteoSwiss

Independent components that contribute to data quality

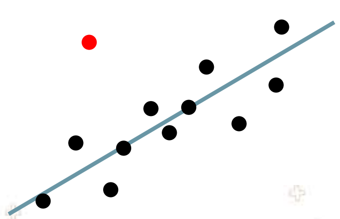
Automated QC testing in **several stages** of data processing chain:

1. Instrument
2. Collection
3. DB import
4. Post-import: hourly, daily, …, to seasonal

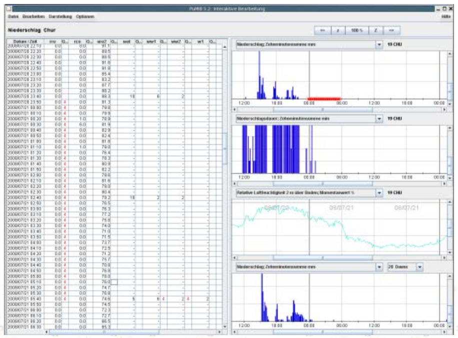
Logical rules:



Data driven:



Expert inspection:



Legacy Quality Information

Our legacy data model records **what happened**: measurement failed test of category

1. Physical impossibility
2. Climatologically unlikely
3. Inconsistent to another parameter
4. Spatially inconsistent

But **what does that imply**: how serious is a test error?

Categories 2 – 4 are difficult to interpret -> most customers only make use of physical impossibility information.

Goal: Make quality information **usable**. For example “If a measurement failed this test in the past, it was deemed implausible by the expert 2 out of 3 times.”

Overview

1. Introduction
2. QC at MeteoSwiss
3. Probabilistic plausibility:
    - Prior
    - Test likelihood
    - Posterior
    - Combining multiple test outcomes
    - Integrating expert inspections
4. Toy examples
5. Discussion

Test Outcome Evidence

Automated tests are incomplete and create false alarms.

How do test outcomes contribute evidence about measurement plausibility?

	Test passed	Test failed
Data plausible	True negative (TN)	False positive (FP)
Data implausible	False negative (FN)	True positive (TP)

Plausibility

Definition of plausibility:

- A measurement is plausible if it would pass expert inspection.
- An implausible measurement would either be rejected or corrected by the expert.

Expert inspection is our gold standard and by definition has full evidence strength: We assume that experts don't commit false negative and false positive errors.

Prior Plausibility $p(q)$

Probabilistic plausibility **before** automated testing and inspection:

$$p(q = 1) = 1 - p(q = 0)$$

$q \in \{1,0\}$: measurement is plausible or implausible

Example: $p(q = 1) = 0.99$ corresponds to 1 in 100 chance that measurement would fail expert investigation.

Interpretation depends on:

- Measurement frequency: $p(q = 1) = 0.99$ considered «okay» for daily observations, but «terrible» for automated 10 min measurements
- Customer application

Estimating the Prior Plausibility

Estimated by counting:

$$\hat{p}(q = 1) = 1 - \frac{|\mathcal{I}|}{|\mathcal{M}|}$$

$\mathcal{M}$: set of all tested measurements

$\mathcal{I} \subseteq \mathcal{M}$: implausible measurements

Subjective estimates are also possible.

Test Likelihood $p(t|q)$

Likelihood of test outcome given the plausibility of the measurement:

$$p(t|q)$$

$t \in \{1,0\}$: test outcome «passed» or «failed»

- Tests with low false positive rate $p(t=0|q=1)$ provide strong evidence for implausible measurements.
- Tests with low false negative rate $p(t=1|q=0)$ provide strong evidence for plausible measurements.

Estimated by counting inspected test outcomes

Posterior plausibility $p(q|t)$

Probabilistic plausibility **after** automated testing and/or inspection:

$$p(q|t)$$

Computed from prior and test likelihood using **Bayes' rule**:

$$p(q|t) \propto p(t|q)p(q)$$

- Test outcome either increases or decreases the prior plausibility
- Normalization of posterior is trivial to compute (in our case)

Multiple Test Outcomes

Compute posterior plausibility given multiple test outcomes as

$$p(q|t_1, t_2) \propto p(t_1, t_2|q)p(q)$$

Naïve Bayes assumption: Test outcomes are conditionally independent

$$p(t_1, t_2|q) = p(t_1|q)p(t_2|q)$$

→ Posterior computed from product of individual test likelihoods:

$$p(q|t_1, t_2) \propto p(t_1|q)p(t_2|q)p(q)$$

Overview

1. Introduction
2. QC at MeteoSwiss
3. Probabilistic plausibility:
    - Prior
    - Test likelihood
    - Posterior
    - Combining multiple test outcomes
    - Integrating expert inspections
4. Toy examples
5. Discussion

Toy Example: Physical Limit Test

Prior $p(q)$:

	Plausible	**Implausible**
	0.98	0.02

Assumption: 2 % of all measurements are implausible

Likelihood $p(t_1|q)$:

	Plausible	**Implausible**
Fail	0	0.001
Pass	1	0.999

Assumption: 0.1 % of all implausible values fail physical limits

Posterior:

Plausibility after test failure: $p(q = 1|t_1 = 0) = 0$

Plausibility after test pass: $p(q = 1|t_1 = 1) = 0.98002$

Toy Example: Climatological Limit Test

Prior $p(q)$:

	Plausible	Implausible
	0.98	0.02

Likelihood $p(t_2|q)$:

	Plausible	Implausible
Fail	0.01	0.1
Pass	0.99	0.9

Assumptions:

- *1 % false positive rate (by design)*
- *10 % of all implausible measurements fail climatological limits*

Posterior:

Plausibility after test failure: $p(q = 1|t_2 = 0) = 0.83$

Plausibility after test pass: $p(q = 1|t_2 = 1) = 0.982$

Toy Example: Combining Test Outcomes

Likelihood of climatological limit test $p(t_2|q)$:

	Plausible	**Implausible**
Fail	0.01	0.1
Pass	0.99	0.9

Likelihood of weaker test $p(t_3|q)$:

	Plausible	**Implausible**
Fail	0.02	0.07
Pass	0.98	0.93

Posterior:

Plausibility after test failures: $p(q = 0|t_2 = 0, t_3 = 0) = 0.58$

Plausibility after test passes: $p(q = 1|t_2 = 1, t_3 = 1) = 0.983$

Toy Example: Expert Inspection

Expert corrects a false positive generated by climatological limit test:

Before expert inspection:

Plausibility: $p(q = 1|t_2 = 0) = 0.83$

Likelihood of expert inspection $p(t_4|q)$:

	Plausible	Implausible
Fail	0	1
Pass	1	0

After expert inspection:

Plausibility: $p(q = 1|t_2 = 0, t_4 = 1) = 1$

Overview

1. Introduction
2. QC at MeteoSwiss
3. Probabilistic plausibility:
   - Prior
   - Test likelihood
   - Posterior
   - Combining multiple test outcomes
   - Integrating expert inspections
4. Toy examples
5. Discussion

Practical Concerns

Storage:

- Unknown or irrelevant test outcomes can be safely omitted

Computation:

- Posterior is a multiplication of a few terms
- New tests can be introduced without recomputing existing posterior probabilities

Inference:

- Prior and test likelihoods estimated by simple counting of proportions
- Conditional independence assumption of Naïve Bayes works well

Conclusions and Further Work

Probabilistic Plausibility:

- Quantitative representation of measurement quality between 0 and 100 %
- Combines prior information, multiple outcomes from automated tests and expert inspection
- Each test outcome contributes according to its evidence strength
- Efficient computation, scales to our surface DB (currently ~ 17 billion records)

Project status:

- Detailed concept complete
- Preparation of new DB schema on-going
- Plan to have probabilistic plausibility available in summer 2018

MeteoSchweiz
Operation Center 1
CH-8058 Zürich-Flughafen
T +41 58 460 91 11
www.meteoschweiz.ch

MeteoSvizzera
Via ai Monti 146
CH-6605 Locarno-Monti
T +41 58 460 92 22
www.meteosvizzera.ch

MétéoSuisse
7bis, av. de la Paix
CH-1211 Genève 2
T +41 58 460 98 88
www.meteosuisse.ch

MétéoSuisse
Chemin de l'Aérologie
CH-1530 Payerne
T +41 58 460 94 44
www.meteosuisse.ch