Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

MeteoSwiss

AI for the Quality Control of Surface Data

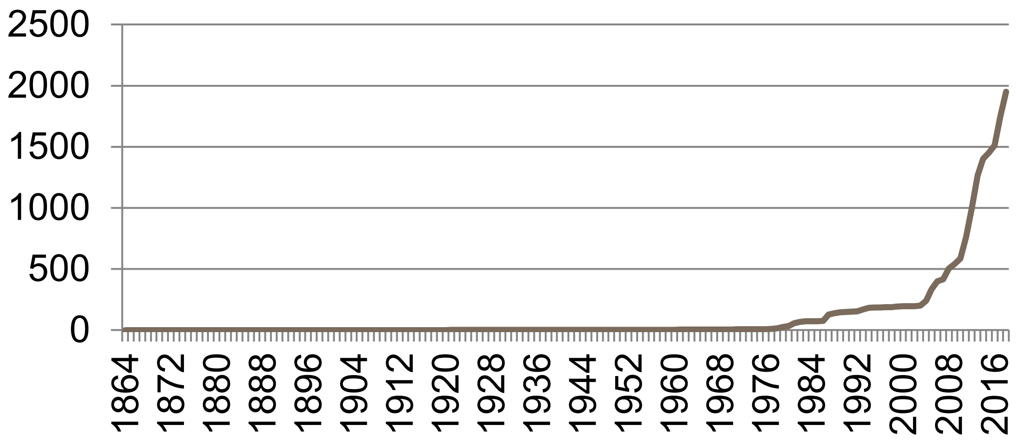
EUMETNET Workshop on Artificial Intelligence for Weather and Climate
Brussels – February 26th, 2020

Christian Sigg, Markus Abbt, Martina Caseri, Claudine Hotz, Valentin Knechtl, Frank Lumpert, Marc Musa and Deborah van Geijtenbeek

christian.sigg@meteoswiss.ch

In a nutshell

The growing volume of surface data is both a **challenge** and an **opportunity**

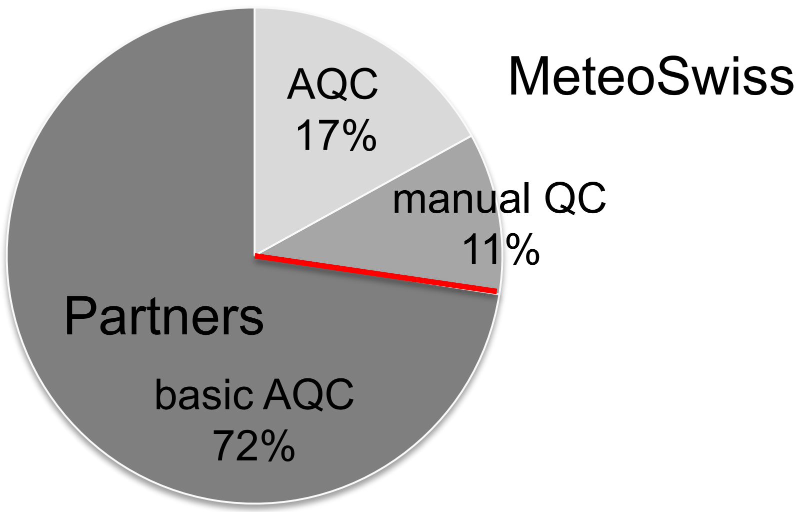


surface data records [millions]

In a nutshell

The growing volume of surface data is both a **challenge** and an opportunity:

- Only a tiny fraction can be inspected manually → automated QC must act as a powerful filter

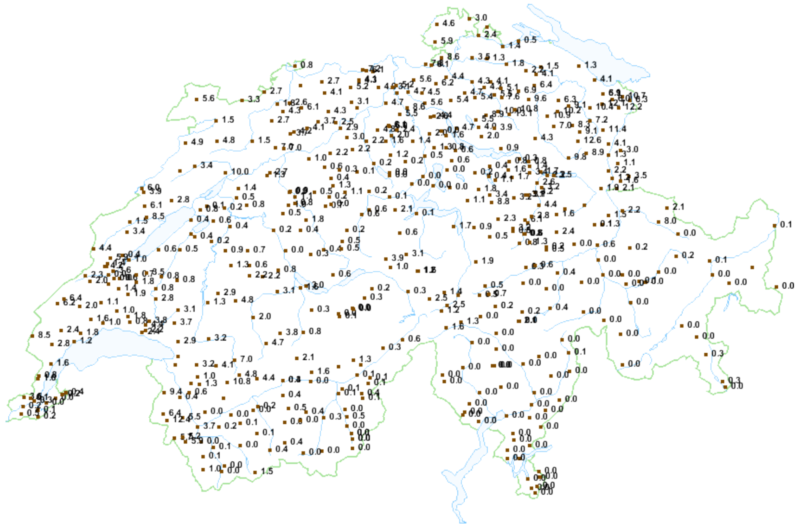


~ 400 suspect values receive daily manual inspection (0.006 % of all surface records)

In a nutshell

The growing volume of surface data is both a challenge and an **opportunity**:

- Only a tiny fraction can be inspected manually → automated QC must act as a powerful filter
- Data-driven quality control works better if more data is available

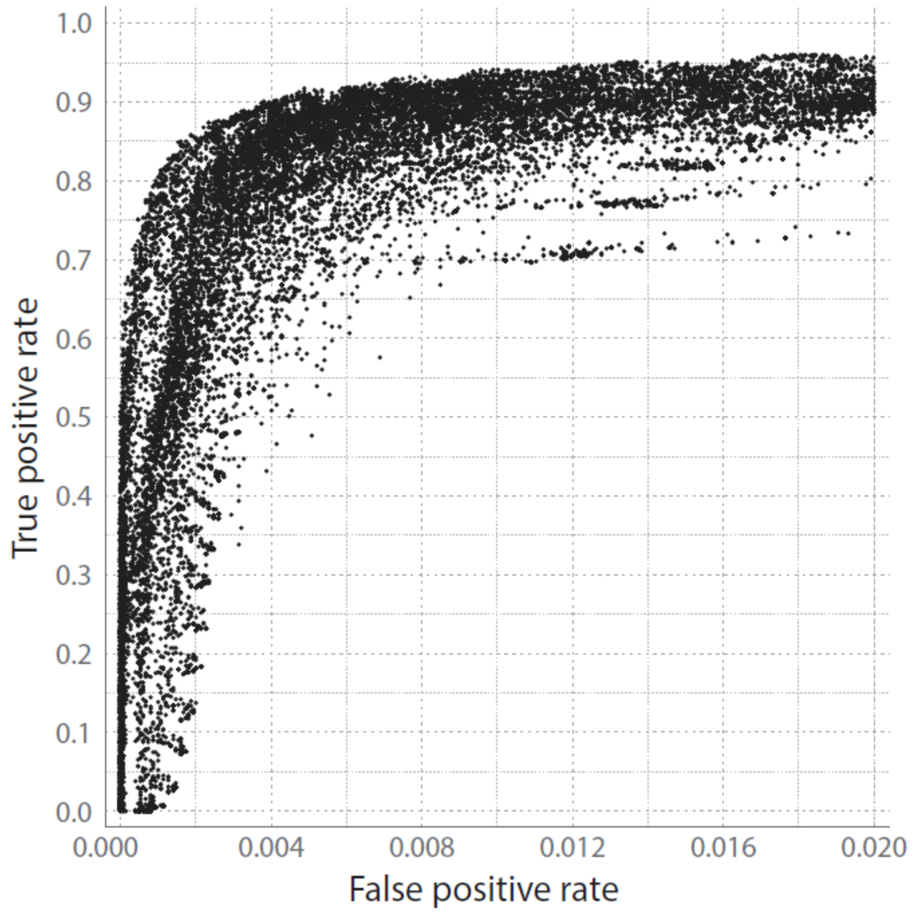


510 sites measuring daily precipitation

In a nutshell

We use AI techniques:

- To develop domain specific QC tests with an optimized cost-benefit ratio



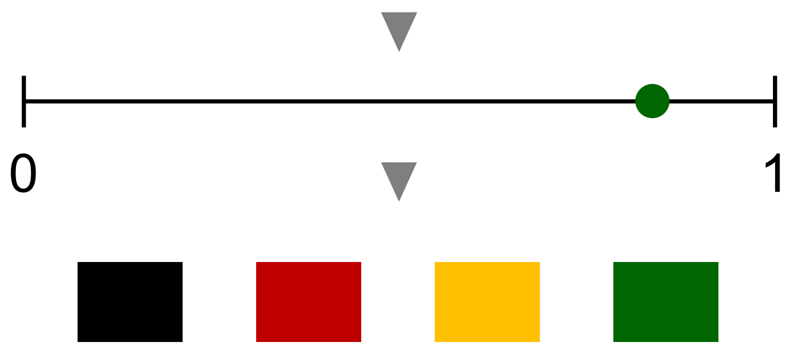
ROC for detecting spurious precipitation measurements in a weighing rain gauge

In a nutshell

We use AI techniques:

- To develop domain specific QC tests with an optimized cost-benefit ratio
- To provide a summary of all available quality information (QI) that is simple, well-defined and relevant to the user

Measurement	Test	Passed
4614406274	8	N
4614406274	112	Y
4614406274	236	Y

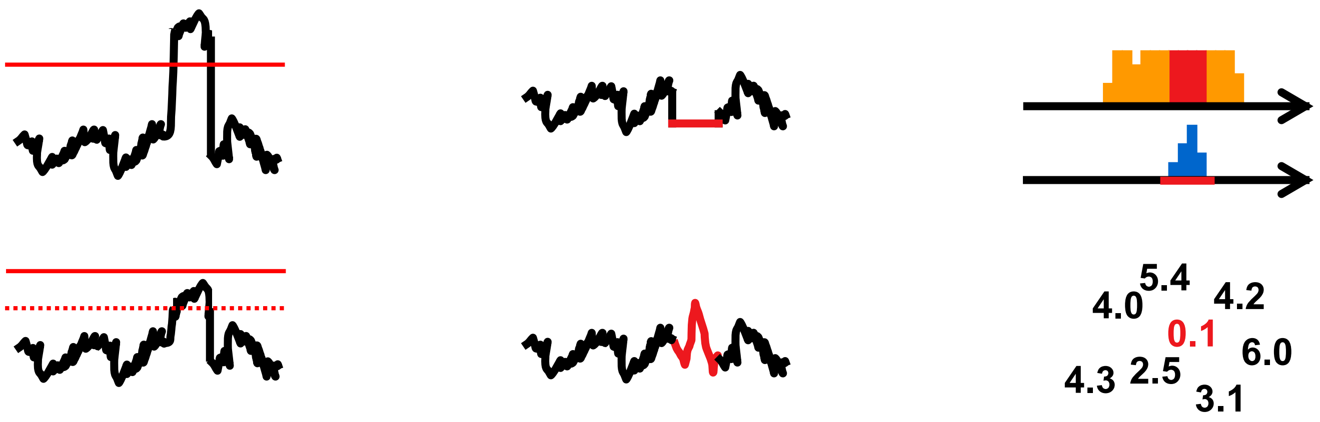


Outline

- From rule design to data-driven models
- Examples
- Probabilistic plausibility
- Running data-driven models in production
- Other AI activities in our group

Rule Design for Automated QC

We employ a rule-based expert system of currently 943 different tests, following WMO guidelines: WMO (2012)

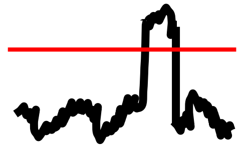


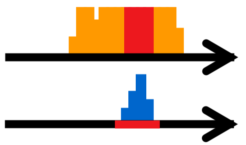
Hard and soft limits *Variability limits* *Consistency*

Strengths and Weaknesses

Evaluation of our rule-set:

- "Simple" rules achieve a good TP to FP ratio, but miss many implausible values
- Consistency rules generate an inacceptable number of FPs, even though they look sensible on paper
- Aggregate complexity: rule-set specification spans > 60,000 table rows
- Redundancy: only 35 % of rules generated test failures in 2019

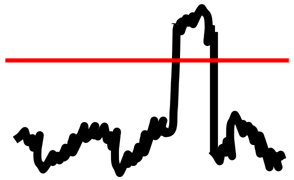
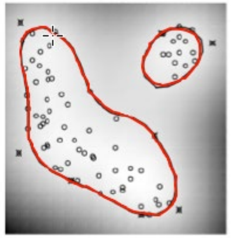




→ Combine simple rules with complex but data-driven models

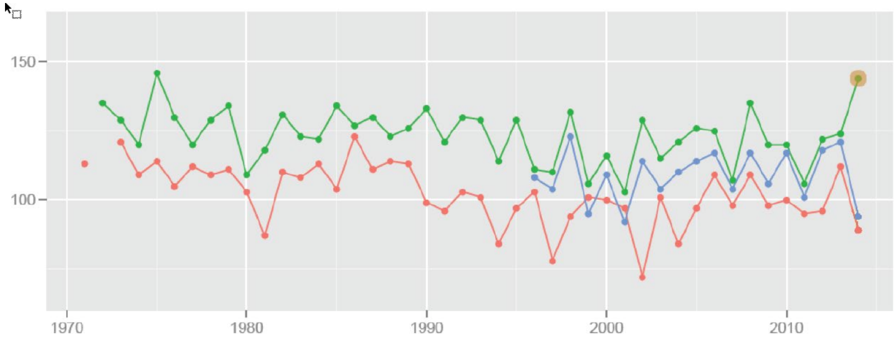
3 Information Sources for Data-Driven QC

1. Relative frequency of occurrence
Principle: “Rough errors are rare”
Model: Outlier detection

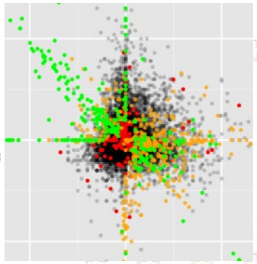



Schölkopf *et al.* (2001)

2. Relationships to other measurements
Principle: “Implausible values are inconsistent”
Model: Continuous regression



3. Expert feedback
Principle: “Model imitates expert”
Model: Discrete classification



Outline

- From rule design to data-driven models
- Examples
- Probabilistic plausibility
- Running data-driven models in production
- Other AI activities in our group

Detecting Spurious Precipitation

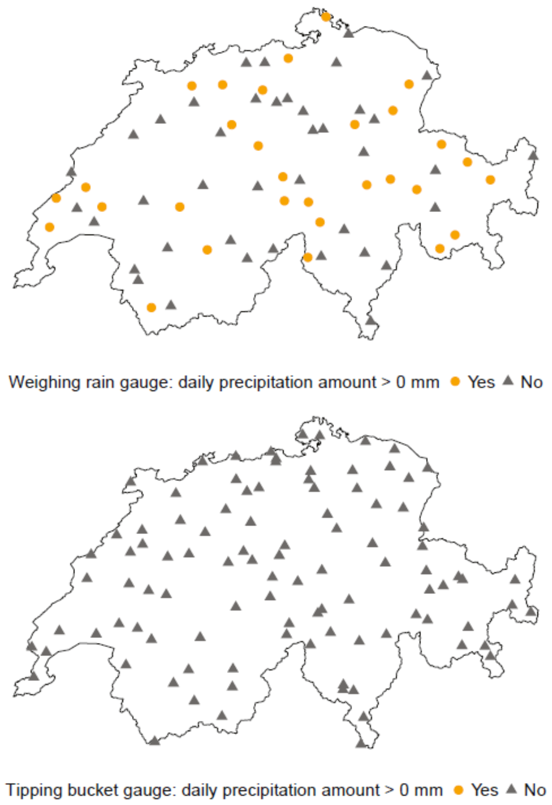
Automated precipitation network:

- 100 Lambrecht 15188/1518H3 tipping bucket
- 131 OTT Pluvio^2 weighing gauges

Spurious precipitation at weighing gauges:

- Isolated measurements of 0.1 to 0.5 mm / 10 min
- Have negative impact on climatological indices, gridded products and NWP verification

We established a QC regime to manually correct offending measurements to zero and performed a systematic review of all events in 2015.



Spurious precipitation events on February 11, 2015

From Analysis to Quality Control Knechtl *et al.* (2019)

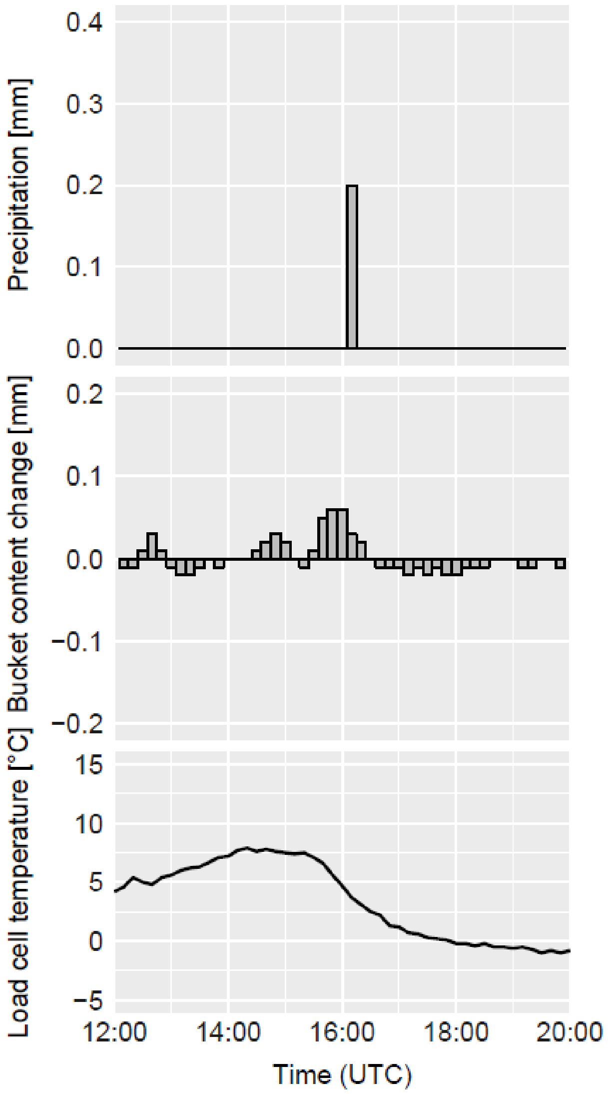
Hypothesis: Spurious precipitation is induced by rapid temperature changes of the load cell

Analysis: Training of SVM classifier on primary and auxiliary measurements and expert corrections

Classifier achieves high specificity and sensitivity

Use for Quality Control:

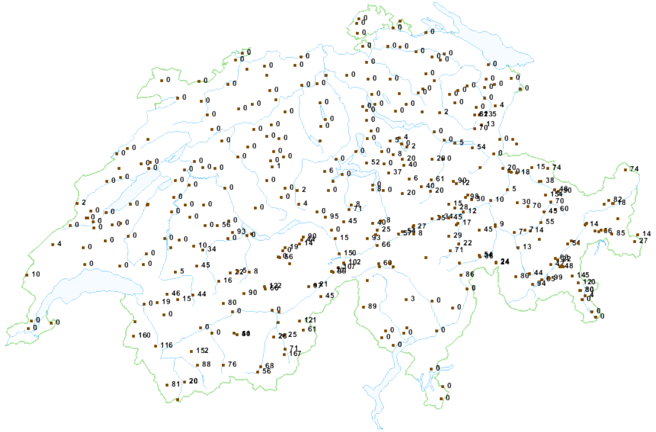
- Classifier is readily deployed as an AQC test
- Site-independent, near real-time
- Classifier relies only on instrument data
  → on-going collaboration with manufacturer



QC of Manual Snow Measurements

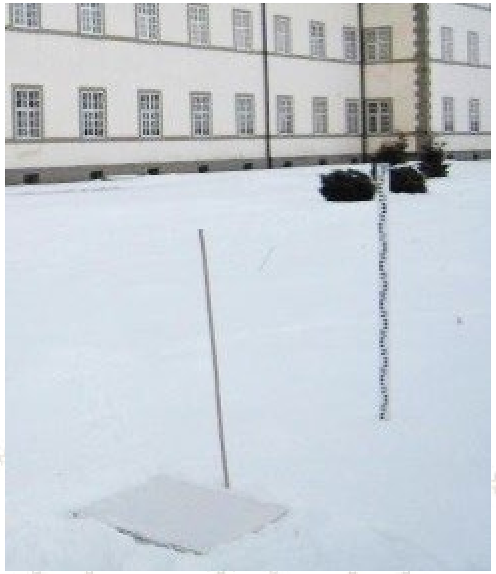
MCH and SLF together operate 372 manual snow measurement sites:

- For avalanche warnings, climatology and hydrology
- Daily measurements of total and new snow depth
- Reported per SMS text



Challenges for QC:

- Drifting snow
- Typos during manual entry
- Observers not following protocol
- Spatial correlation between sites can be low



Predicting Presence and Magnitude

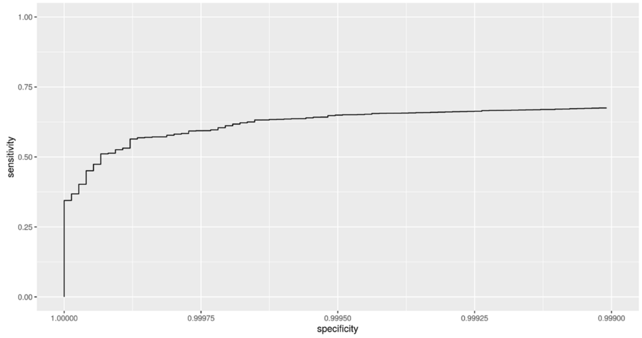
XGBoost classification and regression models for:

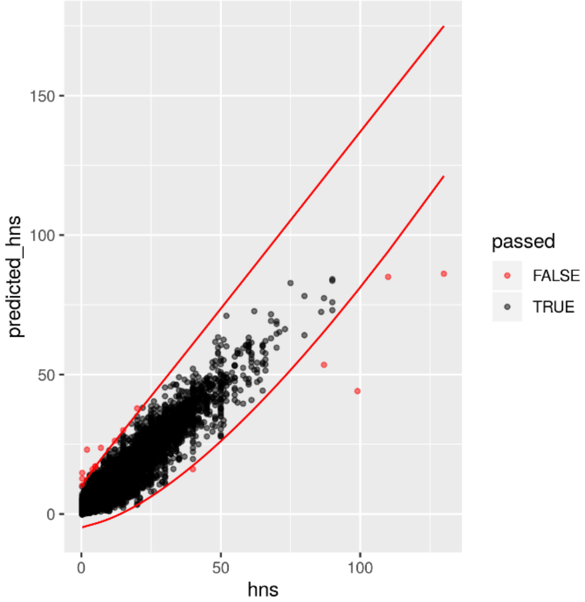
- Presence / absence of new snow and snow cover
- Depth of total and new snow

Features used for prediction:

- Past and future snow measurements
- Temperature, precipitation, global radiation
- Substitution of missing values with grid product estimates

QC pass / fail decision based on ROC curves and quantile regression





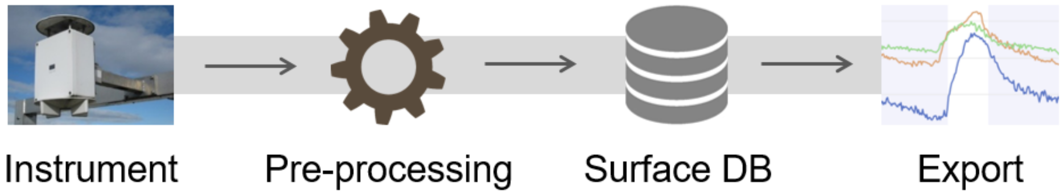
Outline

- From rule design to data-driven models
- Examples
- Probabilistic plausibility
- Running data-driven models in production
- Other AI activities in our group

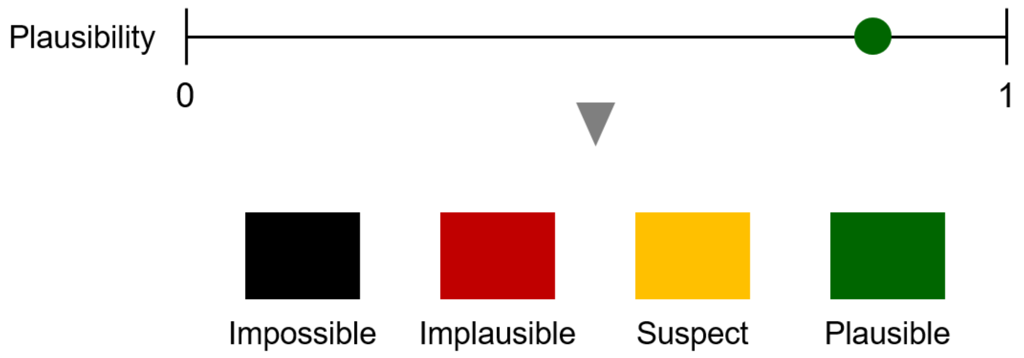
Probabilistic Plausibility

Probabilistic plausibility addresses two challenges:

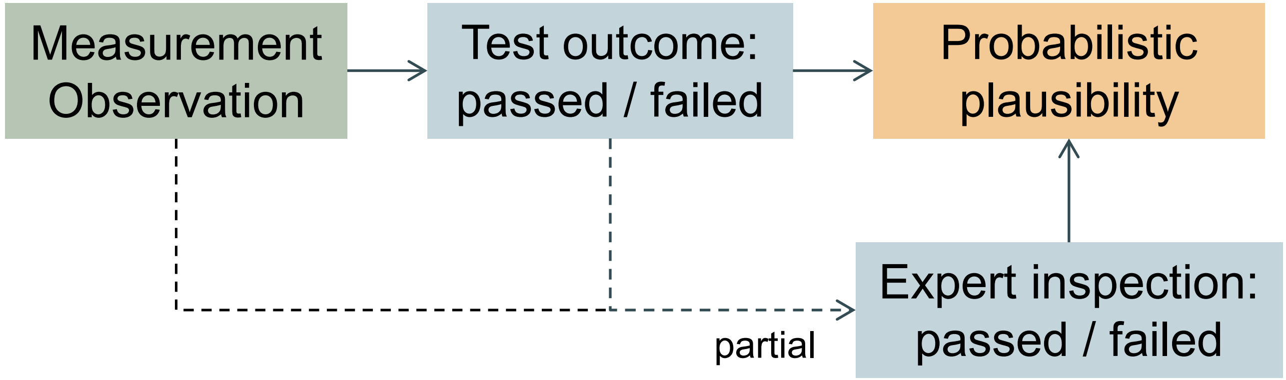
1. How to **combine** QI generated by multiple independent QC systems along the data processing chain



2. How to provide a **summary** of the QI that is simple, well-defined and relevant to the user



Probabilistic plausibility



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

Computation

$$p(q) \rightarrow \boxed{\begin{array}{c}\text{Likelihood}\\ p(t_1|q)\end{array}} \rightarrow p(q|t_1) \rightarrow \boxed{\begin{array}{c}\text{Likelihood}\\ p(t_2|q)\end{array}} \rightarrow p(q|t_1, t_2) \rightarrow \dots$$

Sequential Naive Bayes updates of the plausibility: e.g. Bishop (2006), Sec. 8.2

1. Prior plausibility $p(q)$ before QC
2. Posterior plausibility $p(q|t_1)$ after test result t_1 becomes available
3. Updated posterior $p(q|t_1, t_2)$ after second test result t_2

- Method scales to size of our surface DB (currently 22.8 billion records)
- New tests and QC systems can be introduced without recomputation

Outline

- From rule design to data-driven models
- Examples
- Probabilistic plausibility
- Running data-driven models in production
- Other AI activities in our group

Lessons Learned

Risk of *technical debt*: Sculley *et al.* (2014)

- Feeding models with more data typically increases their performance during training
- But creates dependencies on additional data sources in production

Missing data: naive data imputation can degrade performance in unexpected ways → train cascade of models with fewer features

Erroneous data: induce false positives / negatives → outlier classification of inputs, aggregating over models with split input

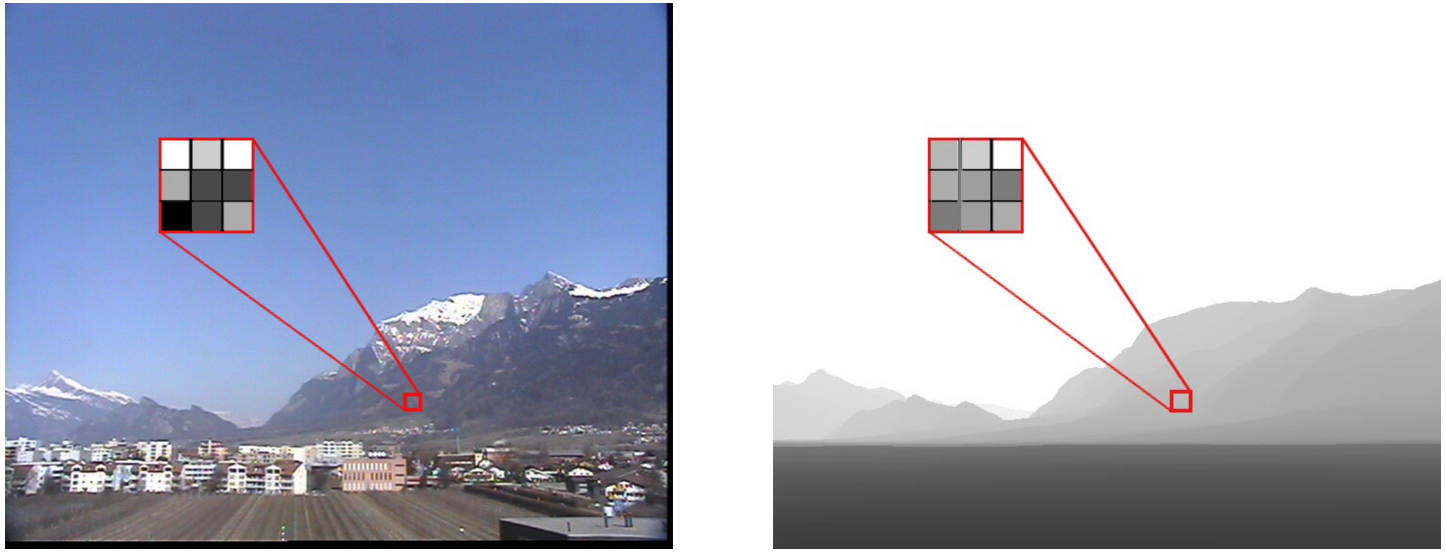
Changing data: e.g. instrument replacement, product version upgrade, new site → simple models reduce re-training need

Outline

- From rule design to data-driven models
- Examples
- Probabilistic plausibility
- Running data-driven models in production
- Other AI activities in our group

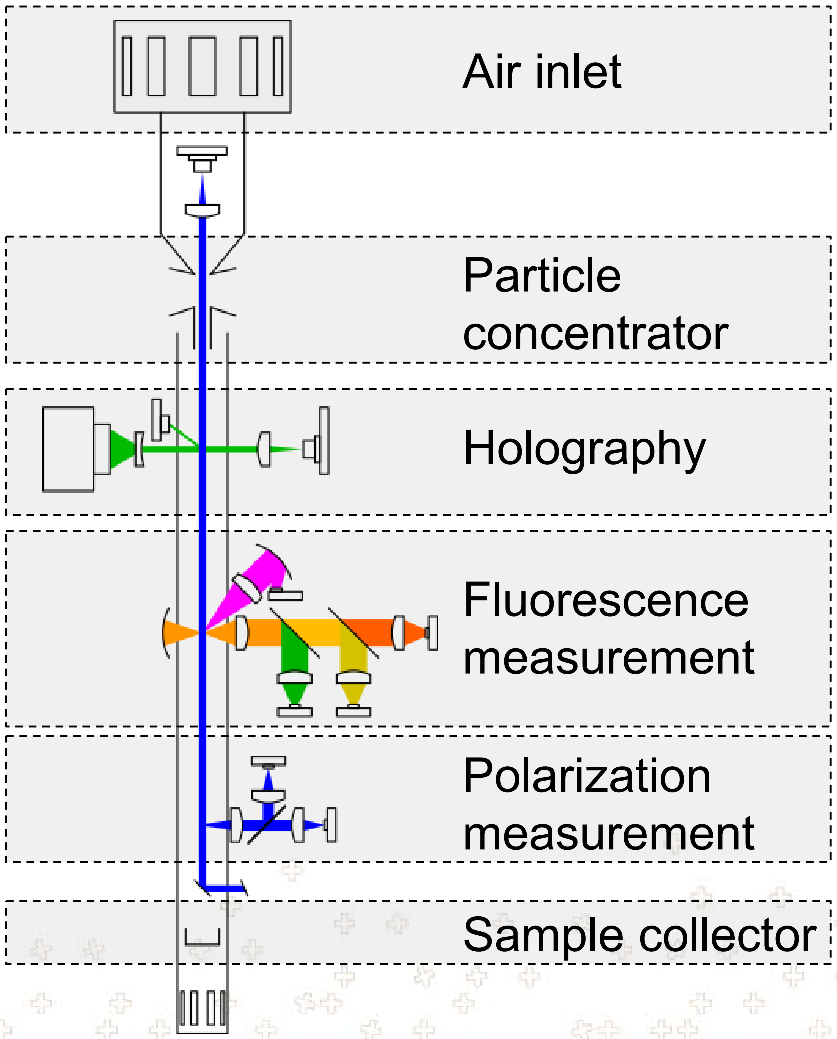
Camera-Based Visibility Estimation

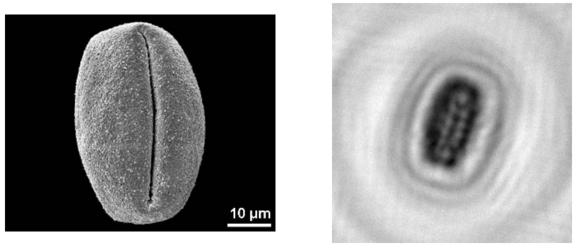
Use CNN classifier to estimate meteorological visibility:



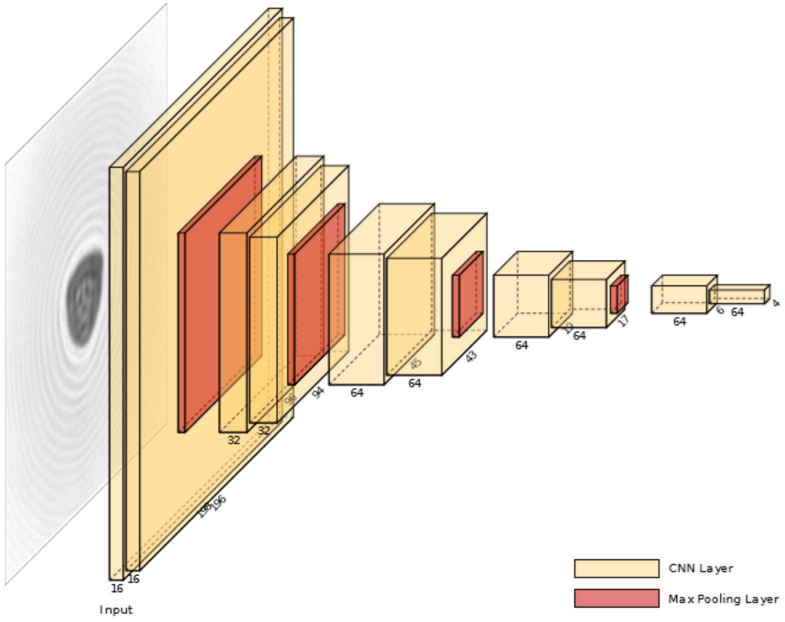
1. Classify image patches as before or beyond the visibility limit
2. Combine with depthmap to estimate prevailing visibility

CNN for Automated Pollen Classification Sauvageat *et al.* (2019)





Hologram of oak pollen



cGAN for Photo-Realistic Forecast Visualization

Use conditional Generative Adversarial Networks Isola *et al.* (2017)
to visualize a weather forecast:

- Generator transforms input image to match future weather conditions
- Discriminator tries to distinguish between real and generated images

Current weather

G

Forecast visualization

Ground truth

Summary

AI models make use of the rapidly growing data volume:

1. Model-based testing for AQC
2. Combining and summarizing QI for the user
3. Analysis and transformation of high-dimensional data domains

Operational use of AI models poses new **challenges**:

- Minimizing technical debt during training
- Keeping model decisions interpretable

AI models clearly realize their potential for the QC of surface data.

Bibliography

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Criel, J. and E. Tsiporkova (2006). *Gene Time Expression Warper: a Tool for Alignment, Template Matching and Visualization of Gene Expression Time Series*. Bioinformatics, 22(2), 251-252.

Gandin, L. S. (1988). *Complex Quality Control of Meteorological Observations.* Monthly Weather Review, 116(5), 1137-1156.

Isola, P., J. Zhu, T. Zhou and A. Efros (2017). *Image-to-Image Translation with Conditional Adversarial Nets*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1125-1134.

Knechtl, V., M. Caseri, F. Lumpert, C. Hotz and C. Sigg (2019). *Detecting Temperature Induced Spurious Precipitation in a Weighing Rain Gauge.* Meteorologische Zeitschrift, 28(3), 215-225.

Sauvageat, E. *et al.* (2019). *Real-time Pollen Monitoring using Digital Holography.* Atmos. Meas. Tech. Discuss., in review.

Sakoe, H. and S. Chiba (1978). *Dynamic Programming Algorithm Optimization for Spoken Word Recognition.* IEEE Trans. ASSP, 26(1), 43-49.

Schölkopf, B., J. C. Platt, J. Shawe-Taylor, A. Smola, and R. C. Williamson (2001). *Estimating the Support of a High-Dimensional Distribution.* Neural Computation, 13(7), 1443-1471.

Sculley, D., G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary and M. Young (2014). *Machine learning: The high interest credit card of technical debt.* In: SE4ML: Software Engineering for Machine Learning (NeurIPS Workshop).

WMO (2012). *Guide to the Global Observing System*. WMO No. 488.