

Automatic data quality control

for understanding extreme climate events

The understanding of extreme events strongly depends on knowledge gained from data. Data integration of multiple sources, scales and earth compartments is the focus of the project Digital Earth, which also join efforts on the quality control of data. Automatic quality control is embedded in the ingest component of the O2A, the observation-to-archive data flow framework of the Alfred-Wegener-Institute. (Fig.1).

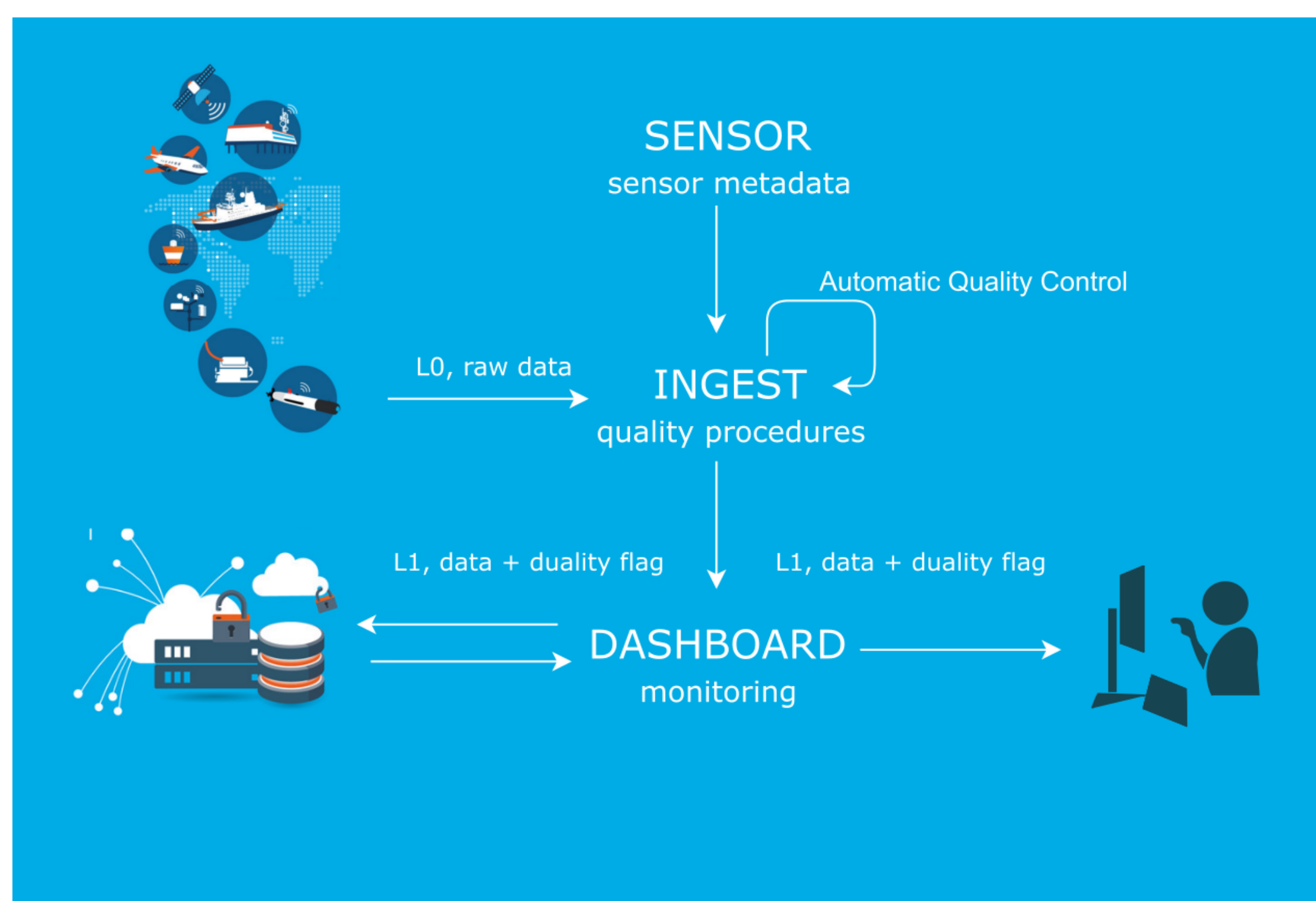


Figure 1. The automatic quality control is embedded in the O2A-Ingest and request for observation properties from the O2A-Sensor to deliver quality-flagged data.

The automatic quality control scans the ingesting data in near-real-time (NRT) format, builds a table of devices, and search - either by absolute or derivative values - for correctness and validity of observations. The availability of observation properties at O2A-Sensor¹, for instance operation ranges, triggers the automatic quality control, which in turn iterates through the table of devices to set the quality flag for each sample and observation. Quality flags are monitored using the O2A-dashboard² (Fig. 2)

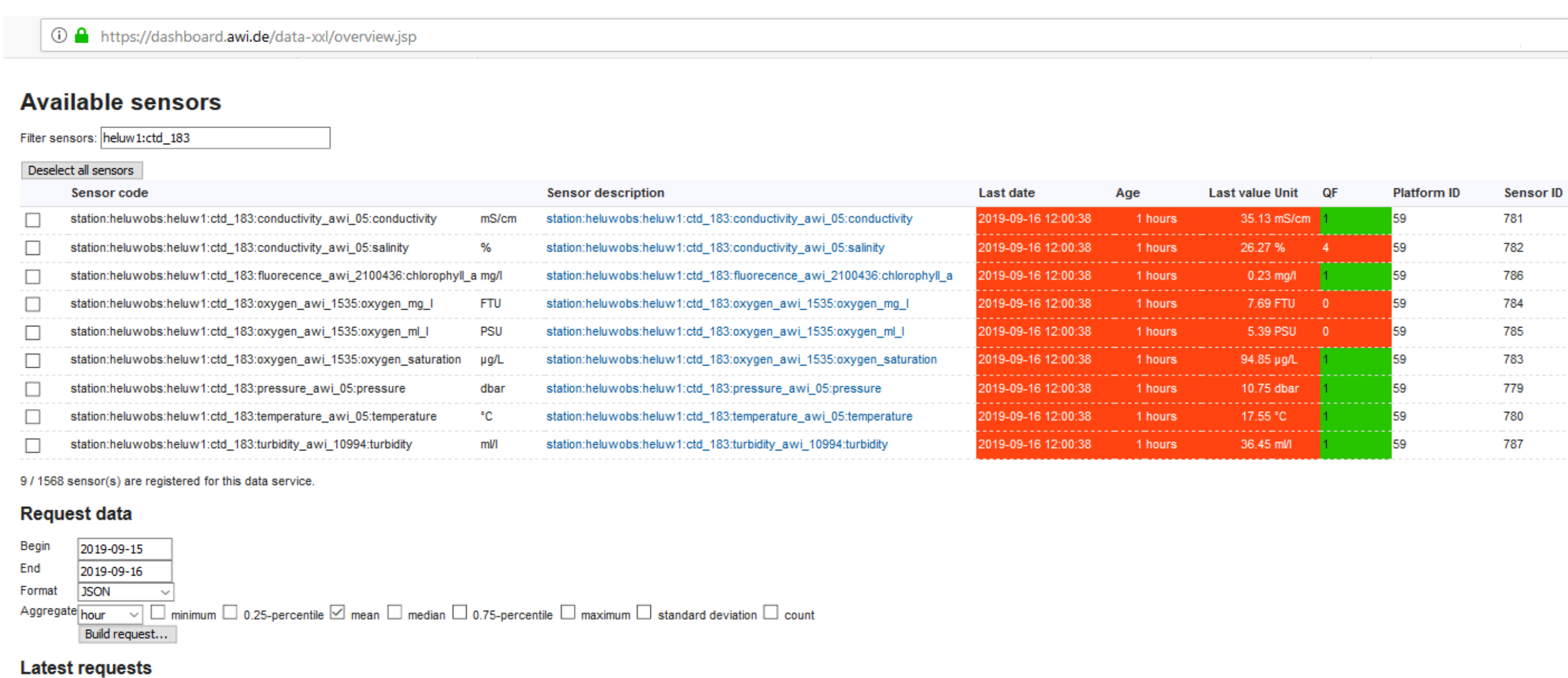


Figure 2. The automatic quality control is embedded in the O2A-Ingest and request for observation properties from the O2A-Sensor to deliver quality-flagged data.

To date, the quality flags in use are sequential and qualitative, i.e. it describes a level of quality in the data. A new flagging system is under development to include a descriptive characteristic that will comprise technical and user interpretation.

At present the O2A-automatic quality control follows a procedural approach to implement formulations found in the literature and other observatory networks. A set of plausibility tests including range, spike and gradient tests are currently operational (Fig 3.).

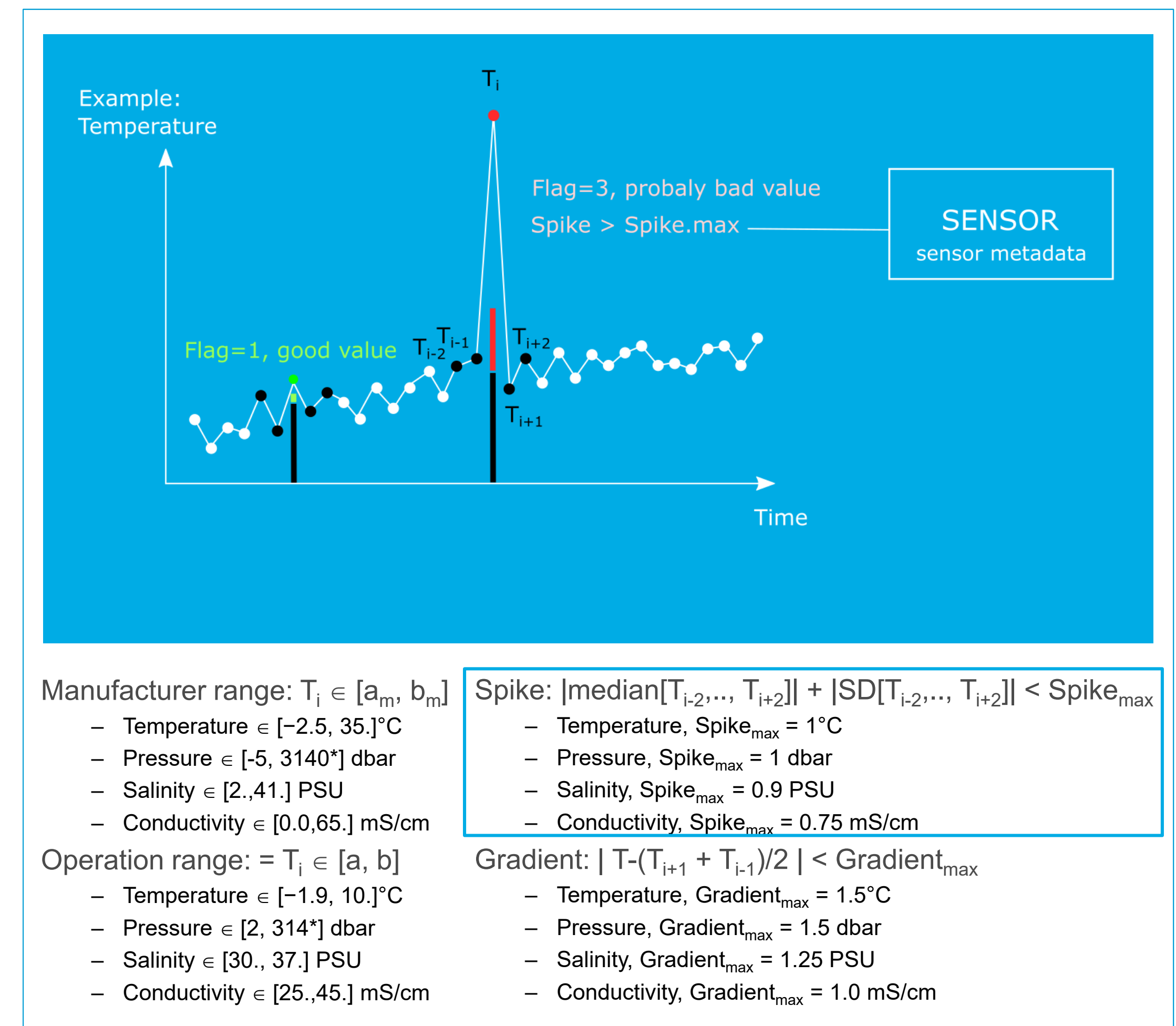


Figure 3. Some of the formulations currently operational in the O2A.

In the technical documentation, equally named quality tests differ in their algorithms. For instance, where “spike” is a deviant from the adjacent values, OOI³ uses ranges, while ARGO⁴ uses arithmetic mean, and FZJ⁵ also limits to differences of same signal. To standardize such procedures a review of existing formulations is under construction, that should also help to overcome limitations of the current approach (Fig 4.).

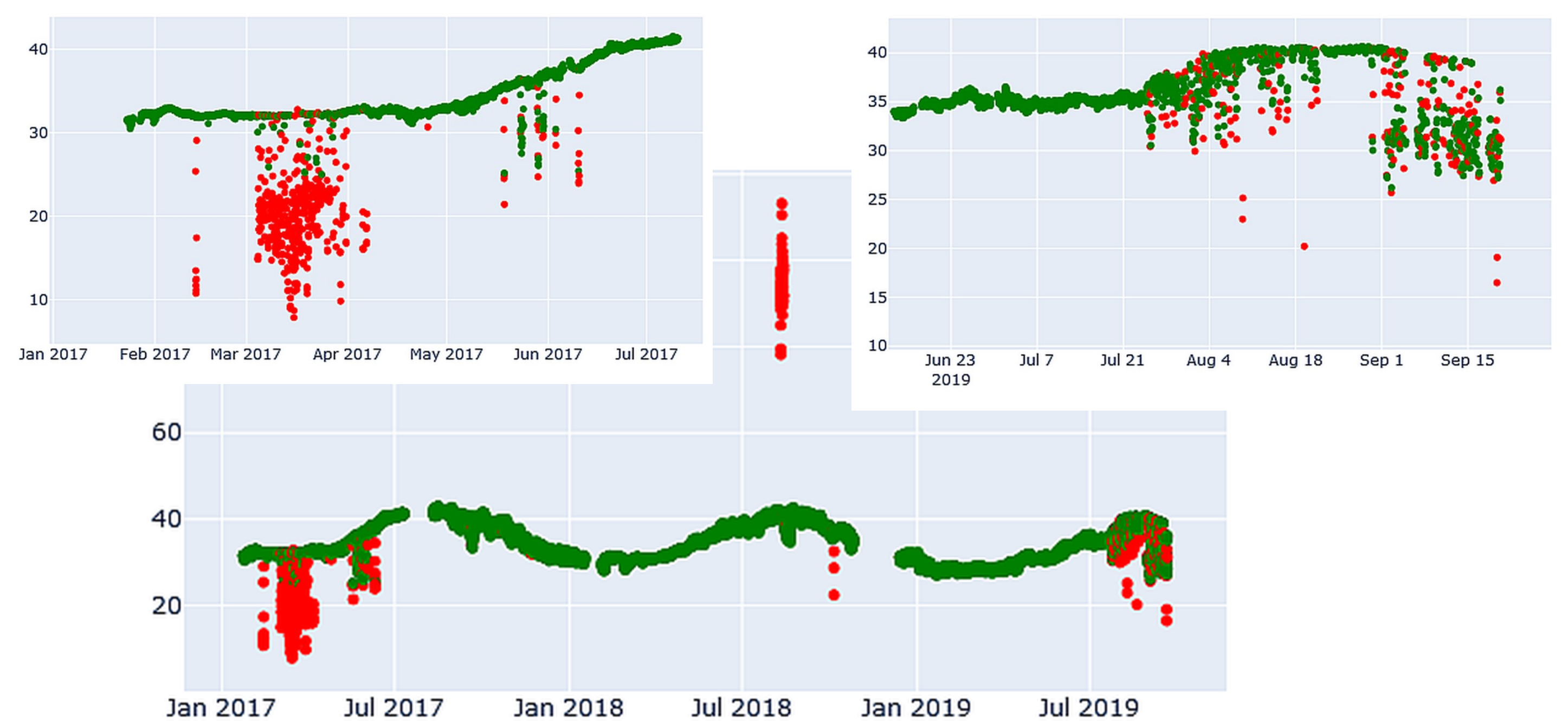


Figure 4. Example of results applied to conductivity (mS/cm) using currently operational tests at O2A-ingest

Within Digital Earth, data on flood and drought events along the Elbe River and methane emissions in the North Sea are to be reviewed using automatic quality control. Fast and scalable automatic quality control will disentangle uncertainty raised by quality issues and thus improve our understanding of extreme events in those cases.