

Quality Control Process for High-Frequency Water Quality Data

The data displayed in the high-frequency water quality dashboard has passed through an automatic Quality Control (QC) process (Roberts, C, 2022).

The process identifies anomalies in environmental sensor data by applying a series of checks:

Instrument limit-based: Detects values falling outside manufacturer-specified sensor limits. These include the value of the parameter of interest, and for more complex instruments can also incorporate the values of other measurements taken by the instrument at the same time, e.g. different raw wavelength measurements and an internal QC score. Note that data just outside the limits may be useful, and may be included in the platform, it is marked as Uncertain.

Impossible values: Identifies values that are not possible for that instrument to measure, such as negative values from a water level sensor which can only measure positive measurements

Recurring values: Identifies repeated or constant values using rolling standard deviation over a defined time window. This is often associated with errors in digital sensors whereby an unsuccessful measurement results in reporting the last value in the memory

Spike detection: Detects sudden deviations from the rolling median centred on a predefined time window. The rolling median is then compared to a threshold calculated by the rolling standard deviation of the same window.

At the end of the process, a quality code is applied to each observation, providing an initial pass over the incoming dataset, allowing end users and applications to omit suspect data. Data that is not flagged by any of the checks gets a default code of 1020 - Auto – Good.

Data that has passed through this process is sufficiently robust to fulfil the primary objective of the program, that is to allow stakeholders to view trends in water quality parameters in near real-time. A secondary objective of the program is to generate a long-term record of values for those parameters. For this, further work is required to identify anomalies such as site-specific impossible values, sensor drift, calibration errors, highly variable values, and readings affected by maintenance or other factors. Additionally, valid spikes which are the result of actual environmental changes need to be preserved. This work is currently performed manually, although work to automate detection of these anomalies is ongoing.

Most data displayed on the application are presented as recorded; however, water level data is aggregated to four hourly averages. Data flagged with a Bad quality code is not aggregated, and the aggregated value is assigned the lowest quality code included in the aggregation.

Any quality code which does not include the word Auto was applied manually.

These are the codes commonly appearing in the application:

Code	Name
1000	Good - Verified
1020	Good - Auto QC
1500	Good - Values manually adjusted
2000	Uncertain - Unchecked
2011	Uncertain - Auto (Above parameter limits)
2012	Uncertain - Auto (Below parameter limits)
2013	Uncertain - Caution spectral quality
2061	Uncertain - Auto (Caution spectral quality)
2230	Uncertain - Value above sensor limit
2500	Uncertain - Values manually adjusted

References

Roberts, C. (2022). waterQUAC: An R Library for Automated Water Quality Data Validation (v1.0.0) [Source code]. Github. <https://github.com/UncleCamsWaterPlans/waterQUAC>

Last updated: 30/04/2026