Check for updates

SOFTWARE TOOL ARTICLE

REVISED SentemQC - A novel and cost-efficient method for quality assurance and quality control of high-resolution frequency sensor data in fresh waters

[version 2; peer review: 3 approved]

Sofie Gyritia Madsen van't Veen[1,2], Brian Kronvang[2], Joachim Audet[2], Thomas Alexander Davidson[2], Erik Jeppesen[2], Esben Astrup Kristensen[1], Søren Erik Larsen[2], Jane Rosenstand Laugesen[1], Eti Ester Levi[2], Anders Nielsen[3], Peter Mejlhede Andersen[2]

[1]Nature and water environment, Envidan A/S, Silkeborg, DK-8600, Denmark
[2]Department of Ecoscience, Aarhus University, Aarhus, DK-8000, Denmark
[3]WaterITech Aps, Skanderborg, DK-8660, Denmark

v2 **First published:** 07 Nov 2024, **4**:244
https://doi.org/10.12688/openreseurope.18134.1
Latest published: 29 May 2025, **4**:244
https://doi.org/10.12688/openreseurope.18134.2

Abstract

The growing use of sensors in fresh waters for water quality measurements generates an increasingly large amount of data that requires quality assurance (QA)/quality control (QC) before the results can be exploited. Such a process is often resource-intensive and may not be consistent across users and sensors. SentemQC (QC of high temporal resolution sensor data) is a cost-efficient, and open-source Python approach developed to ensure the quality of sensor data by performing data QC on large volumes of high-frequency (HF) sensor data. The SentemQC method is computationally efficient and features a six-step user-friendly setup for anomaly detection. The method marks anomalies in data using five moving windows. These windows connect each data point to neighboring points, including those further away in the moving window. As a result, the method can mark not only individual outliers but also clusters of anomalies. Our analysis shows that the method is robust for detecting anomalies in HF sensor data from multiple water quality sensors measuring nitrate, turbidity, oxygen, and pH. The sensors were installed in three different freshwater ecosystems (two streams and one lake) and experimental lake mesocosms. The explored sensor data revealed anomaly percentages ranging from 0.1% to 0.8% of the cleaned datasets by SentemQC. While the sensors in this study contained relatively few

Open Peer Review

Approval Status ✓ ✓ ✓

	1	2	3
version 2 (revision) 29 May 2025		✓ view	✓ view
version 1 07 Nov 2024	✓ view	? view	? view

1. **Andrea Lami**, National Research Council, Water Research Institute (CNR IRSA),, Verbania Pallanza, Italy

2. **Hongwei Guo**, Nankai University, Tianjin, China

3. **Arno Hooijboer**, National Institute for Public Health and the Environment, RIVM, Bilthoven, The Netherlands
Elma Tenner, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

anomalies (<2%), they may represent a best-case scenario in terms of use and maintenance. SentemQC allows the user to include the individual sensor uncertainty/accuracy when performing QC. However, SentemQC cannot function independently. Additional QA-QC steps are crucial, including calibration of the sensor data to correct for zero offsets and implementation of gap-filling methods prior to the use of the sensor data for determination of final real-time concentrations and load calculations.

Plain language summary

SentemQC is an open-source Python tool designed to ensure the quality of high-frequency sensor data. The purpose of the method is to clean and perform quality assurance (QA) and quality control (QC) on large volumes of data such as high-frequency data with data points down to every minute. SentemQC is computationally efficient and user-friendly, making it suitable for processing extensive datasets quickly. At the same time, SentemQC features a straightforward five-step setup, making it accessible even for non-experts to perform anomaly detection. The method identifies anomalies using five moving windows calculating the mean or median in each window which is used to detect and mark the anomalies in the dataset. Our results show that the method works well for water quality data (nitrate, turbidity, oxygen, and pH) in various freshwater ecosystems such as lakes, streams, and experimental ponds. This means that the method is robust. However, SentemQC is not standalone as additional steps like calibration and gap-filling in the dataset are necessary for the final cleaning of the data to perform for example load calculations. In summary, SentemQC is a valuable tool for maintaining data quality, however, it should be considered as a part of a broader QA-QC process.

Keywords

SentemQC, quality assurance, quality control, sensor data, high-frequency data, python tool

H2020 This article is included in the Horizon 2020 gateway.

HE This article is included in the Horizon Europe gateway.

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Sofie Gyritia Madsen van't Veen (svv@ecos.au.dk)

Author roles: van't Veen SGM: Conceptualization, Data Curation, Formal Analysis, Methodology, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kronvang B**: Conceptualization, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Validation, Writing – Review & Editing; **Audet J**: Conceptualization, Investigation, Methodology, Supervision, Writing – Review & Editing; **Davidson TA**: Data Curation, Project Administration, Resources, Writing – Review & Editing; **Jeppesen E**: Project Administration, Resources, Supervision, Writing – Review & Editing; **Kristensen EA**: Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Larsen SE**: Methodology, Software, Supervision, Validation, Writing – Review & Editing; **Laugesen JR**: Funding Acquisition, Project Administration, Supervision, Writing – Review & Editing; **Levi EE**: Resources, Writing – Review & Editing; **Nielsen A**: Conceptualization, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Writing – Review & Editing; **Andersen PM**: Conceptualization, Formal Analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101060020 (Nitrogen and phosphorus load reduction approach within safe ecological boundaries for the Nordic-Baltic region [NORDBALT-ECOSAFE]), grant agreement No 731065 (Network of Leading European AQUAtic MesoCOSM Facilities Connecting Mountains to Oceans from the Arctic to the Mediterranean [AQUACOSM]), and grant agreement No 871081 (Network of Leading Ecosystem Scale Experimental AQUAtic MesoCOSM Facilities Connecting Rivers, Lakes, Estuaries and Oceans in Europe and beyond [AQUACOSM-plus]), and from the Innovation Fund Denmark from the Industrial Researcher programme under grant agreement No 0153-00078B (Sensors application for high temporal resolution monitoring in Danish streams [SENTEM]), and by AnaEE, Denmark.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 van't Veen SGM *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: van't Veen SGM, Kronvang B, Audet J *et al.* **SentemQC - A novel and cost-efficient method for quality assurance and quality control of high-resolution frequency sensor data in fresh waters [version 2; peer review: 3 approved]** Open Research Europe 2025, **4**:244 https://doi.org/10.12688/openreseurope.18134.2

First published: 07 Nov 2024, **4**:244 https://doi.org/10.12688/openreseurope.18134.1

REVISED **Amendments from Version 1**

In this revised version, we have implemented several improvements inspired by the valuable feedback from the three review reports. To the reviewers, we extend our sincere thanks.

In general, additional information about the SentemQC method has been incorporated into the revised manuscript. In particular, the sections "Implementation" and "Parameterization of the SentemQC Model" have been revised to include more detailed descriptions of the five different moving windows used in SentemQC. This includes equations for the calculations within the moving windows, expanded details on the method's parameterization, and information about the validation step. A new Table 3 has been added to explain the five moving windows and their respective purposes. Furthermore, more information regarding the uncertainty of the sensors included in the method has been provided.

Several new high-quality references have been added to the revised manuscript. The Discussion section has been both expanded to include new, relevant aspects and streamlined to better focus on the study's core arguments. For example, we have added specifications regarding the technical feasibility of implementing potential future improvements, such as the integration of electronic logs and handling of missing data in SentemQC.

Revised versions of Figure 1, Figure 3, and Figure 5 have been included. Additionally, the legends for Figure 3, Figure 5, Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, and Figure 11 have been updated to be more descriptive and smaller changes has been made for Figure 2 and Figure 4.

Finally, the description file for SentemQC published on Zenodo has also been revised to include further details on parameterization and validation. The updated version is titled "SentemQC - Description1.2.pdf.

Any further responses from the reviewers can be found at the end of the article

Introduction

Eutrophication of coastal marine waters and freshwater ecosystems is of global concern as these aquatic systems continue to face significant pressure from human activities and climate change (European Commission, 2021; Riemann *et al.*, 2016; Schmidt *et al.*, 2023). Therefore, it is crucial to monitor the quality of aquatic ecosystems. Recent advancements in sensor technology have enabled rapid collection of high-resolution sensor data from freshwater systems, offering new opportunities for high-frequency (HF) monitoring (Bieroza *et al.*, 2023; Blaen *et al.*, 2024; Rode *et al.*, 2016). Although some sensor technologies may claim to require minimal maintenance and quality assurance (QA)/quality control (QC), integrity and accuracy of HF data are essential and important to avoid producing unreliable and useless data. Therefore, data QA and QC are important, and there is a pressing need to implement standardized QA-QC applications (Bieroza *et al.*, 2023; Wollheim *et al.*, 2018). In a study by Rozemeijer *et al.* (2025) they provide a guideline for the best practice when working with HF water quality monitoring, including a description of field practices together with considerations about the purpose of deploying a sensor. Their guidelines are based on international experiences and can be used for optimizing QA workflows, contributing to a complete and standardized QA-QC process when working with HF water quality monitoring.

HF sensor monitoring can be useful both regarding scientific understanding of transport pathways of solutes in catchments, hydrochemistry, and aquatic ecology (Bieroza *et al.*, 2023) and concerning more management aspects through national monitoring programs. In Europe, the chemical status of water bodies (lakes, streams, etc.) is observed with the goal of achieving good status for Europe's water bodies according to rules in the Water Framework Directive (WFD) (European Commission, 2000). National monitoring programs typically use traditional infrequent discrete grab sampling techniques, but these may miss crucial events and information (Cassidy & Jordan, 2011; Levi *et al.*, 2024). To address this issue, HF sensor technology has the potential for future integration into national monitoring programs (Rode *et al.*, 2016; van Geer *et al.*, 2016) to improve both groundwater and surface water quality management. However, when using HF sensors, there is a need for robust QA and QC, data storage infrastructure, and open data access to enable the investigation of long-term trends. Bieroza *et al.* (2023) advocate for the integration of HF water quality measurements into existing monitoring programs as an important step toward improving water quality, together with increased stakeholder engagement, particularly due to changing environmental conditions and stricter regulatory requirements.

Examples of larger international sensor networks for national monitoring are the U.S. Clean Water Act (CWA), the National Ecological Observatory Network (NEON) in the USA, and the Australian Water Act (AWA) (Bieroza *et al.*, 2023; Chabbi & Loescher, 2017; Dekker & Hestir, 2012; Edmonds *et al.*, 2022; van Geer *et al.*, 2016). In Europe, several countries use sensors to monitor the quality of their freshwater systems. Ireland has pioneered this approach by incorporating sensors into its Agricultural Catchments Program (ACP) for over a decade (Jordan *et al.*, 2012). In 2005, the Global Lake Ecological Observatory Network (GLEON) was launched. It is a voluntary international network connecting researchers working with high-resolution sensor data from lakes and reservoirs with the aim to share and communicate their findings (Read *et al.*, 2016; Weathers *et al.*, 2013).

In most other European countries, apart from Ireland, sensors are not fully implemented in national monitoring programs but are often used in special cases, for example in the regional monitoring in Sweden since 2017 (Fölster *et al.*, 2019) and in a limited number of national river monitoring programs in Norway (Kaste *et al.*, 2022; Skarbøvik *et al.*, 2023b). Another example is Lake Maggiore, which has been part of the Long-Term Ecosystem Research (LTER) network, with continuous monitoring of its physical, chemical, and biological characteristics since the 1980s. In 2020, a HF monitoring system was introduced, featuring a buoy equipped with sensors to complement traditional discrete sampling methods. This HF monitoring program, developed through a cross-border collaboration between Italy and Switzerland, enhances lake quality monitoring and supports effective freshwater ecosystem management (Rogora *et al.*, 2023). QA-QC procedures and other main

features (hardware and software) for such an HF monitoring system of lakes are under development based on the data from Lake Maggiore (Tiberti *et al.*, 2021). The QA-QC of the HF sensor data collected in the different countries varies considerably as shown in a study of turbidity sensor results from 11 different monitoring programs/research projects from six different north-western European countries (Skarbøvik *et al.*, 2023a). Some countries use manual methods for data QC, but this can be time-consuming and makes it difficult to compare not only the data between the countries but also between the different monitoring programs (Lannergård *et al.*, 2019; Skarbøvik *et al.*, 2023b; Skarbøvik *et al.*, 2023a; Yang *et al.*, 2023). In addition, large-scale monitoring programs like NEON and GLEON are continuously expanding their geographic reach and sensor equipment (Schmidt *et al.*, 2023). Furthermore, this enormous amount of HF sensor data requires new smart methods of data quality assurance to ensure that the quality of the collected sensor data is good and acceptable and also follows the FAIR principles (Wilkinson *et al.*, 2016). In some countries, for example Poland, HF sensor data is QA and QC by private companies applying confidential methods. In these cases, it is difficult to understand how the QC has been performed.

Several methods have been developed for the QA and QC of aquatic sensor data from fresh waters (Barnett *et al.*, 2012; Belay *et al.*, 2023; Jones *et al.*, 2022; Read *et al.*, 2016; Schmidt *et al.*, 2023; Schmidt & Kerkez, 2023; Talagala *et al.*, 2021). Some methods such as B3 software, ODM Tools Python, the Great Expectations Framework, and the GCE Data Toolbox are mainly based on manual QC and statistics (Barnett *et al.*, 2012; Gong & Campbell, 2022; Horsburgh *et al.*, 2015; Read *et al.*, 2016; Sheldon, 2008), while others, such as pyhydroqc and SaQC, use more advanced machine learning methods (Belay *et al.*, 2023; Jones *et al.*, 2022; Leigh *et al.*, 2019; Schmidt *et al.*, 2023; Schmidt & Kerkez, 2023).

Belay *et al.* (2023) performed a numerical assessment of 13 deep-learning algorithms on two time-series of sensor datasets and found that there are still challenges in anomaly detection techniques. This is primarily due to the lack of a universal approach as each method is specifically designed for a particular application (Belay *et al.*, 2023). Moreover, these machine learning (ML) algorithms can be difficult to follow for the user and are time-consuming to set up and train to fit a specific dataset, and the computational costs involved in training/testing are large (Belay *et al.*, 2023). Schmidt & Kerkez (2023) also emphasized the need for further research into their QC machine learning method to evaluate the performance of the ML model over longer periods and to assess the generalizability and transferability of ML models trained on various sensor data monitoring identical environmental processes.

Schmidt *et al.* (2023) developed the SaQC method, which includes 32 different QC functions to flag anomalies. They divided the methods into three overall parts: processing, basic QC tests, and advanced QC tests, where the third part includes machine learning algorithms. This means that the user can choose which QC test they want to use for a specific sensor. For example, they included the Search and TRace AnomalY (STRAY) algorithm for multivariate spike detection. The STRAY algorithm is unsupervised and uses k-nearest neighbor distances for score calculations for anomaly detection (Talagala *et al.*, 2021). Leigh *et al.* (2019) introduced a ten-step approach for automated detection of anomalies in HF water quality data from in situ sensors. They reviewed and compared different anomaly detection methods such as automated classification rules and several regression and feature-space methods and advocated for a QC approach, that highlighted the importance of understanding end-user requirements and data characteristics to determine the specific type of anomalies and their importance for each sensor type (Leigh *et al.*, 2019). This framework also allows users to select the most appropriate anomaly detection methods for their specific sensor data.

Anomaly detection concerning HF data has different definitions depending on the focus of the study and input data (Talagala *et al.*, 2021). Goldstein and Uchida (2016) identify three different types of anomalies: i) Global anomalies, which are easy to detect since they are very different than the rest of the dataset, ii) local anomalies, which can be difficult to detect since they can only be identified as an anomaly when compared to its neighborhood data points, and iii) clusters of anomalies, which can be very difficult to detect as these types of anomalies can look like the normal dataset when comparing each datapoint to its neighborhood data points. The third anomaly type has received less attention compared to the other two categories (Goldstein & Uchida, 2016; Talagala *et al.*, 2021). In this study, we define an anomaly closely related to the general definition in (Talagala *et al.*, 2021) that states that an anomaly significantly deviates from the majority of the data, with a much larger distance between typical observations and anomalies compared to the distances among typical observations.

Quality assurance (QA) procedures for environmental sensor data ensure that sensors measure reliable concentrations and meet established requirements. This is achieved by following procedures when using in situ sensors. Quality control (QC) involves testing the validity of the sensor data, thereby validating and controlling the data after collection. In this study, we developed a five-step workflow outlining the overall steps for working with environmental sensor data in freshwater environments. This includes essential steps such as sensor deployment, maintaining an electronic logbook, and managing a database for the large volume of data. As described above, Rozemeijer *et al.* (2025) provides QA workflows with practical guidelines for sensor deployment, contributing to a complete QA-QC process.

This study aimed to explore the feasibility of implementing a simplified robust QC approach, called SentemQC, to HF sensor data from multiple freshwater systems. The focus was on developing a method that was computationally efficient, transparent, and had a user-friendly setup/interface. The method was designed to identify and mark anomalies and detect errors or patterns in the raw sensor data. SentemQC identifies a

data point as an anomaly if it significantly deviates from the other data points within a predefined "moving window". This deviation is characterized by a much larger distance to the other data points in the moving window, based on statistical calculations such as average, standard deviation, and/or mean. Using SentemQC, all three defined types of anomalies by (Goldstein & Uchida, 2016) can be detected.

The objectives were i) to develop a simple, open access, and computationally efficient method for QC procedure of HF sensor data called SentemQC; ii) to investigate the robustness of SentemQC by testing it on different sensor measurement types from different aquatic systems and locations; iii) to explore the advantages and disadvantages of the SentemQC program compared to other available programs.

Study sites

In this study, we examined water quality sensor data from three different freshwater systems (Table 1). The higher measuring frequency in the case of streams is chosen because headwater streams are known to be extremely dynamic systems, which necessitates a higher monitoring frequency to capture the changes in water quality parameters. In contrast, lakes generally respond more slowly to environmental changes compared to small headwater streams, allowing for a lower measuring frequency.

The two stream sensor stations were established in Horndrup Stream and Lyby-Grønning Stream in Jutland, Denmark (Figure 1). The Horndrup stream station covers a catchment area of 5.48 km^2, of which 70% is agricultural land and 19% forests (Levin *et al.*, 2017). The soils are primarily dominated by moraine deposits: fine sand mixed with clay (59%) and fine clay mixed with sand (27%). The stream station is positioned in a part of the stream covered by forest. The Lyby Grønning Stream sensor station covers a catchment area of 11.3 km^2 where 84% is agricultural land (Levin *et al.*, 2017). The soils are dominated by fine sand mixed with clay (47%) and fine clay mixed with sand (52%). Both sensor stations are part of the Danish national monitoring program of the aquatic environment, NOVANA (Petersen *et al.*, 2021; Windolf *et al.*, 2012).

The lake sensor station is placed in Lake Ormstrup in Jutland, Denmark. The lake is 11 ha and has an average depth of 3.4 m and a hydraulic retention time of >1 year. The lake catchment is small (ca 1 km^2), and the lake only has outflow during heavy rainfall events, often in the winter period. The lake is eutrophic with high concentrations of chlorophyll and total phosphorus (TP) (Davidson *et al.*, 2024; Søndergaard *et al.*, 2023).

The lake warming mesocosm experiment (LWME) sensor stations are located in Jutland, Lemming, Denmark. Here, 24 flow-through outdoor freshwater mesocosms were established in 2003 and are still running in 2024. The overall aim of the mesocosms is to experimentally study how increased temperatures affect shallow lake systems at contrasting nutrient concentrations. A description of the mesocosm setup can be found in Liboriussen *et al.* (2005).

Data and methods

Sensor data

This study applied various sensors and data parameters (see Table 1). At the stream stations, nitrate sensors from HACH. In-Situ Nitratax™ plus UV (HACH, Düsseldorf, Germany) was employed, and at Horndrup an additional turbidity sensor from HACH was installed. These sensors gathered data continuously for nearly two years (Table 1). The sensors were manually cleaned in the field with acid every two weeks and installed with a measuring frequency of 1 minute. The nitrate sensors were *in-situ* Nitratax plus UV purely photometrical sensors that are based on the principle that nitrate and nitrite ions absorb ultraviolet (UV) light at a wavelength of 210 nm. The Nitratax plus UV sensors had the manufacturing series number 1640674 for the nitrate sensor installed in Horndrup Stream, and 2060341 for the nitrate sensor installed in Lyby-Grønning Stream. The sensor sends a light signal every minute. However, there is a built-in automatic function that averages the last three measurements to stabilise the measurements. The sensors

Table 1. Overview of the examples of high-frequency sensor data included in this study. This Table illustrates the data examples in the study used for quality assurance and quality control (QA-QC) using SentemQC of high-frequency sensor data.

Type of waterbody	Field site	Data parameters	Measuring period	Measuring frequency
Stream	Lyby-Grønning	Nitrate	2021-05-05 to 2023-01-03	1 min
Stream	Horndrup	Turbidity, nitrate	2021-03-10 to 2023-01-02	1 min
lake	Ormstrup	Dissolved oxygen	2023-07-01 to 2023-10-01	15 min
Lake warming mesocosm experiment (LWME)	Lemming	Dissolved oxygen, pH	2014, 2015, and 2016	30 min

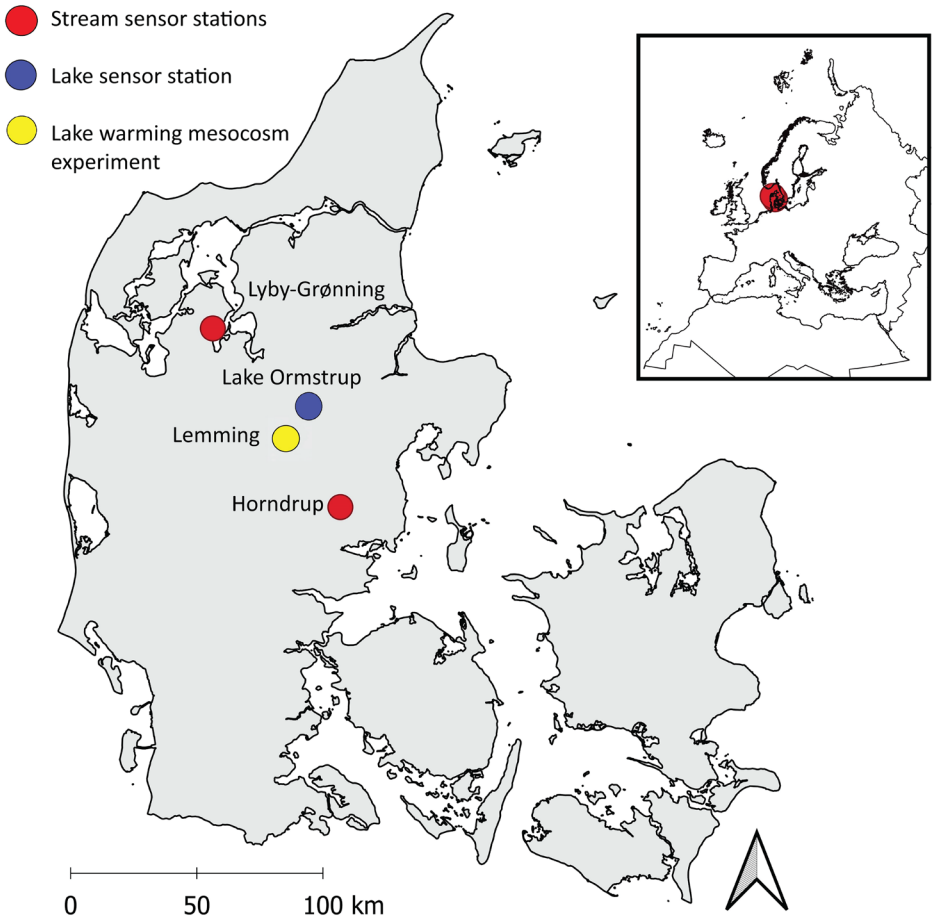

Figure 1. Study sites. The location of the different sensor stations in Denmark. Red dots represent the stream sensor stations Lyby-Grønning Stream and Horndrup Stream, the blue dot represents the lake sensor station Lake Ormstrup, and the yellow dot represents the Lake Warning Mesocosms experiment station in Lemming, Denmark.

also have a built-in wiper technology (double Wiper) for lens cleaning. The double wiper is designed to keep the optical system free of blockages and growth of e.g. algae and deposits (HACH, 2012; HACH, 2014). The nitrate sensors have a measurement range between 0.1 and 100.0 mg N/L. The precision/uncertainty is 3% or ±0.5 mg N/L (HACH, 2012; HACH, 2014). Additionally, the nitrate sensor measurements were before using the SentemQC method corrected for zero offsets, which occur linearly for the nitrate sensors. This correction was achieved by using the sensor measurements taken before and after manual cleaning, along with linear interpolation.

The turbidity sensor installed at the Horndrup stream station were in-situ optical sensor called Solitax (HACH, Düsseldorf, Germany) with the manufacturing series number 1942974. These sensors determine turbidity by an infrared light measuring technique, which quantifies the intensity of light back-scattered by the water at 90° where the light source is infrared light (860nm). The turbidity was measured in the unit FNU (Formazin Nephelometric Unit), which complies with the European ISO 7027. The turbidity sensor has a measurement range between 0.001 to 4000 FNU and a measuring precision/uncertainty of <5% or ±0.01 FNU (Table 2) as well as a built-in function stabilizing the measurements. The turbidity sensor has an inferred light source that measures constantly and provides a 1-minute measurement that is an average of the values recorded during the last minute. Additionally, it has a built-in wiper technology (double Wiper) for lens cleaning (HACH, 2017; HACH, 2018).

Dissolved oxygen (DO) and pH sensors were installed in LWME mesocosm named G2 in Lemming. The dissolved oxygen sensor used was an OxyGuard® two-wire sensor, model 420, and the pH sensor was an OxyGuard light-duty submersion sensor connected to a Manta pH measurement system. Both sensors had a measuring interval of 30 minutes, and all sensors were manually calibrated every week. Further description of the sensors and the mesocosm setup can be found in Liboriussen *et al.* (2005), and the uncertainty is provided in Table 2.

The dissolved oxygen sensors (DO) at Lake Ormstrup station 100011 were Aqua TROLL 500 (In-Situ, Fort Collins, CO,

Table 2. Overview of sensor uncertainty/accuracy for the high-frequency sensor data included in this study. The uncertainty/accuracy of the sensor equipment used to collect the high-frequency sensor data in this study are provided by the manufactories and can be found following the sources in this Table for the different high-frequency sensor data.

Field site	Data parameters	Uncertainty/accuracy	Source
Lyby-Grønning and Horndrup	Nitrate-N	3% of the measured value or ±0.5 mg N/L	(HACH, 2012; HACH, 2014)
Horndrup	Turbidity	Turbidity to 1000 FNU: without calibration <5% of the measured value or ±0.01 FNU	(HACH, 2006)
Ormstrup	Dissolved oxygen	±0.1 mg/L (from 0 to 20 mg/L) ±2% of reading (from 20 to 60 mg/L)	(In-Situ, 2021)
Lemming	Dissolved oxygen	< ±2% of actual value	(OxyGuard International A/S, 1994)
Lemming	pH	0.1 of the actual value is used as a constant. This was chosen because no accuracy is provided in the sensor manual. However, we must be aware that the calibration itself is somewhat uncertain; pH is a log scale and is rounded to 2 decimal places.	(OxyGuard International A/S, 2020)

USA) multisondes installed near the surface (ca. 1.0 m) and at a deeper-water depth (ca. 3.8 m). The measuring frequency was 15 minutes, and the sensor had a wiper function that automatically cleaned the sensor every 15 minutes before measuring. The uncertainty is ±0.1 mg/L (from 0 to 20 mg/L) and ±2% of reading (from 20 to 60 mg/L) (Table 2) (In-Situ, 2021). In addition, the sensors were manually cleaned and calibrated every week. Further description of the sensor setup in Lake Ormstrup can be found in Davidson *et al.* (2024).

QA-QC method

When working with HF sensor data, we recommend following a specific workflow as outlined in Figure 2, ensuring QA-QC (quality assurance and quality control) of the HF sensor data. However, QC using the method described in this study (Step 4) is just one step of the process and stands out for its robustness and simplicity. The SentemQC method can be applied to various HF sensor data by fine-tuning the parameters as described in the following sections.

Implementation

The SentemQC method was developed as a freely available Python code to make the method available and easy to apply for future users. The method is free to download from Zenodo at: https://zenodo.org/doi/10.5281/zenodo.13830624 (Andersen *et al.*, 2024) as an open-source package that aims to develop a smart, simple, open-access, and cost-efficient method for the QC procedure of HF sensor data.

The method is based on simple mathematical algorithms using averages, median, and standard deviation calculations. The algorithm uses moving windows of different widths to flag different data patterns as anomalies. To flag large continuous periods of anomalous data, very wide moving windows are applied, yielding a more stable moving average, while individual anomalies and anomalies are flagged with narrow windows. The narrow moving windows can only be used when the large continuous periods of anomalous data are flagged. The mean, median, and standard deviation were computed for the moving windows using the Pandas library in Python (The pandas development team, 2024). Specifically for each mowing window, rolling statistics were calculated using the Rolling object from pandas.core.window.rolling, with the standard deviation calculated via the Rolling.std() method. Similarly, the moving mean and median were calculated using the Rolling.mean() and Rolling.median() methods, respectively. The moving windows are all centered, and the calculations for a given data point within the window will therefore have the same number of data points before and after. When a moving window passes through data, calculations are only made on the data points that are not flagged. If a given data point lies outside the window, it will be flagged and excluded from the subsequent window calculations. The width of the individual windows is defined by the user as a given number of data points or a time interval (Table 3), while the height of the window measurements changes dynamically depending on the data included in the moving window. The upper limit of the moving window for a given data point is the sum of the moving average or median, the product of the moving standard deviation, a predefined factor, and a constant (Equation 1). The lower limit is the difference between the moving average or median, the product of the moving standard deviation, a predefined factor, and a constant (Equation 2).

The upper boundary for the moving windows for a given sensor measurement is defined by Equation 1:

$$topwin = avg + topadd + std_factor * std \quad (1)$$

The lower boundary of the moving windows for a given sensor measurement is defined by Equation 2:

$$botwin = avg - bottomsub - std_factor * std \quad (2)$$

Where topwin is the top of window/ upper window boundary, botwin: bottom window/ lower window boundary, avg: average

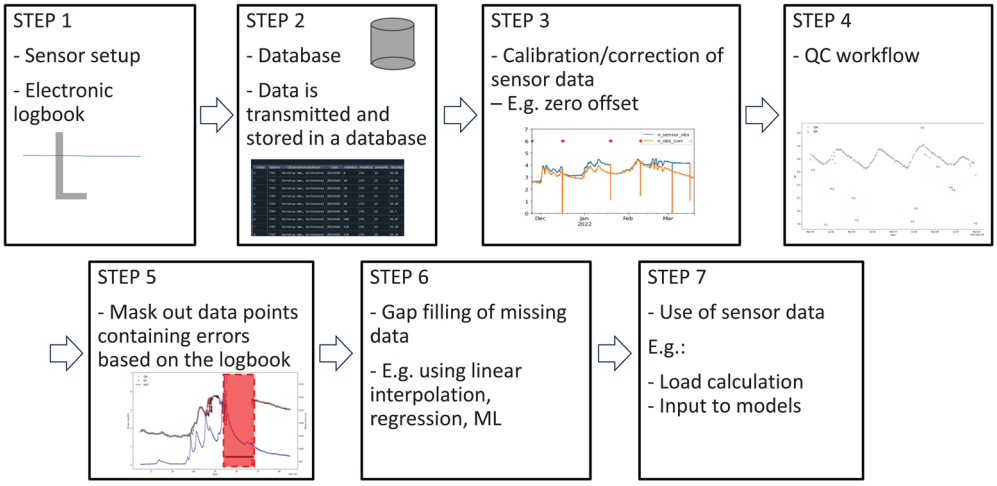

Figure 2. Conceptual model of a hypothetical workflow for data preprocessing high-frequency (HF) sensor data. Step 1 includes an illustration of a sensor installed in a waterbody. Step 2 represents establishing a database containing all HF sensor data. Step 3 represents the calibration of the sensor data with an example showing correction for zero offset of a nitrate sensor. The y-axis is the nitrate-N, and the x-axis is the datetime. The blue line represents the raw data, and the orange line is the corrected sensor data. Step 4 represents and illustrates the quality control (QC) workflow. An example from SentemQC is shown in a Figure where the y-axis is the pH and the x-axis is the datetime. The gray dots are the approved data points and the red dots are the datapoints marked as anomalies. Step 5 represents the use of the electronic logbook to mask out anomalies (red box). Step 6 represents the gap filling of missing data, and step 7 represents the use of the sensor data to for example load calculations.

Table 3. Information about the five different moving windows in the SentemQC method.

Moving windows in SentemQC	Number of datapoints	Calculations	Purpose of window
1	960	Average	To capture clusters of anomalies
2	960	Average	To capture clusters of anomalies
3	48	Median	To include some robustness in the method
4	5	Average	To capture single anomalies, both local and global anomalies
5	5	Average	To capture single anomalies, both local and global anomalies

of the nearest measurements, std: standard deviation of the nearest measurements, std_factor: a constant representation of the factor by which the standard deviation is multiplied (sensor dependent), topadd: a constant that expands the top height of the first moving window (sensor dependent), bottomsub: a constant that expands the bottom height of the first rolling window (sensor dependent).

In total, the method uses five windows, and for the fifth moving window, these calculations are weighted with a triangular window, which means that the data points in the middle of the window are given greater weight than those at the beginning and end. Window three does not calculate moving averages but moving medians to lend some robustness to the method. The specific purpose and characteristics of the five different windows can be found in Table 3. The five windows are chosen to ensure that all types of anomalies are flagged, both global and local individual anomalies, and clusters of anomalies as defined in the Introduction section. In addition, warm-up and cool-down periods are defined in SentemQC to include 480 records for both warm-up and cool-down, however, this number of records can be adjusted by the user. This is implemented in the method to ensure enough data points for average and mean calculations in each moving window. SentemQC will warn if there are insufficient data points for the warm-up and cool-down periods.

Information about the five different moving windows in the SentemQC method. The moving windows are calculating either moving averages or medians based on the input data. Each

window serves a unique purpose in identifying clusters of anomalies or single anomalies, both on a global and local scale.

The uncertainty of the sensor measurements related to the sensor technology can be defined in the configuration when conducting the parameterization of the QC method as ± in the first two moving windows. The uncertainty can be included as a percentage called "uncertainty_pct" in the parameterization or as a constant called "uncertainty_con", depending on how the uncertainty is described by the sensor company, see extended data for the final parameterizations (vant Veen *et al.*, 2024). For example, for the NITRATAX sensor from HACH, the uncertainty is 3% of the measured value or ±0.5 mg N/L (HACH, 2012; HACH, 2014). The different uncertainties used in this study are given in extended data 1(vant Veen *et al.*, 2024) and Table 2.

The method identifies anomalies in the data by comparing the average concentration within a moving window with the established uncertainty range. If a data point falls outside the upper or lower limits of this uncertainty window (represented by ±), it gets flagged as an anomaly. We assume that the uncertainty is found by comparing discrete water grab samples with sensor measurements under controlled conditions in the laboratory and is equally distributed around the measurement point. More specifically, when there is an uncertainty factor associated with the sensor parameterization, SentemQC evaluates whether the sensor measurement plus or minus the uncertainty falls outside the calculated moving window. If the measurement remains within the bounds defined by the uncertainty range, it is not flagged as an anomaly. This is because the SentemQC model is not permitted to outperform the defined uncertainty threshold. A conceptual figure illustrating the QC method is seen in Figure 3.

Operation

SentemQC requires a recent version of Python installed. An environment file that includes the Python versions and packages to run SentemQC is available for download at: https://zenodo.org/doi/10.5281/zenodo.13830624

All files including SentemQC software and the test data for this paper are free to download from Zenodo at: https://zenodo.org/doi/10.5281/zenodo.13830624.

SentemQC operation flow consists in i) Setting up the SentemQC model, and ii) Running the SentemQC model. In total, SentemQC includes six steps illustrated in Figure 4. Steps 4 and 5 are described in more detail in the following section. SentemQC is set up and controlled by a Python file called Main.py, in which the data period, sensor parameter, and input data are defined. This file loads the QC module and parameterization file. It is important and necessary to have electronic logbooks from the sensor station for the calibration and validation/evaluation of the QC method. In the logbook, each field visit, manual cleaning, calibration, and physical distribution should be noted. Therefore, collaboration between researchers and technicians is important. The QC method produces three outputs: 1) Data Frame. This table contains detailed information about data points identified as anomalies. It includes columns for: The raw data points and whether the data point was flagged as an anomaly or not. If flagged, the specific window that identified it as an anomaly is recorded. Additionally, it shows all valid data points. 2) Time Series Plot. This visual representation displays the entire dataset as a time series. Anomalies are marked out to provide a quick overview of their occurrence. 3) An output file. This file illustrates the total number of flagged anomalies and valid data points in each window. Please download and read the file: SentemQC – Description.pdf for a more detailed description of the operation and folder structure of SentemQC available at https://zenodo.org/doi/10.5281/zenodo.13830624.

Parameterization of the SentemQC model

The SentemQC method passes through the sensor data five times in total, as different moving windows cover different amounts of data. The windows are different, however, two are repetitions of the previous window in relation to the number of datapoints included in the window (Table 3). For example, window 1 has the same number of datapoints (width) as window 2. The windows are moving and begin directly after each other, which means that the windows do not interact with each other and are not connected. An example of the parameterization of the oxygen sensor at the Lemming field site from the library file (*RollWindowStdConfig.json*) and explanations of the components in the parameterization of different windows can be seen in the extended data (vant Veen *et al.*, 2024).

Parameterization of the SentemQC model's windows (step 4 in Figure 4) was done individually for the different sensor types installed at the field sites by fine-tuning the model to the specific sensor and field site. The optimization of the parameterization was performed by changing the constants in Equation 1 and Equation 2, which changes the heights of the moving windows (the parameters: topadd, bottomsub, and std_factor). Equation 1 and Equation 2 are the equations used for windows 1, 2, 4, and 5 for the sensor parameterization (window 3 calculates the median instead of the average). For each sensor measurement, the standard deviation and average were calculated using the temporally nearest neighboring measurements, depending on the windows number of measurements (for example, 960 for windows 1 and 2), centered on the measurement. As described above, the average, median, and standard deviation were calculated for the moving windows using the Pandas library in Python (The pandas development team, 2024).

For example, in Lyby-Grønning Stream, the nitrate concentration was relatively high compared to the concentration in Horndrup Stream, and some nitrate peaks occurred with a short duration on a daily basis, probably due to point sources. Therefore, it was important that the QC model did not mark out these high peaks as anomalies but still marked out real anomalies. For the parameterization of SentemQC, a shorter calibration period was used for each individual dataset compared to the entire dataset. This approach helps identify the

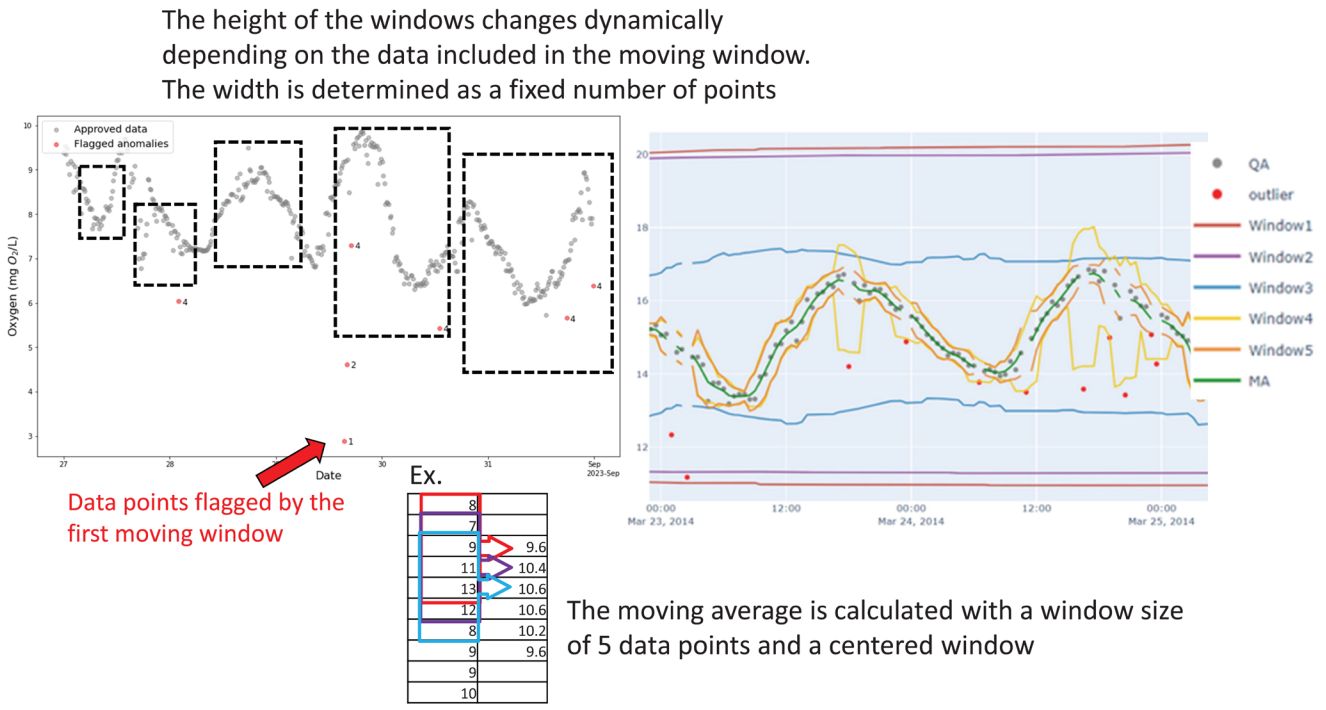

Figure 3. Conceptual model showing the principle behind the SentemQC method. The y-axis is the oxygen while the x-axis is the datatime in both figures. The scatted black boxes represent the different moving windows in the SentemQC program. The gray dots are the approved datapoints, while the red dots are the anomalies flagged by SentemQC. The calculation of the moving average in each window is simplified by the red, blue, and purple boxes.

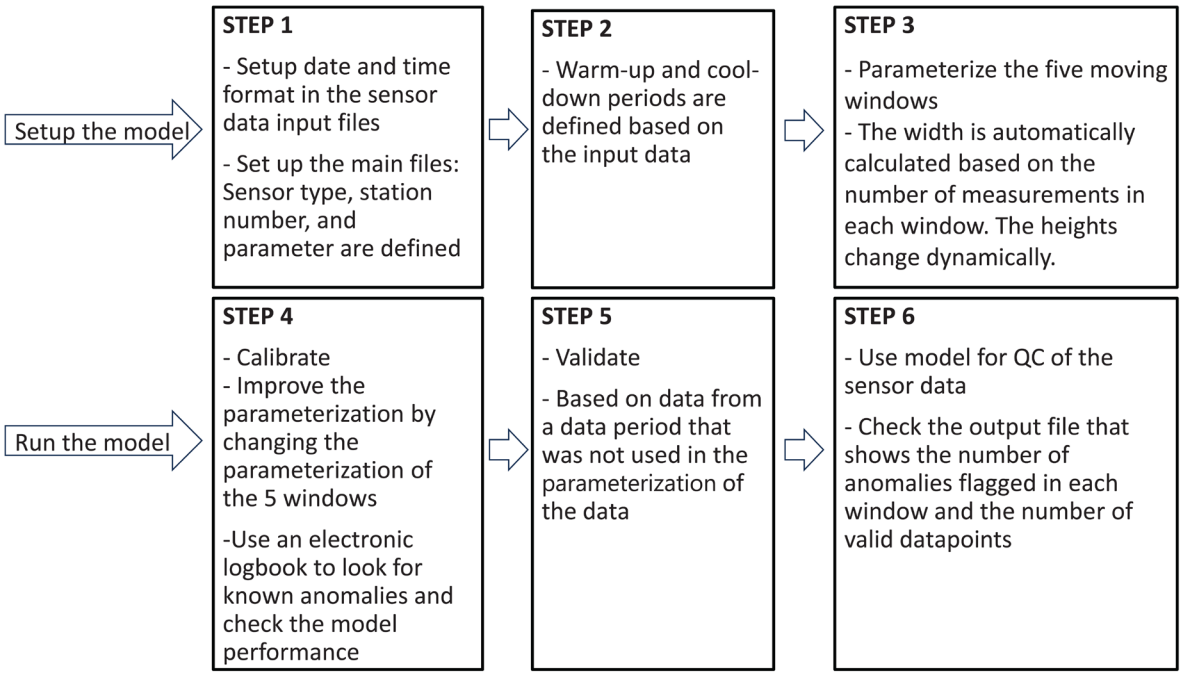

Figure 4. The different steps in the SentemQC quality control (QC) method. The boxes represent the six steps and workflow for data preprocessing in SentemQC.

most optimal parameterization for the different windows across individual sensors and field sites. Following the parameterization of SentemQC for the specific datasets, the model's performance was validated (step 5 in Figure 4). This validation process involves reviewing datapoints excluded from the parameterization of SentemQC to ensure the model accurately flags anomalies during periods with known issues using the electronic logbook. Consequently, the validation heavily relies on the electronic logbook.

The size of the parameterization/calibration and validation datasets depends on the available data points from the specific dataset. For Lemming, the year 2014 was chosen as the calibration period because the mesocosms were exposed to some extreme temperature experiments in 2014. Consequently, the parameterization was calibrated to handle extreme conditions.

See the final parameterization and the calibration periods for the individual sensors in extended data 1 (vant Veen *et al.*, 2024). For more information about the parameterization read the file: SentemQC – Description.pdf available here: https://zenodo.org/doi/10.5281/zenodo.13830624.

Results

The results of using the SentemQC method show the robustness and adaptability of various sensor data types. Adaptability is achieved through model parameterization fitted to the specific characteristics of each sensor data set. Several snapshot examples illustrating the sensor data cleaning process enabled by the QC method are presented in the following sections.

Lake Warming Mesocosm Experiment (LWME) in Lemming

The QC model demonstrates its robustness by its effectiveness in cleaning the oxygen sensor measurements from LWME (Figure 5). The program identifies and flags various anomalies using different "moving-window rolls" (Figure 5). It was able to flag different anomaly types like: i) individual isolated anomalies, which were the most frequent and occurred in most months, ii) sensor error, which was observed in October 2015, indicating a specific issue with the sensor itself, iii) clumps of anomalies, which occurred in March, April, September, and December 2015, suggesting that external factors potentially affected the measurements, iv) calibration uncertainty, which was found in the April and July measurements, possibly due to inaccurate calibration of the sensor, and v) unstable sensor anomalies, which appeared around November 2015, indicating sensor malfunction.

Remarkably, the oxygen sensor data revealed more anomalies near the lower limit compared to the upper limit, which is also flagged by the program. This highlights its ability to handle anomalies and distortions, which contributes to its overall robustness. By using the automatic QC program instead

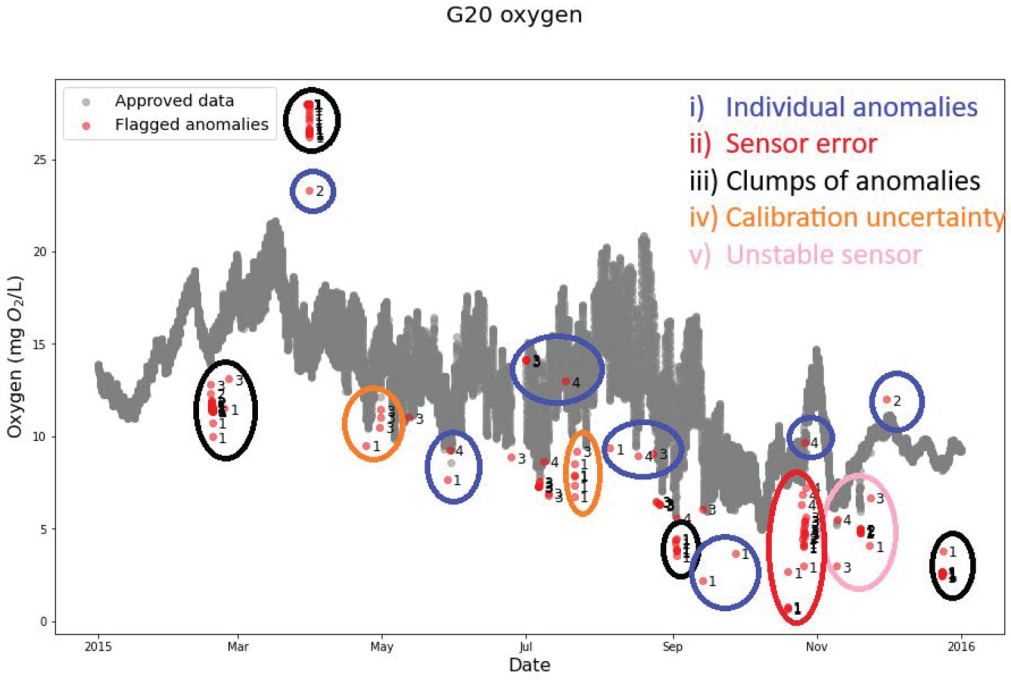

Figure 5. Example of quality control (QC) of oxygen sensor measurements in the lake warming mesocosm experiment using SentemQC. The example illustrates quality control (QC) of Oxygen sensor measurements in the lake warming mesocosm experiment (LWME) in Lemming 2015 using SentemQC and shows that the program flagged different types of anomalies such as: i) individual anomalies (marked with blue circles), ii) sensor error (red), iii) clumps of anomalies (black), iv) calibration uncertainty anomalies (orange), and v) unstable sensor anomalies (pink). The gray dots (Approved data) are the valid data points. The red dots (Flagged anomalies) are the flagged anomalies by the SentemQC program, and the numbers close to the red dots indicate the moving window flagging the anomaly.

of manual cleaning, 10 to 15% more valid measurements were obtained, demonstrating the efficiency and value of the program. Additionally, the calculation time for the QC algorithm for 1.3 million data points was 40 seconds (LWME, Lemming 2015 all mesocosms and parameters).

The QC model also demonstrates its robustness by effectively cleaning the pH sensor measurements from Lemming (Figure 6). Thus, the program identified and flagged various anomalies using different "moving windows" dominated by windows 1 and 3 in the period from March 18 to April 12, 2020 (Figure 6). As for the oxygen sensor data, the pH sensor data also contained more anomalies near the lower limit compared to the upper limit, which were flagged by the program. This also highlights its ability to handle anomaly distortions, which contributes to the QC programs' overall robustness (Figure 6).

However, when zooming in, some individual data points in the data series appeared as potential anomalies although they were not flagged as anomalies by the QC program. This intentional design avoids overly restricting the QC algorithm, which could lead to mistakenly labeling valid data points as anomalies. Therefore, we accepted that anomalies falling inside the upper or lower limits of the uncertainty with values close to the overall trend might not be identified.

Lyby-Grønning stream and Horndrup stream

The QC model cleaning the nitrate sensor measurement from Lyby Stream is parameterized to include nitrate concentration peaks as valid even if the peak is brief (Figure 7). We know from discrete samples in the stream that the peaks around September 12, 2021, are valid. This evidences the model's robustness because it is possible to make adjustments to include such peaks and not mark them as anomalies. At the same time, the QC model could mark the real anomalies that were not real changes in the nitrate concentration measurements in the stream. The high and low concentrations on 7 September 2021 reflect that the sensor was visited and manually cleaned with acid and checked for zero offsets that affect this nitrate sensor before and after manual sensor cleaning (Figure 7). The SentemQC model perfectly marked out these measurements as anomalies (Figure 7).

The QC model marked out single anomalies during the manual cleaning of the nitrate sensor measurement from Horndrup Stream on February 15, March 3, 15, and 30, and April 20 (Figure 8). From February 21 to 23, 2022, the sensor was blocked by branches, leaves, and sand from a heavy rainfall event that transported a lot of sediment to the stream. The QC program marked out these not valid measurements as anomalies. Also, the four single high anomalies around the March 8, 2024, were perfectly marked out as anomalies. In the same period there was no change in the water level (Figure 8).

The use of the QC model cleaning turbidity sensor measurement also showed that the model marked out and captured anomalies and not-valid measuring periods (Figure 9). The model marked out anomalies such as single noise but also not-valid

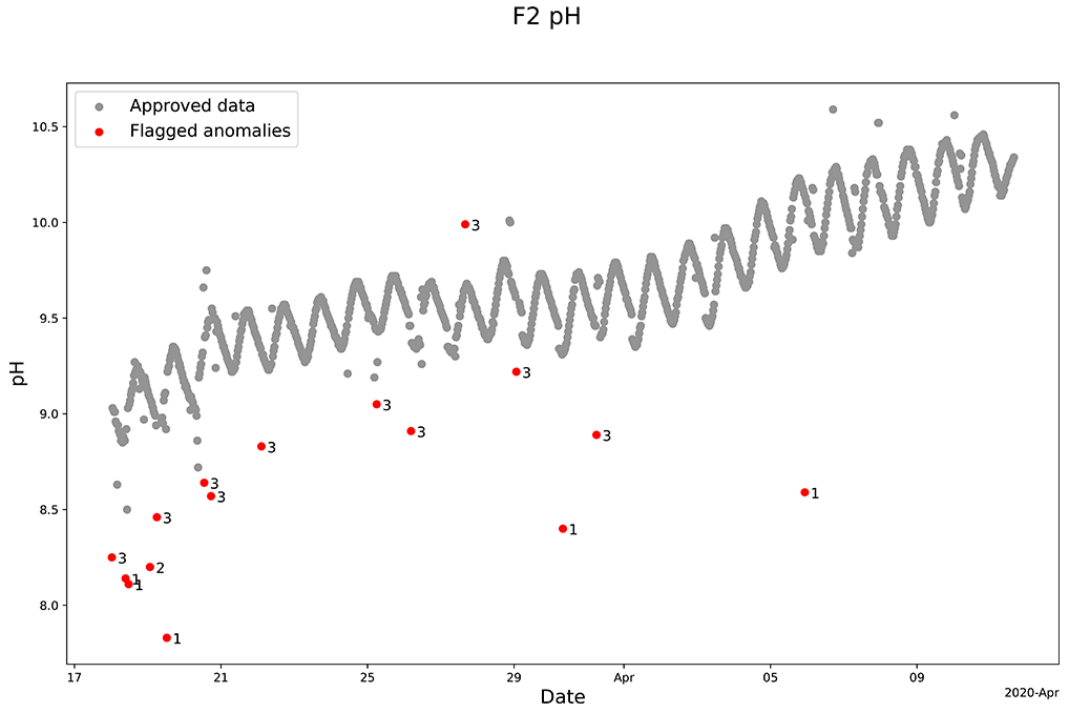

Figure 6. Example of quality control (QC) of pH sensor measurements in the lake warming mesocosm experiment using SentemQC. The example illustrates quality control (QC) of pH sensor measurements in the lake warming mesocosm experiment (LWME) in Lemming 2020 using SentemQC. The gray dots (Approved data) are the valid data points. The red dots (Flagged anomalies) are the flagged anomalies, and the number indicates the moving window flagged the anomaly.

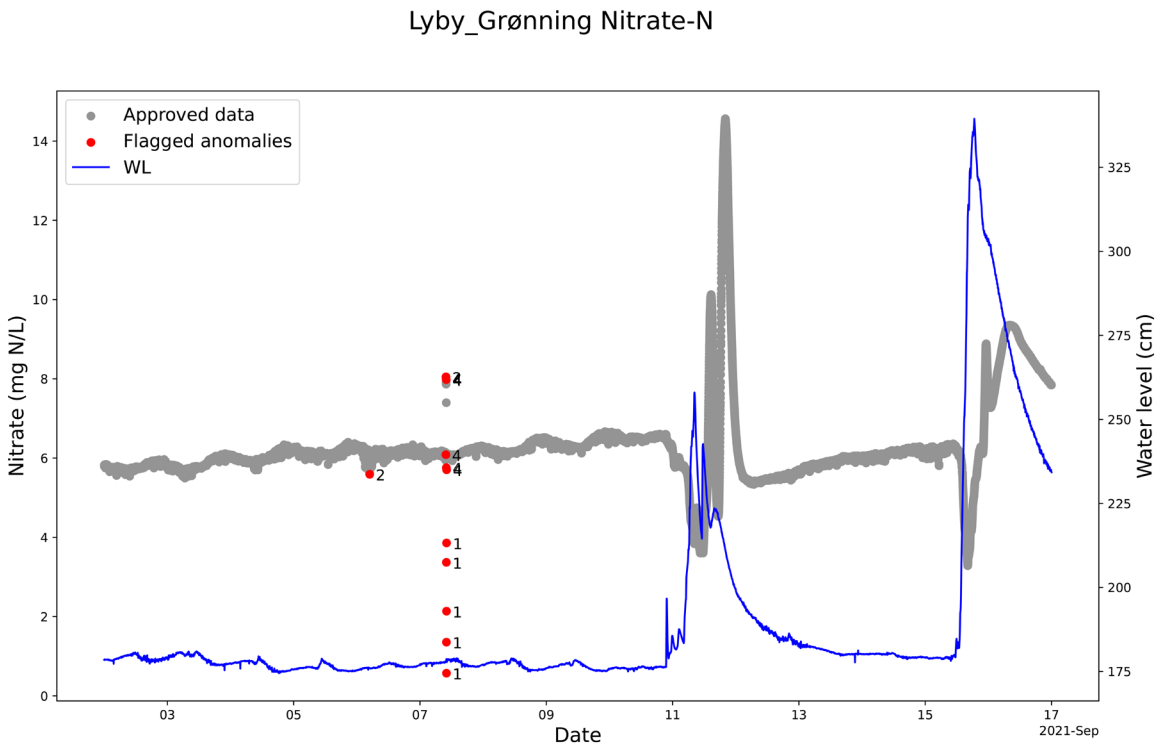

Figure 7. Example of quality control (QC) of nitrate sensor measurements in Lyby-Grønning Stream using SentemQC. The example illustrates quality control (QC) of nitrate sensor measurements in Lyby-Grønning Stream in September 2021 using SentemQC. The gray dots (Approved data) are the valid data points. The red dots (Flagged anomalies) are the flagged anomalies, and the number indicates the moving window flagging the anomaly. The blue line represents the water level in the stream.

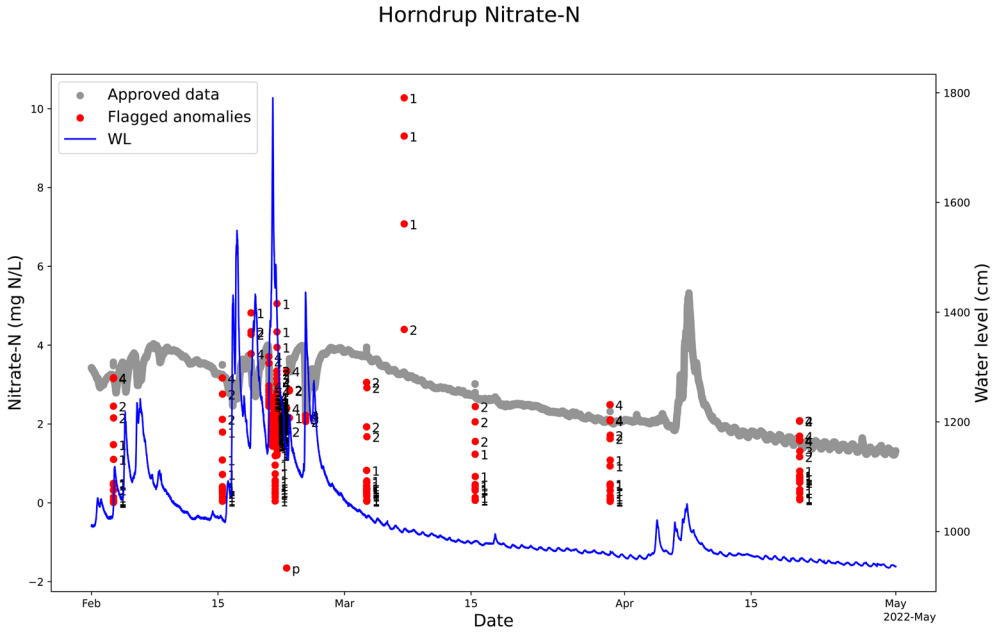

Figure 8. Example of quality control (QC) of nitrate sensor measurements in Horndrup Stream using SentemQC. The example illustrates quality control (QC) of nitrate sensor measurements in Horndrup Stream from February to May 2022 using SentemQC. The gray dots (Approved data) are the valid data points. The red dots (Flagged anomalies) are the flagged anomalies, and the number indicates the moving window that flagged the anomaly. The blue line represents the water level in the stream.

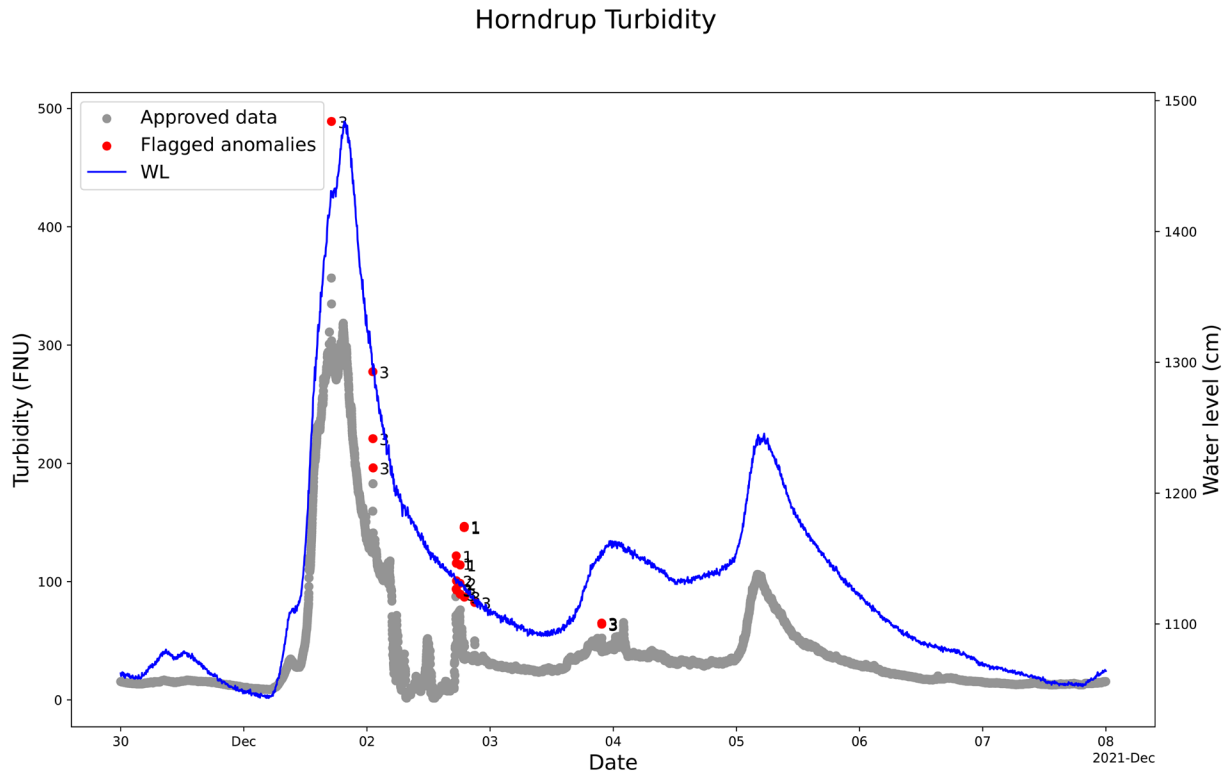

Figure 9. Example of quality control (QC) of turbidity sensor measurements in Horndrup Stream using SentemQC. The example illustrates quality control (QC) of turbidity sensor measurements in Horndrup Stream during a heavy rainfall event at the beginning of December 2021 using SentemQC. The red dots (Flagged anomalies) are the flagged anomalies, and the number indicates the moving window that flagged the anomaly. The gray dots (Approved data) are the valid data points. The blue line represents the water level in the stream.

measurements during a field visit and during manual cleaning of the sensor for the nitrate sensor measurements in Horndrup and Lyby.

The model parameterization was set to include peaks in the turbidity measurement during, for example, heavy rainfall events where the turbidity is increasing. An example is December 2, 2021, where the model did not mark out the peak but only the single very high concentrations as anomalies (Figure 9). Window 3 marked out anomalies during this heavy rainfall event, showing that the model can easily be adjusted for cleaning turbidity sensor measurements (Figure 9).

Lake Ormstrup

The SentemQC model effectively cleaned the oxygen sensor data from Lake Ormstrup (Figure 10). It identified anomalies between August 21 and September 1, 2023 (Figure 10), showing the ability of the model to handle data distortions and improve the data quality. For this oxygen sensor, the SentemQC model also flagged more anomalies near the lower limit than the upper limit (Figure 10). Some data points might look like anomalies, for example around September 25. However, these are likely caused by wind-induced partial mixing with lower oxygen water, which means that the data points are real.

Total amount of anomalies in the different sensor data

Table 4 summarizes the number of anomalies flagged by the different moving window settings during QC in SentemQC. The table also shows QC results excluding uncertainty to show the difference and importance of including this parameter (Table 4). Sensor data from the stream stations yielded a low anomaly rate of 0.1%, 0.1%, and 0.2% including uncertainty, and 0.2%, 0.3%, and 0.5% without uncertainty (Table 4). This was lower than the anomaly rates observed for sensors in Lake Ormstrup and LWME, which were 0.5%, 0.6%, and 0.8% with uncertainty and 1.6%, 1.4%, and 1.2% without uncertainty, respectively (Table 4). The differences in anomaly percentages between the different sensor types may be because the sensors installed in the lake and ponds measure oxygen and pH that are considerably more variable parameters than nitrate and turbidity (Table 4). As expected, SentemQC flagged more anomalies when uncertainty was not included in the program (Table 4). Additionally, SentemQC also flagged data points as anomalies due to sensor errors that may occur, for example, when sensors are undergoing calibration or maintenance in the laboratory (vant Veen *et al.*, 2024). Although SentemQC captured and flagged these points as anomalies, this does not reflect the software's performance in identifying anomalies, and sensor errors were therefore excluded in Table 4.

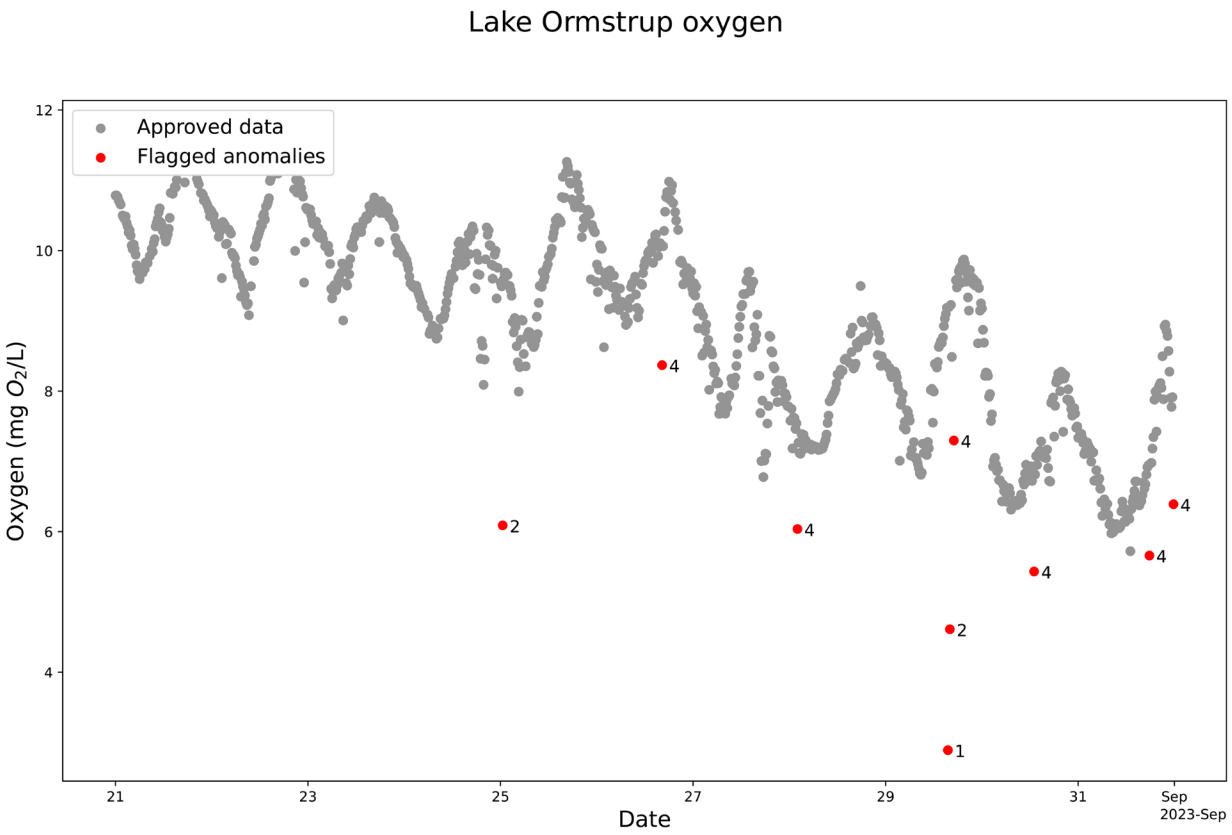

Figure 10. Example of quality control (QC) of oxygen sensor measurements in Lake Ormstrup using SentemQC. The example illustrates SentemQC quality control (QC) of oxygen sensor measurements from Lake Ormstrup from August 21 to September 1, 2023. The red dots (Flagged anomalies) are the flagged anomalies, and the number indicates the moving window that flagged the anomaly. The gray dots (Approved data) are the valid data points.

Table 4. Summary of the number of sensor data points flagged as anomalies using SentemQC. This table illustrates a summary of the number of sensor data points flagged as anomalies in the different windows using the quality control (QC) method SentemQC with and without uncertainty included in the method. The number of datapoints flagged and the percentages of anomalies flagged are illustrated.

Sensor/ parameter	Station	Number of data points flagged as anomalies		Percent flagged as anomalies	
		With uncertainty	Without uncertainty	With uncertainty	Without uncertainty
Nitrate-N (mgN/L)	Horndrup Stream	1399	4818	0.2	0.5
Turbidity (FNU)	Horndrup Stream	1298	1944	0.1	0.2
Nitrate-N (mgN/L)	Lyby-Grønning Stream	1025	2663	0.1	0.3
Oxygen (mg/L)	Lake Ormstrup	56	123	0.6	1.4
Oxygen (mg/L)	Lemming LWME	449	643	0.8	1.2
pH	Lemming LWME	84	264	0.5	1.6

There was no general pattern in the number or percentage of flagged anomalies across the different months for any of the sensors investigated, see extended data (vant Veen *et al.*, 2024). The nitrate and turbidity sensors almost consistently showed an anomaly rate below 0.4% throughout the year. The sensor measurements therefore seem to be robust and not highly affected

by seasonal disruptions during the year. However, the nitrate sensor in Horndrup Stream showed a higher anomaly rate during the winter months (December – February), perhaps due to a heavy bed sediment load, which might have buried the sensor and thus affected its measurements. Throughout the year, sensor height was therefore adjusted in Horndrup Stream based on the amount of rain in the catchment area to ensure that it remained submerged. Unfortunately, in February 2022, the sensor was buried by sediment, and in December 2022 it was positioned too high. In December, the water levels dropped, and the sensor ended up measuring air instead of water, which led to a significant anomaly spike for that specific month (vant Veen *et al.*, 2024). Additionally, both the number of anomalies flagged for the pH sensor in Lemming in February and March 2020 at 1.1% and 2.3% and for the oxygen sensor in Lemming in February and March 2015 at 2.9% and 3.3% were relatively high compared to the other months (vant Veen *et al.*, 2024).

Discussion

The main strengths of SentemQC are its simplicity, robustness, and cost-efficiency concerning time resources, low computer power, and the fact that it is freely available for download and use (Andersen *et al.*, 2024). In SentemQC, all data points are connected because the algorithm compares the data points with nearby data points ("neighboring points") using narrow windows and then compares them with points further away using wider moving windows. By examining both nearby and distant data points, this approach ensures that both individual data points (global and local anomalies) and potential clusters of anomalies are effectively flagged (Figure 5). SentemQC is working even if the dataset contains periods with gaps, however, it does not account for missing data. Consequently, the program calculates the moving average or median in the defined moving windows, including the subsequent datapoints following a gap. This can result in the first datapoint after a period of missing data being flagged as an anomaly if the concentration, for example, has changed significantly. However, since the program only flags the data, it is possible to remove such flagging afterward. Additionally, SentemQC can provide a QC of 1.3 million sensor data points in 40 seconds. In general, methods such as the rate of change or recursive methods are often used for anomaly detection in dataset (Barnett *et al.*, 2012; Géron, 2019). However, the rate of change method does not work well for our HF sensor data because the anomalies in the sensor data are sometimes grouped in clusters as described in the Results section. Therefore, this method only identifies and marks out the first and last data points of each cluster as anomalies, thereby missing the remaining anomalies within the group. Using recursive methods was also ineffective for the QC of our sensor data, they were simply too slow and very computer RAM demanding. Although we have not tested SentemQC on datasets where it fails to perform, exploring its limitations in highly complex environments, such as those with frequent biological disturbances or abrupt events, for example, gas bubble events for methane (Davidson *et al.*, 2024), could be an interesting direction for future work. Importantly, SentemQC does not discard data but flags it, making it a valuable tool for identifying periods of change. This is especially interesting if supporting parameters are included in the method.

The B3 QA-QC software from GLEON (Barnett *et al.*, 2012; Read *et al.*, 2016) allows the user to investigate the sensor data semi-automatedly and correct the data periods in the QA-QC tools provided by the program (Barnett *et al.*, 2012; Read *et al.*, 2016). This can be time-consuming as the user has to manually tap and go through the different QA-QC tools and choose which test to use. Additionally, the user must choose between different QC tools and may select a moving average. However, the moving average is only calculated once compared to our five windows, which means that the data points in B3 are only compared to the neighboring data points a single time. This can be inefficient for identifying clusters of anomalies.

Leigh *et al.*'s (2019) ten step-framework also allows users to select the most appropriate anomaly detection methods for their specific sensor data. The first steps involve a collaborative visual examination of the water quality sensor data with end users and identification of the most important anomalies for the specific sensor data. However, this approach can be very resource intensive. In the framework, the user can choose between regression-based and feature-based methods. It requires some level of expertise to choose the right methods, which highlights the importance of clear communication between end-users and anomaly detection developers (Leigh *et al.*, 2019). Additionally, regression-based methods are semi-supervised and require training, just like ML methods.

The SaQC method allows the user to choose between 32 different QC functions. As described in the Introduction, some of these functions include ML algorithms (Schmidt *et al.*, 2023). These methods also require high-level expertise in the specific sensors and the QC methods. Moreover, there may be differences in how the same type of sensor data, such as turbidity, is cleaned because it is based on the user's individual choice, and this can result in non-harmonized QC approaches. This can render a comparison between the different turbidity data difficult (Skarbøvik *et al.*, 2023a). Furthermore, the SaQC method's first part involves preprocessing of the data input as the data needs to be preprocessed before performing the actual QC. This process includes aligning timestamps as well as smoothing and transforming the data (Schmidt *et al.*, 2023). This is not needed in SentemQC as it can perform QC of all, also shifting time stamps. The input data only needs to contain the correct date-time format. This is an advantage as information can get lost when shifting from 1-minute measuring frequency to 15 minutes.

Additionally, as described in the Introduction, the use of ML approaches for QC often requires substantial computer power and RAM, and it can be difficult to understand the underlying mathematical principles even when spending considerable time on understanding the programming (Belay *et al.*, 2023). Furthermore, it also takes time to train the ML model compared to the simple SentemQC. Often the responsibility

for conducting QC of sensor data quality falls on specialists with in-depth knowledge of the field (domain experts) but less experience with IT technology (Schmidt *et al.*, 2023). Therefore, the transparency and simplicity of our method are a great strength, making it ideal for specialists, even with limited IT knowledge. Furthermore, the user is not required to choose between different methods and can use SentemQC for QC of all types of aquatic sensor data.

Moreover, as pointed out by Belay *et al.* (2023), ML approaches are all very specific for the type of sensor for which they are trained and require input data without noise. This means that it can be difficult to use ML approaches for different types of sensors. Belay *et al.* (2023) consequently concluded that ML methods are promising but that challenges remain as to the current state-of-the-art multivariate time series anomaly detection techniques. They concluded that the challenges are primarily due to the lack of a universal approach, as each method is specifically designed for a particular application (Belay *et al.*, 2023).

The studies of Leigh *et al.* (2019) and Schmidt *et al.* (2023) both provide a general workflow for conducting QC of HF sensor data (Figure 4). Our simple automatic SentemQC could potentially be integrated as an additional QC test within the SaQC program or in Leigh *et al.'s* (2019) framework. As described in the Methods section, the SentemQC program only requires filling in one step to obtain good data quality (step 4, Figure 2). After the QC step (step 4, Figure 2), steps such as implementation of the electronic logbook are needed to mark out periods where the sensor measurement should not be included in the final data time series.

Another step in processing HF sensor data prior to use is gap-filling (step 6, Figure 2), e.g. calibration of the sensor data based on gap samples for correction of zero offsets (step 3, Figure 2), which SentemQC cannot presently perform. For the nitrate sensors in the streams used in this study, a zero-offset of the sensor data needs to be performed based on measurements from before and after manual cleaning of the sensor in the field. These zero offsets are most probably caused by algae, biofilm, and dirt that settle on the sensor lenses when installed in the streams (van't Veen *et al.*, 2020), and for these sensors, this offset is linear but depends on both stream-specific and season-specific conditions. Schmidt *et al.* (2023) corrected for zero offset of a dissolved organic carbon (DOC) sensor due to biofilm growth and dirt accumulation at the sensor lenses using the SaQC program. The offset exhibited an exponential development, and they used regression to correct for the drift. Therefore, we assume that offsets are site-specific for each sensor and stream.

In future work, both the electronic logbook and methods for gap-filling and calibration against grab samples can be implemented in SentemQC. Incorporating these important steps is a promising direction for the future development of SentemQC. From a technical perspective, integrating electronic logbooks could be included as a next step after the five moving windows have masked out anomalies. It would require users to adopt a standardized format for the logbook, particularly for date-time conventions, to ensure compatibility with SentemQC. This step would enable the tool to automatically identify and mask periods as anomalies marked as errors in the logbooks. For gap-filling, several methods could be considered as a final step in SentemQC. Rozemeijer *et al.* (2025) describe different options for gap-filling in water quality sensor time series. A simple approach, such as the linear interpolation method could be implemented as an option. Alternatively, more advanced techniques could be offered, including machine learning-based methods such as Random Forest, which seems like a promising method (Barcala *et al.*, 2023). Further investigation is needed to determine the most suitable approach for integration into SentemQC. The different methods for gap-filling could be implemented as optional modules in SentemQC, allowing users to select the most appropriate strategy for their specific dataset. However, this would affect the complexity and universality of the SentemQC approach between different sensor data types.

The use of supporting sensor data parameters could be included in our QC method, e.g. water level, discharge, temperature, or conductivity. For example, Schmidt *et al.* (2023) used water level and water temperature as supporting parameters when performing QC of the spectral absorption coefficient at 254 nm (SAC254), which served as a proxy for measuring the dissolved organic carbon (DOC) concentration using the SaQC program. However, this required more transformation to normalize the input data. Leigh *et al.* (2019) also tested the use of supporting parameters for QC of environmental data, applying feature-based models. They also tested dynamic regression models (RegARIMA models) to forecast the concentration of one parameter based on multiple water quality variables. Yet, both methods can be difficult to follow and require transformation of the input data, and both regression methods and feature-based methods require great computer power for classification (Leigh *et al.*, 2019). Using water level to QC of turbidity datasets could be useful. During high rainfall events, this should not lead to excessive flagging in the event of a rapid increase in turbidity. Similarly, for nitrate, programming the system to be less sensitive during rainfall events is possible, but comes with a delay, as seen for Horndrup and Lyby-Grønning streams. However, determining the appropriate delay is challenging due to site-specific and seasonal variations, rendering, as described earlier, the method more complicated. Another obvious co-comparison is oxygen and pH, which often follow each other, in lakes and mesocosms. Using supporting parameters would make SentemQC stronger but requires more computational resources. Furthermore, it would make the program less simple and transparent to follow when performing the QC of the sensor data, depending on the type of sensor data and supporting parameters. In addition, Lannergård, Fölster and Futter (2021) found that turbidity sensor measurements were not only correlated with an increase in discharge (Q). Thus, for low-flow events (mean Q < 2 m3/s), the turbidity of the water showed short, repeating increases that were largely independent of changes in the discharge and high turbidity concentrations occurred during periods with decreasing discharge after a high-discharge

event. This implies that users should be careful when choosing supporting parameters, for example when performing QC of turbidity sensor data.

As mentioned in the Introduction, there is a need for harmonized and automated QA-QC methods for use on aquatic HF sensor data to allow comparison of data between countries, and because the use of manual methods for QC of large amounts of data is time-consuming (Bieroza *et al.*, 2023; Skarbøvik *et al.*, 2023a). SentemQC has been tested and used for QC of different sensor data (nitrate, pH, oxygen, turbidity) collected from three different freshwater systems such as streams, lakes, and LWMEs. The method is thus robust, and our study shows that the method can be a sufficient solution for standardizing QA and QC of aquatic sensor data as it is simple, cost-efficient, and open source.

However, SentemQC could also be used on other types of HF sensor data technologies such as lab-on-a-chip, nanosensors, DNS-based biosensors, or molecular biosensors that all measure with different measuring frequencies and offer different advantages and limitations (Bieroza *et al.*, 2023), and on sensors applied in wastewater treatment plants where sensor drift is a known problem that may generate high economic costs and energy usage (Hansen *et al.*, 2022). Former studies have shown that QC of sensor data from wastewater treatment plants can be difficult and sometimes even not possible due to insufficient data and electronic logbooks with missing logs and low data resolution (Hansen *et al.*, 2022). As pointed out by Scarisbrick-Hauser and Rouse (2007) and Hansen *et al.* (2022), insufficient data quality is a problem in multiple industrial cases. Companies invest in data collection to gain valuable insights, but insufficient data quality often creates a disconnection between intention and reality. We believe that SentemQC, with its simplicity and robustness, can be of great relevance for gaining more reliable data.

In our study, the use of SentemQC on the sensor data collected at the different sensor stations identified anomalies, flagging from 0.1% to 1.6% of the data as described in the Results section (Table 4). While the sensors used here contained relatively few anomalies (<2%), they may represent a best-case scenario in terms of use and maintenance. Sensors used for monitoring and/industrial applications may not be maintained as intensively as under research applications, and this may increase the potential for anomalies and the need for QC. By applying SentemQC, we ensured that only reliable and accurate data was considered. For comparison, Lannergård, Fölster and Futter (2021) identified anomalies and removed 0.4% of their sensor turbidity measurements in streams, recorded with a frequency of 10 or 15 minutes during six years (2012–2017). They used a manual approach to identify the anomalies by calculating different statistical elements such as standard deviation, mean, maximum, and minimum. The sensor observations were log-transformed, and all turbidity measurements falling outside three standard deviations of the daily mean were identified. However, they also followed several steps during the QC of their sensor data, including calibration using discrete manual samples (Lannergård *et al.*, 2021).

In SentemQC, we include the uncertainty related to the sensor itself. This uncertainty is given by the manufacturer or company, for example, HACH. As described in the Methods section, uncertainty is supposed to be equally distributed around the measurement point. The uncertainty given by the companies is usually stated as plus/minus a percentage to be understood as the relative standard deviation of a large number of analyses where the correct concentration of the chemical substance is known. Normally, analysis results are assumed to have a normal distribution around the true concentration; therefore, it can be discussed whether the uncertainty interval should be extended to include plus/minus three standard deviations. If we assume that the sensor measurements exhibit a normal distribution around the true concentration, it covers 96.6% of all outcomes compared to only 68.2% with plus/minus one standard deviation.

The difference between including and excluding uncertainty when performing SentemQC is given in Table 4, and, as expected, SentemQC flagged more anomalies when excluding the uncertainty (Table 4). For November 2021, the nitrate measurement in Lyby-Grønning Stream is displayed in Figure 11 as an example. More specifically, the SentemQC flagged 58 anomalies including uncertainty compared to 106 without uncertainty out of 43,200 data points in November 2021 (0.1% vs 0.2%). While the program can identify anomalies, flagging them within the uncertainty range presents challenges, as this requires absolute certainty within the uncertain area. However, this is not the case, and the SentemQC program cannot be better than the uncertainty. Therefore, it is important to include the uncertainty. Additionally, none of the currently explored QC methods consider the uncertainty of sensor measurements provided by the sensor companies when performing QC on sensor data.

Conclusion and perspectives

SentemQC is a cost-efficient and simple program designed to identify and mark out anomalies and detect errors or error patterns in HF sensor data, and this is needed as sensors are neither free of errors nor maintenance-free. SentemQC has, by testing and application, proven effective for QC of sensor data (such as nitrate, pH, oxygen, and turbidity) from various freshwater ecosystems, including streams, lakes, and lake warming mesocosm experiments (LWME). This broad application shows the robustness of SentemQC.

Given its simplicity, cost efficiency, robustness, and open-source availability, SentemQC is a promising solution for standardizing QC of different aquatic sensor data. The SentemQC concept that connects all the data points in the input data by examining both nearby and distant data points ensures that both individual anomalies and potential clusters or groups of anomalies are effectively flagged.

There is a high demand for standardized and automated QA-QC procedures for HF aquatic sensor data to ensure cross-national comparability, and here, SentemQC can be useful. Moreover, SentemQC saves workforce compared with the time-intensive manual QC methods, which are impractical for handling enormous amounts of sensor data.

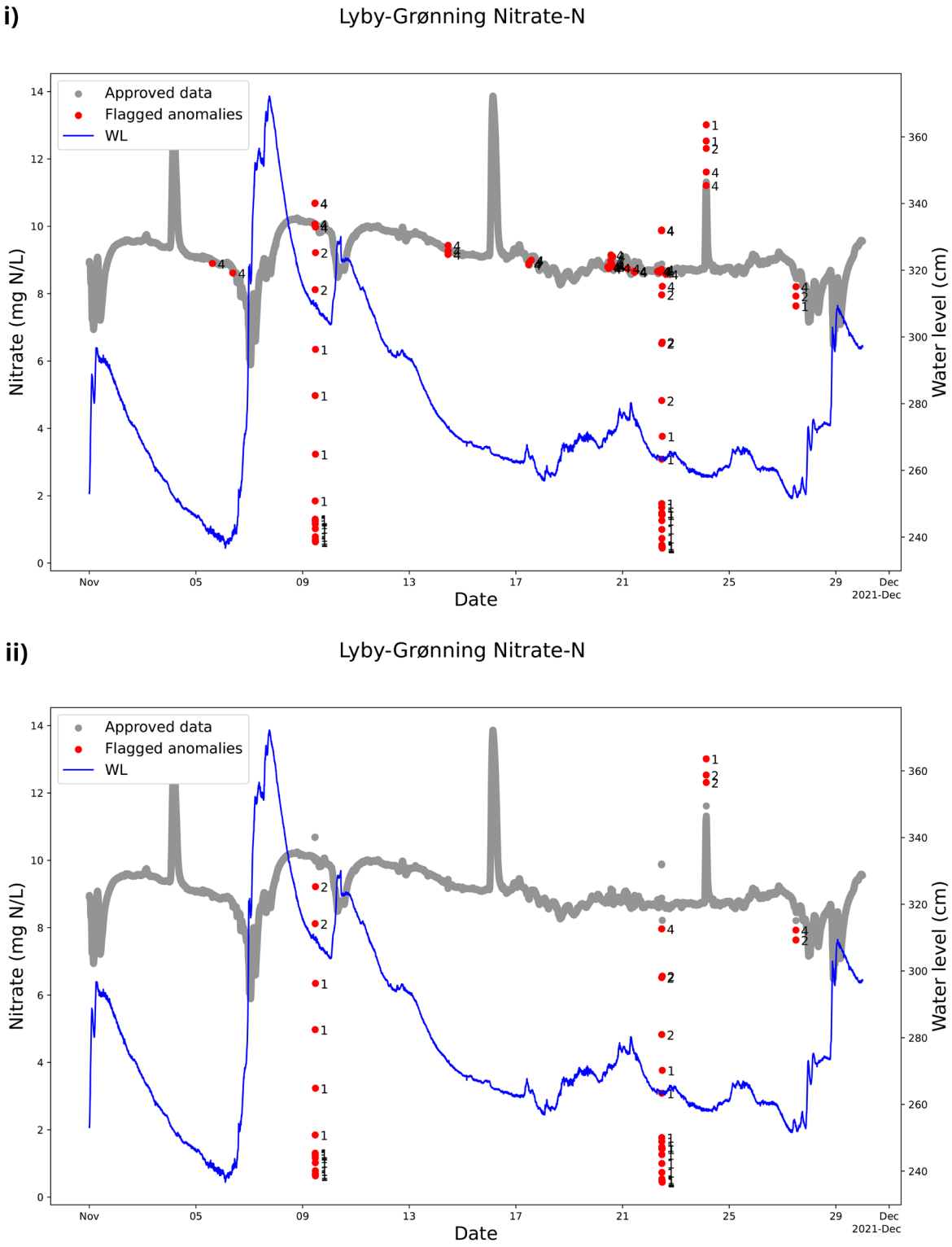

Figure 11. Example of quality control (QC) of data from Lyby-Grønning Stream using SentemQC with and without uncertainty included. These two examples illustrate SentemQC quality control (QC) of nitrate sensor measurements from Lyby-Grønning Stream in November 2021 with and without uncertainty included in the method. I) shows the QC of the data without including uncertainty, and Ii) shows the QC of the data including uncertainty. The gray dots (Approved data) are the valid data points. The red dots (Flagged anomalies) are the flagged anomalies, and the number indicates the moving window that flagged the anomaly. The blue line represents the water level in the stream.

The future development of SentemQC includes integrating the electronic logbook and implementing a calibration procedure for sensor data. Additionally, establishing a mathematical correlation between the dimensions of the moving windows (height and width) for various sensor types, together with supporting parameters would improve SentemQC. Furthermore, future developments of SentemQC might include an ML approach, a method to include classification of anomaly types within the method, and a gap-filling method also using ML. This would enable SentemQC to provide the QC in real time and to do forecasting, which is needed in connection with future warning systems and sensor maintenance. However, this requires considerable effort, complicates the method, and increases the computational demand, as earlier discussed. Currently, we do not have sufficient data to optimize SentemQC by including an ML approach and to retain the universality of the method between different sensor data types. Additionally, a more detailed analysis of how sensor measurement frequency influences data characteristics and the performance of SentemQC across different sensor types is a promising direction for future work.

Ethics & consent

Ethical approval and consent were not required.

Data availability

SentemQC source code, documents, and test datasets examples for this paper are available free of charge from:

Zenodo: SentemQC, https://zenodo.org/doi/10.5281/zenodo.13830624

This project contains following data:

Python code and test dataset for SentemQC,

Data is available under the license Creative Commons Attribution 4.0 International license.

Extended data

Extended data are available free of charge from

Zenodo: Extended data for the paper: "SentemQC - A novel and cost-efficient method for quality assurance and quality control of high-resolution frequency sensor data in fresh waters", https://doi.org/10.5281/zenodo.13884940

Data is available under the license under Creative Commons Attribution 4.0 International license.

Software availability

Source code available free of charge from: Zenodo: SentemQC: Python code and test dataset for SentemQC,

URL: SentemQC (zenodo.org) DOI: https://zenodo.org/doi/10.5281/zenodo.13830624

Archived software available from:

URL: SentemQC (zenodo.org) DOI: https://zenodo.org/doi/10.5281/zenodo.13830624

License: SentemQC is under Creative Commons Attribution 4.0 International license.

Acknowledgments

We would like to thank the technicians Uffe Mensberg, Emma Damgaard Lempert, Henrik Stenholt, and Carsten Brian Nielsen for their effort and help on collecting data in the field. Thanks to Dorte Nedergaard, Lene Tofteberg, Karina Bomholt Oest, and Marlene Skjærbæk analyzing data in the lab. Thanks to Pia Grewy and Carsten Brian Nielsen for helping to handle the high-frequency data in the database and providing online access to the nitrate sensor measurements. Thanks to Rasmus Rumph Frederiksen for quality assurance and optimization of the SentemQC Python code. Thanks to Anne Mette Poulsen for the appreciated proofreading of the paper.

The Generative Artificial Intelligence (GAI) tool "Microsoft Copilot" which is based on GPT-4 architecture has been used in this article to improve the language of the article. This includes alternative wordings and proofreading.

References

Andersen PM, van't Veen SGM, Rumph Frederiksen R, *et al.*: **SentemQC: python code and test dataset for SentemQC.** *Zenodo.* 2024.
Publisher Full Text

Barcala V, Rozemeijer J, Ouwerkerk K, *et al.*: **Value and limitations of machine learning in high-frequency nutrient data for gap-filling, forecasting, and transport process interpretation.** *Environ Monit Assess.* 2023; **195**(7): 892.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Barnett L, *et al.*: **B3 data QAQC software user's guide.** University of Waikato, New Zealand, LERNZ, 2012.

Belay MA, Blakseth SS, Rasheed A, *et al.*: **Unsupervised anomaly detection for IoT-Based multivariate time series: existing solutions, performance analysis and future directions.** *Sensors (Basel).* 2023; **23**(5): 2844.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Bieroza M, Acharya S, Benisch J, *et al.*: **Advances in catchment science, hydrochemistry, and aquatic ecology enabled by high-frequency water quality measurements.** *Environ Sci Technol.* 2023; **57**(12): 4701–4719.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Blaen PJ, Khamis K, Lloyd CEM, *et al.*: **Automated sensing methods for dissolved organic matter and inorganic nutrient monitoring in freshwater systems.** In: Krause, S., Hannah, D., Grimm, N. (Eds.), *Ecohydrological Interfaces.* Wiley, 2024; 213–233.
Publisher Full Text

Cassidy R, Jordan P: **Limitations of instantaneous water quality sampling in surface-water catchments: comparison with near-continuous phosphorus time-series data.** *J Hydrol.* 2011; **405**(1–2): 182–193.
Publisher Full Text

Chabbi A, Loescher HW: **Terrestrial ecosystem research infrastructures: challenges and opportunities.** Boca Raton, FL: CRC Press, 2017.
Publisher Full Text

Davidson TA, Søndergaard M, Audet J, *et al.*: **Temporary stratification promotes large greenhouse gas emissions in a shallow eutrophic lake.** *Biogeosciences.* 2024; **21**(1): 93–107.
Publisher Full Text

Dekker AG, Hestir EL: **Evaluating the feasibility of systematic inland water quality monitoring with satellite remote sensing.** CSIRO: Water for a healthy country national research flagship, 2012; 105. (Accessed: 14 December 2021).
Publisher Full Text

Edmonds JW, King KBS, Neely MB, *et al.*: **Using large, open datasets to understand spatial and temporal patterns in lotic ecosystems: NEON case studies.** *Ecosphere.* 2022; **13**(5): e4102.
Publisher Full Text

Fölster J, Lannergård E, Valley S, *et al.*: **Sensorer för vattenkvalitet i miljöövervakning av vattendrag - Hur användbara är de i praktiken?** Vatten och miljö: Rapport 2019: 10. Institutionen för vatten och miljö, SLU, 2019; 35.
Reference Source

Géron A: **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems.** Second edition. Beijing [China]; Sebastopol, CA: OReilly Media, Inc, 2019.
Reference Source

Goldstein M, Uchida S: **A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data.** Edited by D. Zhu, *PLoS One.* 2016; **11**(4): e0152173.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Gong A, Campbell J: **Great expectations.** *Zenodo.* 2022.
Publisher Full Text

HACH: **SOLITAX sc Betjeningsvejledning DOC023.58.03232.** Ed 3b; 12/2022 (2006), 2006; 34. (Accessed: 5 May 2024).
Reference Source

HACH: **Onlinemåling af nitrat - Nitratsonder med UV- og ISE-teknologi, DOC033.58.30142**. 2012.
Reference Source

HACH: **NITRATAX sc Betjeningsvejledninger, udgave 6A, DOC023.58.03211**. 2014.
Reference Source

HACH: **SOLITAX sc, User Instructions: SOLITAX sc, Addendum documentation - Operation with the SC1500 Controller, DOC273.98.90593**. 2017.

HACH: **Tender Text Process turbidity probe with predictive diagnostics, DOC353.52.35047_Solitax**. 2018.
Reference Source

Hansen BD, Hansen TB, Moeslund TB, *et al.*: **Data-driven drift detection in real process tanks: bridging the gap between academia and practice.** *Water.* 2022; **14**(6): 926.
Publisher Full Text

Horsburgh JS, Reeder SL, Jones AS, *et al.*: **Open source software for visualization and quality control of continuous hydrologic and water quality sensor data.** *Environ Model Softw.* 2015; **70**: 32–44.
Publisher Full Text

In-Situ: **Aqua TROLL multiparameter sondes**. 2021; (Accessed: 23 May 2024).
Reference Source

Jones AS, Jones TL, Horsburgh JS: **Toward automating post processing of aquatic sensor data.** *Environ Model Softw.* 2022; **151**: 105364.
Publisher Full Text

Jordan P, Melland AR, Mellander PE, *et al.*: **The seasonality of phosphorus transfers from land to water: implications for trophic impacts and policy evaluation.** *Sci Total Environ.* 2012; **434**: 101–109.
PubMed Abstract | **Publisher Full Text**

Kaste Ø, Gundersen CB, Sample J, *et al.*: **The Norwegian river monitoring programme 2021 – water quality status and trend.** NIVA Report 7760. Norwegian Institute for Water Research and the Norwegian Environment Agency - NIVA, 2022; 45.
Reference Source

Lannergård EE, Fölster J, Futter MN: **Turbidity-discharge hysteresis in a meso-scale catchment: the importance of intermediate scale events.** *Hydrol Process.* 2021; **35**(12): e14435.
Publisher Full Text

Lannergård EE, Ledesma JLJ, Fölster J, *et al.*: **An evaluation of high frequency turbidity as a proxy for riverine total phosphorus concentrations.** *Sci Total Environ.* 2019; **651**(Pt 1): 103–113.
PubMed Abstract | **Publisher Full Text**

Leigh C, Alsibai O, Hyndman RJ, *et al.*: **A framework for automated anomaly detection in high frequency water-quality data from *in situ* sensors.** *Sci Total Environ.* 2019; **664**: 885–898.
PubMed Abstract | **Publisher Full Text**

Levin G, Iosub CI, Jepsen MR: **Basemap02. Technical documentation of a model for elaboration of a land-use and land-cover map for Denmark.** Technical Report from DCE – Danish Centre for Environment and Energy 95. Aarhus University, DCE – Danish Centre for Environment and Energy, 2017; 64.
Reference Source

Levi EE, Jeppesen E, Nejstgaard JC, *et al.*: **High resolution chlorophyll-a *in-situ* fluorescence sensors versus *in-vitro* chlorophyll-a measurements in mesocosms with contrasting nutrient and temperature treatments [version 1; peer review: 2 approved].** *Open Res Eur.* 2024; **4**: 69.
Publisher Full Text

Liboriussen L, Landkildehus F, Meerhoff M, *et al.*: **Global warming: design of a flow-through shallow lake mesocosm climate experiment: shallow lakes and global warming.** *Limnol Oceanogr Methods.* 2005; **3**(1): 1–9.
Publisher Full Text

OxyGuard International A/S: **OxyGuard model 420 - dissolved oxygen probes with built-in transmitter**. 1994; (Accessed: 23 May 2024).
Reference Source

OxyGuard International A/S: **OxyGuard pH manta and redox manta. Single-channel measurement transmitters.** 2020.
Reference Source

Petersen RJ, Blicher-Mathiesen G, Rolighed J, *et al.*: **Three decades of regulation of agricultural nitrogen losses: experiences from the Danish agricultural monitoring program.** *Sci Total Environ.* 2021; **787**: 147619.
PubMed Abstract | **Publisher Full Text**

Read JS, Gries C, Read EK, *et al.*: **Generating community-built tools for data sharing and analysis in environmental networks.** *Inland Waters.* 2016; **6**(4): 637–644.
Publisher Full Text

Riemann B, Carstensen J, Dahl K, *et al.*: **Recovery of Danish coastal ecosystems after reductions in nutrient loading: a holistic ecosystem approach.** *Estuaries Coasts.* 2016; **39**(1): 82–97.
Publisher Full Text

Rode M, Wade AJ, Cohen MJ, *et al.*: **Sensors in the stream: the high-frequency wave of the present.** *Environ Sci Technol.* 2016; **50**(19): 10297–10307.
PubMed Abstract | **Publisher Full Text**

Rogora M, Cancellario T, Caroni R, *et al.*: **High-frequency monitoring through *in-situ* fluorometric sensors: a supporting tool to long-term ecological research on lakes.** *Front Environ Sci.* 2023; **10**: 1058515.
Publisher Full Text

Rozemeijer J, Jordan P, Hooijboer A, *et al.*: **Best practice in high-frequency water quality monitoring for improved management and assessment; a novel decision workflow.** *Environ Monit Assess.* 2025; **197**(4): 353.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Scarisbrick-Hauser A, Rouse C: **The whole truth and nothing but the truth? The role of data quality today.** *Direct Market: Int J.* 2007; **1**(3): 161–171.
Publisher Full Text

Schmidt JQ, Kerkez B: **Machine learning-assisted, process-based quality control for detecting compromised environmental sensors.** *Environ Sci Technol.* 2023; **57**(46): 18058–18066.
PubMed Abstract | **Publisher Full Text**

Schmidt L, Schäfer D, Geller J, *et al.*: **System for automated Quality Control (SaQC) to enable traceable and reproducible data streams in environmental science.** *Environ Model Softw.* 2023; **169**: 105809.
Publisher Full Text

Sheldon WM: **Dynamic, rule-based quality control framework for real-time sensor data.** In: *Proceedings of the Environmental Information Management Conference.* Citeseer, Department of Marine Sciences, University of Georgia, Athens, Georgia, USA, 2008; 145–150.
Reference Source

Skarbøvik E, Isidorova A, *et al.*: **Monitoring water quality with sensors - based on experiences from streams in Northern Europe.** Nordbalt Ecosafe, 2023b.
Reference Source

Skarbøvik E, vant Veen SGM, Lannergård EE, *et al.*: **Comparing *in situ* turbidity sensor measurements as a proxy for suspended sediments in North-Western European streams.** *CATENA.* 2023a; **225**: 107006.
Publisher Full Text

Søndergaard M, Nielsen A, Skov C, *et al.*: **Temporarily and frequently occurring summer stratification and its effects on nutrient dynamics, greenhouse gas emission and fish habitat use: case study from Lake Ormstrup (Denmark).** *Hydrobiologia.* 2023; **850**(1): 65–79.
Publisher Full Text

Talagala PD, Hyndman RJ, Smith-Miles K: **Anomaly detection in high-dimensional data.** *J Comput Graph Stat.* 2021; **30**(2): 360–374.
Publisher Full Text

The European Commission: **Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy**. 2000; (Accessed: 24 October 2019).
Reference Source

The European Commission: **REPORT FROM THE COMMISSION TO THE COUNCIL AND THE EUROPEAN PARLIAMENT on the implementation of Council Directive 91/676/EEC concerning the protection of waters against pollution caused by nitrates from agricultural sources based on Member State reports for the period 2016–2019. COM(2021) 1000 final.** 2021.
Reference Source

The pandas development team: **pandas-dev/pandas: pandas.** 2024.
Publisher Full Text

Tiberti R, Caroni R, Cannata M, *et al.*: **Automated high frequency monitoring**

of Lake Maggiore through *in situ* sensors: system design, field test and data quality control. *J Limnol.* 2021; **80**(2).
Publisher Full Text

van Geer FC, Kronvang B, Broers HP: **High-resolution monitoring of nutrients in groundwater and surface waters: process understanding, quantification of loads and concentrations, and management applications.** *Hydrol Earth Syst Sci.* 2016; **20**(9): 3619–3629.
Publisher Full Text

vant Veen SGM, Kronvang B, Andersen PM, *et al.*: **Extended data for the paper: "SentemQC - a novel and cost-efficient method for quality assurance and quality control of high-resolution frequency sensor data in fresh waters".** *Zenodo.* 2024.
http://www.doi.org/10.5281/ZENODO.13884939

van't Veen SGM, Holm H, Kronvang B: **Undersøgelse af anvendelse af sensorer i overvågningen- og test af en nitratsensor i vandløb.** 1. Aarhus, Denmark: Aarhus Universitet, DCE – Nationalt Center for Miljø og Energi, 2020; 79.
Reference Source

Weathers KC, Hanson PC, Arzberger P, *et al.*: **The Global Lake Ecological Observatory Network (GLEON): the evolution of grassroots network science.** *Limnol Oceanogr Bull.* 2013; **22**(3): 71–73.
Publisher Full Text

Wilkinson MD, Dumontier M, Aalbersberg IJJ, *et al.*: **The FAIR guiding principles for scientific data management and stewardship.** *Sci Data.* 2016; **3**: 160018.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Windolf J, Blicher-Mathiesen G, Carstensen J, *et al.*: **Changes in nitrogen loads to estuaries following implementation of governmental action plans in Denmark: a paired catchment and estuary approach for analysing regional responses.** *Environ Sci Policy.* 2012; **24**: 24–33.
Publisher Full Text

Wollheim WM, Bernal S, Burns DA, *et al.*: **River network saturation concept: factors influencing the balance of biogeochemical supply and demand of river networks.** *Biogeochemistry.* 2018; **141**(3): 503–521.
Publisher Full Text

Yang X, Zhang X, Graeber D, *et al.*: **Large-stream nitrate retention patterns shift during droughts: seasonal to sub-daily insights from high-frequency data-model fusion.** *Water Res.* 2023; **243**: 120347.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ✔

Version 2

Reviewer Report 30 June 2025

https://doi.org/10.21956/openreseurope.22256.r54867

© **2025 Hooijboer A et al.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Arno Hooijboer**
National Institute for Public Health and the Environment, RIVM, Bilthoven, The Netherlands
Elma Tenner
Environmental quality, National Institute for Public Health and the Environment, Bilthoven, Utrecht, The Netherlands

The authors made significant improvements in the text and incorporated all our comments on the first version. The article is now much clearer on the working and parametrization of the method.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Monitoring water quality on farmlands, high frequency water quality monitoring

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 20 June 2025

https://doi.org/10.21956/openreseurope.22256.r54868

© **2025 Guo H.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Hongwei Guo**
Nankai University, Tianjin, China

The authors have carefully revised the manuscript and provided satisfactory responses to my comments raised in the previous round of review. The revisions have substantially improved the

quality and clarity of the manuscript. I recommend this manuscript for indexing.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Water quality of remote sensing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 12 December 2024

https://doi.org/10.21956/openreseurope.19603.r46097

© 2024 Hooijboer A et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Arno Hooijboer
[1] National Institute for Public Health and the Environment, RIVM, Bilthoven, The Netherlands
[2] National Institute for Public Health and the Environment, RIVM, Bilthoven, The Netherlands
Elma Tenner
[1] Environmental quality, National Institute for Public Health and the Environment, Bilthoven, Utrecht, The Netherlands
[2] Environmental quality, National Institute for Public Health and the Environment, Bilthoven, Utrecht, The Netherlands

The paper describes an automate system in a Python script for QA/QC on high frequent water quality data. The paper is very interesting and easy to read. We endorse the need to do a proper QA and QC to HF data from water quality sensors. It is well-described that for large amount of HF-data quality control should be automated by standardized methods. There are many methods that uses different techniques for QA/QC. It is useful to have an overview of different methods and the authors convince the reader to keep it simple for different reasons.

Implementation of the software to the supplied data was easy (within less than one hour).

We have some suggestions for improvement:

In general: we got confused about the difference between QA and QC. In our opinion, the data is cleaned afterwards, so it is mainly Quality Control of data (the software is therefore called SentemQC). A definition of both terms and an explanation on which part of the QA and QC the system is useful would make it clear.

In the text
Page 6 Figure 1: In our print the figure has a low resolution. It would me neater to make it high-res.

page 7: “measuring frequency of 30 minutes” should be measuring interval (or frequency of once per 30 minutes).

page 8: “the upper limit of the mw for a given data point is the sum” etc. We would appreciate if you provide the formulas, which is clearer than explanation in text.

page 9 Figure 4. “Step 4 improve the parametrization by etc”. The power of automated QC is that is performed along standard rules and not randomly. By changing the parametrization doesn't this introduce a human eye which make the QC arbitrary?
Step 5: The validation is not in the text, shouldn't that be so? what is the size of validation dataset compared to the whole dataset?

Page 10: “The QA/QC model passes through the data five times in total”... etc. Since this is the core of the paper and the model we would appreciate it if the author would explain it more in detail.
- why is 5 windows chosen and not 4 or 6?
- what does each window do?
- how do the windows interact with each other?

“A smaller parametrization/calibration period” etc: do you mean that the window-sizes are smaller? Please explain how to optimize the parametrization.

“the program identifies and flags various anomalies” etc. it would make it more sound if the different examples that are given in the text (i, ii etc) also appear in the figure with arrows or circles.

Figure 5: what does QA and stand ER for? In the figure is space enough to explain it more in the legend of the figure (for example, approved data (grey) and flagged anomalies (red).

page 12: “the high and low concentrations on 7 September” etc: it is a good sign that the method flags the data during maintenance but in our opinion those data should be disposed form the dataset using the logbook, before doing a QC cleaning of data.

In the last sentence before paragraph “Lake Ormstrup”, it refers to Figure 10, which should be Figure 9 in our opinion.

Discussion: "In SentemQC, all data points are connected because the algorithm compares the data points with nearby data points". Please explain what happens when data gaps appear, is the method also working with missing data?

Downloaded files
We downloaded the software and the datafiles, however we couldn’t find the YML-file and the VST-level data.

We recommend to put the Python script file also on GitHub

In the .pdf with information of the software more explanation should be provided on the parameterization. In the pdf only the parameters are explained of the first window (RollWindowStdConfig.json) we suggest to explain all parameters (and how to define the parameterization). For us it wasn't clear what settings to use on our own sensor data.

In the text the authors write that std_factor1 is set to 2.5 but in the example it is 3.

We look forward to use this approach on our own HF water quality data!

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Partly

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Monitoring water quality on farmlands, high frequency water quality monitoring

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 28 May 2025

Sofie G. M. van't Veen

Responses to Reviewer Report 3 - Arno Hooijboer and Elma Tenner The paper describes an automate system in a Python script for QA/QC on high frequent water quality data. The paper is very interesting and easy to read. We endorse the need to do a proper QA and QC to HF data from water quality sensors. It is well-described that for large amount of HF-data quality control should be automated by standardized methods. There are many methods that uses different techniques for QA/QC. It is useful to have an overview of different methods and the authors convince the reader to keep it simple for different reasons.

Implementation of the software to the supplied data was easy (within less than one hour). We have some suggestions for improvement: *Response: We sincerely thank the reviewers for their thoughtful and encouraging feedback on our manuscript. We appreciate the time and effort invested in reviewing our work, and we are pleased to hear that the paper was found to be both interesting and accessible. We are especially glad to hear that the software has been successfully implemented, which shows the practical value of our approach. In response to the reviewers' comments and suggestions, we have revised the manuscript accordingly. The updated version (Version 2) incorporates improvements based on the recommendations provided in all three review reports. We hope these changes further enhance the clarity and impact of the paper. The Responses specific to review 3 are seen below. The reviewer's original comments are presented in* **bold** *below, followed by our detailed responses in italic.* ***Thank you for your review, it is much appreciated.***

In general: we got confused about the difference between QA and QC. In our opinion, the data is cleaned afterwards, so it is mainly Quality Control of data (the software is therefore called SentemQC). A definition of both terms and an explanation on which part of the QA and QC the system is useful would make it clear.
Response: Thank you for your comment. We are glad you find the paper relevant and well-described, and we are very happy to hear you successfully implemented the software.
In response to your general question, we have provided definitions for both terms and an explanation of QA-QC, highlighting the aspects of QA and QC where SentemQC is applicable in the introduction section (see "Quality assurance (QA) procedures for environmental sensor data…"). We agree that SentemQC primarily focuses on the QC of sensor data. Consequently, we have revised parts of the text in the revised paper to clear this.

In the text
Page 6 Figure 1: In our print the figure has a low resolution. It would me neater to make it high-res.
Response: Thank you for pointing this out. Figure 1 has been revised and improved to a high-resolution version, including an insect showing Denmark's location in Europe and a new scalebar.

page 7: "measuring frequency of 30 minutes" should be measuring interval (or frequency of once per 30 minutes).
Response: Thank you for this comment. This has been corrected in the revised manuscript to "…measuring interval of 30 minutes".

page 8: "the upper limit of the mw for a given data point is the sum" etc. We would appreciate if you provide the formulas, which is clearer than explanation in text.
Response: Thank you for this comment. This is a very good idea. We have enhanced the 'Implementation' section by including formulas for SentemQC and the moving windows (equations 1 and 2). Additionally, we have expanded the sections "Implementation" and "Parameterization of the SentemQC model" with additional text and a table that provides detailed information about each of the five moving windows. This table includes the calculations, the purpose of the window, and the number of data points included in the window (width).

page 9 Figure 4. "Step 4 improve the parametrization by etc". The power of automated

QC is that is performed along standard rules and not randomly. By changing the parametrization doesn't this introduce a human eye which make the QC arbitrary?
Response: Thank you for this comment. We agree that introducing automatic parameterization in SentemQC would significantly enhance the program. This could be achieved by incorporating supporting parameters, such as water level in future versions of SentemQC, which we are presently working on. In the present version, parameterization is user-dependent, as we calibrate it based on the specific site's characteristics, the type of data measured, and information from the electronic logbook. We acknowledge that this process should be automated in future versions.
Step 5: The validation is not in the text, shouldn't that be so? what is the size of validation dataset compared to the whole dataset?
Response: Thank you for pointing this out. We have updated the "Parameterization" section to include the validation process see page 10, " Following the parameterization of SentemQC for the specific datasets, the model's performance is validated (step 5 in Figure 4). This validation process involves reviewing datapoints excluded from the calibration of the parameterization to ensure the model accurately flags anomalies during periods with known issues using the electronic logbook. Consequently, the validation heavily relies on the electronic logbook. ". For example, if the stream and data are subject to high nitrate peaks occurring over short periods, the moving windows should not mark these peaks as anomalies.

The size of the validation dataset depends on the available data points from the specific stream, sensor, and data type. We included all available data points for validation, excluding the calibration period. For example, for oxygen measurements in LWME, the calibration period was one year (2014) out of the three-year dataset (2014 to 2016). The electronic logbook is extremely important for the validation as we use it to ensure that SentemQC marks the anomalies we know for sure are anomalies. See the following explanation at page 10:
"The size of the validation dataset depends on the available data points from the specific dataset. For Lemming, the year 2014 was chosen as the calibration period because the mesocosms were exposed to some extreme temperature experiments in 2014. Consequently, the parameterization was calibrated to handle extreme conditions.".
Information about the specific periods used for calibration of the test data and the total period of data used for validation of SentemQC is found in the extended data for the paper in the tables in "Extended data 1_parameterizations for SentemQC paper.pdf". Available from: https://doi.org/10.5281/zenodo.13884940 **Page 10: "The QA/QC model passes through the data five times in total"... etc. Since this is the core of the paper and the model we would appreciate it if the author would explain it more in detail.**
- why is 5 windows chosen and not 4 or 6?
- what does each window do?
- how do the windows interact with each other?
Response: Thank you for these relevant questions. We improved the section "Parameterization of the SentemQC model" on page 10 with explanations of the questions above. This includes a new table explaining each of the five windows in detail as described above. This table includes the calculations for each window, the purpose of the window, and the number of data points (width). We added further information in the text on page 10: "The windows are moving and begin directly after each other, which means that the windows do not use each other and are not connected." SentemQC includes five moving windows because, during the development of the model, testing showed that five windows were sufficient to capture the different types of anomalies. However, in the future development of SentemQC, we may expand the model to include six or seven windows if

necessary for optimal performance.

"A smaller parametrization/calibration period" etc: do you mean that the window-sizes are smaller? Please explain how to optimize the parametrization. *Response: Thank you for this relevant question. No, we do not adjust the width of the moving windows during the parameterization. The sentence should clarify that a smaller part of the total dataset is used for calibration/parameterization, while the remainder is used for validation. For further clarification, we have revised the sentence on page 10: "For the parameterization of SentemQC, a shorter calibration period was used for each individual dataset compared to the entire dataset. This approach helps identify the most optimal parameterization for the different windows across individual sensors and field sites…. "*

To explain how to optimize the parametrization, we added additional information to the section "Parameterization of the SentemQC model" page 10.
"The optimization of the parameterization was performed by changing the constants in equations 1 and 2, which changes the heights of the moving windows (the parameters: topadd, bottomsub, and std_factor)…". In addition, more information about parameterization is also added to the description file of SentemQC, available together with the program at Zenodo: https://zenodo.org/doi/10.5281/zenodo.13830624.

"the program identifies and flags various anomalies" etc. it would make it more sound if the different examples that are given in the text (i, ii etc) also appear in the figure with arrows or circles.
Response: Thank you for this very good suggestion. We have improved Figure 5 to illustrate the different types of anomalies with circles in different colors.

Figure 5: what does QA and stand ER for? In the figure is space enough to explain it more in the legend of the figure (for example, approved data (grey) and flagged anomalies (red).
Response: Thank you for this comment. Much appreciated. The legends in Figures 3, 5, 6, 7, 8, 9, 10, and 11 are updated with more descriptive legends (Approved data and Flagged anomalies).

page 12: "the high and low concentrations on 7 September" etc: it is a good sign that the method flags the data during maintenance but in our opinion those data should be disposed form the dataset using the logbook, before doing a QC cleaning of data.
Response: Thank you for this relevant comment. It is possible to remove maintenance periods before QC is performed using SentemQC. However, we prefer to retain periods with maintenance and other periods where we know problems have been identified from the electronic logbook to achieve the most optimal parameterization of the model for the individual dataset, as well as validation, which heavily relies on the electronic logbook. This is presently a good method to ensure that anomalies during problematic periods are flagged as anomalies. In the future, it might be considered to remove such periods when we include supporting parameters to adjust the parameterization.

In the last sentence before paragraph "Lake Ormstrup", it refers to Figure 10, which should be Figure 9 in our opinion.
Response: Thank you for pointing this out. We agree, it should be Figure 9. This has been changed

in the revised version of the paper.

Discussion: "In SentemQC, all data points are connected because the algorithm compares the data points with nearby data points". Please explain what happens when data gaps appear, is the method also working with missing data?
Response: Thank you for your question. SentemQC is working even if the dataset contains gaps, but it does not account for missing data.
We updated the text in the discussion: "SentemQC is working even if the dataset contains periods with gaps, however, it does not account for missing data. Consequently, the program calculates the moving average or median in the defined moving windows, including the subsequent datapoints following a gap. This can result in the first datapoint after a period of missing data being flagged as an anomaly if the concentration for example has changed significantly. However, since the program only flags the data, it is possible to remove such flagging afterward".
We are considering how to deal with this in future versions of SentemQC.

Downloaded files
We downloaded the software and the datafiles, however we couldn't find the YML-file and the VST-level data.
Response: Thank you for this comment. The YML-file (SentemQC.yml) is uploaded in the versions 2 and 3 of SentemQC in Zenodo and can be downloaded from Zenodo here: ***DOI: 10.5281/zenodo.13830623****. The Water level (WL) datasets are not provided as they are not a part of the test data and are not used in SentemQC. It is only shown in the examples from the streams in the paper to illustrate the turbidity and nitrate measurements in relation to the water levels measured in the streams***.**

We recommend to put the Python script file also on GitHub
Response: Thank you for this comment. We agree. We are working on a new version of SentemQC that will be available on GitHub.

In the .pdf with information of the software more explanation should be provided on the parameterization. In the pdf only the parameters are explained of the first window (RollWindowStdConfig.json) we suggest to explain all parameters (and how to define the parameterization). For us it wasn't clear what settings to use on our own sensor data.
Response: Thank you for this comment. We have revised the SentemQC – Description_V12.pdf file to include more explanations about the parameterization and which parameters we recommend to fine-tune for optimization of the parameterization. The settings for the nitrate test data are available and can be used as a guide. SentemQC was originally developed based on the oxygen data which is not very stable and we found out that these settings when using SentemQC on the other test data available were suitable as guidance. However, it is important to know the dataset and set the parameterization for the specific sensor type and stream as described in the paper. In addition, the formula explaining the parameters used in the windows is now included in the paper in the section "QC-QC method".

In the text the authors write that std_factor1 is set to 2.5 but in the example it is 3.
Response: Thank you for this comment and thanks for pointing this out. It has been changed in the revised version of SentemQC – Description_V12.pdf file.

We look forward to use this approach on our own HF water quality data!
Response: Thank you very much. We highly appreciate that the reviewers find SentemQC useful and want to use it on their own HF water quality data! Thank you for your review of our paper. We think it improved the paper and this is much appreciated.

Competing Interests: No competing interests were disclosed.

Reviewer Report 04 December 2024

https://doi.org/10.21956/openreseurope.19603.r46103

© 2024 Guo H. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Hongwei Guo
[1] Nankai University, Tianjin, China
[2] Nankai University, Tianjin, China

This paper introduces SentemQC, a novel and efficient QA-QC method for water quality high-frequency sensor data, highlighting its simplicity, robustness, and time efficiency. The paper is well-structured and the method is applicable to a variety of water quality scenarios and parameters, showing strong versatility. Specific improvement suggestions:

1. It is recommended to provide a more detailed algorithm description in the methods section, particularly regarding the rules for multi-window comparison and the detailed logic for anomaly labeling.

2. Although the article mentions considering sensor measurement uncertainty, there is a lack of detailed discussion on the theoretical basis and practical application of this aspect. It is suggested to add relevant statistical background explanations.

3. In the results section, it would be helpful to provide a categorical analysis of different types of anomalies (e.g., sensor drift, random noise, clustered anomalies) and their specific manifestations.

4. SentemQC is described as suitable for various sensor types, but there is insufficient in-depth analysis of the relationship between sensor measurement frequency and data characteristics. It is recommended to supplement this analysis.

5. The discussion mentions potential future improvements, such as incorporating electronic logs and missing data imputation. It is advised to further specify the technical feasibility of these improvements.

6. While the simplicity of SentemQC is emphasized, the article lacks a demonstration of the actual

operation process. It would be beneficial to include this.

7. It is suggested to discuss potential scenarios or boundary conditions where SentemQC may not be applicable, such as challenges encountered in complex environments like high biological disturbance waters.

8. The discussion section suggests combining supporting parameters (e.g., water level, temperature), and it would be valuable to include a case study that validates the improvement in anomaly detection accuracy when incorporating these parameters.

9. Some references are concentrated, so it is recommended to expand on recent high-quality relevant literature to provide a more comprehensive background to the study.

10. It is suggested to include a sensitivity analysis of the impact of different threshold settings on the results to validate the robustness of the method.

11. Quantifying the performance of SentemQC in anomaly detection using metrics such as accuracy, recall, etc., would provide a clearer evaluation of its effectiveness.

12. Although the discussion section is comprehensive, some parts appear somewhat lengthy. It is recommended to condense and summarize these sections, focusing more on the core arguments.

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Water quality of remote sensing

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 28 May 2025

Sofie G. M. van't Veen

Responses to Reviewer Report 2 - Hongwei Guo This paper introduces SentemQC, a novel and efficient QA-QC method for water quality high-frequency sensor data, highlighting its simplicity, robustness, and time efficiency.

The paper is well-structured and the method is applicable to a variety of water quality scenarios and parameters, showing strong versatility. Specific improvement suggestions:
Response: We sincerely thank the reviewer for the thoughtful and encouraging feedback on our manuscript. We are grateful for the time and effort dedicated to reviewing our work and are pleased to hear that the paper was found to be well-structured and that the SentemQC method was well-received. We particularly appreciate the recognition of its simplicity, robustness, and adaptability across various water quality parameters. In response to the reviewer's suggestions, as well as those from the other two review reports, we have revised the manuscript accordingly. The updated version (Version 2) reflects these improvements, and we hope the changes further strengthen the clarity and impact of our work. The answers specific to review 2 are seen below. The reviewer's original comments are presented in **bold** *below, followed by our detailed responses in italic.* ***Thank you for your review, it is much appreciated.***

1. It is recommended to provide a more detailed algorithm description in the methods section, particularly regarding the rules for multi-window comparison and the detailed logic for anomaly labeling.
Response: Thank you for your valuable comment. This is a very good idea. We have revised the 'Implementation' section by including formulas for SentemQC and the moving windows (equations 1 and 2). Additionally, we have expanded the section "Implementation" with a table (Table 3) that provides detailed information about each of the five moving windows. This table includes the calculations, the purpose of the window, and the number of data points included in the window (width).

2. Although the article mentions considering sensor measurement uncertainty, there is a lack of detailed discussion on the theoretical basis and practical application of this aspect. It is suggested to add relevant statistical background explanations.
Response: Thank you for your valuable comment. We acknowledge the need for a more detailed explanation regarding the sensor measurement uncertainty. In our study, the uncertainty values used in the method are those specified by the manufacturers of the sensors. These are industrial-grade products, and the uncertainty is documented in the respective sensor manuals. While we rely on these manufacturer-provided values, we agree that it could be relevant in future work to investigate whether the stated uncertainty remains valid over time. For instance, after approximately three years of use, wear on the sensor lens may affect measurement accuracy. However, such an investigation falls outside the scope of the present study.
Regarding the practical implementation of the sensor uncertainty in the SentemQC method, we have now added a more detailed description of how the uncertainty is incorporated in the five moving windows used in the method. Please refer to the section "Implementation" for this additional explanation.
3. In the results section, it would be helpful to provide a categorical analysis of different types of anomalies (e.g., sensor drift, random noise, clustered anomalies)

and their specific manifestations.
Response: We appreciate the reviewer's suggestion and agree that a categorical analysis of anomaly types is a valuable addition. While the present version of SentemQC does not explicitly distinguish between different types of anomalies during runtime, we have expanded Figure 5 to illustrate examples of various anomaly types, such as individual anomalies, clustered anomalies, and calibration uncertainty, that the method is capable of flagging. This helps clarify the range of anomalies that SentemQC can detect.
We acknowledge that implementing a classification of anomaly types within the method itself would be a meaningful extension. This idea is included in the "Conclusion and perspectives" section in the revised manuscript, where we outline potential directions for future development. However, this functionality is not yet implemented in the present version of SentemQC.

4. SentemQC is described as suitable for various sensor types, but there is insufficient in-depth analysis of the relationship between sensor measurement frequency and data characteristics. It is recommended to supplement this analysis.
Response: Thank you for your valuable comment. We agree that an in-depth analysis of the relationship between sensor measurement frequency and data characteristics is an interesting and relevant direction. However, we think that such an analysis is beyond the scope of this study. It is worth noting that the datasets used in this study include data from different sensor types with varying measurement frequencies. To perform the suggested analysis thoroughly, we would need to conduct a dedicated iteration for each dataset from each sensor type, which would require a significantly expanded study. We appreciate the suggestion and will consider it for future work, where a more systematic investigation of how measurement frequency influences data characteristics and the performance of SentemQC could be explored.

5. The discussion mentions potential future improvements, such as incorporating electronic logs and missing data imputation. It is advised to further specify the technical feasibility of these improvements.
Response: Thank you for your comment. We agree that the suggested future improvements, incorporating electronic logs and implementing missing data imputation, are important steps for improving SentemQC. In response to your comment, we have revised the discussion section to include a more detailed consideration of the technical feasibility of incorporating electronic logs and missing data imputation into SentemQC. We now clarify that integrating electronic logs would require a standardized format across users for the electronic logbooks, particularly with respect to datetime formatting, to enable automated masking of known error periods as anomalies. Additionally, we outline several potential approaches to missing data gap-fill, ranging from simple linear interpolation to more advanced methods such as concentration–discharge relationships and machine learning techniques. This highlights the need for further investigation to determine the most suitable implementation strategy. We appreciate your input, which has helped us strengthen the discussion of future improvements.

6. While the simplicity of SentemQC is emphasized, the article lacks a demonstration of the actual operation process. It would be beneficial to include this.
Response: Thank you for your comment and for pointing this out. We have added supporting information in the 'Implementation' section and the "Parameterization of the SentemQC model" section, which includes supporting information about the different steps in SentenQC and how to optimize the parameterization of the model, the validation of the model, etc. In addition, more

information about the operation process and parameterization are also added to the description file (SentemQC – Description_V12.pdf) of SentemQC available together with the program at Zenodo: https://zenodo.org/doi/10.5281/zenodo.13830624.

7. It is suggested to discuss potential scenarios or boundary conditions where SentemQC may not be applicable, such as challenges encountered in complex environments like high biological disturbance waters.
Response: Thank you for your comment. We appreciate the reviewer's suggestion to discuss potential limitations of SentemQC for more complex environmental datasets. As for now, we have not evaluated datasets where SentemQC failed to perform adequately. We have added a short discussion of the exploration of its limitations to the "Discussion" section. One of the key advantages of SentemQC is that it does not discard any data; instead, it flags data points, allowing for flexible post-processing and interpretation. However, we agree that it would be interesting to investigate its performance on high-frequency datasets with abrupt disturbances, such as those containing bubble-induced signals from gases like methane. These datasets could present unique challenges due to their sudden and irregular nature. That said, we see strong potential for SentemQC in analyzing such datasets that include periods of change or disturbance. In such cases, the method could be particularly valuable when supported by other parameters that could help identify periods with high activity. This could furthermore enhance the robustness of the data cleaning and quality control of the complex environmental sensor datasets.

8. The discussion section suggests combining supporting parameters (e.g., water level, temperature), and it would be valuable to include a case study that validates the improvement in anomaly detection accuracy when incorporating these parameters.
Thank you for this insightful comment. We agree that incorporating supporting parameters such as water level and temperature has strong potential to enhance the anomaly detection accuracy. In response to your suggestion, we conducted preliminary tests using water level as a supporting parameter. These tests showed good performance. We adjusted the method so that during periods of significant discharge variability, the method dynamically adjusted the standard deviation thresholds within the moving windows. As a result, fewer data points in the dataset were flagged as anomalies in such periods, indicating improved contextual sensitivity. This dynamic adjustment where the upper and lower boundaries of the moving window adapt based on the supporting parameters demonstrates the potential of this approach. However, further testing and validation are needed before full implementation. We are currently developing version 2 of SentemQC, which will include the possibility to include supporting parameters. We expect that this enhancement will significantly improve performance, particularly for datasets like methane, which exhibit rapid fluctuations that make it more challenging to identify anomalies.

9. Some references are concentrated, so it is recommended to expand on recent high-quality relevant literature to provide a more comprehensive background to the study.
Response: Thank you for the suggestion. In response, we have expanded the literature in the revised manuscript to include several recent high-quality studies. For example, some are focusing on high-frequency sensor monitoring in lakes and other relevant advancements in the field. These additions aim to provide a more comprehensive and up-to-date background for the study.

10. It is suggested to include a sensitivity analysis of the impact of different threshold settings on the results to validate the robustness of the method.

Response: Thank you for this valuable suggestion. A sensitivity analysis was conducted during the development of SentemQC as a trial-and-error approach. Specifically, we stress-tested the method using the oxygen dataset from Lemming by changing the number of data points included in the five moving windows. For example, the chosen setting of 960 data points in windows 1 and 2 corresponds to a 20-day time series from Lemming, with 48 data points per day. Through this process, we identified a configuration that provided robust performance across different scenarios. When applying SentemQC to other sensor data types from various freshwater environments, we found that the selected window settings gave reliable results. This indicates that the method is not overly sensitive to threshold variations and supports its robustness across the different datasets. We believe that a more in-depth sensitivity analysis of the threshold value (the number of data points in each window defining the width of the window) would add limited value to the study. This is because the "optimal" number of data points is likely to vary depending on the characteristics of each dataset. Our primary objective was to ensure that the SentemQC method is robust and adaptable across a wide range of dataset types. Applying SentemQC, which was originally developed using the oxygen dataset as described above, to other sensor datasets from different freshwater environments demonstrated that the selected parameters delivered reliable quality control, thereby supporting the method's robustness.

11. Quantifying the performance of SentemQC in anomaly detection using metrics such as accuracy, recall, etc., would provide a clearer evaluation of its effectiveness.
Response: Thank you for this thoughtful comment. While we agree that performance metrics such as accuracy or recall are commonly used to evaluate anomaly detection methods, they are not directly applicable to SentemQC. This is because SentemQC is not a predictive model in the traditional sense, but rather a rule-based quality control framework designed to flag anomalies. A potential approach to quantitatively assess its performance could involve deploying multiple identical sensors (e.g., five of the same type) in the same aquatic environment (e.g., a lake or stream). By comparing the anomalies flagged across these parallel datasets, one could estimate the consistency and reliability of the SentemQC method. However, we believe that such an experimental setup is beyond the scope of the current study, though we acknowledge it would be a valuable direction for future work.

12. Although the discussion section is comprehensive, some parts appear somewhat lengthy. It is recommended to condense and summarize these sections, focusing more on the core arguments.
Response: Thank you for this helpful comment. We have carefully reviewed the discussion section and revised several parts to improve clarity and focus in the revised manuscript. Specifically, we have condensed sections that were overly detailed and streamlined the text to better highlight the core arguments and key findings.

Competing Interests: No competing interests were disclosed.

Reviewer Report 29 November 2024

https://doi.org/10.21956/openreseurope.19603.r46101

© 2024 Lami A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Andrea Lami
[1] National Research Council, Water Research Institute (CNR IRSA),, Verbania Pallanza, Italy
[2] National Research Council, Water Research Institute (CNR IRSA),, Verbania Pallanza, Italy

This paper deals with an increasing difficulty associated with the technical development of automatic sensor to monitor significative water quality parameters.

In fact, associated with technological development there is an increasing need to find out a protocol to clean and perform quality assurance (QA) and quality control (QC) on large volumes of data such as high-frequency data with data points down to every minute.

I found the discussion about the difficulties to filter out outlier well described and appreciate the effort to keep as simple as possible to be user-friendly, making it suitable for processing extensive datasets quickly

One point that it was not fully clear to me is if you keep track of the number of points that are retained at the end of the filtering process to calculate the mean. Do you consider a lower limit below which mean was not calculated?

I have some minor comments or suggestions listed here below:

In the Abstract:
they write “Sensor data from the stream stations yielded anomaly percentages of 0.1%, 0.1%, and 0.2%, which were lower than the anomaly percentages of 0.5%, 0.6%, and 0.8% for the…”
It is not clear to me to what these percentage are referred. They are the different parameters? You mention 4 parameters and listed 3 percentages.

In the Introduction:
a) where they report some examples of HF monitoring in Europe, ***Authors may consider to include the references 1 and 2***

b) you wrote “32 different CQ functions to flag anomalies” ***It is correct or should be QC***

Study site:
In Fig. 1 I would add an inset to show the location of Denmark in Europe.
Is the scale at the bottom of the figure, correct? 140000.00 Km?

In the data and method
There is any reason why you chose a different frequency in acquisition for DO and pH between LWME and lake Ormstrup?
The higher frequency in case of streams respect to lake is because you expect river to be a more dynamic system.

Results
In Fig. 3. ***The plot shows pH and Temperature data, but the legend refers only to pH.***
In the left plot it is stated the moving window have a constant width, but form the plots it seems variable…From the text seems that it is constant within each window and decreasing form W1 to W5? But those shown in the plot do not seems to follow this logic (the first is narrow and the last is wider).

References
1. Tiberti R, Caroni R, Cannata M, Lami A, et al.: Automated high frequency monitoring of Lake Maggiore through in situ sensors: system design, field test and data quality control. *Journal of Limnology*. 2021; **80** (2). Publisher Full Text
2. Rogora M, Cancellario T, Caroni R, Kamburska L, et al.: High-frequency monitoring through in-situ fluorometric sensors: A supporting tool to long-term ecological research on lakes. *Frontiers in Environmental Science*. 2023; **10**. Publisher Full Text

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: limnology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 28 May 2025
Sofie G. M. van't Veen

Respond to Reviewer report 1 – Andrea Lami *We sincerely thank Reviewer#1 for the thoughtful and encouraging feedback on our manuscript. We would like to thank Reviewer#1 for the positive reception of our manuscript. We are very grateful for the reviewer's time and suggestions for the improvement of our paper. We have adjusted the paper based on the recommendations from the three review reports received (see version 2 of the manuscript). The responses specific to Review 1 are seen below. The reviewer's original comments are presented in*

bold, *followed by our detailed responses in italic/green.* ***Thank you for your review and time, it is much appreciated.***

Review:
This paper deals with an increasing difficulty associated with the technical development of automatic sensor to monitor significative water quality parameters.

In fact, associated with technological development there is an increasing need to find out a protocol to clean and perform quality assurance (QA) and quality control (QC) on large volumes of data such as high-frequency data with data points down to every minute.

I found the discussion about the difficulties to filter out outlier well described and appreciate the effort to keep as simple as possible to be user-friendly, making it suitable for processing extensive datasets quickly

One point that it was not fully clear to me is if you keep track of the number of points that are retained at the end of the filtering process to calculate the mean. Do you consider a lower limit below which mean was not calculated?

Response: *Thank you for your feedback. We are pleased that you found the discussion on outlier filtering clear and appreciated our efforts to maintain a user-friendly approach suitable for processing large datasets efficiently. Regarding your question about the number of data points retained after filtering: currently, we do not enforce a strict lower limit for calculating averages or medians. However, to ensure the reliability of statistical calculations within the moving windows, we have implemented a warm-up and cool-down period. Each of these periods includes 480 records by default, though this parameter can be adjusted by the user. This mechanism helps prevent calculations based on an insufficient number of data points, particularly at the beginning and end of the dataset. We have now clarified this aspect in the "Implementation" section of the manuscript, including additional details about the five different moving windows used in SentemQC. We also appreciate your suggestion to consider a minimum threshold for retained data points when computing summary statistics. We agree this would enhance the robustness of the tool and plan to incorporate this feature in a future version of SentemQC.*

I have some minor comments or suggestions listed here below:

In the Abstract:
they write “Sensor data from the stream stations yielded anomaly percentages of 0.1%, 0.1%, and 0.2%, which were lower than the anomaly percentages of 0.5%, 0.6%, and 0.8% for the...” *It is not clear to me to what these percentage are referred. They are the different parameters? You mention 4 parameters and listed 3 percentages.*

Response: Thank you for this comment. We have revised the abstract's description to indicate that: “The explored sensor data revealed anomaly percentages ranging from 0.1% to 0.8% of the cleaned datasets by SentemQC." Since the abstract is limited to 300 words, a more detailed description of the detected anomaly percentages from the various sensor stations can be found in the results section "Total amount of anomalies in the different sensor data," as well as in Extended Data 2

available on Zenodo: vant Veen et al., 2024.

In the Introduction:

a) where they report some examples of HF monitoring in Europe, ***Authors may consider to include the references 1 and 2***

Response: Thank you for the reference about high-frequency sensor monitoring in lakes. We have included the two suggested references in the introduction and some other new references.

b) you wrote "32 different CQ functions to flag anomalies" *It is correct or should be QC*

Response: Thank you for bringing this to our attention. It should be QC. It has been corrected in the revised manuscript.

Study site:

In Fig. 1 I would add an inset to show the location of Denmark in Europe.

Is the scale at the bottom of the figure, correct? 140000.00 Km?

Response: Thank you for this important comment. The correct unit is 140 km. We have created a new, enhanced version of Figure 1 that includes an inset showing Denmark's location in Europe, a new scale bar, and a higher resolution of the figure.

In the data and method

There is any reason why you chose a different frequency in acquisition for DO and pH between LWME and lake Ormstrup?

The higher frequency in case of streams respect to lake is because you expect river to be a more dynamic system.

Response: Thank you for your question. The difference in acquisition frequency for DO and pH between LWME and Lake Ormstrup is primarily due to the nature of the monitoring projects involved.

Yes. The higher measuring frequency in the case of streams is because headwater streams are known to be extremely dynamic systems, which necessitates a higher frequency of data to capture the changes in water quality parameters. In contrast, lakes generally respond more slowly to environmental changes compared to small headwater streams, allowing for a lower measuring frequency. This explanation is included in the "Study sites" section in the manuscript. In addition, the monitoring at the stream stations was conducted as part of a PhD project aimed at testing the sensors. We hope this clarifies the rationale behind the different monitoring frequencies.

Results

In Fig. 3. ***The plot shows pH and Temperature data, but the legend refers only to pH.***

In the left plot it is stated the moving window have a constant width, but form the plots it seems variable…From the text seems that it is constant within each window and decreasing form W1 to W5? But those shown in the plot do not seems to follow this logic (the first is narrow and the last is wider).

Response: Thank you for this comment and for pointing this out. We have revised the conceptual Figure 3 to display only oxygen data. Additionally, windows 1 to 5 now follow the method's logic with decreasing widths (the same width for windows 1 and 2, and for windows 4 and 5, illustrating the principle in SentemQC). Furthermore, the legends have been revised, and "QA" and "ER" have been changed to "Approved datapoints and Flagged anomalies".

Competing Interests: No competing interests were disclosed.