KTH ROYAL INSTITUTE
OF TECHNOLOGY

Degree Project in Biotechnology

Second cycle 30 credits

Particle analysis of drinking water – an online, early warning system approach

LOVA LUNDQUIST BAUMGARTNER

Stockholm, Sweden 2023

Abstract

With emerging challenges in ensuring safe supplies of drinking water to consumers worldwide, there is a need for innovative technologies for water quality monitoring. In this project, an instrument which can detect and classify particles in drinking water using machine-learned models was investigated. The aim was to assess its usefulness as an online, early warning system in the drinking water treatment plant and distribution system at Norrvatten. The assessment was conducted as two separate parts: (1) an analysis of data previously collected by three instruments located around Norrvatten's plant and distribution system, with the aim of finding trends and creating baselines, and (2) by conducting spiking experiments in the instrument using known contaminants in a lab environment. The contaminants tested were *E. coli*, *B. megaterium*, humic acids, cyanobacteria *Synechocystis* PCC 6803, and biofilm. Flow cytometry was performed on the same contaminants to enable comparison. It could be concluded from the data analysis that the instrument can detect seasonal trends in particle levels. In addition, there are classes of particles which are not subject to these fluctuations, making them promising as independent parameters in a warning system. There were however indications that the instrument can make unexpected classifications of particles due to differences in composition in the training data water and the water at Norrvatten. The lab experiments showed that the instrument could detect all contaminants tested, even cell numbers of a few hundred cells/mL. Comparing with flow cytometry, the novel instrument could detect concentrations of cyanobacteria below the detection limit of the flow cytometer, indicating a high sensitivity. It was concluded that the instrument has properties desired in an early warning system, but its usefulness at Norrvatten is limited in its current state due to the unexpected classifications of particles in their water.

***Keywords*—** Drinking water, online monitoring, machine learning, early warning system, baseline.

Sammanfattning

Med nya utmaningar för att tillgodose behoven av tjänligt dricksvatten hos konsumenter över hela världen krävs innovativa tekniker för övervakning av vattenkvalitet. I det här projektet undersöktes ett nytt instrument som detekterar och klassificerar partiklar i dricksvatten med hjälp av maskininlärda modeller. Målet var att utvärdera dess användbarhet som ett onlinesystem för tidig varning på Norrvattens reningsverk och ledningsnät. Utvärderingen utfördes som två separata delar: (1) en översiktilig analys av data som tidigare samlats in av tre instrument placerade på Norrvattens reningsverk och ledningsnät med målet att hitta trender och definiera tröskelnivåer, och (2) genom att uföra spikningsexperiment i instrumentet med kända föroreningar i laboratoriemiljö. Föroreningarna som undersöktes var *E. coli*, *B. megaterium*, humussyror, cyanobakterier av stam *Synechocystis* PCC 6803, och biofilm. Flödescytometri genomfördes på samma föroreningar för att möjliggöra jämförelser. Dataanalysen visade att instrumentet kan upptäcka säsongsvariationer i partikelnivåer. Dessutom har det partikelklasser som inte varierade med dessa fluktuationer vilket gör dem lovande som oberoende parametrar i ett varningssystem. Det fanns dock indikationer på att instrumentet kan göra oförutsedda klassifikationer av partiklar utifrån skillnader i sammansättningen mellan träningsdatats vatten och vattnet på Norrvatten. De laborativa experimenten visade att instrumentet kunde detektera alla föroreningar som testades, även vid cellantal på några få hundra celler/mL. Det kunde jämföras med flödescytometern, där det nya instrumentet kunde upptäcka halter av cyanobakterier under detektionsgränsen för flödescytometern, vilket indikerar en hög känslighet. Därför drogs slutsatsen att instrumentet har potential som tidigt varningssystem, men dess användbarhet hos Norrvatten är begränsad i dess nuvarande tillstånd på grund av de oförutsedda klassificeringarna av partiklar i deras vatten.

Nyckelord— Dricksvatten, online-övervakning, maskininlärning, system för tidig varning, tröskelnivå.

Contents

Abbreviations

DWTP	Drinking water treatment process
DWDS	Drinking water distribution system
EWS	Early warning system
DIHM	Digital in-line holographic microscopy
CFU	Colony forming unit
FCM	Flow cytometry
HPC	Heterotrophic plate count
SGI	SYBR Green I
PI	Propidium iodide
TCC	Total cell count

1 Introduction

1.1 Background

Water scarcity is a global issue recognized as the 6th of the United Nations 17 Global Sustainability Goals: Clean Water and Sanitation. As of today, there are 2 billion people globally who lack access to safely managed drinking water services [1]. At the same time, there are emerging challenges to the drinking water supplies even in countries where the availability is often taken for granted. Growing populations lead to increasing demands for safe water, simultaneously as climate change and human activities bring new issues to the existing supply chains [2]. There are several examples over the years when contamination of the drinking water has affected the consumers and resulted in illness, hospitalizations, and deaths. In 1993, 400000 people in Milwaukee, USA, were affected by a pathogen outbreak in the water. Similarly, microbial contamination of the drinking water in Walkerton, Canada, resulted in 7 deaths and 2300 cases of serious illness. Chemical leaks into drinking water have also occurred, for example aluminium sulphate spillage in Camelford, UK, in 2001, ammonium leakage in Tel Aviv, Israel in 2001, and crude 4-methylcyclohexanemethanol into the Elk River in USA in 2014 [3].

In Sweden, similar accidents have happened. There were 78 waterborne outbreaks between the years 1992 and 2011, according to the Public Health Agency of Sweden, (Folkhälsomyndigheten). The outbreaks with the highest number of infected people could be linked to surface water, although polluted groundwater resulted in a higher number of separate incidents. The water originated in a treatment plant in 53% of the cases, and the polluting agent was unknown in 54% of the accidents. Largest and most well-known are the *Cryptosporidium hominis* outbreaks in Östersund and Skellefteå, in 2010 and 2011, of which a total of 47000 people were affected. The consequences were long-lasting – for months on end the inhabitants had to boil their water before consumption [4].

Accidents like these are what drinking water producers aim to prevent. Norrvatten is Sweden's 4th largest drinking water producer and an association of 14 municipalities north of Stockholm. The treatment plant Görvälnverket treats surface water from Lake Mälaren [5]. A picture of their treatment process can be seen in Figure 1. The treatment steps are (1) water intake, (2) water sifting, (3) pumping, (4) mixing with $Al_2(SO_4)_3$ addition, (5) flocculation, (6) sedimentation, (7) sand filtration, (8) carbon filtration, (9) UV treatment, (10) monochloramine and lime addition, and (11) pumping from the reservoir into the distribution network [6]. Like other drinking water producers around the world, Norrvatten face emerging challenges with ensuring a reliable supply of safe water to their consumers. There are many novel methods on the market to target the challenges, just as there are many possible types and sources of contamination that have to be considered. At Norrvatten, new technologies are being tested and evaluated continuously, and this project concerns one such innovation.

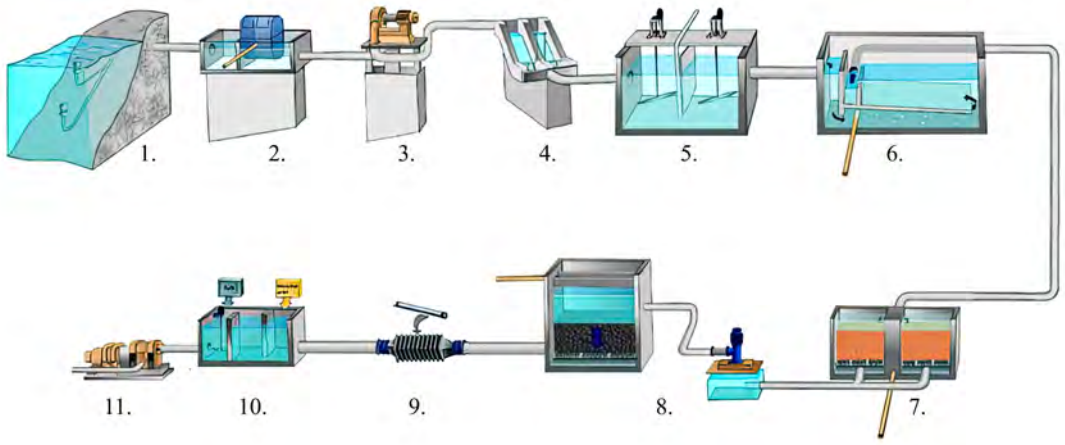


Figure 1: The treatment process at Görvälnverket. (1) Water intake, (2) water sifting, (3) pumping, (4) mixing with $Al_2(SO_4)_3$ addition, (5) flocculation, (6) sedimentation, (7) sand filtration, (8) carbon filtration, (9) UV treatment, (10) monochloramine and lime addition, (11) pumping from the reservoir into the distribution network [6].

1.2 Contamination

Contaminants can be categorised in different ways. Common divisions are between anthropogenic or natural, or whether they are inorganic, organics, biological, or radiological [7]. This project focuses on the distinction between dissolved and particulate compounds , since the instrument investigated detects only particulate matter [8].

Particulate contaminants can originate from many different sources, and end up in drinking water by various routes. If the raw water source is contaminated, the drinking water treatment process (DWTP) acts as a barrier to prevent the pollutants from reaching the consumers. Should this barrier fail, it could affect all consumers connected. The water could also get contaminated as it flows through the drinking water distribution system (DWDS) [9]. This could happen by, for example, under-pressure in the pipes [10].

Sometimes the drinking water is at risk from contamination by wastewater. Wastewater pipes sometimes happen to run in close connection to drinking water pipes. Since wastewater has animal and human origins, it contains many types of particles which could be dangerous to drink. The most well-known risk is that of pathogen outbreaks [10].

Another well-known threat to water quality is buildup of biofilm within the pipes of the DWDS, which could create a beneficial environment for pathogens and even aid in the spread of antibiotics resistance. Other unwanted effects include corrosion, foul taste, odour, clogging and changed hydraulics. A sudden change in flow rate or air bubbles in the water can cause biofilms to detach from the pipes and cause contamination [11] [12].

Humic substances could be brought into the drinking water from water bodies nearby the treatment plant or DWDS. These compounds are dark, amorphous, organic materials resulting from biological and chemical processes. They are so highly transformed that it is no longer possible to identify what they were originally constituents of. Less humic substances are formed in water than in soil, but the particles formed on land can easily

migrate into nearby water bodies [13]. While humic substances in themselves are safe to drink, they can result in disinfection by-products (DBPs) if they react with chlorine in the water treatment process, which have proven to cause many detrimental health effects [14]. They can also cause visible discoloration of the water [15].

An emerging threat that is believed to get worse with a warmer climate and eutrophication of water bodies is outbreaks of cyanobacterial blooms [16]. Cyanobacteria pose a risk to drinking water quality since certain species produce toxins which can be harmful or even lethal to humans as well as animals. There are indications that the toxin-producing strains may be even more favored by rising temperatures and eutrophication than the non-toxin producing ones. To further worsen the issue, it is complicated to remove the variety of toxins in the DWTP. There is usually some method available to remove a certain toxin, but not one that works for all kinds [16].

1.3 Conventional water quality monitoring methods

It is desired to ensure that no contaminants are present in the drinking water. In Sweden, the drinking water producers must deliver water that fulfills the quality criteria stipulated by the Swedish Food Agency (Livsmedelsverket). These criteria concern the entire production chain, from raw water all the way through the DWTP and DWDS to the consumer's taps [17]. According to the regulations, there are limit values for microbial and chemical parameters which must not be exceeded, and in order to ensure this, the producers need to test the water regularly. As highlighted in the descriptions above of known water contamination accidents, pathogen outbreaks seem to pose the greatest risk to human health in Sweden. Detecting specific harmful microorganisms at low levels is rarely possible, instead so-called indicator organisms are often utilised to prove the presence of contamination [18]. For example, *E. coli*, intestinal enterococci, *Clostridium perfringens*, or somatic coliphages can indicate fecal contamination [19] [15]. Finding microorganisms cultivable at 22°C can signal that the drinking water treatment is insufficient, or if detected at the homes of consumers, that there could be issues in the DWDS caused by for example leakage or cross-connections. Testing for slow-growing microbes could indicate growth within the treatment plant or the DWDS [15]. Norrvatten has an accredited lab at Görvälnverket where they conduct all the analyses stipulated by the Swedish Food Agency.

Since the 19th century, microbial water quality control has been based on monitoring the numbers of heterotrophic cultivable bacteria in the water by a method called heterotrophic plate count (HPC). It takes several days from sampling to yield results and is often conducted with spot-checks, meaning that rapid changes of the water quality may not be detected on time, or missed entirely. Additionally, it has been deduced over the years that drinking water contains an abundance of microorganisms, of which only an undefined fraction of heterotrohps is cultivable. This means that the remaining species will not be quantified in the HPC results [20]. The HPC method also does not give any clues about which genera are detected in the test. Even though the HPC methods have been developed over the years to give a better indication specifically of pathogenic presence, there is a widespread perception that HPC analysis has a limited applicability, and that additional methods are required to provide a more comprehensive picture of the microbial

composition in drinking water [21].

Detecting microorganisms by cultivating them is a direct way of proving their presence. There are also methods of proving them indirectly, by monitoring physical parameters like conductivity, turbidity, pH, and so on [22].

1.4 Early warning systems

A common denominator between most contamination incidents is that they could not be detected early enough by the conventional methods to enable preventative measures [3]. In recent years, online monitoring methods have been developed with the intention of providing complementary tools to the conventional methods. The idea with these novel technologies is primarily to provide surveillance of some parameter in the water, for example the microbial composition, as close to real-time as possible [23]. These online measurements could then be part of a mechanism for detection, recognition, and notification, commonly called an early warning system (EWS) [24]. Such a system could warn the producers when potentially harmful compounds are present in the water and enable measures to be taken in time [22].

There are several online monitoring technologies on the market. The most simple ones are automations of earlier mentioned parameters like turbidity, pH, and conductivity [22]. As mentioned they can provide an indirect detection of microbes, often measuring dissolved rather than particulate content in the water [25]. In addition to these standard measurements, novel instruments have been developed based on innovative ideas. They are often specific for a certain type of pollutant [24]. One category of instruments are those based on enzymatic kits to detect the presence of bacteria. They can either give an indication of the total microbial number, like the ColiMinder (VWMs GmbH, Austria), or target specific enzymes for a certain type of bacteria, like the BACTControl (microLAN, The Netherlands). Online flow cytometers have been developed, which quantify the microbial content based on nucleic acid staining, for example the BactoSense (bNovate, Switzerland) [23].

There are also instruments based on optics that capture images of the cells or particles in the water and categorise them according to some form of algorithm. For example, an optical online sensor based on dark-field microscopy, BACMON (GRUNDFOS, Denmark), has been developed which can divide photographed particles into either biotic or abiotic matter [23]. There is also an instrument called AQUATRACK (AQUA-Q, Sweden) which detects deviations in water quality using an online sensor system based on an optical laser. In addition, AQUATRACK samples the water at the time of detected contamination, enabling further analysis of the variation [26].

These methods require various levels of maintenance, and come in a range of costs. Depending on the where the instrument is intended to be used, and how critical the water quality is at that site, these factors may be more or less crucial [23].

1.5 Qumo instrument

This project concerns a similar monitoring device, called Qumo (Uponor corporation, Finland). A picture of the instrument can be seen in Figure 2. It is an online particle detector that is based on a technology called digital in-line holographic microscopy (DIHM) [25]. DIHM, like other types of digital holographic microscopy, requires no lens and builds on the principle that a digital sensor records the diffracted wavefront stemming from a sample illuminated by coherent light from a point source, and generates a hologram from the scattered rays. The light source and the sample are placed close together, while the recording sensor is located farther away, and a magnification factor is introduced [27]. Larger water volumes can be analysed using DIHM as compared to microscopy with a lens. This is due to the capacity of capturing many images in succession, each image having a deep field of view, and because the field of view is larger. The Qumo instrument generates 45 holograms per second, analysed on 0.5-1.5 dL water/min, and produces a statistical average of the data points every 3 minutes [25].

Figure 2: The Qumo instrument.

The holograms generated are analysed using a supervised machine learning model which separates them into different categories depending on features learned from the training data [28]. Particles are distinguished between particles that are not normally present in clean drinking water, and those that are [25]. Distinguishing factors are the size, between 1 to 50 µm, and shape, from asymmetric to symmetric. Five classes are defined in the latest software update. Normal particles, **N-particles**, are those normally present in clean water, not indicating any contamination. **Small particles** are simply particles of a size 1-3 µm. **B-particles** indicate larger, irregularly shaped particles that could be for example biofilm fragments, flocs, protozoa, etc. They are said to be rare in drinking water. **C-particles** are indicators of contamination by sewage or stormwater, while **F-particles** indicate fibres, or simply put particles that are long, but thin. It is claimed that neither F- nor C-particles are normally present in drinking water. Particles could be classified into more than one class simultaneously. Lastly, the **Total particles** are presented as the sum of all particles detected. The units of the particles classes are particles/mL (pcs/mL). In the instrument software, the levels of Total, Small, B-, C-, and F-particles are displayed for the user to monitor [29] [30]. The accompanying software enables the user to set alarm limits of particle thresholds for when the instrument should signal a warning for detection of of a certain class. [25].

1.6 Data handling and alarm limits

Vast amounts of data is generated by an instrument like Qumo and must as a consequence be handled, processed, and understood in order for the technology to be beneficial for the user. As mentioned, alarm limits can set by the user in the Qumo software. A complication in defining alarm limits in any EWS is the fact that drinking water of good quality has fluctuations in composition – there are many harmless particles of varying amounts which are normally present in the water [25]. Water quality fluctuations can be divided into two categories. First, there are event which can trigger irregular rises in particle levels, for example maintenance events where the flow is suddenly changed and particles are detached by shear stress. The second type of events are regular and follow cycles over time, corresponding to water usage by consumers and seasonal variations in the raw water source [31]. Seasonal fluctuations are particularly important in a lake like Mälaren, whose ecological processes are governed by thermal stratification. Located in a climatic zone, the yearly variations in temperature affect water density. During summer and winter, the lake is stratified, meaning that there are layers of water with different densities which inhibit exchange of nutrients and oxygen throughout the lake. In spring and autumn, the entire lake will achieve the same temperature, enabling complete mixing of the water body. This phenomenon, called lake turnover, means that oxygen will be supplied to the deep parts of the lake, while the plentiful nutrients in the deep waters will be spread to the shallower, sunlit regions. As a consequence of lake turnover, a growth spurt will take place in the lake, leading to an increase in organisms, and thus, particles to be detected [32] [33].

The challenge for an EWS is to detect water contamination despite this background noise, which has proven difficult [31]. Normal fluctuations can vary substantially in magnitude and it is desired to detect contaminants at minute concentrations among those variations. The risk of getting false alarms must be weighed against the probability of missing an actual contamination [34]. For this purpose, a baseline signal is commonly established, upon which the alarm limits can be based. Ideas to combat this include finding parameters which only respond to contamination events, and defining a dynamic baseline rather than a static ones for the parameters which do vary with normal quality fluctuations [31]. It should be noted that the terms *baseline* and *alarm limit* will be used interchangeably in this text, both meaning a threshold in particle levels which if exceeded should indicate deteriorating water quality. This is not the case in all texts on the topic.

For the Qumo instrument, it is suggested to set alarm limits for each particle class by first monitoring how they are affected by the normal water quality fluctuations over a few weeks time. According to Uponor, the B-particles levels are usually less than 1 pcs/mL in water coming directly from a DWTP, while levels of up to 2-3 pcs/mL could be acceptable further out in a DWDS. No corresponding guidelines are given for C- or F-particles, but since it is stated that they should not be present in clean water, it might be feasible to assume that their levels should be almost 0 pcs/mL in water coming out of the DWTP, and preferably stay as low as possible throughout the DWDS [29] [30].

1.7 Project objectives

The aim of this study was to investigate the usefulness of the Qumo instrument as an EWS in a drinking water treatment plant or DWDS setting. The project was divided into two

parts: data analysis and practical laboratory experiments. Data analysis was conducted to explore trends over time in the data, as well as methods for defining baselines/alarm limits, based on data previously collected by three Qumo instruments at Norrvatten. The three instruments are called N1, N2, and N3. N1 was located right after the complete DWTP at Görvälnverket, before the water goes into storage. N2 was placed at a water tower in the DWDS which is regularly filled and emptied. N3 was placed at a pump station further out into the DWDS, where it takes several days for the water to reach. The three instruments are placed in the same distribution line, meaning that some of the water that goes through N3 first has passed through N2 and N1. The data analysis was primarily intended to show an overview of what trends and baselines could be, rather than a complete account of all events, since that would be outside of the scope of this project.

The lab experiments were conducted with the intention of testing how the instrument characterised 5 different contaminants, chosen to represent particles which are known to pollute drinking water. The contaminants were grown in the lab or purchased, and prepared to known concentrations. They were *Escherichia coli*, *Bacillus megaterium*, cyanobacteria *Synechocystis* PCC 6803, humic acids, and biofilm. *E. coli* and *B. megaterium* were tested as a representation of bacterial pollution. Humic acids were tested as an example of organic particles that could be present in case of pollution from e.g., lake water. Biofilm was investigated in the Qumo since it is a category of particle that is known to occur in a DWDS. Cyanobacteria were tested since it is of interest to monitor their presence in drinking water due to their toxins.

2 Materials and methods

2.1 Data analysis tools

The data from the three instruments N1, N2, and N3 was initially studied using the software aCurve (Gemit solutions, Sweden), with the intention of finding general trends and extraordinary events. Around 21 months of data was available at the time of analysis. The newest categories, C- and Small particles, were implemented on March 31th 2021. Therefore, the analysis focused on the time period 2021-03-31 to 2023-01-01.

Due to the previously mentioned seasonal fluctuations of particles caused by lake turnover, it may be necessary to define baselines for each season separately. Sweden's meteorological and hydrological institute (SMHI), reports on the onset and duration of spring, summer, autumn, and winter in Sweden, based on the meteorological definitions of the seasons, see Table 1 for the start and end of each season during year 2021 and 2022 [35]. These dates were used when testing seasonal baselines on Qumo data.

Table 1: The start and end of the seasons during 2021 and 2022 [35].

	Spring	**Summer**	**Autumn**	**Winter**
2021	19th of February	10th of May	11th of October	26th of November
2022	6th of March	9th of May	18th of October	3rd of December

In aCurve, data from other instruments and sensors was also available, which made it possible to make comparisons with the data from N1, N2, and N3. There was a flow

cytometer called BactoSense (bNovate, Switzerland) located at the same water tower as N2. The BactoSense and Qumo data overlaps from March 2021 to April 2022. At the same tower, there were also sensors monitoring the water level in the tower. Starting from May 2021, flow cytometry (FCM) measurements were also available from the site where the N1 instrument was located, measured with an Accuri C6 Plus flow cytometer (BD Biosciences, Belgium) in the lab at Görvälnverket. Data on algae counts in the raw water at Görvälnverket was also available for the entire time period 2021-03-31 to 2023-01-01.

Data was exported from aCurve in Excel format with an hourly resolution. Since the instrument provides a measurement every third minute, it is in theory possible to have a higher resolution in the data. However, since the analysis was intended to give an overview it was deemed to be detailed enough. The data was analysed in Python. The libraries Matplotlib, pandas, and NumPy were used to make calculations and graphs.

2.1.1 Baseline theory and methodology

The first baseline method tested was one based on an approximation of how much of the instrument run-time consists of unusual events, and based on that calculate a baseline on a certain percentile [25]. To approximate an alarm limit based on trends in previously collected data is also what it recommended by Uponor for the Qumo instrument [29]. The 97th percentile was calculated and plotted for previously collected data to visualise what a moderately high, static baseline would look like on actual data.

The second method tested was to calculate the mean of the data over a time period, and add three standard deviations calculated over the same time period to the mean [23]. This is a well-established definition of the detection limit of a sensor [34]. Just like calculating a certain percentile based on previous data, this baseline is static, which is compatible with the Qumo software.

The third baseline strategy was a simple moving window analysis, also called rolling analysis. A benefit with a moving window analysis is that the continuous changes in water quality are incorporated into the baseline, which is not static but rather updates over time with a resolution of a set ”time window” corresponding to a subset of data points of a certain size [23] [36]. A window size of 1 week was used to approximate a baseline for the data. For each window, the mean of the data points within the window was calculated. To that mean, three standard deviations was added for the entire time period, meaning the common standard deviation over all windows. These weekly sums correspond to the third baseline variant.

After having analysed the general trends of each particle class, it could be determined which of the baseline methods would be most applicable on the respective categories. Examples of baselines applied on data frorm N1 and N2 are shown in this report. Since an alarm limit in the Qumo software would have to be based on previously collected data, the baseline examples calculated from 2021 data were also applied on data from 2022 to give a sense of how well it predicted the trends of the following year.

2.2 Instruments

For this project, the primary instrument investigated in the lab was a Qumo. In addition, two Accuri C6 Plus flow cytometers (BD Biosciences, Belgium) were used in the lab at Görvälnverket to provide complementary analyses on some of the samples once they had been tested in the Qumo. The preparation protocol for the FCM samples can be found in Appendix 8.1. SYBR Green I (SGI) and Propidium Iodide (PI) were used as fluorescent dyes to stain the samples. Both live and dead cells can be stained by SGI, while PI cannot permeate cell membranes and thus stains only broken cells. Together, they enable a quantification of the number of viable cells [37].

2.2.1 Qumo set-up and maintenance

The Qumo instrument experiments were conducted at the Royal Institute of Technology (KTH) in Stockholm, Sweden. The setup for the experiments consisted of a Qumo instrument connected to silicon tubes (4X8 mm and 6X10 mm) and a pump. Duran flasks of different sizes were used as containers for the water and contaminants to be pumped through the instruments. See Figure 3 for pictures of the setup. The flow rate was controlled regularly using a measuring cylinder and a watch. It was kept at around 68 mL/min. To ensure that the particles were uniformly distributed in the test liquid, a magnetic stirrer was used. Between each experiment, the instrument was flushed with deionised water for at least 20 minutes.

The instrument flow cell was cleaned regularly between experiments, at least once a week. Through the camera view, the windows could be manually controlled for fouling. Unusual particle levels were also interpreted as indications of flow cell fouling. After a number of experiments had been conducted, the flow was redirected through the instrument in order to avoid fouling of the sensors and the flow cell.

(a) **(b)**

Figure 3: The experimental setup. (a) An example of how the bottle with liquid to be analysed was connected to the silicone tubes on the magnetic stirrer. (b) Shows how the tubes were connected to a pump and to the Qumo instrument.

2.3 Contaminants preparation and experiment execution

To test the contaminants, spiking experiments were conducted. A spiking experiment entails analysing a sample containing known concentrations of a certain compound [38]. Each contaminant was tested in spiking experiments at least three times. All were diluted in sterilised deionised water. For the spiking mixtures intended to be run in the FCM, samples were taken out of the liquid tested in the Qumo.

2.3.1 *Escherichia coli*

E. coli have a length of around 1.5 µm and width of 0.5 µm [39]. The *E. coli* of strain DSM498 were grown in nutrient broth. Their approximate number of colony forming units (CFUs) were calculated from an overnight culture. The culture was diluted to an OD_{600} of 0.1, which was further diluted in series to dilution factors of 10^5 and 10^6. 100 µL of the 10^5 and 10^6 dilutions were plated in duplicate on nutrient agar. The CFU/mL was used to dilute the cells to the intended concentrations for the spiking experiments. The concentrations chosen were based on what had previously been tested in the same instrument [25].

Concentrations of 10^3 and 10^4 *E. coli* cells/mL were diluted from overnight cultures which had been diluted and measured to have OD_{600} of 0.1. Each concentration was tested in triplicate spiking experiments. Each concentrations was run for around 60-90 minutes by circulating the contaminated liquid through the instrument. First, the lower concentration of 10^3 cells/ml was tested, and then without flushing in between, the higher concentration of 10^4 cells/ml. The lower concentration was run with a total volume of 1.5 L. Once it had been analysed, 0.5 L of the starting volume was emptied from the bottle. To the remaining 1 L of liquid, additional bacterial cells were added to achieve the desired higher concentration of contaminant. The *E. coli* cells were tested in a fourth spiking experiment where a lower cell count was aimed for, around 10^2 cells/mL. This sample mixture was left overnight, circulating through the Qumo.

For FCM analysis, the *E. coli* were stained using SGI and PI.

2.3.2 *Bacillus megaterium*

B. megaterium have a size of around 4 µm in lengths, and 1.5 µm in width [40]. The *B. megaterium* of strain DSM1668 were grown in nutrient broth. Their CFU/mL were calculated in the same manner as for the *E. coli*, except that the OD_{600} 0.1 mixture required only to be diluted by factors 10^4 and 10^5. The concentrations of *B. megaterium* cells intended to be tested were 10^2/mL and 10^3/mL. They were diluted from overnight cultures which had been diluted and measured to have OD_{600} of 0.1.

Similarly as *E. coli*, the *B. megaterium* were tested in triplicate spiking experiments. The procedure was the same as described for *E. coli*, where the lower concentration was tested first, and the higher after, once the bottle had been emptied of 500 mL. This procedure was foregone for the last *B. megaterium* spiking replicate, where no liquid was taken out other than the FCM sample liquid of around 30 mL. Additional cells were simply added directly to the remaining volume.

For FCM analysis, the *B. megaterium* were stained using SGI and PI.

2.3.3 Humic acids

Humic acids were ordered from Merck (product number 53680). 30 mg of humic acids were dissolved in 1 L sterilised deionised water. To remove the bigger particles, they were allowed to settle for around 15 minutes before the liquid was carefully decanted into another bottle. This was done in order to avoid clogging the instrument during the spiking experiment. The humic acid mixture was circulated through the instrument for approximately 90 minutes. This spiking experiment was repeated thrice. For the FCM analyses, the humic acid was stained using SGI.

2.3.4 Cyanobacteria (*Synechocystis* PCC 6803)

The cyanobacteria *Synechocystis* PCC 6803 have a diameter of around 2 µm and are a unicellular form of fresh water living cyanobacteria [41] [42]. *Synechocystis* PCC 6803 cells were obtained from the Department of Protein Science at KTH. An approximate number of cells in the obtained sample was given, around 10^9 cells/mL. This cell count was used to dilute the *Synechocystis* PCC 6803 to around 10^2 cells/mL in a test volume of 1.5 L. The spiking experiment was run for 4-5 hours, and repeated on three separate occasions.

The cyanobacteria were analysed in the FCM based on their chlorophyll and phycocyanin content. As a test, they were also stained with SGI in a separate run.

2.3.5 Biofilm

Biofilm was grown on plastic in E-flasks in laboratory conditions. For the first replicate, plastic pieces cut from 50ml falcon tubes were used, and in the second and third replicates K5 carriers were used instead. A mixed culture of 5 different strains of bacteria were inoculated in the flasks, *Enterococcus, Pseudomonas, Salmonella, Serratia and Comomonas denitrificans 110*. Two different concentrations of nutrient broth were tested, one with undiluted broth and one with nutrient broth diluted 10 times with sterilised deionised water. After 4-14 days, varying for the replicates, when the bacteria had visibly grown, the liquid culture was decanted and the plastic carriers were rinsed with sterile deionised water. The biofilm was detached by adding additional water and shaking vigorously, as well as using magnetic stirrers at high speeds. The detached biofilm was studied under the microscope.

The biofilm mixtures were tested in the Qumo first in a volume of 1 L, of which 500 mL consisted of water with detached biofilm pieces. After 1-2 hours, another 500 mL of water was added to dilute the mixture. This diluted mixture was run through the instrument for at least 45 minutes. Samples of the mixture was taken before and after the dilution, and their optical densities were measured at OD_{600}. This procedure was repeated in triplicate experiments. The biofilm spiking mixtures were not tested in the FCM.

3 Results

3.1 Data analysis

3.1.1 Disregarded data and notable events

There were a few known instrument malfunctions which made periods of the data unavailable. The N3 instrument had stopped functioning during the summer 2021 and been replaced. The same instrument was also moved to a sand filter within the DWTP in the summer 2022, making all data thereafter incomparable to the earlier data.

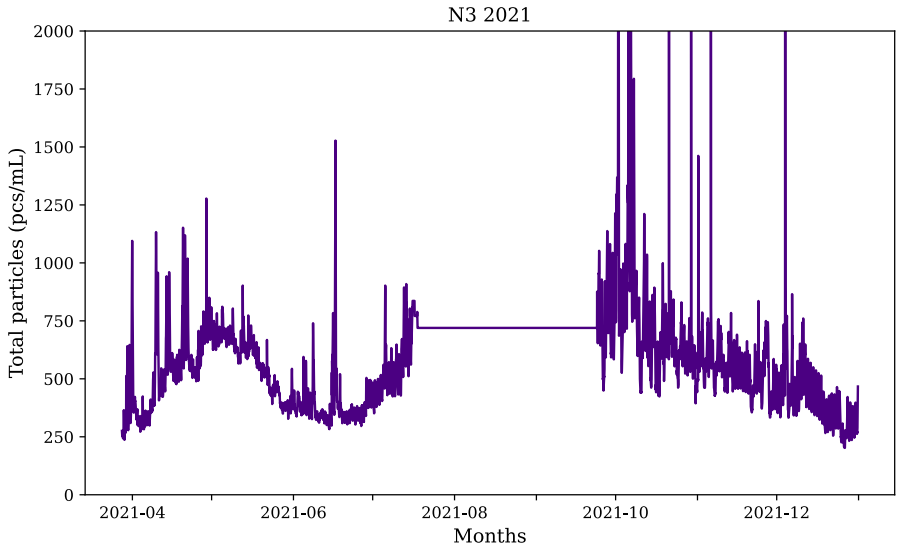


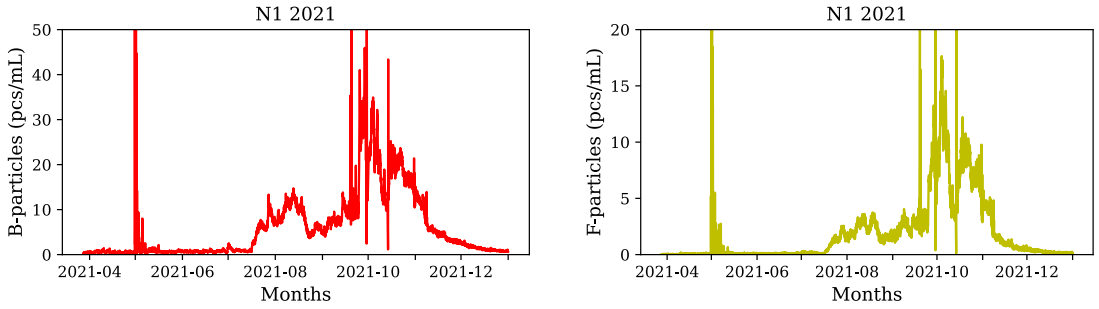
Figure 4: Example of what a period of flatline particle levels can look like. Here from the N3 instrument data on Total particles during the summer of 2021. This was due to an instrument malfunction. Note that outlier y-values in have been cut off to make the graph easier to interpret.

During some time periods, there were what can be called "flatlines" in the data. One example of such a period was the aforementioned malfunction of the N3 instrument in the summer of 2021, see Figure 4 where the Total particles are shown as an example. This meant that all particle categories lay flat at the same level for a prolonged time. Flatlines can be a sign of faulty connection where the measurements are not accurately reported to the software, something which also happened during the course of this project when the cloud component that handles data from the Qumo devices broke. Other longer-lasting flatlines were N1 2022-11-20 to 11-23 and N2 2022-10-20 to 10-23.

The N2 instrument at the water tower recorded daily fluctuations of particles which could be explained by the sensors measuring water levels in the tower, see Figure 23a in Appendix 8.2.2 for an example. In contrast, there were a few events where the water tower was completely emptied, followed by a sharp rise in particle levels, see Figure 23b in Appendix 8.2.2 for an example of such an event.

During the entire time period analysed, there were no known extensive contamination accidents reported by Norrvatten. Despite this fact, there were instances in the data where

heightened levels of B- and F-particles could be observed, the particle classes which according to Uponor should be scarce or absent in clean drinking water, as mentioned in the introduction. One example can be seen in Figure 5, where B-particles and F-particles were continuously detected in several pcs/mL by N1 from summer all the way into the winter months of 2021. Heightened levels of C-particles over 1 pcs/mL over prolonged periods of time could however not be observed.



(a) **(b)**

Figure 5: (a) Heightened levels of B-particles detected by N1 during 2021. (B) Heightened levels of F-particles detected by N1 during 2021. Note that outlier y-values in have been cut off to make the overall trends easier to see.

3.1.2 Trends in the data

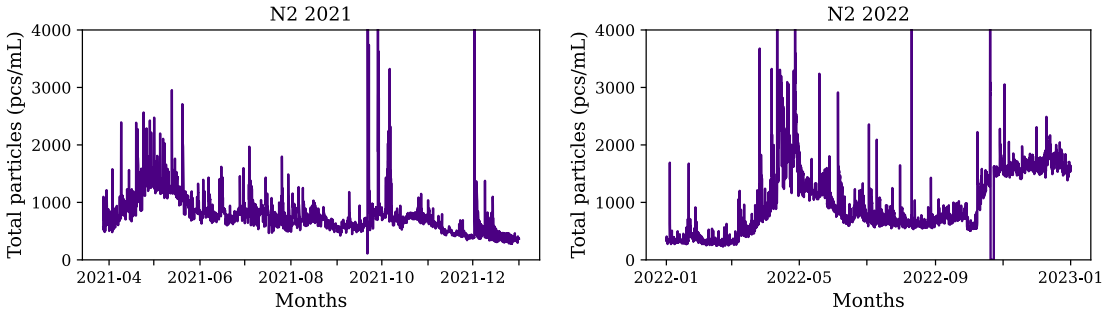
The seasonal trends for the different instruments had certain similarities in Small and Total particles, particularly a noticeable increase in Total and Small particles during April-June, see Figure 6 for an example where the Total particles detected by N2 in the water tower 2021 and 2022 are shown. The trends in Total particles detected by N2 could be compared with the corresponding data collected by N1 and N3, see Figure 16 in Appendix 8.2.1. Similarly, the trends for Small particles detected by all three instruments are shown in Figure 17 in Appendix 8.2.1, indicating a similar trend.

The N1 instrument data could be compared to FCM measurements and algae counts from the lab at Görvälnverket during the same time period, and the trends looked similar particularly in Spring, see Figures 21 and 22 in Appendix 8.2.2.

The N2 data could be compared to data from the BactoSense FCM, see Figure 24 for a comparison between total cell counts detected by the BactoSense, and the Small particle levels detected by Qumo during the same time period. Both instruments detected higher number of particles in the spring, but the Qumo curve did not follow the same rise in the autumn as the BactoSense did, when comparing with the summer measurements for each instrument.

To provide numerical comparison, the means of Total and Small particles detected by N1, N2 and N3 each season were calculated. Before calculations were made, the flatline data points were removed from the data set. The means calculated for Total and Small particles detected by N2 during the time period analysed can be seen in Table 2, and in Tables 7 and 8 in Appendix 8.2.1 for N1 and N3 data. Increased particle levels during spring correlate

with the lake turnover. However, while lake turnover is also known to happen during autumn, that is less obvious in the Qumo data as compared with the increases of particles in springtime.



(a) **(b)**

Figure 6: Total particles detected by the N2 instrument during the time periods (a) 2021-03-28 to 2022-01-01, and (b) 2022-01-01 to 2023-01-01. Note that outlier y-values in have been cut off to make the overall trends easier to see.

Table 2: The means of Total and Small particles detected by N2 in the water tower during each season of 2021 and 2022.

	Total particles (pcs/mL)	**Small particles (pcs/mL)**
Spring 2021	1177.48	300.67
Summer 2021	820.32	224.92
Autumn 2021	639.18	207.81
Winter 2021	359.20	112.34
Spring 2022	1300.50	311.29
Summer 2022	845.85	184.42
Autumn 2022	1616.99	326.93

Looking at B-particles, yearly trends were less obvious, particularly considering the sudden rise shown in Figure 5a detected by N1 in 2021. The other B-particle trends detected by N1 2022, and N2 and N3 in 2021 and 2022 can be seen in Figure 18 in Appendix 8.2.1. Perhaps with longer time periods of data, a more definite statement on B-particle seasonal detections could be provided.

While there were no seasonal trends to be discerned in the C-particles, it appeared like the levels were higher and more fluctuating in the N2 and N3 data than for the N1 instrument, see Figure 19 in Appendix 8.2.1. This could indicate a difference in the C-particle levels in the treatment plant and in the DWDS.

No clear seasonal trends could be seen in F-particle data either, see Figure 20 in Appendix 8.2.1.

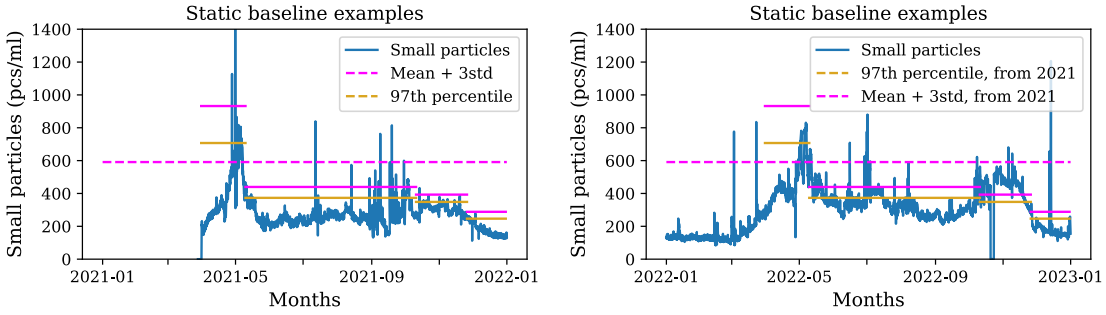
3.1.3 Baselines

As could be seen in the previous section, the C-, F-, and B-particle categories appear less sensitive to the normal water quality fluctuations, and it may thus be simpler to define a

baseline for those categories. Particularly for C-particles, a static baseline would most likely suffice. The same could be true for B- and F-particles, but the heightened levels shown in Figure 5 are a complicating factor.

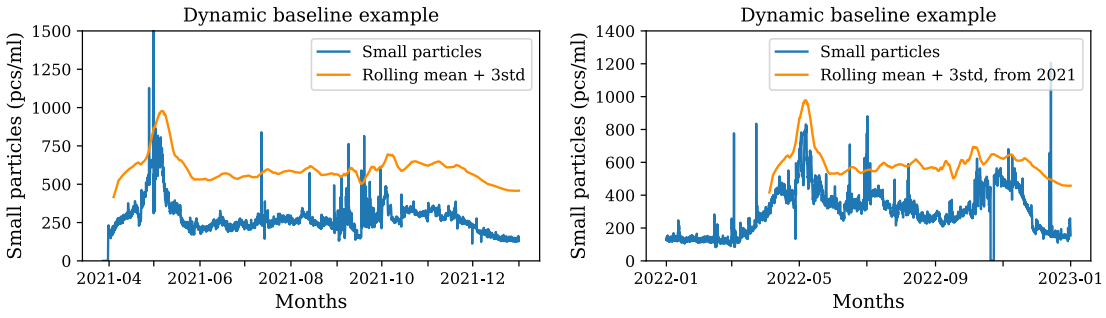
Since it could be observed that Small and Total particle categories fluctuate over the seasons, defining a baseline will likely be most challenging for those categories. As an example, the previously explained methods for defining alarm limits were tested on Small particle data collected by the N1 and N2 instruments during 2021 and 2022.

In Figure 7, the static baseline approaches of the 97th percentile, and the mean plus three standard deviations are shown. The baselines are calculated both for the entire year, as well as for each season separately. Figure 7a shows what the baselines look like applied on the data which they were calculated from, and Figure 7b shows what the same baselines look like applied on data collected the year after.



(a) **(b)**

Figure 7: Examples of what the static baselines approaches could look like applied on real Small particles data collected by the N1 instrument. (a) The baselines calculated from 2021 data, plotted together with the data they were calculated from. (b) The same baselines calculated from 2021 data, plotted together with data from 2022 instead. Note that outlier y-values in (a) have been cut off to make the overall trend easier to see.



(a) **(b)**

Figure 8: Examples of what the moving window baseline approaches could look like applied on real Small particles data collected by the N1 instrument. (a) The baseline calculated from 2021 data, plotted together with the data it was calculated from. (b) The same baseline based on 2021 data plotted together with data from 2022 instead. Note that outlier y-values in (a) have been cut off to make the overall trend easier to see.

The moving window baseline approach with a data window of one week can be seen in Figure 8. Figure 8a shows what the baseline look like applied on the data which it was calculated from, and Figure 8b shows what the same baseline looks like applied on data collected the year after.

The static and dynamic baseline examples can also be seen applied on Small particle data from the N2 instrument in Figures 25 and 26 in Appendix 8.3 for comparison. There, it can be seen that the dynamic baseline did not fit as well on the data of 2022, as it did for N1.

3.2 Lab experiments

Looking at the results from the lab experiments, note that particle levels reported by the instrument could sometimes get notably high while the system was empty of water or the water was still in the system while the pump was turned off. These occasions often happened in connection with the experiments, as preparations were made. The duration of each experiment is highlighted so as not to be confused with these events.

3.2.1 Sterilised deionised water as background

Since sterilised deionised water was used to dilute the spiking contaminants, the background noise from deionised water was tested in the Qumo. The flow cell was cleaned before starting. The Total, Small, B-, C-, and F-particle levels when pumping only deionised water through the Qumo can be seen in Figure 9.

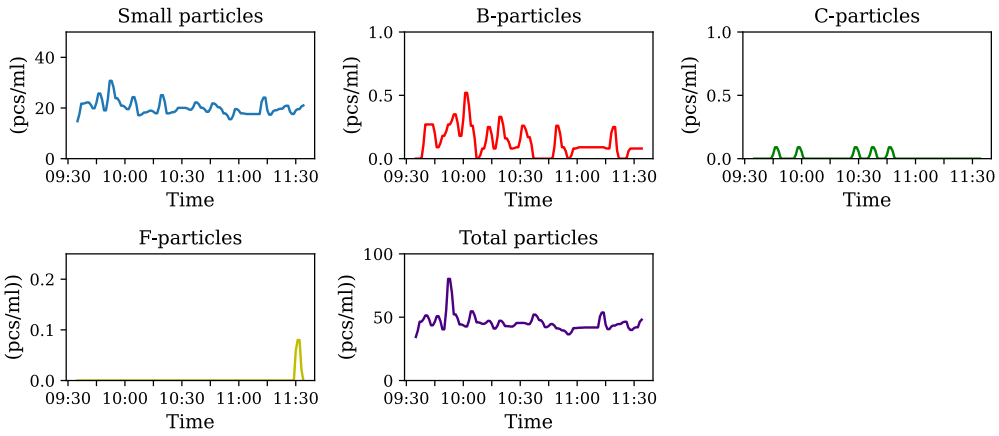
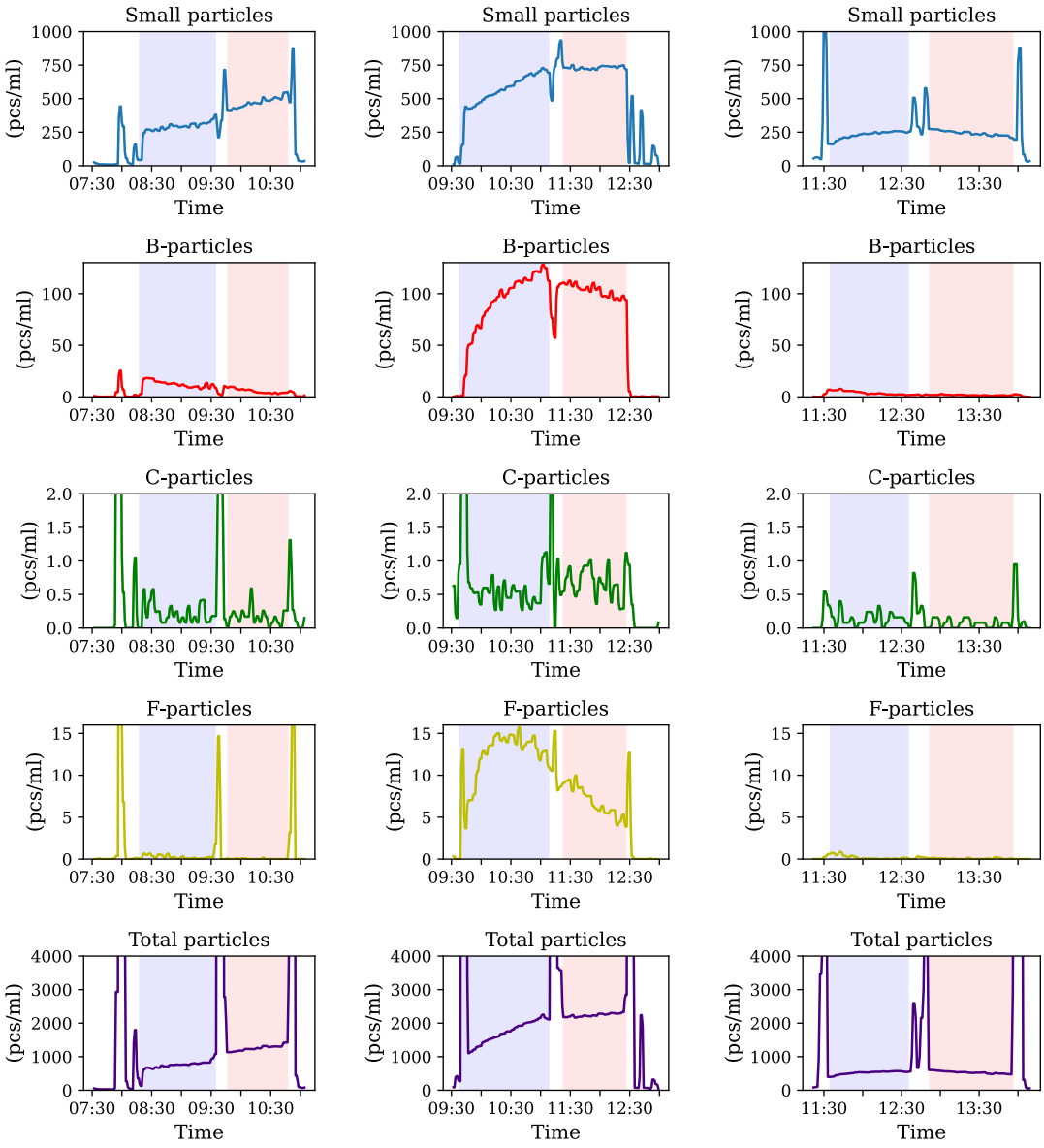


Figure 9: The Total, Small, B-, C-, F-particles levels detected testing deionised water in the Qumo for approximately 2 hours.

3.2.2 *E. coli*

The levels of Total, Small, B-, C- and F-particles detected during the first three *E. coli* spiking experiments can be seen in Figure 10, where (a) shows replicate experiment 1, (b) replicate 2, and (c) replicate 3. All three were detected as substantial increases in Total particles, of which a substantial fraction consisted of the Small particle class.



(a) Experiment 1 **(b)** Experiment 2 **(c)** Experiment 3

Figure 10: (a)-(c) show the Small, B-, C-, F-, and Total particle levels detected during the three replicate *E. coli* spiking experiments. The blue fields show duration of the lower concentration test of 10^3 cells/mL, while the red field show the duration of the 10^4 cells/mL concentration tested after. Note that outlier y-values have been cut off in some graphs to make the overall trends easier to see.

The Total particle levels differed for the first three spiking experiments despite the fact they were all prepared using the same procedure. The profile also looked different between the replicates, sometimes the levels increased or decreased during the experiment. For experiment 1 and 2, the Small particle level rose after addition of more cells, while no such increase could be discerned for experiment 3.

The B-particles profiles differed between the replicates. For experiment 1, the B-particle profile showed a decreasing trend in contrast to the Small particle levels during the same experiment, with a maximum of around 17-18 pcs/mL. During experiment 2, the B-particles were substantially higher, at most around 120 pcs/mL. In the third replicate experiment, the B-particles were started at around 6-7 pcs/mL, then decreased down to a

somewhat stable level of around 1-2 pcs/mL, see Figure 10.

As can be seen, the F- and C-particles were very low during experiment 1 and 3. During experiment 2, the C-particles were low, but the F-particles were much higher than in 1 and 3, increasing to levels over 10 pcs/mL.

The fourth *E. coli* experiment was run overnight. In Figure 11, the particle categories Total, Small, C-, B-, and F- can be seen during that experiment. The Total and Small particle levels were stable during the first few hours, but then started increasing, continuing all the way until the termination of the experiment the following morning. The B-particles were highest in the beginning at around 4-5 pcs/mL, and then decreased to stay below 1 pcs/mL for the rest of the experiment. The C- and F-particles were below 1 pcs/mL for the entirety of the experiment.

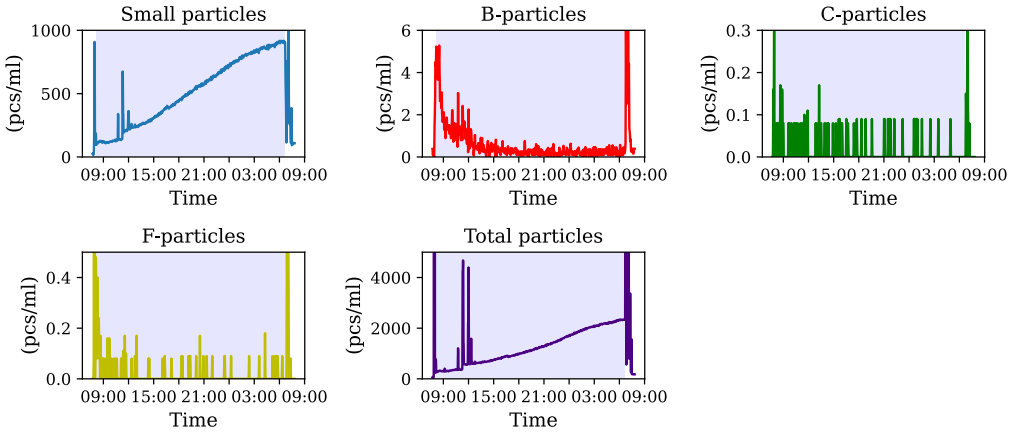


Figure 11: The Small, B-, C-, F-, and Total particle levels detected during the fourth *E. coli* spiking experiment, which was run overnight. The blue field shows the duration of the experiment. A starting cell concentration of 10^2 cells/mL was tested. Note that outlier y-values have been cut off in some graphs to make the overall trends easier to see.

In Table 3 the results of the FCM measurements conducted at Görvälnverket are shown. As can be seen, the total cell counts/mL (TCC/mL) differed a bit from each other, but were in the same magnitude as was aimed for, with a possible exception being experiment 3 where the cell count was significantly lower than intended. As can be seen from comparing the SGI results to the PI+SGI numbers, the cells where unbroken in most of the samples at the time of measurement. The cell number in experiment 4 increased overnight, indicating growth. Note that for experiment 3 testing the concentration 10^4 cells/mL, there were outliers in replicates both for the SGI and PI+SGI samples. Therefore, the PI+SGI average is higher than the SGI average, which cannot happen.

Table 3: The results of testing the *E. coli* spiking mixtures in the flow cytometer at Görvälnverket.

	SGI average TCC/mL	PI+SGI average TCC/mL
Experiment 1, 10^3 cells/ml	1240	967
Experiment 1, 10^4 cells/ml	17380	8007
Experiment 2, 10^3 cells/ml	1733	1127
Experiment 2, 10^4 cells/ml	7580	6187
Experiment 3, 10^3 cells/ml	393	373
Experiment 3, 10^4 cells/ml	2640	2760
Experiment 4, 10^2 cells/ml	580	480
Experiment 4, next day	8447	6033

3.2.3 *B. megaterium*

The levels of Total, Small, B-, C- and F-particles detected during the three *B. megaterium* spiking experiments can be seen in Figure 12, where (a) shows replicate experiment 1, (b) replicate 2, and (c) replicate 3. The *B. megaterium* spiking experiments resulted in substantial increases in the Total particles, and much of it were Small particles. Similarly as for the *E. coli* tests, the Small and Total particle profiles differed for the three replicates despite the fact they were all prepared using the same procedure. In experiment 2 and 3, the Small and Total particle profiles reflected the transition from the lower cell concentration to the higher, while in experiment 1 it was less obvious.

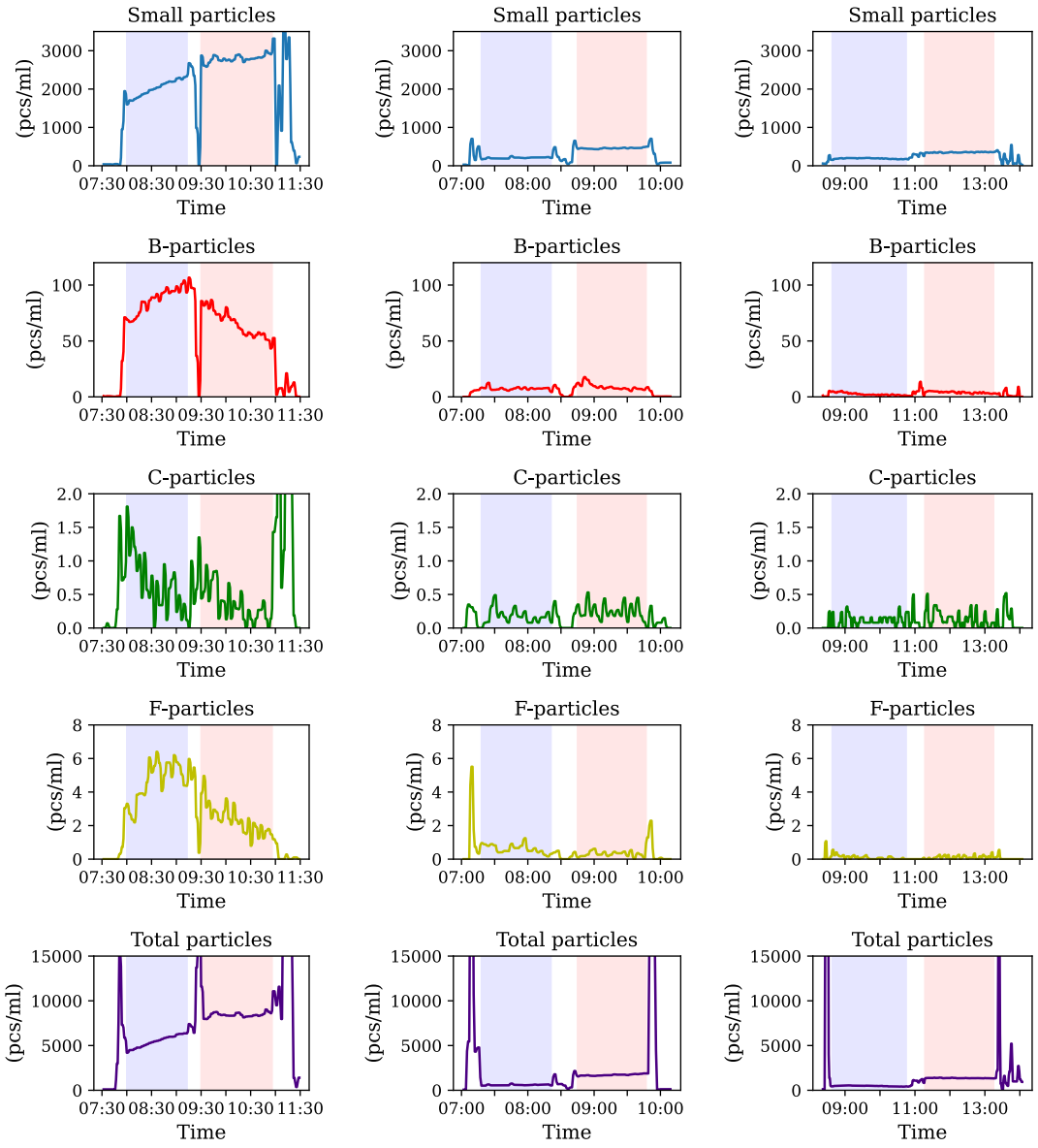
When it came to B-particles, the levels were highest during experiment 1, at most around 100 pcs/mL, which correlated with the higher Total particle level as well. In experiments 2 and 3, the levels were significantly lower.

C- and F-particles were below 1 pcs/mL during the entirety or majority of experiments 2 and 3. During experiment 1 however, they were somewhat higher, particularly the F-particles.

Table 4 shows results of the FCM measurements conducted at Görvälnverket for each *B. megaterium* spiking experiment. The TCC/mL differ a bit from each other as well as from the intended number. Comparing the SGI and PI+SGI results, it can be seen that a substantial fraction of the *B. megaterium* cells were broken.

Table 4: The results of testing the *B. megaterium* spiking mixtures in the flow cytometer at Görvälnverket.

	SGI average TCC/mL	PI+SGI average TCC/mL
Experiment 1, 10^2 cells/ml	1650	620
Experiment 1, 10^3 cells/ml	4247	1460
Experiment 2, 10^2 cells/ml	527	307
Experiment 2, 10^3 cells/ml	947	320
Experiment 3, 10^2 cells/ml	333	100
Experiment 3, 10^3 cells/ml	773	547



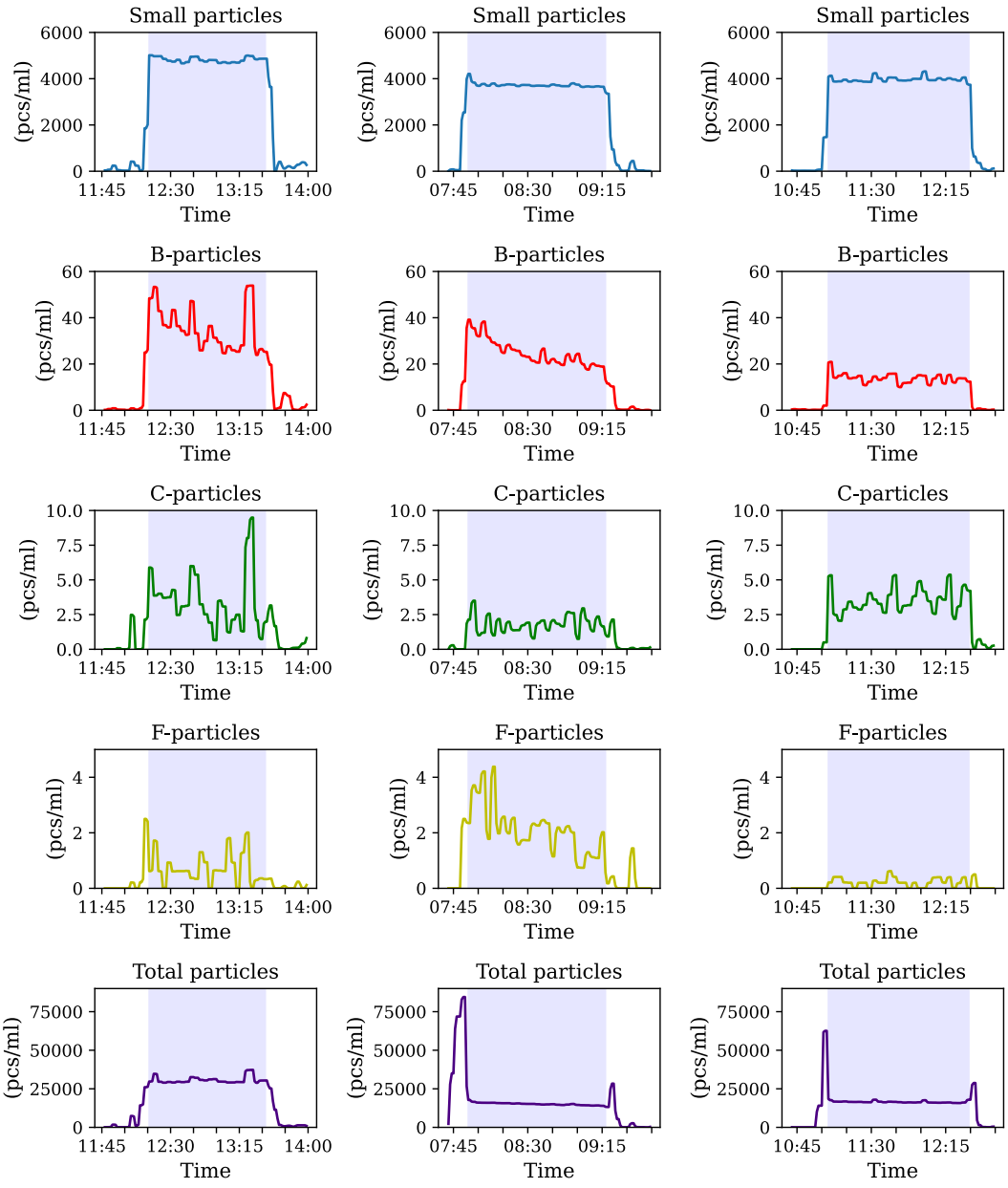
(a) Experiment 1 **(b)** Experiment 2 **(c)** Experiment 3

Figure 12: (a)-(c) show the Small, B-, C-, F-, and Total particle levels detected during the three replicate *B. megaterium* spiking experiments. The blue fields show the duration of the lower concentration test of 10^2 cells/mL, while the red field show the duration of the 10^3 cells/mL concentration tested after. Note that outlier y-values have been cut off in some graphs to make the overall trends easier to see.

3.2.4 Humic acids

The levels of Total, Small, B-, C- and F-particles detected during the three humic acid spiking experiments can be seen in Figure 13, where (a) shows replicate experiment 1, (b) replicate 2, and (c) replicate 3. The humic acid spiking experiments resulted in quite even profiles of Total and Small particles. As compared with the *E. coli* and *B. megaterium* experiments, a smaller fraction of the particles were classified as Small, 1-3 µm. B-particles increased substantially in number in each replicate experiment. In experiments 1 and 2, the levels increased to around 40-50 pcs/mL at the highest, while they never went above 15-20 pcs/mL in the third experiment. When it came to C- and F-particles, the C-particle level increased in all three experiments and was at least above 1 pcs/mL for most

of the duration, occasionally rising to several times that level. The F-particles were lower in experiment 1 and 3, most of the time below 1 pcs/mL. In experiment 2 the F-particles were somewhat higher, around the same levels as the C-particles.



(a) Experiment 1 **(b)** Experiment 2 **(c)** Experiment 3

Figure 13: (a)-(c) show the Small, B-, C-, F-, and Total particle levels detected during the three replicate humic acid spiking experiments. The blue fields show the duration of the experiments.

The humic acid FCM results are shown in Table 5. The number of detections were quite similar in each sample, in agreement with the Qumo results.

Table 5: The results of testing the humic acid spiking mixtures in the flow cytometer at Görvälnverket.

	SGI average detections
Experiment 1	308707
Experiment 2	333513
Experiment 3	352240

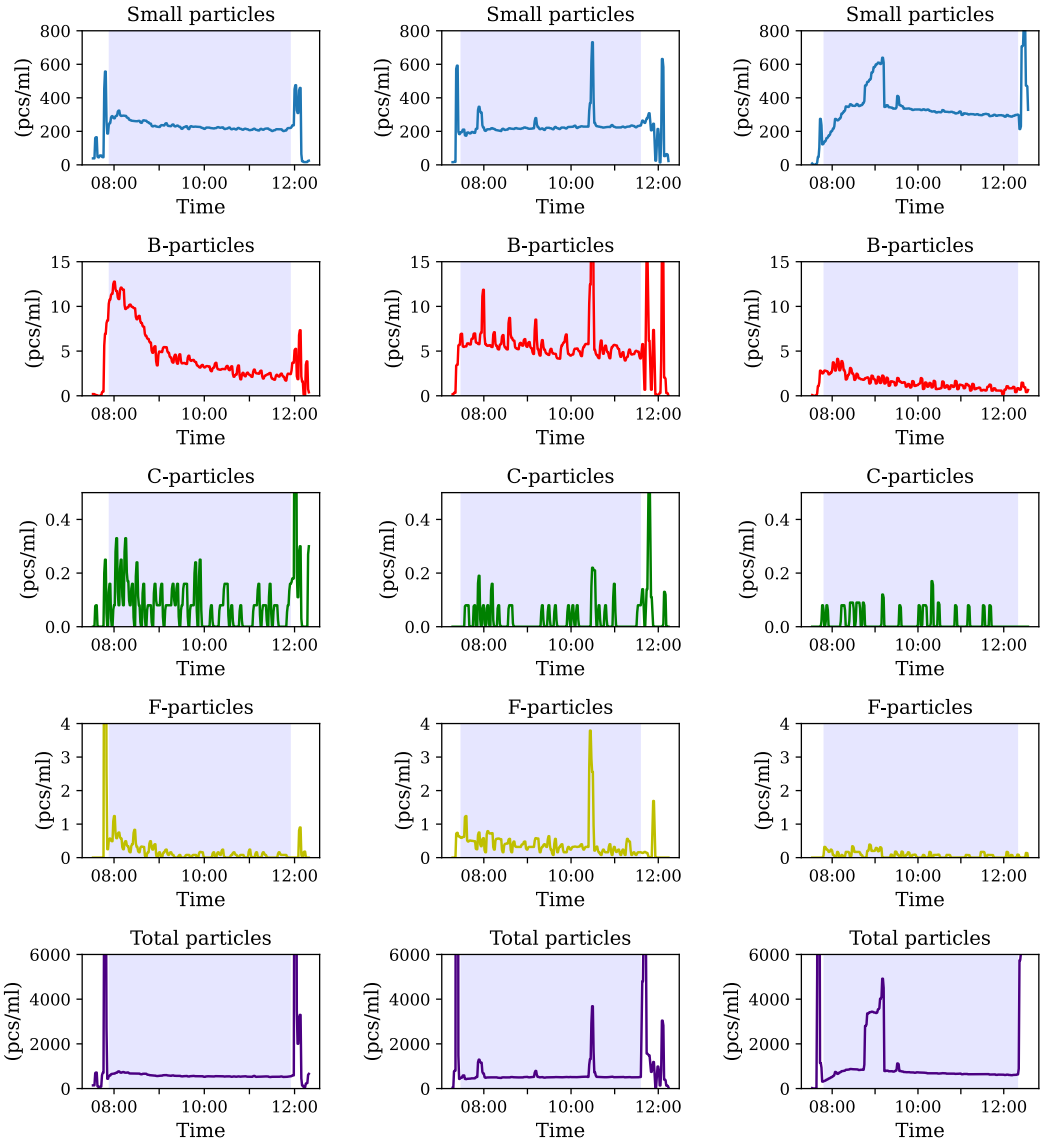
3.2.5 Cyanobacteria (*Synechocystis* PCC 6803)

The levels of Total, Small, B-, C- and F-particles detected during the three cyanobacteria spiking experiments can be seen in Figure 14, where (a) shows replicate experiment 1, (b) replicate 2, and (c) replicate 3. The Total and Small particles were quite similar in number during each of the three cyanobacteria experiments, and their profiles rather stable. There were however a few sudden spikes in experiment 2, as well as one prolonged increase observed during experiment 3. The B-particle levels were less uniform between the replicates. In experiment 1, the B-particle levels reached around 12 pcs/mL to then decrease over the course of the experiment. In experiment 2, the levels were around 5 pcs/mL during the entire test. In experiment 3 the levels were lower, starting at around 2.5 pcs/mL and decreasing over time. The C- and F-particles were always very low during all three experiment. The only exception was a notable peak of F-particles during experiment 2.

When testing the cyanobacteria spiking mixtures in the FCM, it was not possible to detect any cells. A dilution series of the *Synechocystis* PCC 6803 was tested with concentrations of 10^6, 10^5, 10^4, and 10^3 cells/mL. The FCM could detect each of these concentrations, although the 10^3 cells/mL gave a very low result. This could explain why the spiking mixture gave no output, since it was diluted to be in the 10^2 cells/mL range. As a test, the cyanobacteria were also stained with SGI. However, that yielded unreasonably high results.

Table 6: The results of testing the cyanobacteria spiking mixtures in the flow cytometer at Görvälnverket.

	Average TCC/mL
10^6 cells/ml	1007600
10^5 cells/ml	88880
10^4 cells/ml	7780
10^3 cells/ml	840



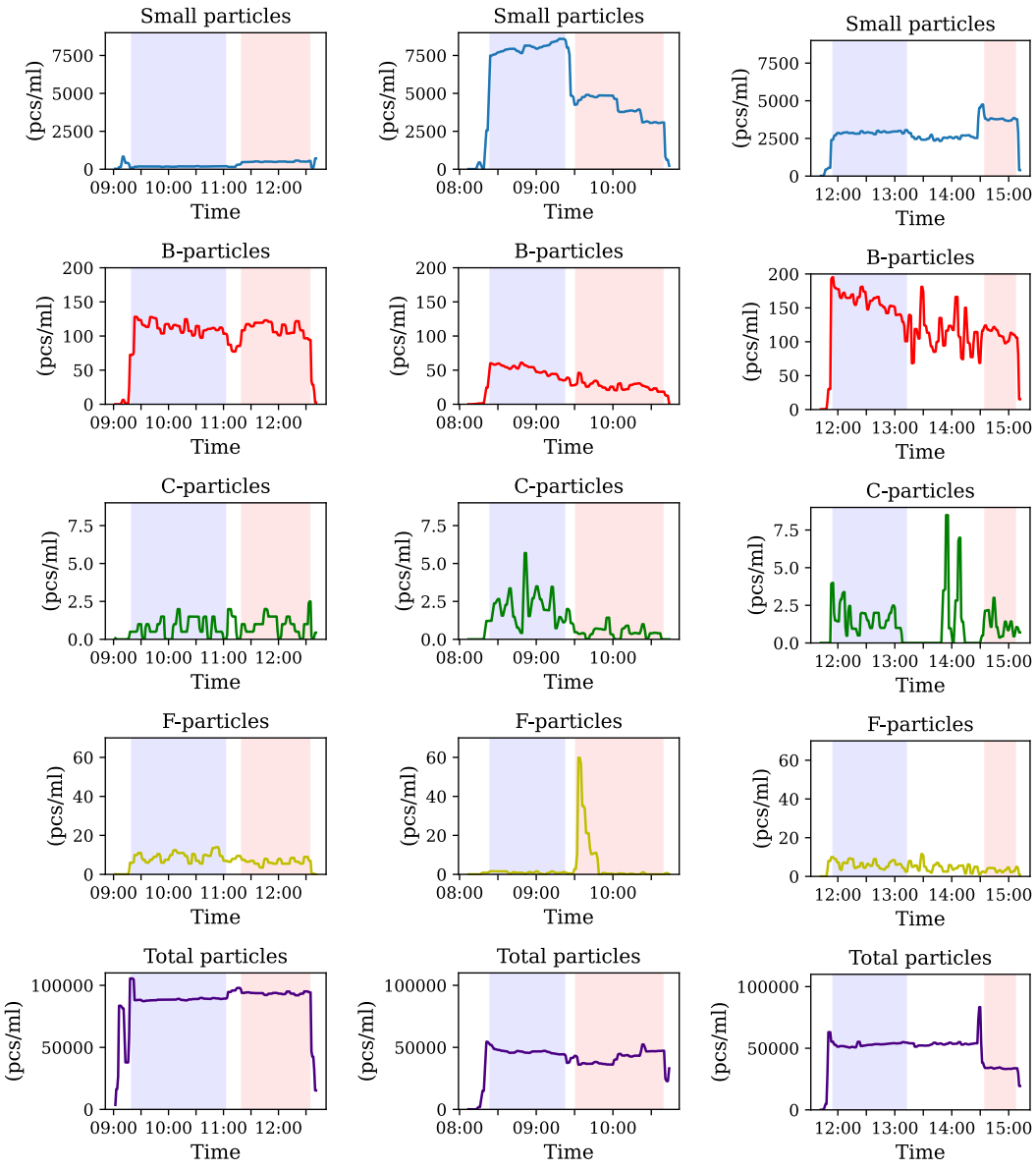
(a) Experiment 1 **(b)** Experiment 2 **(c)** Experiment 3

Figure 14: (a)-(c) show the Small, B-, C-, F-, and Total particle levels detected during the three replicate cyanobacteria spiking experiments. The blue fields show the duration of the experiments. A cell count of 10^2 cells/mL was tested. Note that outlier y-values have been cut off in some graphs to make the overall trends easier to see.

3.2.6 Biofilm

The levels of Total, Small, B-, C- and F-particles detected during the three biofilm spiking experiments can be seen in Figure 15, where (a) shows replicate experiment 1, (b) replicate 2, and (c) replicate 3. The Total particles were quite similar in experiment 2 and 3, and substantially higher in experiment 1. Small particles differed substantially in number between all three experiments. The Small particles increased in number in experiments 1 and 3 despite the fact that water was added, while they decreased in experiment 2. The B-particle levels differed between each experiment, particularly when comparing experiment 2 with experiments 1 and 3. In the second experiment, the B-particles only reached levels of around 60 pcs/mL, while they were over 100 pcs/mL, even approaching

200 pcs/mL in the first and third experiments respectively. In experiments 2 and 3 it was possible to discern the dilution of the biofilm in the B-particles levels. The F-particles increased noticeably during experiments 1 and 3, while they were lower with only one distinct peak during experiment 2. The C-particles were lower but did fluctuate around a few pcs/mL during each replicate.



(a) Experiment 1 **(b)** Experiment 2 **(c)** Experiment 3

Figure 15: (a)-(c) show the Small, B-, C-, F-, and Total particle levels detected during the three replicate biofilm spiking experiments. The blue fields show the duration of the higher biofilm concentration, while the red fields show the duration of testing once water had been added and the biofilm was diluted.

4 Discussion

Analysing the data collected by the N1, N2, and N3 instruments, it does appear like the Qumo instruments are able to detect seasonal variations in water composition. It is particularly obvious for the Small and Total categories. This was further confirmed when comparing with other data available for the project, the FCM measurements and algae counts. The lake turnover in spring resulted in significant increases in particles detected by Qumo, as well as algae numbers and total cell counts detected by FCM. However, the autumn lake turnover was less evident in Qumo data as compared to the other analyses, for example when comparing the Small particle levels in the water tower with the BactoSense measurements at the same site in Figure 24 in Appendix 8.2.2. It could also be seen in the seasonal averages that the Total and Small particle levels detected by Qumo were higher in summer than in autumn for some of the time periods analysed. The discrepancy between the different analysis methods may be caused by the fact that they do not measure exactly the same parameters, even though there are overlaps.

The B-, C-, and F-particle classes did not show obvious seasonal trends. This was particularly true for C- and F-particles which makes them good candidates for an early warning system, since their signals would not have to be discerned from background noise to provide a warning. This is something which has been confirmed in a prior study as well [25]. However, an important result from analysing the historical data was the heightened levels of B- and F-particles detected by the N1 instrument during late 2021. The N1 instrument was located directly after the treatment process at Görvälnverket, and there were no known quality deficiencies in the water during the time period. This is an indication that the Qumo models encountered particles which were not present in the training data and classified them as B- and F-particles, even though these particles were evidently not harmful. According to an Uponor representative, these heightened levels of B- and F-particles could be because of algae or cyanobacteria, which due to their variety in shapes and sizes can be classified in different ways. The Qumo models were trained on data from treated water in which B-, C-, and F-particles were sparse, which is why it is stated that they should be absent from clean drinking water. At Norrvatten it appears like some harmless algae can pass through the DWTP and thus be present in the treated water [28]. If the composition of the water at Norrvatten can give rise to these unexpected classifications, it raises questions about the usefulness of Qumo as an EWS there. An EWS should ideally provide rapid warnings of deteriorating quality, and if such a warning will always requires additional analyses in the lab, the speed will be lost.

Out of the three baseline methods tested, the two static approaches are most plausible to use in the Qumo software at present. For C- and F-particles, a static baseline approach appear to be a reasonable alternative, and may not even have to be updated according to seasonal fluctuations. B-particles seem a little more sensitive to fluctuations, but could perhaps also be compatible with a static baseline. The harmless algae and cyanobacteria being classified as F- or B-particles are a complication for designing a baseline for those categories, however. For Total and Small particles, calculating an alarm limit for each season based on older data will likely be a more reliable approach than using the same limit throughout the year. A static baseline could mean many false alarms during the lake turnover, and undetected real contamination events during winter. A moving window ap-

proach on the other hand seem to follow the trends better and would possibly be the most reliable alternative in terms of probability of detecting unusual events, but would require continuously updating the alarm limits in the Qumo software. Thus, it is not as useful for a drinking water producer at the moment, where resources may not be available for such maintenance requirements. An EWS would ideally mean less work for the operators, not more. Also, it can be seen when applying the dynamic baseline on the N2 data in Figure 26 in Appendix 8.3, that the historical data is not always a reliable approximation of the particle levels of the upcoming year.

The results of the lab experiments can be discussed in the light of how the particle classifications are described by Uponor. All contaminants tested in the spiking experiments were detected by the instrument, even the lowest bacterial concentrations with only a few hundred cells/mL. As expected, the bacterial tests of *E. coli* and *B. megaterium* resulted in levels of Small particles which made up a large fraction of the Total particles detected in the experiments. This corresponds well with their known sizes, and the range of the Small particles class of 1-3 µm. Increases of B-, C-, and F-particles during the *E. coli* and *B. megaterium* experiments were not consistent for each replicate and thus less reliable for analysis. Most consistent throughout the tests of both organisms was that the C-particles were low in all tests. The cell counts differed noticeably between replicates for both *E. coli* and *B. megaterium*, which was confirmed by the FCM measurements. This is possibly due to human errors, particularly when diluting the cell cultures to the low cell numbers desired. It difficult to replicate cell numbers in repeat experiments. The cell cultures were also measured by optical density, which is an inexact technique. Additionally, the cells may have formed lumps or chains formations, giving rise to differing particle classifications.

Analysing the humic acid results, it is clear that at the concentration tested, there were thousands of particles detected, out of which the great majority were not classified as either B-, C-, or F-particles, the ones that are supposed to indicate contamination. Had a much lower amount of humic acids been tested, it is likely that the levels of B-, C-, and F-particles would have been too low to be significant, even if the Total and Small particle levels would still be high enough to be detected. In other words, the greater part of the humic acid particles seem to be categorised as Normal particles. This is in line with the fact that humic acids are not in themselves harmful to consume.

The cyanobacteria also resulted in quite a large fraction of Small particles out of the total number. Interestingly, when testing the cyanobacteria spiking mixture in the FCM, the cell count was too low to be detected, indicating that the Qumo instrument is quite sensitive even at very low cell concentrations. However, it must be considered that the background noise was sterilised deionised water, which is far from the reality when monitoring the water quality of drinking water.

When testing the biofilm, a rise in B-particles was expected. This could indeed be observed for all three replicates, serving as confirmation of the claims stated by Qumo regarding the B-particle category. There were much higher values of Small particles during replicates 2 and 3, which is hypothesised to be because of unsufficient rinsing before the biofilm was detached. In other words, there were likely residual cells left in the liquid, detected by Qumo as Small particles. The B-particle levels also fluctuated between each

replicate, likely because of the biofilm detachment method which made it difficult to accurately repeat the process each time.

A difficulty with the instrument is that it is not possible to discern what is what out of the particles it detects. As was shown in the spiking experiments, there were increases in certain categories regardless of which contaminant and what kind of particles were detected. For example, all bacteria spiking tests gave increases in Small particles as expected, but so did the humic acid. Since microbial contamination has been the cause of the major contamination accidents from drinking water in Sweden, and those organisms can be present at such low levels, it seems uncertain whether the Qumo instrument can provide enough benefits to be worth the investment with its present abilities. Despite the sensitivity displayed by the Qumo as compared to the FCM, the difficulty in establishing reliable baselines for the particle categories, and the unexpected classifications of particles, makes it complicated for an operator to draw benefit from that sensitivity. Even if the Qumo can detect minuscule concentrations of bacteria in the water, if that bacterial presence cannot be discerned from the background noise they will go undetected nonetheless.

In conclusion, with its great sensitivity, and the classifications of C- and F-particles which are independent from normal quality fluctuations, the Qumo instrument has characteristics which are wanted in an early warning system. However, since it can make unexpected classifications of particles at Norrvatten, its usefulness as an online early warning system there is limited in its present state.

5 Future perspectives

With many types of particles being of interest in drinking water, an instrument that can detect a multitude of compounds could provide great aid for the detection and prevention of accidents. For this purpose, using artificial intelligence and machine learning appears promising. There are ideas to develop the Qumo instrument further, to enable analysis of particles smaller than 1 µm, and introduce new categories of particle classifications. The model could be trained to recognise cyanobacteria and algae and assign them to new categories [28]. Perhaps an instrument based on machine learning, Qumo or future alternatives, could be trained to characterise the water quality at the specific site where it is intended to function, to avoid contradictory classifications like the ones observed in this project.

In the future, an EWS might consist of instruments with in-built software for baseline constructions and normal fluctuations filtering, so that the operators would not have to analyse the data and set alarm limits on their own. Such advanced baseline algorithms could for example be based on adaptive algorithms, which have an advantage in that they continuously update the alarm detection based on the incoming data, forming a more dynamic monitoring capacity [34].

In an even bigger perspective, the systems could be integrated, where the data from many instruments are used together to detect changes in water quality. For example, hydraulic changes resulting in detached particles could then be disregarded as “anormal fluctuations”, and so on [34].

6 Acknowledgements

I would like to thank Linda and Mikael for being so welcoming and positive. You made every visit at Norrvatten a delight, and the project feel less daunting. Linda, thanks for dropping me off at the bus stop every time, and for helping me get that job! Thank you Guna, for tutoring me and for dedicating time to this project. Your vast knowledge brought many new insights to my work. Thank you to Esa Hämäläinen for sharing your expertise about the Qumo instrument, and answering so quickly to my emails! Your help was invaluable. To personnel at Norrvatten and the floor 1 lab at AlbaNova, thank you for your helpful and friendly attitudes. To the Department of Protein Science at KTH, thank you for donating the cyanobacteria. Thank you Emma, for the company in the lab and on the way to Norrvatten. You never fail to make me laugh. Jennifer and William – I am so glad you helped me with the big, scary autoclave! Yosef, thank you for inspiring me to study biotechnology in the first place, and for being the most wonderful friend one could ever wish for. 8 years and counting!

Och sist, till min allra käraste person: tack för allt. För ditt tålamod, din värme, din matlagning, och oändligt många andra saker jag skulle vilja säga. Potatoes and molasses.

7 References

[1] United Nations, "Clean water and sanitation." Available at: https://www.un.org/sustainabledevelopment/water-and-sanitation/. Accessed: 2023-02-27.

[2] L. G. Valdiviezo Gonzales, F. F. García Ávila, R. J. Cabello Torres, C. A. Castañeda Olivera, and E. A. Alfaro Paredes, "Scientometric study of drinking water treatments technologies: Present and future challenges," *Cogent Engineering*, vol. 8, no. 1, p. 1929046, 2021.

[3] S. Dejus, A. Nescerecka, G. Kurcalts, and T. Juhna, "Detection of drinking water contamination event with Mahalanobis distance method, using on-line monitoring sensors and manual measurement data," *Water Supply*, vol. 18, no. 6, pp. 2133–2141, 2018.

[4] Folkhälsomyndigheten, "Exempel på dricksvattenburna utbrott i Sverige." Available at: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/smittsamma-sjukdomar/vattenburen-smitta/exempel-pa-dricksvattenburna-utbrott-i-sverige/. Accessed: 2023-04-10.

[5] Norrvatten, "Reningsprocessen." Available at: https://www.norrvatten.se/om-norrvatten/. Accessed: 2023-02-27.

[6] Norrvatten, "Om Norrvatten." Available at: https://www.norrvatten.se/dricksvatten/dricksvattenproduktion/reningsprocessen/. Accessed: 2023-05-20.

[7] S. Sharma and A. Bhattacharya, "Drinking water contamination and treatment techniques," *Applied water science*, vol. 7, no. 3, pp. 1043–1067, 2017.

[8] S. Zgheib, R. Moilleron, M. Saad, and G. Chebbo, "Partition of pollution between dissolved and particulate phases: What about emerging substances in urban stormwater catchments?," *Water Research*, vol. 45, no. 2, pp. 913–925, 2011.

[9] Folkhälsomyndigheten, "Sjukdomsinformation om vattenburna infektioner och utbrott." Available at: https://www.folkhalsomyndigheten.se/smittskydd-beredskap/smittsamma-sjukdomar/vattenburen-smitta/. Accessed: 2023-04-10.

[10] M. J. Focazio, D. W. Kolpin, K. K. Barnes, E. T. Furlong, M. T. Meyer, S. D. Zaugg, L. B. Barber, and M. E. Thurman, "A national reconnaissance for pharmaceuticals and other organic wastewater contaminants in the United States—II) Untreated drinking water sources," *Science of the total Environment*, vol. 402, no. 2-3, pp. 201–216, 2008.

[11] I. Douterelo, R. Sharpe, and J. Boxall, "Bacterial community dynamics during the early stages of biofilm formation in a chlorinated experimental drinking water distribution system: implications for drinking water discolouration," *Journal of applied microbiology*, vol. 117, no. 1, pp. 286–301, 2014.

[12] H. Jang, R. Rusconi, and R. Stocker, "Biofilm disruption by an air bubble reveals heterogeneous age-dependent detachment patterns dictated by initial extracellular matrix distribution," *npj Biofilms and Microbiomes*, vol. 3, no. 1, p. 6, 2017.

[13] M. Hayes, "Humic substances: Progress towards more realistic concepts of structures," in *Humic Substances*, pp. 1–27, United Kingdom: Elsevier Ltd, 1998.

[14] A. D. Nikolaou, S. K. Golfinopoulos, T. D. Lekkas, and M. N. Kostopoulou, "DBP levels in chlorinated drinking water: effect of humic substances," *Environmental monitoring and assessment*, vol. 93, pp. 301–319, 2004.

[15] Livsmedelsverket, "Indikatorparametrar." Available at: https://kontrollwiki.livsmedelsverket.se/artikel/792/indikatorparametrar#koliforma-bakterier. Accessed: 2023-05-24.

[16] M. Y. Cheung, S. Liang, and J. Lee, "Toxin-producing cyanobacteria in freshwater: A review of the problems, impact on drinking water safety, and efforts for protecting public health," *Journal of Microbiology*, vol. 51, pp. 1–10, 2013.

[17] Livsmedelsverket, "Dricksvattenproduktion." Available at: https://www.livsmedelsverket.se/foretagande-regler-kontroll/dricksvattenproduktion. Accessed: 2023-05-03.

[18] Livsmedelsverket, "Dricksvattenkvalitet." Available at: https://www.livsmedelsverket.se/livsmedel-och-innehall/dricksvatten/dricksvattenkvalitet. Accessed: 2023-05-03.

[19] Livsmedelsverket, "Parametrar för mikroorganismer." Available at: https://kontrollwiki.livsmedelsverket.se/artikel/379/parametrar-for-mikroorganismer. Accessed: 2023-05-24.

[20] S. Van Nevel, S. Koetzsch, C. R. Proctor, M. D. Besmer, E. I. Prest, J. S. Vrouwenvelder, A. Knezev, N. Boon, and F. Hammes, "Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring," *Water Research*, vol. 113, pp. 191–206, 2017.

[21] M. Danielsson, L. Holmer, D. Hellström, C. Schleich, A. Keucken, J. Barup, L. Meyer-Lind, S. Chan, T. Rosenqvist, C. J. Paul, and P. Rådström, "Mikrobiologisk analys för biostabilt dricksvatten. ledningsnät, vattentorn och monokloramin," Tech. Rep. 2022-9, Svenskt vatten, Stockholm, 2022.

[22] M. H. Banna, S. Imran, A. Francisque, H. Najjaran, R. Sadiq, M. Rodriguez, and M. Hoorfar, "Online Drinking Water Quality Monitoring: Review on Available and Emerging Technologies," *Critical Reviews in Environmental Science and Technology*, vol. 44, no. 12, pp. 1370–1421, 2014.

[23] J. Favere, F. Waegenaar, N. Boon, and B. De Gusseme, "Online microbial monitoring of drinking water: How do different techniques respond to contaminations in practice?," *Water Research*, vol. 202, p. 117387, 2021.

[24] D. Hou, X. Song, G. Zhang, H. Zhang, and H. Loaiciga, "An early warning and control system for urban, drinking water quality protection: China's experience," *Environmental Science and Pollution Research*, vol. 20, pp. 4496–4508, 2013.

[25] M. Koppanen, T. Kesti, M. Kokko, J. Rintala, and M. Palmroth, "An online flow-imaging particle counter and conventional water quality sensors detect drinking wa-

ter contamination in the presence of normal water quality fluctuations," *Water research (Oxford)*, vol. 213, pp. 118149–118149, 2022.

[26] AQUA-Q AB, "R3Water Technology Fact Sheet." Available at: https://aqua-q.se/wp-content/uploads/2013/07/R3Water-technology-fact-sheet_AQUATRACK_final_160128.pdf. Accessed: 2023-05-15.

[27] M. Sanz, M. Trusiak, J. Garcia, and V. Micó, "Variable zoom digital in-line holographic microscopy," *Optics and Lasers in Engineering*, vol. 127, p. 105939, 2020.

[28] E. Hämäläinen. Personal conversation, March 2023.

[29] Uponor, "User Handbook for Uponor Water Quality Measurement System UWQMS." Unpublished document, 2021.

[30] Uponor, "Improvements in Release 2.1 April 2021." Unpublished document, 2021.

[31] M. Koppanen, T. Kesti, J. Rintala, and M. Palmroth, "Can online particle counters and electrochemical sensors distinguish normal periodic and aperiodic drinking water quality fluctuations from contamination?," *Science of the Total Environment*, vol. 872, p. 162078, 2023.

[32] G. Kirillin and T. Shatwell, "Generalized scaling of seasonal thermal stratification in lakes," *Earth-Science Reviews*, vol. 161, pp. 179–190, 2016.

[33] Y. Yankova, S. Neuenschwander, O. Köster, and T. Posch, "Abrupt stop of deep water turnover with lake warming: Drastic consequences for algal primary producers," *Scientific Reports*, vol. 7, no. 1, p. 13770, 2017.

[34] P. Jonsson, D. Lindgren, M. Asplund, H. Stavklint, M. Magounakis, S. Mokhlesi, T. Pettersson, M. Eriksson, F. Winquist, C. Jonasson, D. Ilver, and N. Strömbeck, "Elektronisk tunga och andra onlinesensorer för detektion av föroreningar i dricksvattennätet," Tech. Rep. 2018-15, Svenskt vatten, Stockholm, 2018.

[35] SMHI, "Årstider." Available at: https://www.smhi.se/kunskapsbanken/meteorologi/arstider/arstider-1.1082. Accessed: 2023-03-15.

[36] E. Zivot, J. Wang, E. Zivot, and J. Wang, "Rolling Analysis of Time Series," *Modeling financial time series with S-Plus®*, pp. 299–346, 2003.

[37] T. Habtewold, L. Duchateau, and G. K. Christophides, "Flow cytometry analysis of the microbiota associated with the midguts of vector mosquitoes," *Parasites & vectors*, vol. 9, pp. 1–10, 2016.

[38] V. T. Bowen and H. L. Volchok, "Spiked sample standards; Their uses and disadvantages in analytical quality control," *Environment International*, vol. 3, no. 5, pp. 365–376, 1980.

[39] D. Shiomi, H. Mori, and H. Niki, "Genetic mechanism regulating bacterial cell shape and metabolism," *Communicative & integrative biology*, vol. 2, no. 3, pp. 219–220, 2009.

[40] M. Eppinger, B. Bunk, M. A. Johns, J. N. Edirisinghe, K. K. Kutumbaka, S. S. Koenig, H. Huot Creasy, M. Rosovitz, D. R. Riley, S. Daugherty, *et al.*, "Genome se-

quences of the biotechnologically important Bacillus megaterium strains QM B1551 and DSM319," *Journal of bacteriology*, vol. 193, no. 16, pp. 4199–4213, 2011.

[41] Bionumbers, "Cell diameter." Available at: https://bionumbers.hms.harvard.edu/bionumber.aspx?s=n&v=3&id=114564. Accessed: 2023-05-22.

[42] T. Zavřel, P. Očenášová, and J. Červenỳ, "Phenotypic characterization of *synechocystis* sp. PCC 6803 substrains reveals differences in sensitivity to abiotic stress," *PLoS One*, vol. 12, no. 12, p. e0189130, 2017.

8 Appendices

8.1 Flow cytometry protocol

The SYBR Green I solution must be diluted to a concentration of 100X with dimethyl sulfoxide (DSMO) before use.

To prepare the Propidium iodide staining solution, 250 µL of the SYBR Green I 100X solution is added to 50 µL of Propidium iodide (1 mg/mL), giving a Propidium iodide + SYBR Green I solution.

8.1.1 Staining with SYBR Green I

495 µL of the sample to be analysed is added in triplicate to 1.5 mL Eppendorf tubes. 5 µL of the 100X SYBR Green I solution is added to each replicate sample, giving a final SYBR Green I concentration of 1X. All tubes are vortexed before incubation in 37°C for 15 minutes. Before analysis in the flow cytometer, the tubes are vortexed once again.

8.1.2 Staining with Propidium iodide + SYBR Green I

494 µL of the sample to be analysed is added in triplicate to 1.5 mL Eppendorf tubes. 6 µL of the Propidium iodide + SYBR Green I solution is added to each replicate water sample, giving a final SYBR Green I concentration of 1X, and Propidium iodide of 0.3 mM. All tubes are vortexed before incubation in 37°C for 15 minutes. Before analysis in the flow cytometer, the tubes are vortexed once again.

8.1.3 Settings on the flow cytometer (BD Accuri C6 Plus)

1. Run limits: 50 µL

2. Fluidics: Medium (35 µL/min)

3. Threshold: FL1:500

8.2 Trends in data

8.2.1 Qumo

Table 7: The means of Total and Small particles detected by N1 during each analysed season of 2021 and 2022.

	Total particles (pcs/mL)	**Small particles (pcs/mL)**
Spring 2021	1416.29	392.50
Summer 2021	922.19	261.60
Autumn 2021	916.00	304.14
Winter 2021	533.37	146.62
Spring 2022	1183.23	352.03
Summer 2022	1012.91	339.25
Autumn 2022	976.28	402.39

Table 8: The means of Total and Small particles detected by N3 during each analysed season of 2021 and 2022. Note that the summer of 2021 and autumn of 2022 have been disregarded as motivated in the results section.

	Total particles (pcs/mL)	**Small particles (pcs/mL)**
Spring 2021	564.26	173.24
Autumn 2021	603.64	126.24
Winter 2021	392.83	68.21
Spring 2022	1180.31	182.17
Summer 2022	829.79	124.02

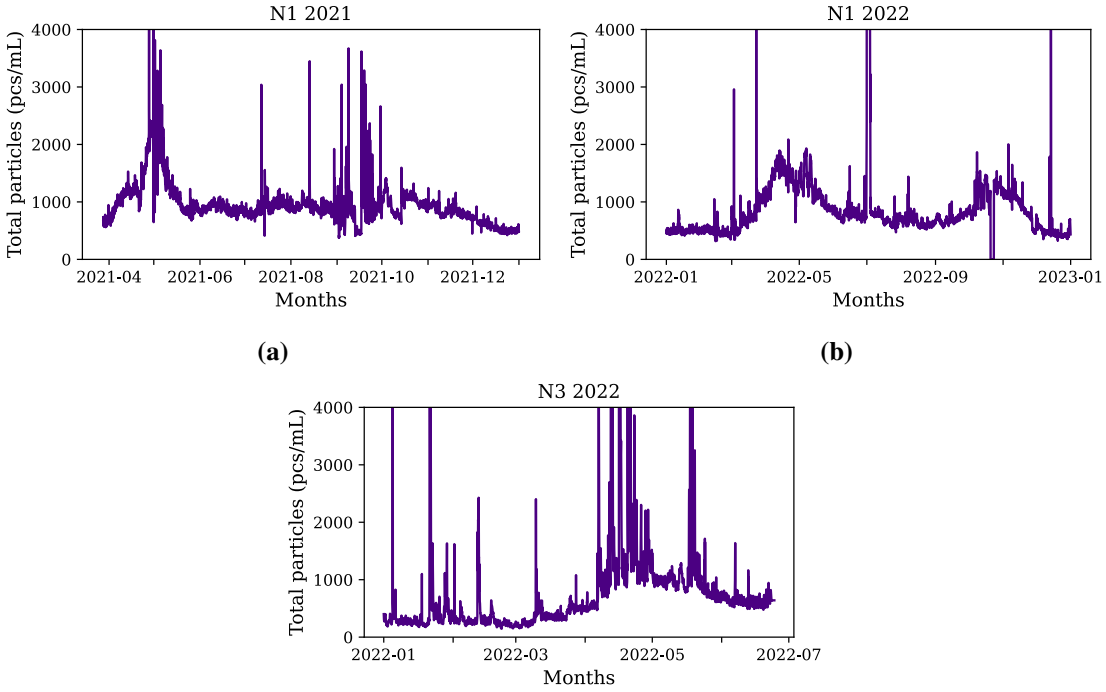


Figure 16: Total particles detected by the N1 instrument during the time periods (a) 2021-03-31 to 2022-01-01, and (b) 2022-01-01 to 2023-01-01. Note that the highest y-values have been cut off to make the graph easier to interpret.

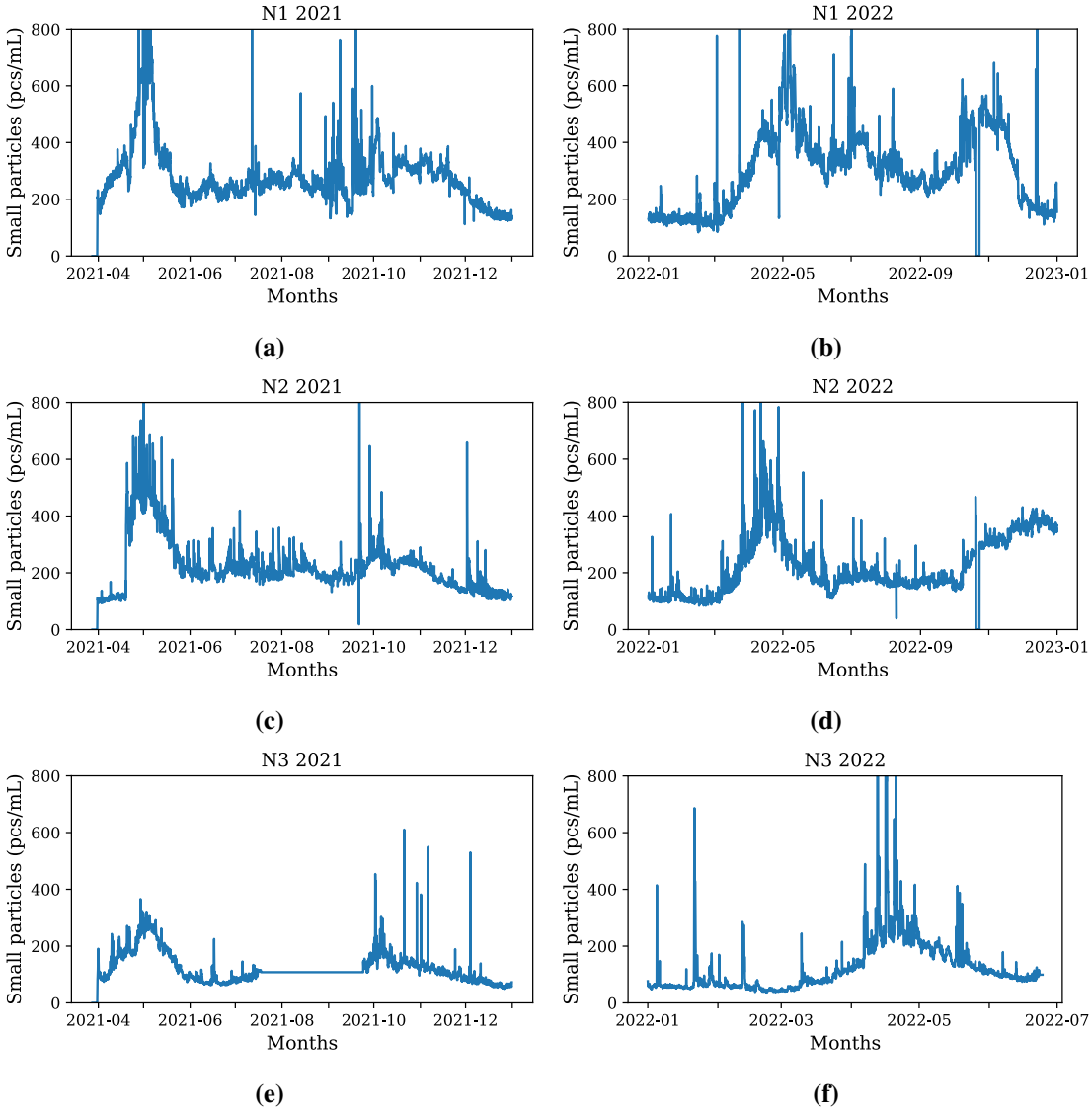


Figure 17: Small particles detected by N3 during the years (a) 2021 and (b) 2022. Note that the highest y-values have been cut off to make the trends easier to discern. The 2022 measurements end at the end of June, when the instrument was moved from the site at the pump station.

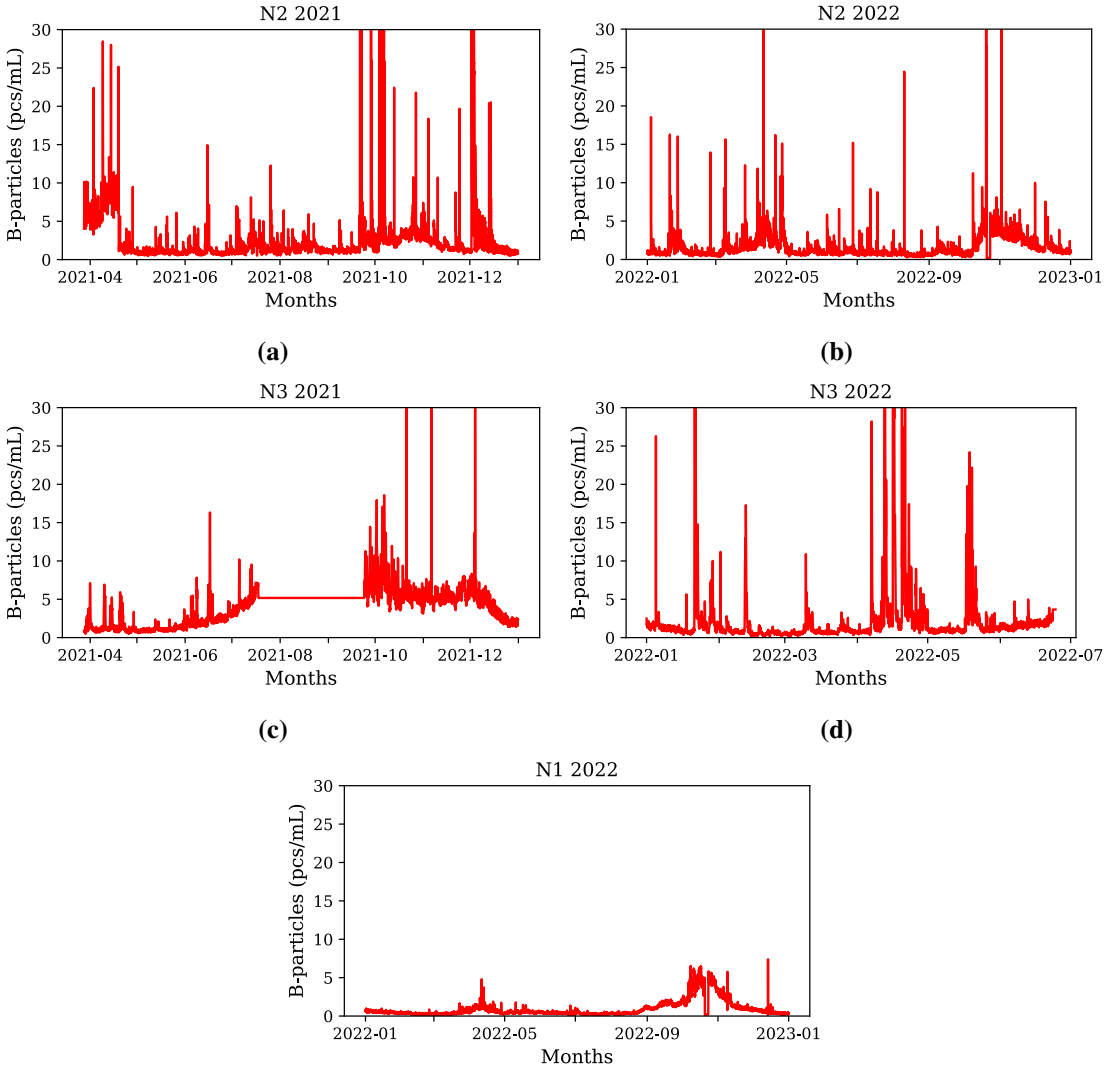


Figure 18: B-particles detected by (a) N2 2021, (b) N2 2022, (c) N3 2021, (d) N3 2022, (e) N1 2022. Note that the highest y-values have been cut off to make the trends easier to discern in (a)-(d). The 2022 measurements of N3 end in June, when the instrument was moved from the site at the pump station.

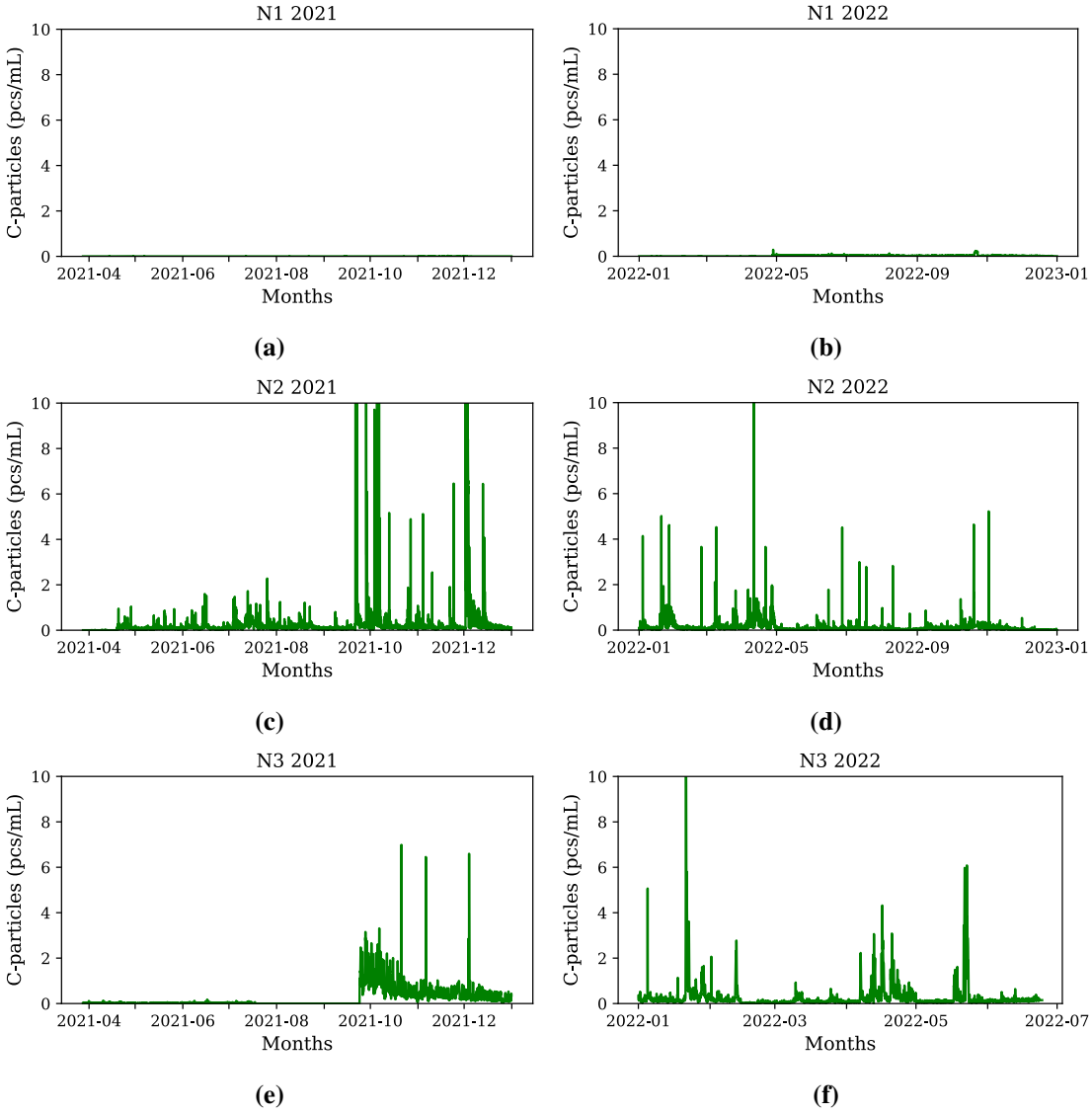


Figure 19: C-particles detected by (a) N1 2021, (b) N1 2022, (c) N2 2021, (d) N2 2022, (e) N3 2021, (f) N3 2022. Note that the highest y-values have been cut off to make the trends easier to discern. The 2022 measurements of N3 end in June, when the instrument was moved from the site at the pump station.

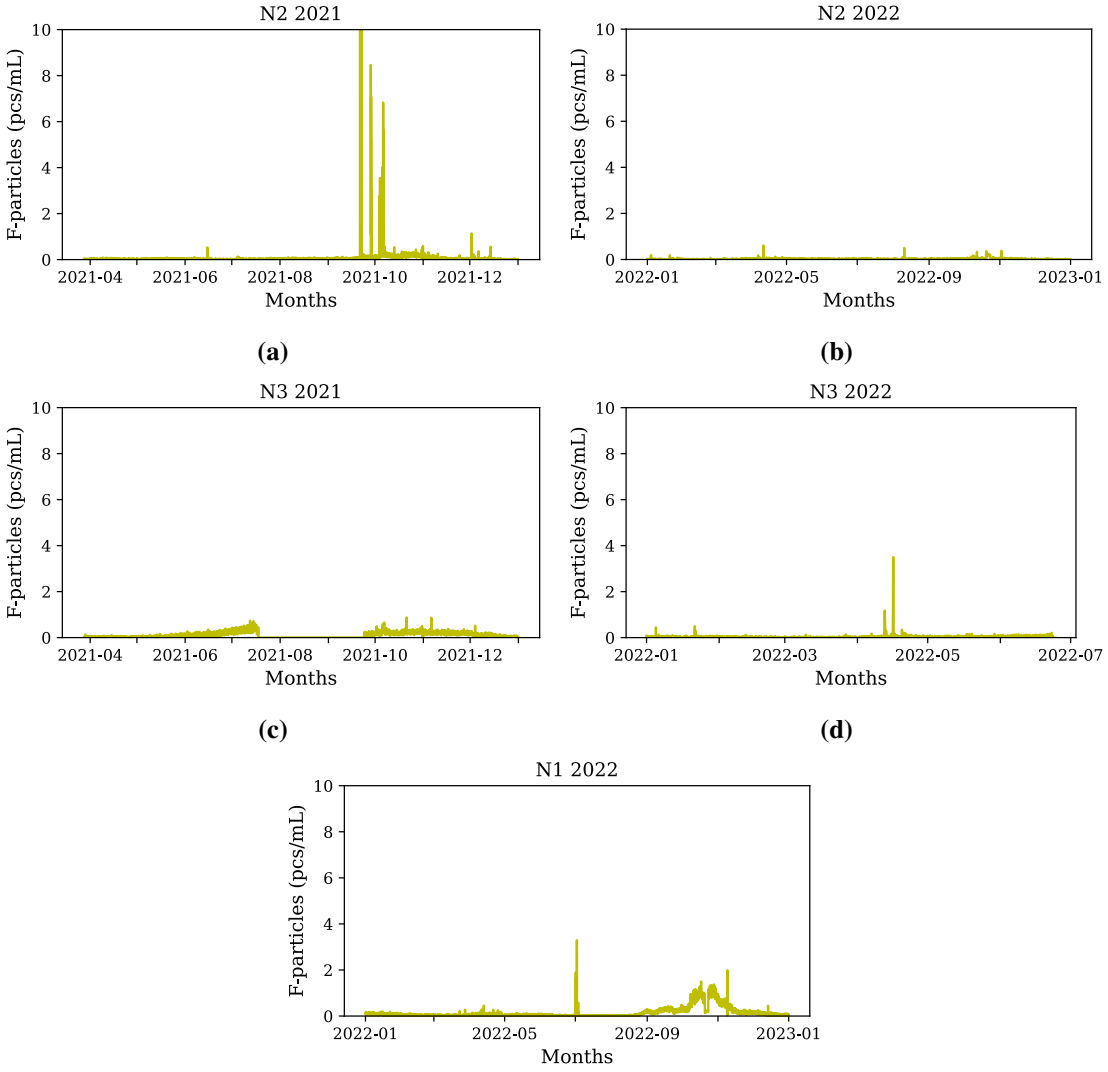
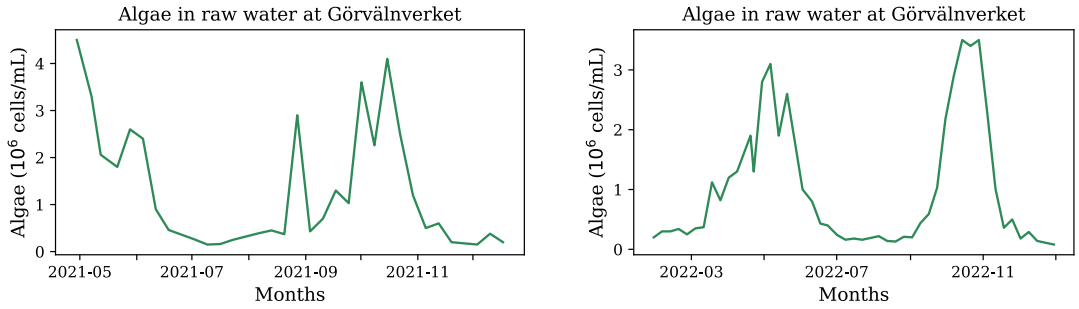


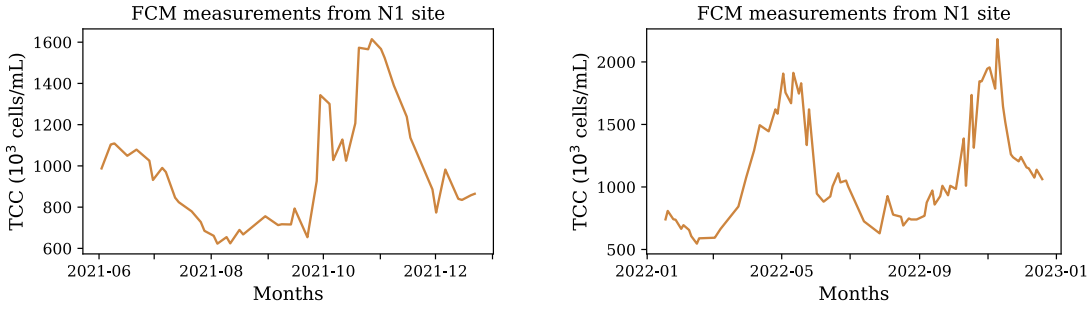
Figure 20: F-particles detected by (a) N2 2021, (b) N2 2022, (c) N3 2021, (d) N3 2022, (e) N1 2022. Note that the highest y-values have been cut off to make the trends easier to discern. The 2022 measurements of N3 end in June, when the instrument was moved from the site at the pump station.

8.2.2 Other instruments and analyses



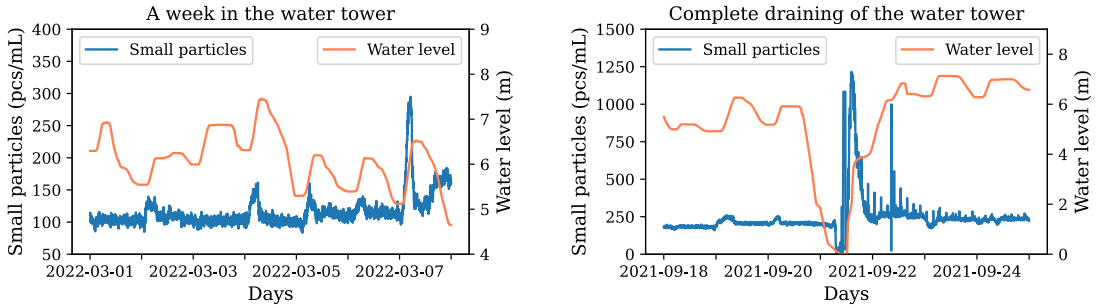
(a) **(b)**

Figure 21: Algae counts in the raw water at Görvälnverket during (a) 2021 and (b) 2022.



(a) **(b)**

Figure 22: FCM measurements from the lab at Görvälnverket during (a) 2021 and (b) 2022. Water samples taken from the same site at which N1 was located.



(a) **(b)**

Figure 23: (a) The Small particle levels detected by N2 during a week in March 2021, plotted together with the water level in the tower over the same time period. (b) The Small particle levels detected in conjunction with a complete draining of the water tower, as can bee seen by the measured water level. A very short flatline in the Qumo data can be seen just after the drain.

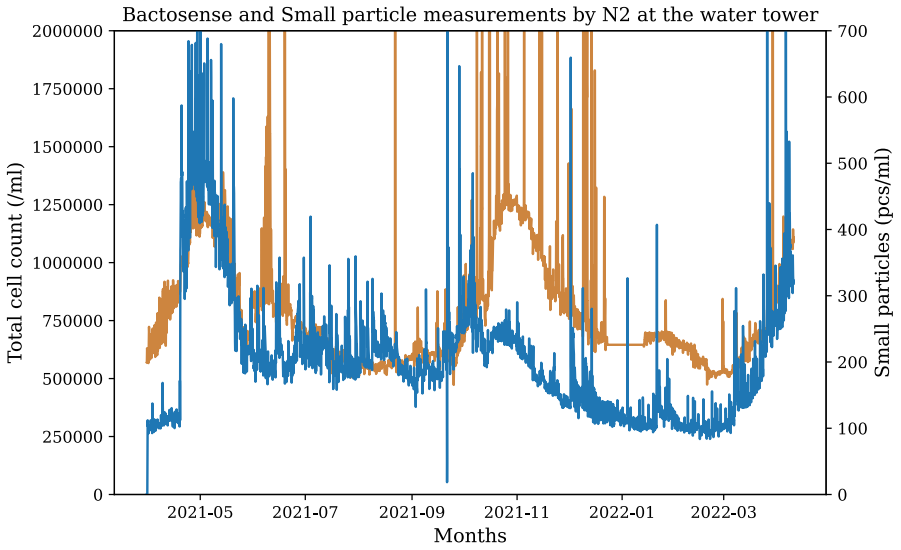
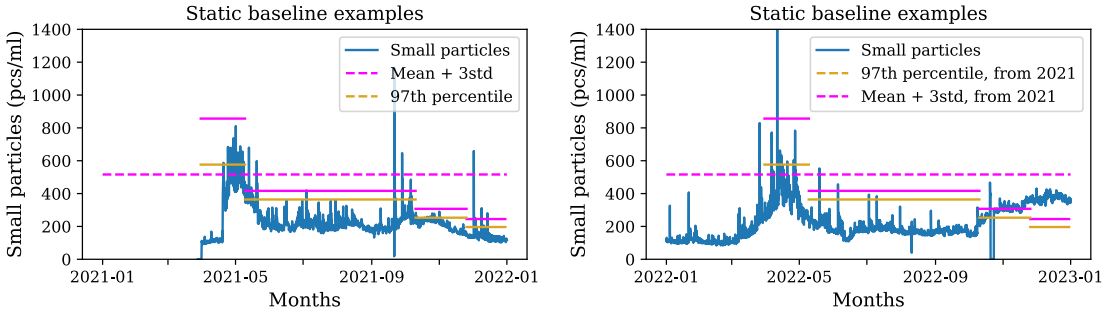


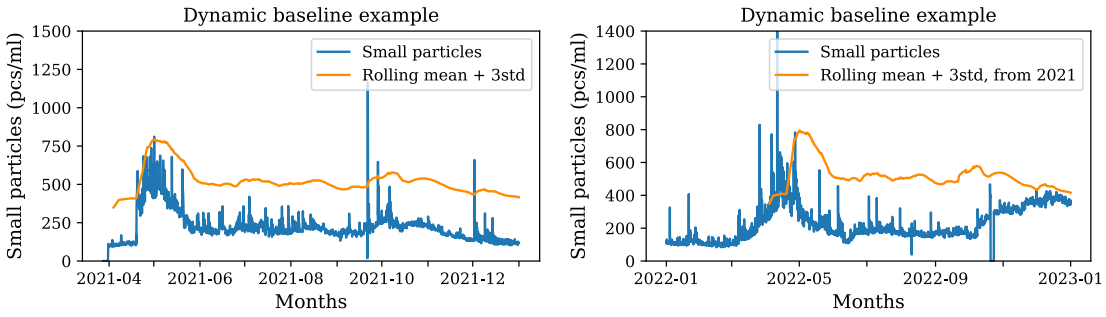
Figure 24: BactoSense measurements from the water tower plotted together with Small particles measured by N2 at the same site during the same time period. Note that the highest y-values have been cut off to make the trends easier to discern.

8.3 Baselines



(a) **(b)**

Figure 25: Examples of what the static baselines approaches could look like applied on real Small particles data collected by the N2 instrument. (a) The baselines calculated from 2021 data, plotted together with the data they were calculated from. (b) The same baselines calculated from 2021 data, plotted together with data from 2022 instead. Note that outlier y-values in (a) have been cut off to make the overall trend easier to see.



(a) **(b)**

Figure 26: Examples of what the moving window baseline approaches could look like applied on real Small particles data collected by the N2 instrument. (a) The baseline calculated from 2021 data, plotted together with the data it was calculated from. (b) The same baseline based on 2021 data plotted together with data from 2022 instead. Note that outlier y-values in (a) have been cut off to make the overall trend easier to see.