Exploring Flow Cytometry for Monitoring of Microbial Water Quality During Maintenance of Drinking Water Pipes

Master Thesis in Applied Microbiology

Lund University, Faculty of Engineering

Måns Zamore

2021-06-22

Abstract

Faster methods of analysing microbial drinking water quality after pipe breaks or new pipe installations would have beneficial implications. This thesis investigates the use of flow cytometry and cultivation-based techniques in connection to maintenance works, and how flow cytometry can aid in the decision-making around putting drinking water pipes in operation. Previous work has not used online flow cytometry to assess microbiological water quality after pipe breaks, nor has the influence of the biofilm formation process on bacteria in the bulk water been well understood. Here, online flow cytometry combined with novel computational methods, including principal coordinates analysis and calculations of the Bray-Curtis dissimilarity metric based on flow cytometric fingerprints, are used to yield nuanced microbiological information. The potential benefit of applying flow cytometry in connection to maintenance or pipe installations has been shown to be substantial.

Popular science summary (in Swedish)

Ny metod kan minska tiden för dricksvattenanalyser

Läckor i dricksvattenledningar kostar stora summor pengar och orsakar mycket besvär för samhället. Nya mikrobiologiska analysmetoder har visat sig kunna korta analystiden betydligt – utan att äventyra kvaliteten.

Mikrobiologiska analyser av dricksvattnet gör ofta att det dröjer länge innan vattnet i vattenledningarna återigen kan användas efter läckor och underhåll. Detta examensarbete har undersökt om en ny analysmetod, flödescytometri, är säker och om den i så fall också kan förkorta analystiden. Svaret är ja: flödescytometri kan vara en mycket lämplig metod för dricksvattenanalys.

När det sker störningar i dricksvattensystemet leder det till stora problem och kostnader för samhället. Många minns kokningsrekommendationen i Lund vintern 2017. Under sex veckor uppmanades invånarna att koka dricksvattnet, på grund av att de mikrobiologiska analyserna visade höga halter av en viss typ av bakterier. För VA SYD ledde det till skadeståndskrav på mångmiljonbelopp.

En av de största flaskhalsarna efter en reparation eller ett underhållsarbete i dricksvattenledningar är idag de mikrobiologiska analyserna. De odlingsbaserade metoderna, där en liten mängd dricksvatten gjuts in i en näringsrik lösning, kan avläsas först efter några dagar. Metoderna har tillämpats i årtionden, men är kanske inte längre de bästa alternativen. På senare år har nya mikrobiologiska analysmetoder utvecklats, och den mest lämpliga för dricksvattenanalys är antagligen flödescytometri – en teknik som med hjälp av laserstrålar och självlysande färg kan räkna, och i viss mån även särskilja, bakterierna i ett prov på bara 15 minuter.

En av utmaningarna med flödescytometri är att nyttja de stora datamängder som genereras. Metoden ger indirekt information om DNA-innehåll i varje bakterie i ett prov, men trots detta har den tidigare oftast använts för att enbart räkna bakterierna. I detta examensarbete har i stället de senaste beräkningsmetoderna som finns i litteraturen tillämpats på mätdata från automatiska flödescytometrar. På så vis har mikrobiologiska förändringar kunnat följas med stor precision. Resultaten visar att flödescytometri kan vara en mycket lämplig metod för att snabbt ta reda på om dricksvattnet är säkert.

Contents

1 Introduction

Pipes in the drinking water system need maintenance occasionally. There may be pipe breaks which require immediate action, or the maintenance can be planned to prevent deterioration. Another type of maintenance, or intervention, is when the pipe network expands to accommodate for new buildings. These events present several microbiological challenges. For instance, pathogenic microorganisms can enter the pipe during interventions. One way to prevent this from happening is to flush the pipes with large volumes of drinking water (Van Nevel et al. 2017a). This water ends up in the wastewater system and is not delivered to any consumers. The waste of water — as well as other factors relevant to drinking water distribution network interventions, such as reduced water supply, disturbed traffic conditions, and increased workload for personnel — leads to substantial costs both for water utilities and other parts of the society.

With the large costs associated with maintenance of water pipes, there is a large potential benefit if maintenance works can be completed faster. A major bottleneck is the time required for microbial analysis. Current microbiological methods require incubation for 1-7 days depending on the method and which microorganisms are of interest (SLVFS 2001). Therefore, faster methods of analysing drinking water have the potential of reducing the time of maintenance with several days. That is one of the focus areas of this thesis.

Another relevant microbiological aspect that could reduce the time spent on maintenance is a more in-depth understanding of the natural microbiological processes that occur in new or repaired pipes. It is well-established that drinking water pipes are lined with a layer of biofilm, usually formed by non-pathogenic microorganisms (Flemming 2002). New drinking water pipes will, however, not have a biofilm initially. This may affect the number of bacteria that are found in the water samples. Søborg et al. (2020) found indications that the number of bacteria measured by heterotrophic plate counts (HPC) in new pipes follows a specific pattern, which could be a result of the formation of biofilm. If such a pattern is generally occurring in new pipes, valuable conclusions about when to take samples in new pipes, timeframe of the work and what to expect could be drawn. That is the second focus area of this thesis.

This thesis is part of a project that investigates faster and safer maintenance operations in the drinking water distribution system. Involved water utilities are VA SYD and Sydvatten, and their joint research and development organisation Sweden Water Research. The data in this thesis has therefore been provided by VA SYD and Sydvatten. The project has received a grant from Svenskt Vatten Utveckling (SVU), and will continue for all of 2021.

The aims of this thesis are (i) to investigate whether faster microbiological analysis methods can be used after maintenance works or new pipe installations, and (ii) to examine the natural microbiological process that occurs in new pipes. The thesis consists of a literature study of the relevant subjects as

well as advanced analysis of collected microbiological data from maintenance works of VA SYD's and Sydvatten's drinking water distribution system in Skåne, Sweden.

1.1 Current methods of assessing drinking water quality

Current methods of assessing microbiological drinking water quality after maintenance or installations of new pipes are based on cultivating bacteria on agar plates. A number of analyses are required by regulations to be performed by water utilities as a part of their self-monitoring (SLVFS 2001). Some of these analyses have strict limits, where the water cannot be used if the limits are exceeded, while other analyses are designed to trigger investigations if the limits are exceeded. For instance, there are strict limits that include no detection of *Escherichia coli (E. coli)* or coliform bacteria per 100 ml of drinking water. These bacteria are referred to as indicator bacteria, due to their functioning as indicators for potentially pathogenic microorganisms. Additional bacteria in this category, which also are analysed, are *Enterococci* and *Clostridium*.

Limits that require investigation for exceeded limits include the that the HPC (a method requiring 3 days of incubation) should be less than 100 colony-forming units (cfu) per milliliter (SLVFS 2001). Additionally, the number of slow-growing bacteria (which requires 7 days of incubation) should be less than 5000 colony-forming units per milliliter. The presence of actinomycetes, yeasts, and mold is also analysed.

However, sampling at or after maintenance operations, or new pipe installations, is not classified as self-monitoring and therefore, the regulations do not apply in these instances. Rather, it is the water utilities' internal regulations that apply. In 2020, VA SYD decided not to perform analysis of slow-growing bacteria after pipe installations or maintenance operations, in order to save resources. The background to this decision is that an internal survey showed that in all cases where the slow-growing bacteria was above accepted limits for a certain sampling point, the HPC also exceeded the limits. Sydvatten, on the other hand, has not made a similar decision and therefore still performs analysis of slow-growing bacteria after maintenance operations.

Furthermore, it is not always that the microbiological water quality is tested after leaks and many minor pipe breaks are left undiscovered. There may not be any need for quality analysis after repairing minor pipe breaks, as long as there is no reason to believe that a contamination has occurred. Thus, it is up to the water utilities to decide if and how analyses are to be done after leaks.

1.2 Biofilms in pipes

Drinking water contains large amounts of bacteria, commonly in the range of 10^4 to 10^5 cells per milliliter (Hammes et al. 2008), usually not exceeding 5 x 10^5 cells per milliliter (Van Nevel et al. 2017b). Most of

the bacteria in the drinking water distribution system (DWDS) are not suspended in the bulk water, but rather attached to the pipe walls as biofilms (Liu et al. 2016). Liu et al. note that biofilms are notoriously ubiquitous and resistant towards measures of controlling bacterial growth, such as the use of disinfectants. A key component of the biofilm is the extracellular polymeric substances (EPS) that the microorganisms produce (Flemming 2002). The EPS resembles glue in that it enables cells to attach to each other and to surfaces (Liu et al. 2016).

1.2.1 Biofilm formation

The formation of biofilm can be divided into four stages, according to Hemdan et al. (2021): adhesion, microcolony/evolution, maturation, and detachment/dispersion. The adhesion step consists of a loose and reversible attachment of planktonic (in the swim-phase) microbial cells to the surfaces. The microcolony/evolution step is characterised by a more permanent adhesion of the microbial cells to the surface, with adhesive agents like lipopolysaccharides. Subsequently, the cells proliferate and produce EPS as a protective wrapping. The maturation stage includes communication with other bacteria in the biofilm through biochemical signals. The signalling system is called quorum sensing (QS) and enables group behaviour of all bacterial species involved in the biofilm. Finally, the detachment/dispersion occurs either due to various biochemical signals or due to physical forces from hydraulic changes near the biofilm surface. During this step bacteria are released from the biofilm.

The biofilm formation process in drinking water pipes has been investigated in an experimental drinking water distribution system (DWDS) by Douterelo et al. (2014). The study gives valuable information on the biofilm formation process in drinking water pipes, both on the timescales that are relevant and the microbiomes. Biofilm development stages were observed for 28 days using fluorescent microscopy and 16s rRNA sequencing. Initially, a few bacterial species colonised the pipe for the first seven days. These are primary colonisers that eventually enable other bacteria to attach. After seven days, other bacterial species — secondary colonisers — were inclined to attach to the biofilm. The authors further found that the coverage of the cells was low and difficult with the microscope to detect on day seven, but increased three-fold from day 14 to 28.

The influence of the pipe material on the microbial populations of the water and biofilm was the subject of another study (Learbuch et al. 2021). The study was performed under “worst-case” conditions, with the water being semi-stagnant and relatively warm (25 °C). Methods used were ATP concentration determination, 16s rRNA gene sequencing and qPCR. The results indicated that PVC-P had the most biofilm, PE having an intermediate amount of biofilm and PVC-C, glass and copper had the lowest amount of biofilm. The microbial communities of the biofilm also varied depending on the pipe material. The development of biofilm was also studied as a function of time. Both the measures of biomass and community composition indicate that the biofilm reaches a stable situation sometime between four and eight weeks after its formation. This study highlights relevant timescales and also

elucidates the complexity of the subject, due to the large impact the pipe material has. The effect of pipe materials was also investigated by Douterelo et al. (2016) who reported that the material of the pipes highly influenced the composition of the biofilm communities. Both cast iron and plastic pipes were shown to be suitable surfaces for biofilm formation, but the types of bacteria in the biofilms varied significantly. Surface roughness and flow hydrodynamics also impact the formation process (Cowle et al. 2020), with roughness of plastic materials having a significant effect on the biofilm formation, that is likely a larger effect than flow hydrodynamics.

A common practice among water utilities is to flush drinking water pipes with large amounts of water for an extended period after interventions (e.g., maintenance or installations) in the pipes. The effect of flushing on the biofilm was investigated by Douterelo et al. (2014), with results indicating that some bacterial species in the biofilm were effectively removed by flushing while others remained attached to the pipe walls. Douterelo et al. (2016) also investigated the effect of flushing on biofilms and biofilm regrowth after flushing. Drinking water pipes in operation, with an already existing biofilm, were studied. It was found that biofilms that grew after flushing of a pipe had a different community structure than the biofilms that existed before flushing (Douterelo et al. 2016). Two possible explanations are suggested, namely that seasonal variations in the source water affect the subsequent biofilm formation and that the parts of the biofilm that remain after flushing influence the continued biofilm regrowth.

1.3 Flow cytometry

Flow cytometry is a technique which has existed for decades but has begun being used for drinking water analysis since 2008 (Hammes et al. 2008). The technique lets sample pass into a narrow chamber where particles are aligned in a single file. Laser beams are then used to strike each particle, and detectors measure how the light is scattered. For drinking water analysis, the particles of interest are often bacteria. Flow cytometers usually utilise fluorescent stains to better be able to distinguish between various cells. Stains used for microbial analysis include SYBR Green I and Propidium iodide (PI), which fluoresce when they bind to double-stranded DNA (Safford and Bischel 2019). Live cells are impermeable to PI, which enables it to function as a marker for cells with broken membranes. Flow cytometers exist both as laboratory instruments and as online variants.

Flow cytometry has numerous advantages over the cultivation-based methods that are currently used for assessing water quality. Compared to HPC, the method is fast, sensitive, reproducible, and possible to automate (Van Nevel et al. 2017b). It is possible to reach a result after only 15 minutes rather than a number of days that are required for cultivation-based methods. The entire bacterial community in the water can be counted rather than the approximate one percent that is measured by HPC (Van Nevel et al. 2017b). However, flow cytometry has its limitations, including difficulties in clearly distinguishing between live and dead cells, that the data analysis is prone to subjective interpretations and that aggregates or clumps of bacteria are difficult to handle properly (Van Nevel et al. 2017b). The method

is yet to have gained regulatory status for drinking water quality assessment in any country except for Switzerland (Safford and Bischel 2019).

A common output of flow cytometers is scatter plots where various regions of interest can be discerned (fig. 1) (Gatza et al. 2013). The number of total cells can be quantified if using only SYBR Green I as staining, while the number of intact cells can be quantified if also using PI. These cells are counted using gates, which are regions of interest that are defined on the scatter plots from flow cytometers. Clouds within the gates that correspond to bacteria with low or high nucleic aid content (LNA or HNA, respectively) can be defined.

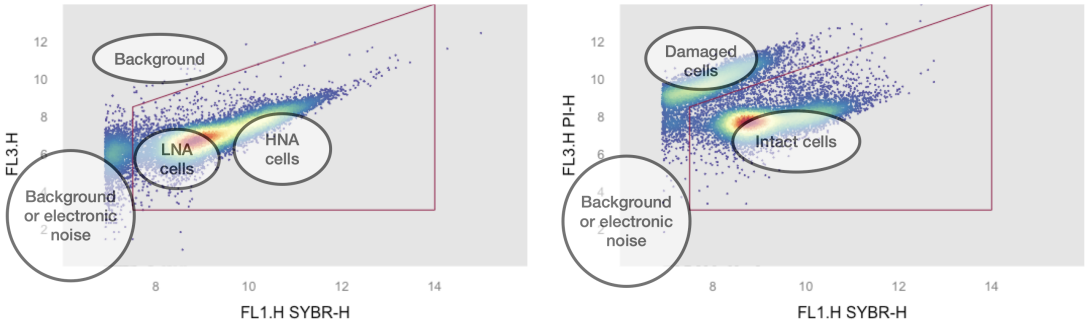

Figure 1: Schematic figure outlining relevant areas of flow cytometric scatter plots. To the left, a sample stained with only SYBR Green I is shown, enabling total cell counting. To the right, a sample stained with both SYBR Green I and Propidium iodide is shown, enabling counting of intact and cells with damaged membranes. FL1-H and FL3-H refer to two different fluorescence detectors.

Water utilities are interested in if flow cytometry can replace methods based on cultivation such as HPC. Studies have been made investigating correlations between HPC and flow cytometry metrics, for instance by Cheswick et al. (2019), Gabrielli et al. (2021), and by Van Nevel et al. (2017b). Cheswick et al. (2019) made a study with a large dataset spanning over three years and 213 water treatment works. No clear link was found between the detection of coliforms and the flow cytometry data. Note, however, that the study only investigated the total cell counts and intact cell counts of the flow cytometric data. More advanced computational methods incorporating all the information present in the scatter plot, so-called fingerprinting, may be more effective than these measures. The study by Gabrielli et al. (2021) used online flow cytometry for several months and compared the results with cultivation-based methods. The variations detected by flow cytometry did not correlate with the cultivation-based methods like the HPC. The authors speculate that the lack of correlations may be due to the HPC only detecting copiotrophic bacteria while flow cytometry detects the entire bacterial community. Van Nevel et al. (2017b) analysed a different large dataset, and found no correlations between flow cytometric cell counts and the HPC. Still, despite the lack of correlations with cultivation-based methods, Van Nevel et al. (2017b) argue that flow cytometry is a meaningful process variable and suitable to replace

HPC. The suggestion is to use the principle of no abnormal change, establishing baseline values of flow cytometry metrics and detecting when these change considerably.

A potential application of flow cytometry for water utilities is to determine when drinking water can be used after maintenance of the pipes. Van Nevel et al. (2017a) conducted a study using flow cytometry bacterial cell counts and fingerprinting for that purpose. Three major supply pipes were emptied, treated with chlorine and then flushed with drinking water. Comparisons were made between the flushed water and reference water, and the time-point of similarity in cell concentrations and fingerprints was deemed as a suitable endpoint for flushing. Notably, the authors based the flow cytometry fingerprinting on the raw data, without any gating, thereby including the background in the analysis of the fingerprints. This is argued to be an advantage since the background can also reflect components of water quality including organic particles (for example free DNA, viruses, fragments of cells) or inorganic particles (such as salt crystals or metal oxides). Taking the background into account is therefore expected to make the measurements more conservative. Background particles can be flushed out later than the bacteria. The results from the study indicated that flow cytometry can be used to determine when a pipe can be brought back into operation, giving more conservative yet faster results than cultivation-based methods.

1.4 Computational methods of analysing microbial flow cytometry data

Microbial flow cytometry and associated computational methods is a rapidly evolving area (Rubbens and Props 2021). Through advanced computational analysis, flow cytometry has been used to track changes in drinking water over time or in response to environmental conditions (Chan et al. 2018). Limited research has focused on comparisons of various computational methods of analysing microbial flow cytometry data. Steps included in typical microbial analysis of flow cytometry data are briefly described below.

1.4.1 Pre-processing

Microbial flow cytometry data has traditionally been transformed with logarithm functions (Rubbens and Props 2021). However, with the logarithm not being able to handle negative values, an alternative is the arcsine hyperbolic function. Functions for transforming flow cytometry data are available in the *flowCore* package (Ellis et al. 2021).

1.4.2 Quality control

There can be deviations during the flow cytometry data acquisition, owing to for instance large particles or air bubbles that end up trapped in the system. To account for these deviations, quality controls

packages such as *flowAI* (Monaco et al. 2016) can be used.

1.4.3 Removal of noise

An inherent characteristic of flow cytometers is that they measure electronic noise (Rubbens and Props 2021), and particles that may not be of interest are measured as well. To select only the cells in a flow cytometry analysis, gating is performed. Gating can also be performed with the *flowCore* package (Ellis et al. 2021).

1.4.4 Fingerprinting

Most advanced methods of analysing microbial flow cytometry data are based on a fingerprinting method (Rubbens and Props 2021). Fingerprinting is a way of dividing the data into many smaller regions, or bins. Cells can thereafter be counted in each region, which results in a higher-resolution and more nuanced model of the cells than merely the total cell count. Multiple fluorescence detectors can be included in the fingerprint, the optimal number was investigated by Rubbens et al. (2017). Various fingerprinting methods have been developed and there is a lack of research focused on comparing different fingerprinting methods. Regions of interest can be defined manually (Koch et al. 2013b). Equal-space binning is another alternative that makes a grid of the data with equal sizes of each region, and is the method utilised in the CHIC package (Koch et al. 2013a) as well as the *Phenoflow* package (Props et al. 2018). More advanced methods divide the cytometric data into regions of varying sizes and shapes (Rubbens et al. 2021b). Regions with larger cell densities are put into smaller grids.

1.4.5 Analysis of fingerprints

Post-fingerprinting, a variety of methods can be applied to the fingerprints to test for differences among samples. Methods include traditional multivariate methods such as PCA analysis, NMDS analysis (Rubbens and Props 2021) or methods that estimate ecological parameters such as alpha diversity (Props et al. 2016). Moreover, flow cytometric fingerprints have been shown to correlate with taxonomic community structures of 16S rRNA gene amplicon sequencing (Props et al. 2016) and predictive models have been developed to identify diseases from gut microbiomes (Rubbens et al. 2021a).

1.4.6 Computational applications for drinking water analysis

Specifically for analysis of online flow cytometry of drinking water, some models have been developed. Sadler et al. (2020) developed a computational model in Python, *bactoML*, which uses flow cytometry data to detect contaminations in drinking water. The model uses fingerprinting and distance-based

outlier calculations. It was found that the model was better at detecting microbiological anomalies than simple flow cytometric measures such as cell concentrations and percentages of high nucleic acid bacteria. The output of the model is an easily interpretable outlier score which ranges between 0 and 1, depending on how deviating the fingerprint of the sample is.

Another study which calculated a metric that ranges between zero and one was made by Favere et al. (2020). The authors suggest that the Bray-Curtis dissimilarity of flow cytometric fingerprints can be used for detecting changes in microbial water quality. The Bray-Curtis dissimilarity will range between zero and one, where two identical fingerprints will have a value of zero and two fingerprints that have no bins in common will have a value of one. It is argued that the metric presents a more nuanced view of the bacterial composition than merely the total cell count.

As noted by Safford and Bischel (2019), most research has focused on two-dimensional data (such as scatter plots). Analysing more dimensions could give richer insights. For instance, forward scatter, side scatter, and multiple fluorescence signals can be computationally analysed simultaneously. Such methods were employed both by Sadler et al. (2020) and Favere et al. (2020) and future research could develop even more advanced models, perhaps including more machine learning or neural networks.

2 Methods

2.1 Microbiological methods

The microbiological analyses described in this section have all been carried out by personnel at VA SYD's laboratory in Malmö, Sweden or Sydvatten's laboratory in Stehag, Sweden. VA SYD's laboratory is certified by SWEDAC and works in accordance with STAFS 2020:1 which includes SS-EN ISO/IEC 17025:2018. Resulting raw data was used for further analysis in this thesis.

2.1.1 HPC and slow-growing bacteria

HPC and slow-growing bacteria were determined according to *SS-EN-ISO 6222:1999 ed. 1* and *SS-EN-ISO 8199:2018 ed. 2*, by mixing 1 ml of water sample with dissolved yeast extract agar in petri dishes and let to solidify. Thereafter, the plate was incubated for three days hours in 22 °C. The number of colonies was counted and yielded the HPC. Thereafter, the plates were incubated for another four days, and the resulting colonies were counted again and yielded the number of slow-growing bacteria.

2.1.2 Detection of indicator bacteria

Detection of indicator bacteria, such as *Escherichia coli* and coliforms, was also performed. The method is also based on cultivation on agar plates. However, the media differs — for instance, for detection of coliforms and *E. coli*, chromogenic coliform agar was used. Furthermore, membrane filtration of 100 mL of sample was used for coliforms and *E. coli.*

Other microorganisms were also detected for the two leaks. For instance, some samples were analysed for molds, yeasts, actinomycetes, *Clostridium perfringens* and intestinal *Enterococci*. These methods are also based on using selective substrates in the petri dishes.

2.1.3 Flow cytometry

The flow cytometry analysis was carried out according to the manufacturers' instructions with two different types of cytometers: BactoSense® cytometers (bNovate, Switzerland) for online monitoring and BD Accuri ™ C6 Plus cytometers (Becton Dickinson UK Ltd., Oxford, UK) for analysis at laboratories.

2.1.3.1 BactoSense The BactoSense® cytometers have a 488-nm solid-state laser, two fluorescence detectors denoted FL1 and FL2 with (525/45 and 715 LP respectively), and a side-scatter detector denoted SSC (488/10). Forward scatter is not measured. Before analysis, the samples were automatically

stained with SYBR Green I and incubated for 10 minutes at 37 °C. SYBR Green I is detected by the FL1 detector.

2.1.3.2 BD Accuri C6 Plus The BD Accuri™ C6 Plus cytometer has a green fluorescence detector and a red fluorescence detector denoted FL1 and FL3 respectively (533/30 and 670 LP). Samples were stained with SYBR Green I for total cell counting, and additionally with Propidium iodide (PI) which enables intact cell counting. SYBR Green I is primarily detected by the FL1 detector, while PI is primarily detected by the FL3 detector.

2.2 Collection of data

Data has been collected from three different occasions: two major leaks denoted the Teckomatorp and Vomb leak respectively, and one new pipe installation in Torna Hällestad that was sampled in March-April 2021. These three cases have been chosen based on availability of data and practical constraints. Sampling was performed by operating personnel at the water utilities. The data consists of flow cytometry analyses made with online flow cytometers (for the Teckomatorp leak) and laboratory cytometers (for the Vomb leak and Torna Hällestad pipe installation), as well as HPC for all sampling sites. Additional microbiological data consisting of plate counts specific for coliforms and similar microorganisms was also available for the two leaks.

2.2.1 Case 1: Teckomatorp pipe break

A major leak occurred between Teckomatorp and Västra Strö in September 2020 (fig. 2). The pipe has the diameter 1000 mm and is made of fiberglass. After the leak occurred, the pipe was closed at Västra Strö and an additional leak occurred in Teckomatorp. Repairs were thereafter made, and the pipes were then flushed with drinking water for approximately two weeks. During this time, samples were taken daily (excluding weekends) at Västra Strö (A), sampling point before the leak (B), sampling point after the leak (C) and Teckomatorp (D) further downstream of the leak (fig. 3). Cultivation-based analyses were made for HPC, slow-growing bacteria, and plating for coliforms and other indicator bacteria. Furthermore, online flow cytometers (BactoSense) were set up at Västra Strö and Teckomatorp. The sampling was started simultaneously as the flushing of the pipe was begun, which was after the reparations were completed. Samples were taken at 4-hour intervals at Västra Strö and hourly at Teckomatorp. The distance between Västra Strö and Teckomatorp is around 10 km.

Figure 2: The pipe break that occurred between Teckomatorp and Västra Strö. Photo by Sydvatten AB.

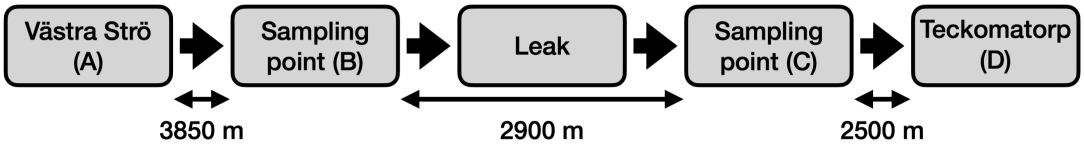

Figure 3: Schematic figure outlining the positions of the relevant sampling points for the Teckomatorp pipe break. Approximate distances between sampling points have been marked below the figure.

2.2.2 Case 2: Vomb pipe break

There was a large leak in a major supply pipe from the drinking water treatment plant Vombverket in August 2020 (fig. 4). Even though the geographical location of the pipe break was not in Vomb, it will henceforth be referred to as the Vomb pipe break. The pipe has the diameter 1100 mm and is made of concrete. After the repairs had been made, the pipes were flushed with drinking water for approximately two weeks. During this time, samples were taken from three different sampling points denoted VA, VB and VC (fig. 5). VA is located before the leak while VB and VC are located after the leak. Samples were taken daily, including weekends. Samples were analysed both with the cultivation-based microbiological methods HPC and analysis for *E. coli*/coliforms, and four of the samples were also analysed with a laboratory flow cytometer at VA SYD's Malmö laboratory (Accuri C6 Plus).

Figure 4: Repair of the pipe break of the major supply pipe from Vombverket. Photo by Sydvatten AB.

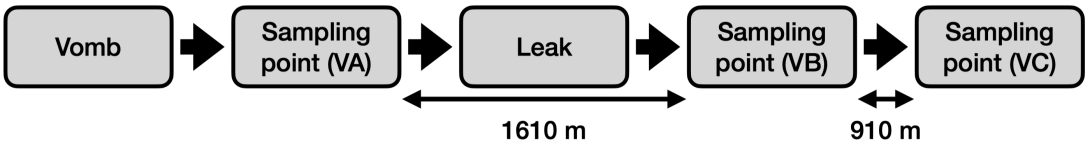

Figure 5: Schematic figure outlining the positions of the relevant sampling points for the Vomb leak. Approximate distances between sampling points have been marked below the figure.

2.2.3 Case 3: Torna Hällestad pipe installation

A new pipe installation in Torna Hällestad was sampled after being put in operation, from 23rd of March to 16th of April 2021. The pipe installation was performed to accommodate for a newly built residential area. The sampling was begun after the new pipe had been installed and approved. The pipe was continuously flushed with drinking water during the sampling period. Samples were taken approximately every two days, excluding weekends and holidays. Sampling points included were a reference point and a sample point at the end of a new section of pipe. The reference point is from a nearby pipe that, according to the projectors, should not have been affected by the new pipe installation. The pipe is of the material polyethylene (PE) and of the diameter 160 mm. The pipe has connections to another older pipe of diameter 80 mm making it possible (albeit unlikely) that water can temporarily flow in both directions, which is typical for smaller pipe systems feeding residential areas.

The sampling was performed via a valve connected to a small hose in plastic. Samples were analysed with HPC and one of two separate laboratory flow cytometers (Accuri C6 Plus) – a few samples were analysed at Sydvatten's laboratory in Stehag but most samples were analysed at VA SYD's laboratory in Malmö.

2.3 Data analysis

Raw, unprocessed data was received from the water utilities in the format of Excel or PDF files for the cultivation-based data, and in the Flow Cytometry Standard format for the flow cytometry data. The data has been analysed with the statistical programming language R (version 4.0.4) (R Core Team 2021).

2.3.1 Flow cytometry data

The flow cytometry data has been analysed with the steps outlined in this section, for all the three datasets.

2.3.1.1 Loading and pre-processing of files Flow cytometry files were loaded in the Flow Cytometric Standard format with the *flowCore* package (version 2.3.2) (Ellis et al. 2021). Data was cleaned from anomalous values owing to instrument variations (e.g., flow rate stabilities, signal acquisition) with the *flowAI* package (version 1.20.1) (Monaco et al. 2016). Raw data was however retained for cell counting, as recommended by Rubbens and Props (2021). Data was transformed using the arcsine hyperbolic function, which unlike the traditionally used logarithmic transformation can also transform negative values.

2.3.1.2 Gating Gating was performed with the *flowCore* package to exclude electronic noise in all analyses. Further gating was also performed to exclude inorganic particles and cell aggregates, and thereby only include cells. This step was however not performed when the background was of interest, as described further in the Results and Discussion sections. Examples of chosen gates are shown in fig. 6.

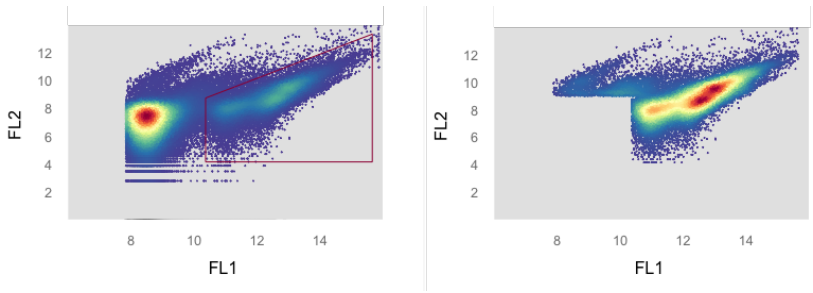

Figure 6: To the left, an example of gating used to select total cells is shown. The red outline is assumed to encompass the cells in the sample. To the right, an example is shown of gating used to select both cells and background, but excluding electronic noise. FL1 and FL2 denote two different fluorescence detectors.

Visualisations of flow cytometric scatter plots were made with *flowViz* (Ellis et al. 2020). The *animation* package (Xie 2013) was used to visually verify that the gate encompassed the cells in all samples.

2.3.1.3 Fingerprinting Advanced fingerprinting was performed with the fingerprinting functions available in the *Phenoflow* package (Props et al. 2018), originally developed by Clement and Thas (2014). The number of bins used for the fingerprints was 256. Thereafter, random resampling to the lowest cell counts was performed with the *Phenoflow* package to remove the impact of varying cell counts. Bray-Curtis dissimilarities of the fingerprints were thereafter calculated with the *vegan* package (Oksanen et al. 2020). The principle coordinates ordination (PCoA) of the dissimilarity matrix was plotted using functions from the *Phenoflow* package.

2.3.1.4 Bray-Curtis dissimilarity metric The Bray-Curtis dissimilarity metric was calculated as by Favere et al. (2020). The chosen fingerprint method was equal-space binning with functions available in the *Phenoflow* package. Bray-Curtis dissimilarities of the fingerprints were calculated with the *vegan* package (Oksanen et al. 2020).

2.3.1.5 Cell counting Number of cells were counted in the cell gate using functions provided in the *flowCore* package.

2.3.2 Plotting

Plotting of flow cytometric data as well as HPC was made using the *tidyverse* package (Wickham et al. 2019) with *ggplot2* (Wickham 2016).

3 Results

3.1 Case 1: Teckomatorp pipe break

3.1.1 HPC

HPC were plotted for the Teckomatorp pipe break (fig. 7). Day 0 was when the flushing began, which was when the repairs had been completed. The reference water (point A) was stable at or slightly above zero while the other sampling points displayed more fluctuating values and shared some peaks (for instance at day nine, with values of 39 and 85 cfu/ml for sampling points C and D). Note that the legal requirement threshold is at a HPC of 100 colony-forming units (cfu)/ml (SLVFS 2001), which no sample reached in this data. However, Sydvatten's internal threshold of 50 cfu/ml was surpassed on certain days. Interestingly, on day nine there were also elevated values (53 cfu/ml) at sampling point B, which was upstreams of the pipe break. Theoretically, upstream sampling points should not be affected by pipe breaks or contaminations that occur downstream of them.

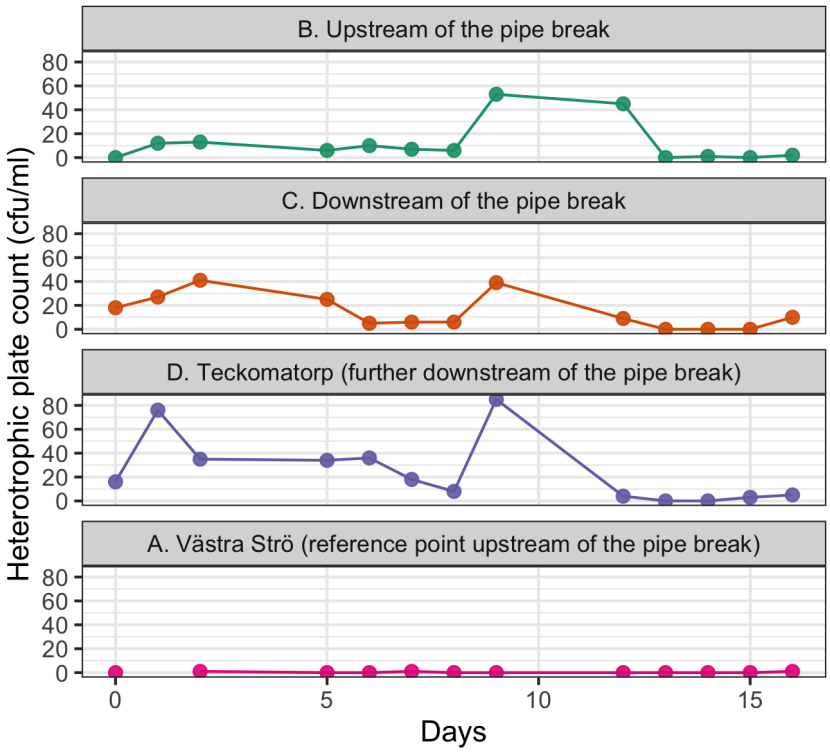

Figure 7: HPC (cfu/ml) for the Teckomatorp pipe break. Day 0 was when the flushing began, after the repairs had been completed. Sampling points have been labeled A-D based on how the water flows (upstream to downstream), with sampling point A not affected by the pipe break and used as a reference point.

3.1.2 Slow-growing bacteria

The slow-growing bacteria were also plotted for the Teckomatorp pipe break (fig. 8). Similarly to what was observed for HPC (fig. 7), the reference water (point A) exhibited lower and more stable values of slow-growing bacteria than the other sampling points, between 0 and 320 cfu/ml. Likewise, the upstream sampling point (B) had lower values (between 52 and 258 cfu/ml) than the two downstream sampling points. The two downstream sampling points after the pipe break showed fluctuations, between 260 and 1120 cfu/ml for point C and between 115 and >5000 cfu/ml for point D. Note that the first sample from Teckomatorp (point D) is not plotted due to its result being greater than 5000 cfu/ml, i.e. above the accepted limit by regulations.

Notably, the peaks in the counts of slow-growing bacteria did not clearly correspond to the peaks of the HPC. The HPC (fig. 7) peaked at day 9 for sampling points B-D and had elevated values for day 1-3 for sampling points C-D. In comparison, the slow-growing bacteria did not peak at day 9. In similarity to the HPC, the initial values were elevated for sampling points C-D.

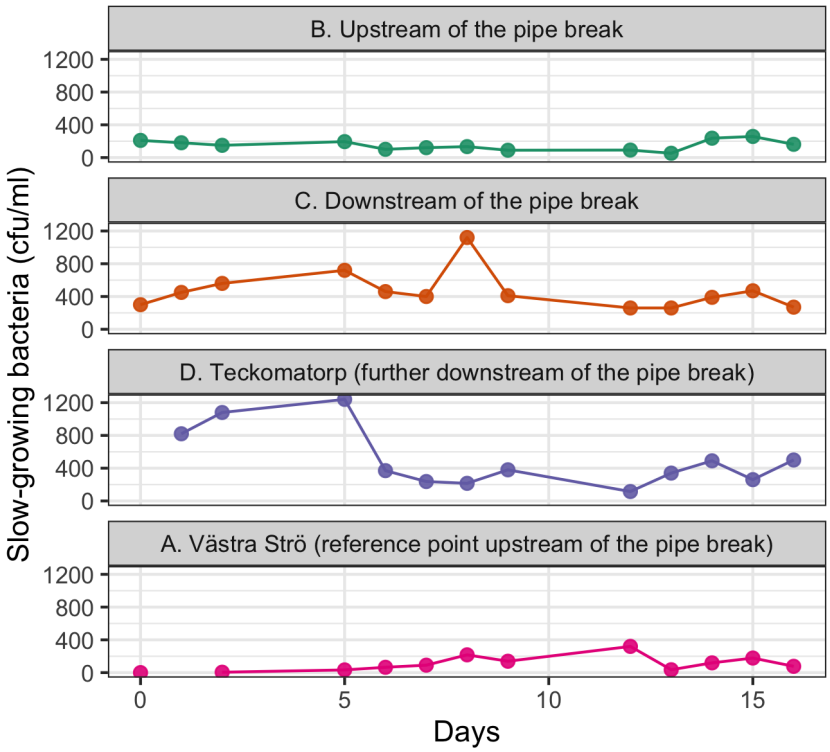

Figure 8: Slow-growing bacteria (cfu/ml) for the Teckomatorp pipe break. Day 0 was when the flushing began, after the repairs had been completed. Sampling points have been labeled A-D based on how the water flows (upstream to downstream), with sampling point A not affected by the pipe break and used as a reference point.

3.1.3 Detection of indicator organisms

Plating for indicator bacteria from sampling points A-D in the Teckomatorp leak detected no coliforms or *Escherichia coli*. However, *Clostridium perfringens* was detected at 1 cfu/ml in 1 sample, on day 8 at sampling point C, downstream of the pipe break. *Clostridium perfringens* is an indicator organism which is thought to indicate potentially pathogenic contamination of the drinking water. This bacterium was only detected once and it is possible that it was a false positive. However, if not, the detection of this bacterium is an example of why the flushing of pipes and microbiological analyses are performed after pipe breaks, since the water utilities cannot deliver water containing such indicator organisms to consumers.

3.1.4 Flow cytometry

To construct an overview of the microbial communities in the samples of the Teckomatorp and Västra Strö sampling points using flow cytometry data, PCoA plots of the Bray-Curtis dissimilarity matrixes were made. PCoA plots were made both resampled to the minimum number of cells in the dataset (fig. 9), and non-resampled (not shown). Resampled plots are thereby independent of changes in the number of cells, unless these changes are also accompanied by microbial community changes. Both the resampled (fig. 9) and non-resampled (not shown) plot had a similar appearance, with two separate clusters for the two sampling points and the first seven samples from Teckomatorp deviating from these clusters. The plot is clear in that the measured fluorescence is different among the Teckomatorp and Västra Strö sampling points. A portion of this difference may however be due to instrumental variations (see Discussion section). Furthermore, it is clear that the samples from the first 7 hours at the Teckomatorp sampling point were different since these samples appear outside the cluster (fig. 9).

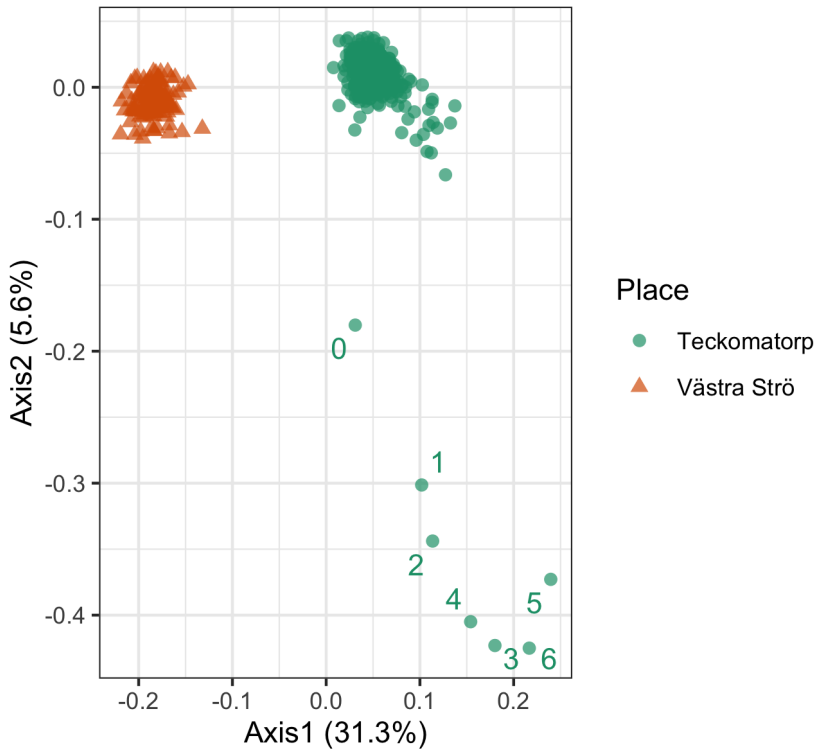

Figure 9: PCoA plot of the Teckomatorp pipe break flow cytometry data, resampled and thereby independent of the sheer number of cells. The Teckomatorp samples were taken from water that has passed the pipe break section, while samples from Västra Strö were used as a reference point before the pipe break. Labels indicate number of hours since start of sampling and flushing. The axes represent principal components that explain a certain part of the variance in the data. The percentage of variance explained by the respective components is stated in the axis labels.

The Bray-Curtis dissimilarity metric was also calculated for each sample in this dataset (fig. 10). The underlying principle behind this figure is similar to the PCoA plot (fig. 9) but by plotting the dissimilarity as a function of time, further insights can be gained. The figure (fig. 10) shows that the first seven samples differed from the other samples at the sampling point, with their dissimilarity metric differing from the subsequent samples. It is evident that elevated values appear around day 6-7 in the Teckomatorp sampling point. The plotted lines for Teckomatorp and Västra Strö do not overlap, which implies that the microbial communities at the two sampling points were not the same. This can be compared with the PCoA plot (fig. 9), which shows two clusters that do not overlap. In other words, the results from the PCoA plot and the Bray-Curtis dissimilarity metric are the same, but the presentation of the dissimilarity metric may be more appropriate when the variable of interest is time.

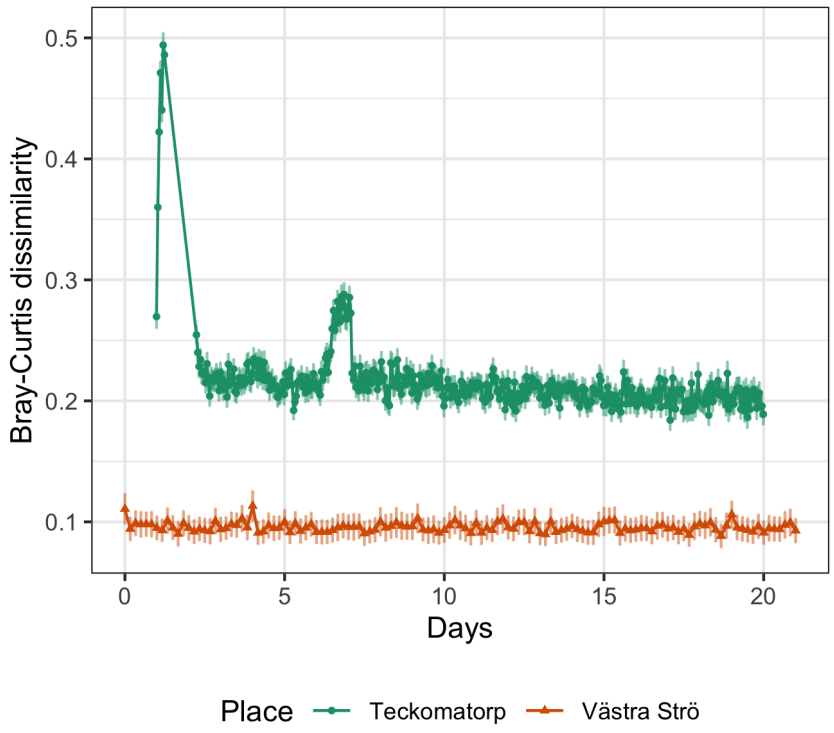

Figure 10: The Bray-Curtis dissimilarity metric calculated from gated data from online flow cytometers set up at the Teckomatorp and Västra Strö sampling points. Day 0 was when the flushing began, after the repairs had been completed. Error bars corresponding to the standard deviation are displayed in semi-transparent colors.

Flow cytometry metrics such as total cell counts (fig. 11) and percentage of cells with HNA (not shown) were also calculated. The total cell counts reiterate the pattern shown in the PCoA plots and the Bray-Curtis dissimilarity metric. The first seven samples from the Teckomatorp sampling point deviate, at one point the cell concentration reaches 1 000 000 cells/ml. It is also apparent that the counts from the two sampling points differ consistently.

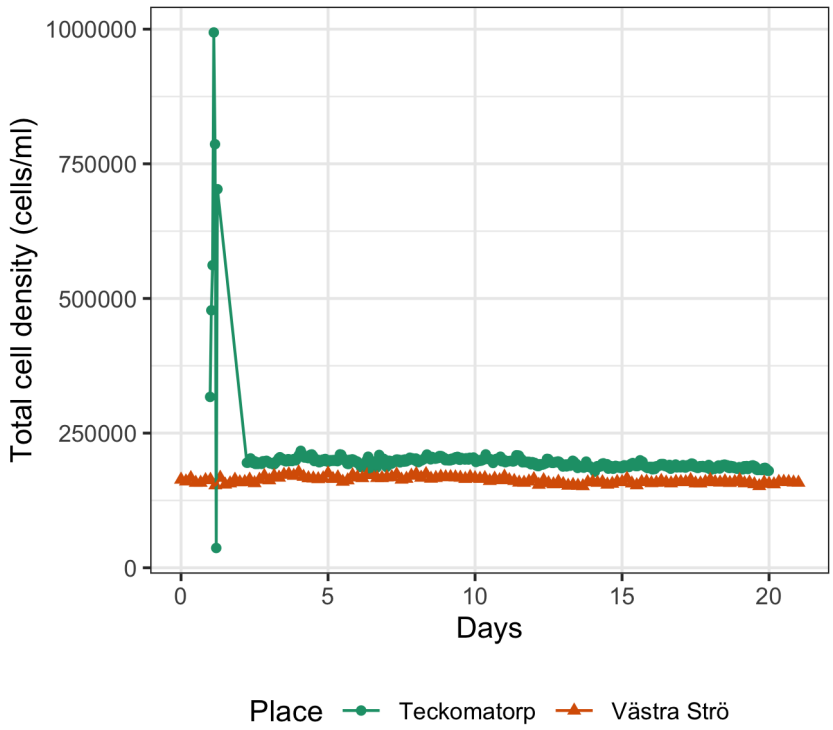

Figure 11: Total cell counts for the Teckomatorp and Västra Strö sampling points (downstream and upstream of pipe break respectively), as measured by online flow cytometers. The sampling interval is hourly for Teckomatorp and every four hours for Västra Strö.

With the first seven samples from the Teckomatorp sampling point differing considerably from the rest, it is difficult to discern any patterns in the total cell counts (fig. 11). Therefore, the limits of the y-axis were adjusted in another figure (fig. 12), where the first seven samples from the Teckomatorp sampling point are outside the limits. The figure shows relatively small fluctuations of up to 30 000 cells/ml in total cell counts, and a consistent difference in cell counts between the two sampling points of around 5000-25 000 cells/ml. The concentrations of cells at the Teckomatorp sampling point are around 180 000-220 000 cells/ml, while the reference sampling point at Västra Strö has concentrations of around 145 000-175 000 cells/ml.

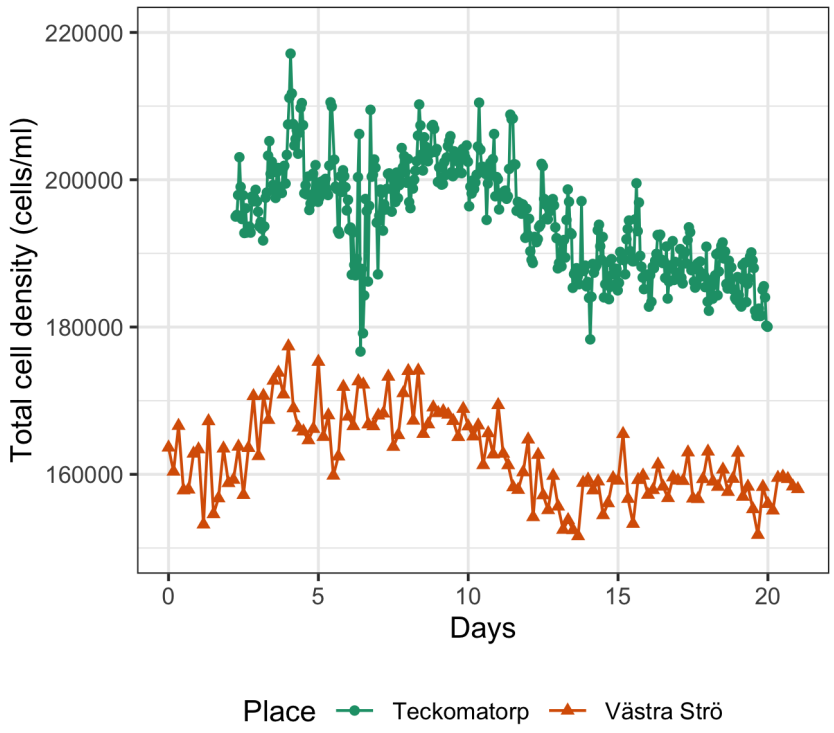

Figure 12: Total cell counts for the Teckomatorp and Västra Strö sampling points (downstream and upstream of pipe break respectively), as measured by online flow cytometers. The sampling interval is hourly for Teckomatorp and every four hours for Västra Strö. The limits on the y-axis have been adjusted and thus the first 7 samples from Teckomatorp are outside the limits.

3.1.5 Influence of background

A possibility with flow cytometry is also to investigate the effect of the background, which can include both inorganic particles such as salt crystals and metal oxides and organic particles such as cell fragments, viruses, free DNA and humic acids (Van Nevel et al. 2017a). The effect of the background was investigated, both by including the background in the PCoA analysis and Bray-Curtis dissimilarity metric and by simply counting particles in the background. The Bray-Curtis dissimilarity metric, calculated including the background, is shown in fig. 13. As can be seen, the pattern is the same as with the Bray-Curtis dissimilarity without the background (fig. 10), but the peaks are more pronounced. This supports the ideas related to including the background presented by Van Nevel et al. (2017a), who argued that including the background would make flow cytometry measurements and comparisons more conservative.

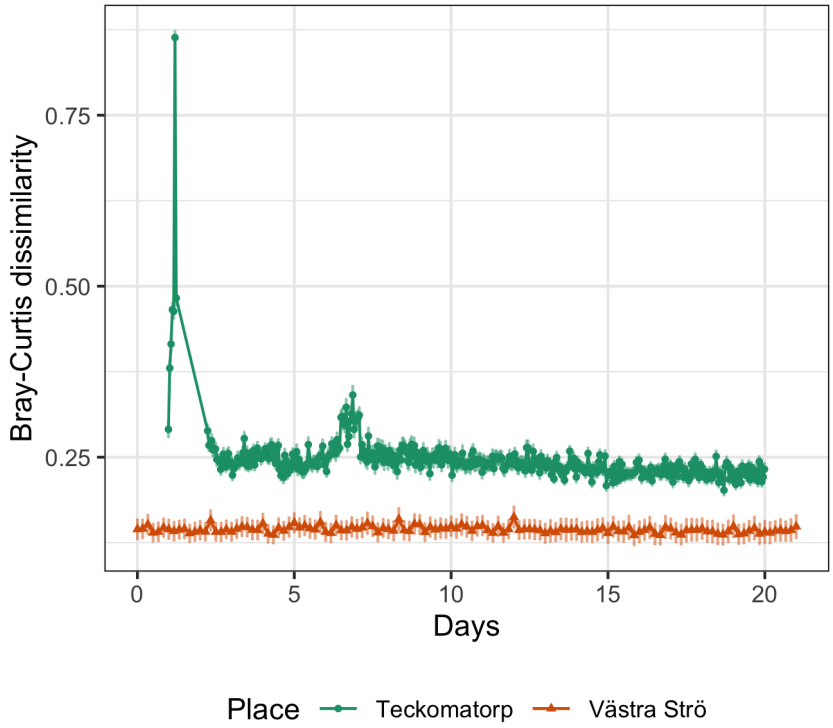

Figure 13: The Bray-Curtis dissimilarity metric calculated from data including the background from online flow cytometers set up at the Teckomatorp and Västra Strö sampling points. Error bars corresponding to the standard deviation are displayed in semi-transparent colors.

3.2 Case 2: Vomb pipe break

3.2.1 HPC

HPC were plotted for the three sampling points relevant for the Vomb pipe break (fig. 14). Detected presence of coliforms (1 cfu/ml) occurred in one sample, from day 2 in sampling point A. Note that both the legal threshold of 100 cfu/ml and Sydvatten's internal threshold of 50 cfu/ml are surpassed for many samples. The gap at day 4 is due to contaminated agar bottles. It is important to note that the water sampled in this case was flushed out and was not delivered to any consumers.

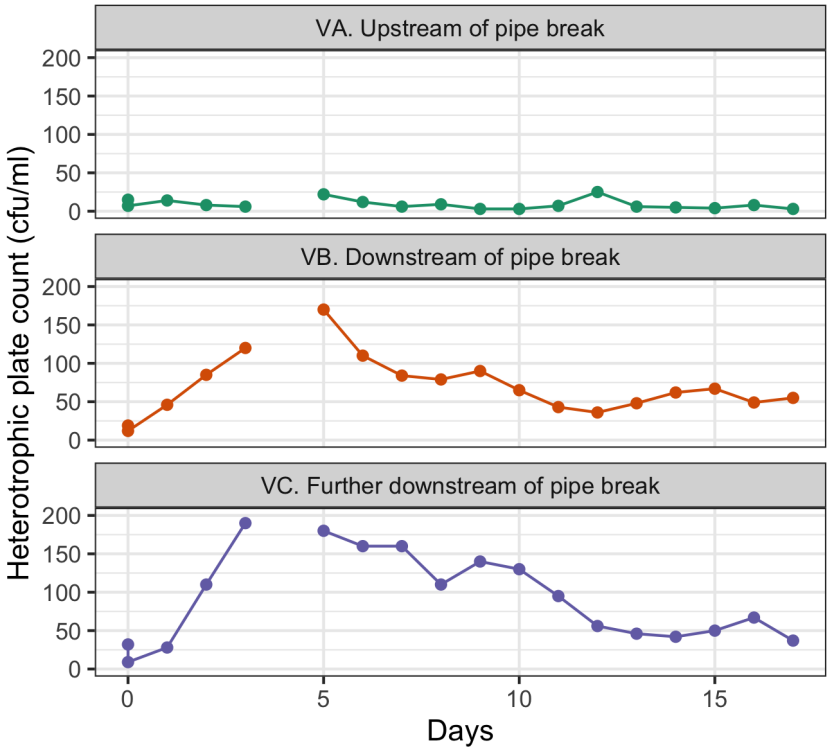

Figure 14: HPC for the Vomb pipe break. The gap at day 4 is due to contaminated agar bottles.

3.2.2 Flow cytometry

Total cell counts were also plotted for the Vomb dataset (fig. 15). However, flow cytometry samples were only taken on four days and the error bars corresponding to the standard deviation are large. It is not possible to deduce a pattern or trend. The upstream sampling point VA appears to show an increased cell count (from 375 000 to 440 000) at day 7, but it should not be affected by the pipe break in theory. Therefore, this difference is likely due to another factor (for example, changes in flow rates or variations in the sampling procedure).

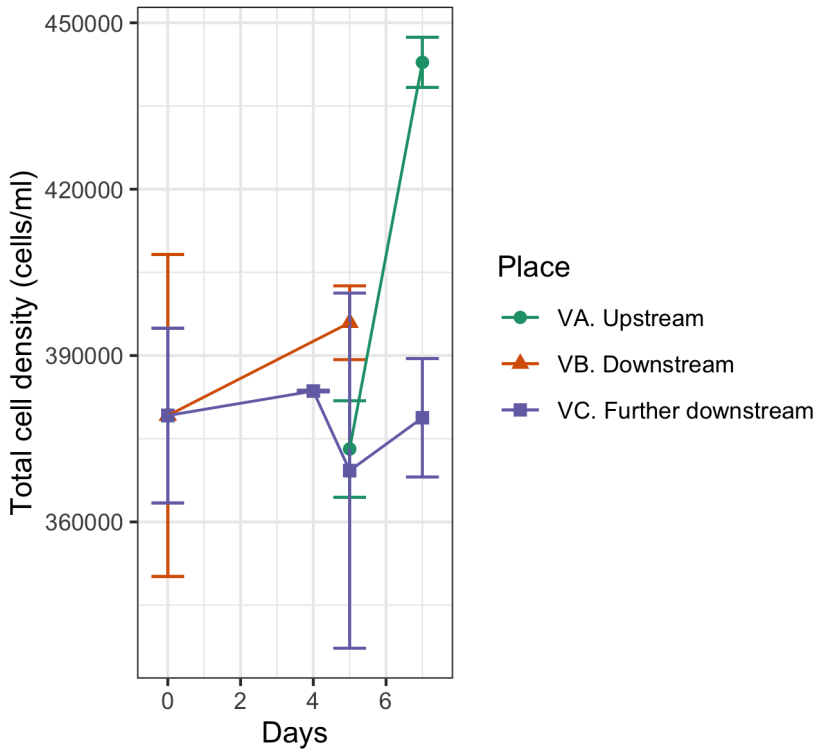

Figure 15: Total cell counts for the Vomb pipe break, as measured with a laboratory flow cytometer. Error bars correspond to the standard deviation.

3.3 Case 3: Torna Hällestad pipe installation

The data from the Torna Hällestad pipe installation consists of flow cytometry samples and HPC from two different sampling points: a reference sampling point and a sample point at the end of a new section of pipe. Both HPC and flow cytometry total cell counts are plotted (fig. 16). Notably, the HPC increased 8-fold after day 14 and doubled in the subsequent sample, reaching values well above the threshold accepted by regulations (100 cfu/ml). For the flow cytometry total cell counts, the values were between 375 000 and 740 000 cells/ml, and the difference between reference and sample was at most two-fold for day 20. The increases in HPC at the sample point were accompanied by increases in flow cytometry total cell counts (for instance, with HPC increasing from 22 cfu/ml to 180 cfu/ml from day 14 to 18 and total cell counts increasing from 525 000 to 575 000 the same days). Note that the total cell counts for the reference point varied considerably between 380 000 cells/ml for day 4 to 520 000 cells/ml for day 14.

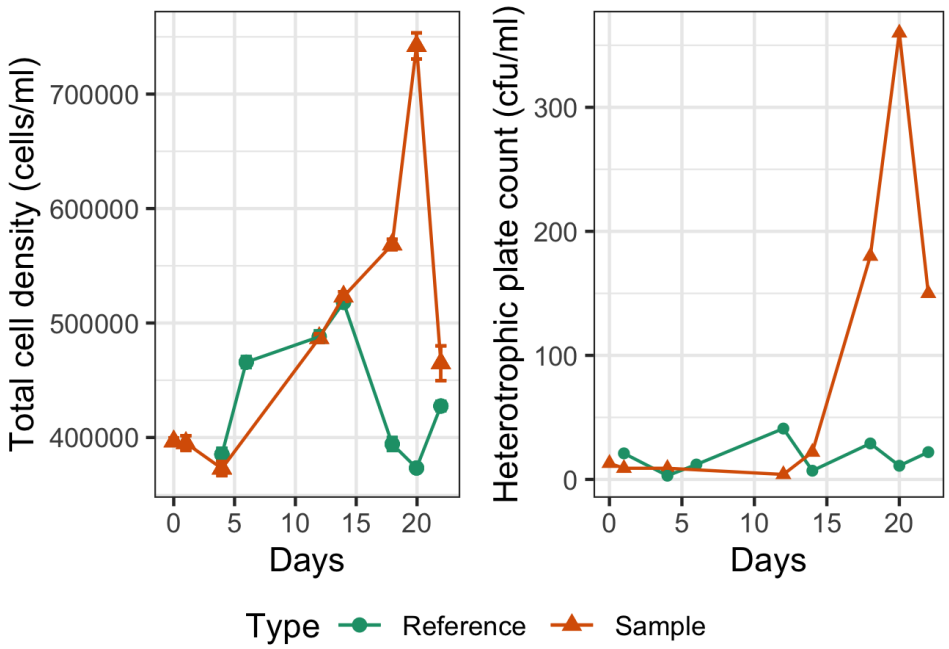

Figure 16: Total cell density and HPC for Torna Hällestad. Total cell densities were measured in triplicates with one of two separate laboratory flow cytometers (either at VA SYD's laboratory or at Sydvatten's laboratory). Error bars correspond to the standard deviation. HPC values were determined by cultivation. The reference point is from a nearby pipe that should not have been affected by the new pipe installation.

The intact cell densities were also plotted (fig. 17). Arguably, this metric may be more relevant for comparisons with HPC since only intact cells will be able to form colonies on the agar plates. This plot gives similar insights as the total cell count (fig. 16). The counts ranged from 200 000 to 350 000 cells/ml, indicating that roughly half of the cells measured by total cell counts were dead (or at least had broken cell membranes). This was also true for the total cell count peak of 740 000 cells/ml (fig. 16), which displayed an intact cell count of 350 000 cells/ml. It is notable that the standard deviations of some samples were high (for instance, for day 14 the standard deviation represented almost 50 000 cells/ml), indicating a larger uncertainty for these samples.

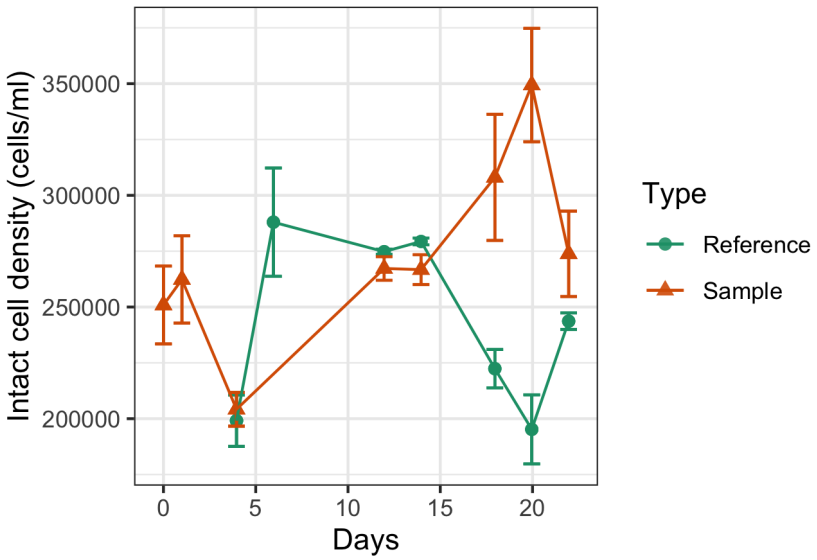

Figure 17: Intact cell densities for the Torna Hällestad pipe installation. Intact cell densities were measured in triplicates with one of two separate laboratory flow cytometers (error bars correspond to the standard deviation), using both SYBR Green I and Propidium iodide as staining. The reference point is from a nearby pipe that should not have been affected by the new pipe installation.

Further analysis was also performed on the flow cytometric data, including calculation of HNA ratios, PCoA plots (not shown) and calculation of the Bray-Curtis dissimilarity metric (fig. 18). All three of these methods may elucidate changes in the microbial communities. Some insights could be derived from this Bray-Curtis dissimilarity metric figure (fig. 18), for instance the microbial communities are clearly different in the last three samples between reference and sample points. These samples also differ considerably in the number of total and intact cells (fig. 16 and fig. 17).

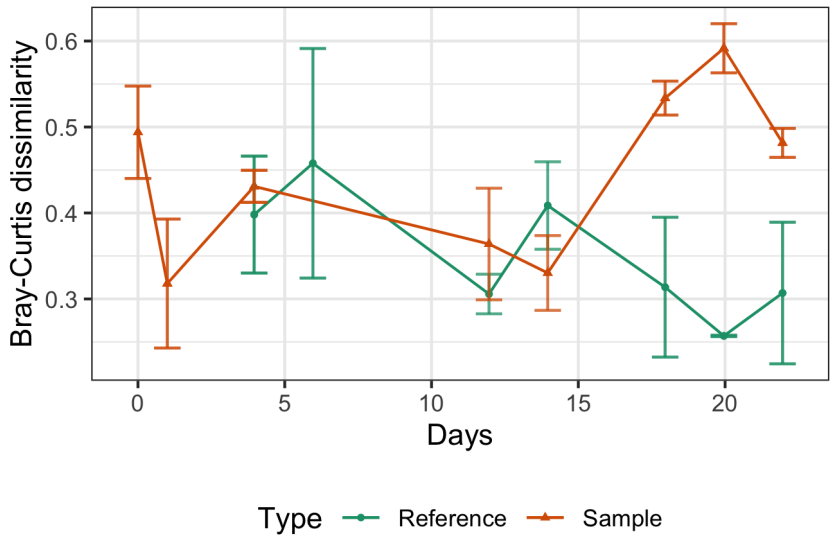

Figure 18: The Bray-Curtis dissimilarity metric calculated from reference samples and samples from the Torna Hällestad pipe installation. Samples were analysed with one of two separate laboratory flow cytometers. Error bars corresponding to the standard deviation of triplicate samples are displayed in semi-transparent colors. The reference point is from a nearby pipe that should not have been affected by the new pipe installation.

4 Discussion and conclusions

4.1 Flow cytometry for monitoring water quality after maintenance

One of the aims of this thesis was to investigate whether faster microbiological analysis methods can replace those currently used after maintenance works or new pipe installations. The investigated method has been flow cytometry, both online flow cytometry and laboratory flow cytometry. Some insights about the use of flow cytometry have been gained. Online flow cytometry has proven to be especially useful due to the short intervals of sampling, and novel computational methods have also proven to be useful for analysing the flow cytometry data.

4.1.1 Case 1: Teckomatorp pipe break

To begin with, both the PCoA plot (fig. 9) and the Bray-Curtis dissimilarity (fig. 10) metric calculation for the Teckomatorp pipe break provide valuable characterisations of changes in the microbial communities. It is clear that the first seven hours of sampling downstream of the pipe break deviate microbiologically. After another 5 days, the dissimilarity metric increased again, and was relatively high for approximately 1 day. The cause of this increase is unknown and was not reflected in the total cell counts (fig. 12), nor in the HPC (fig. 7) or the slow-growing bacteria (fig. 8). The increase of the dissimilarity metric clearly indicates that there was a shift of the microbial community during this time, which then suddenly returned to normal. Possible explanations may be changes in the flow, stagnation of water at the sampling point, contamination with soil or other foreign matter, or malfunctioning of the instrument. The exact cause remains to be further investigated.

The pipe break in Teckomatorp generated a flow cytometry dataset that is interesting to compare with the study by Van Nevel et al. (2017a). In that study, pipes of diameters 1050 mm, 900 mm and 400 mm were monitored with flow cytometry, which can be compared to the Teckomatorp pipe diameter of 1000 mm. A notable difference is that the pipes in the study by Van Nevel et al. (2017a) were subject to shock-chlorination, which was not performed in Teckomatorp. The study compared flow cytometry samples to reference samples and found that the samples reached similarity to the reference samples after 2-4 hours. The timeframe is similar to the initial 7 hours of deviating samples shown in fig. 9 and fig. 10. However, the samples from Teckomatorp do not seem to ever reach similarity with the reference samples. Therefore, it is not possible to draw similar conclusions in this thesis as in the study by Van Nevel et al. (2017a). It is only possible to state that the samples from Teckomatorp appear to stabilise in their microbial communities after seven hours, not that they reach similarity with the reference samples.

A possible explanation to the consistent differences between reference and sampling points in this dataset is the long distance between Teckomatorp and Västra Strö. The distance between Teckomatorp

and Västra Strö is around 10 km, which is significantly longer than the longest distance between reference point and sampling point in the study by Van Nevel et al. (2017a), i.e. 3 km. Even that distance was considered long by the authors and made it impossible to make absolute comparisons between reference water and sampling water. The long distance causes a large difference in time for the two sampling points, since it takes a long time for the water to flow from Västra Strö and Teckomatorp. Furthermore, the microbial communities can be expected to change during its flow in pipes. This may for instance be due to release of bacteria from the biofilm, which has an increasing effect with increasing distances in pipes (Chan et al. 2019). It may also be due to the decreasing effect of chlorination of water farther from the drinking water treatment plant, which would lead to less chlorine being present in the Teckomatorp sampling point than in the Västra Strö sampling point (Schleich et al. 2019).

Computationally, the flow cytometry data has been analysed with similar methods in this thesis and the study by Van Nevel et al. (2017a). The fingerprinting method is the same, with origins from Clement and Thas (2014). The choice of gates differed, with the study by Van Nevel et al. (2017a) using raw flow cytometry data without any gates, with the argument that this should make the analysis more conservative since the area above the gate contains information about loose biofilm, cell aggregates etc. — organic matter that is not desired in drinking water (Van Nevel et al. 2017a). A potential problem with including ungated flow cytometry data is however the electronic noise from the instrument. This electronic noise can vary randomly between samples, and therefore it can be speculated that omitting the electronic noise would lead to a more accurate analysis. High levels of electronic noise could be discerned in the Flow cytometry scatter plots, especially those from online flow cytometers. The electronic noise was therefore omitted using gating in this thesis. The fingerprints were subjected to Fischer discriminant analysis in the study by Van Nevel et al. (2017a), while this thesis uses Principle Coordinates Analysis. The methods are similar in nature in that they try to find linear combinations of the data, although Fischer discriminants seem to be more focused on differences in the data.

Another difference with the study by Van Nevel et al. (2017a) is that the study used the same flow cytometer for analysing all samples, while this thesis analysed data from two separate online flow cytometers for the reference (Västra Strö) and sampling (Teckomatorp) points. Using two cytometers opens up another potential source of errors, namely the instrumental variations. Flow cytometers depend on lasers, and require calibration if the resulting data is to be compared with data from other cytometers. Calibration had not been performed in this case. However, from examining animations and scatter plots from the two online flow cytometers, there is not any obvious shift in fluorescence apparent between the two machines. This suggests that the instrumental variations between these two particular cytometers used is not large. The contrary can be observed for case 3: Torna Hällestad, where two separate cytometers also was used, with a visible shift in FL1 for the two machines. This will be discussed further for case 3.

4.1.2 Case 2: Vomb pipe break

The flow cytometry data from the Vomb pipe break is from only 4 sampling occasions (fig. 15), and therefore it was difficult to deduce any patterns. This dataset reiterates the importance of taking many samples, preferably with short time intervals and with online instruments.

4.1.3 Case 3: Torna Hällestad pipe installation

The Torna Hällestad pipe installation is different from case 1 and 2 primarily in that it is not an example of a major pipe break which is repaired, but rather an installation of a new pipe. This implies that, contrary to in case 1 and 2, no biofilm existed before the start of sampling, although some organisms could be present from the manufacturing and transport of the pipe. Another difference is that the diameter of the pipe was smaller than in case 1 and 2 – at 160 mm compared to 1000 mm and 1100 mm. This leads to an increase in surface-to-volume ratio, i.e. the inner surfaces of the pipe have an increased contact with the bulk water compared to larger diameter pipes. Consequently, it can be expected that the biofilm formation has a larger effect in smaller pipes than larger pipes, whatever that effect may be.

With the pipe being new, it is possible that organic carbon could have leached from the pipe and functioned as nutrient for biofilm formation (Neu and Hammes 2020). However, it is also possible that the leaking organic carbon observed by Neu and Hammes (2020) does not occur in this type of drinking water pipes, since the study was made specifically on pipes in buildings (which may be of lower quality than pipes in the distribution system).

As the cell count appears to vary independently of time, the observed variations in the flow cytometry data from Torna Hällestad did not seem to primarily be a function of time since start of sampling (fig. 16). Factors that may have a larger impact, and which may obscure any patterns that are of interest related to biofilm formation, may include time of sampling, flow rates, stagnation of water etc. (Gabrielli et al. 2021). Further supporting this hypothesis is the fact that the reference point exhibits large variations between 380 000 cells/ml to 520 000 cells/ml. Clearer results may have been gained if such factors could be kept constant and controlled, and if the sampling frequency could be increased to sample for instance hourly with online flow cytometers.

As described previously, the flow cytometry analysis was performed with two separate flow cytometers for this case — either at VA SYD's laboratory or Sydvatten's laboratory. A visible shift in the FL1 axis could be discerned for the two machines, which indicates that the instruments are calibrated differently. Therefore, the choice of gates was adjusted to fit for both cytometers, in an attempt to counteract the effect of instrumental differences. Still, it can be expected that the instrumental differences are a source of error in this dataset.

Notably, the data from the Torna Hällestad pipe installation appears to show a similar pattern for the flow cytometric cell counts and the HPC, with values that increase at the same timepoints (for instance on day 20, fig. 16). This is not in accordance with what has been found in the literature, where the two methods appear uncorrelated (Van Nevel et al. 2017b; Gabrielli et al. 2021). As pointed out by Van Nevel et al. (2017b), the two methods do not provide the same information and should therefore not be used for the same purpose. A possibility is that the apparent correlations between the two metrics observed in Torna Hällestad are coincidental.

4.1.4 Computational methods of analysing flow cytometry data

It has become clear that flow cytometers generate vast datasets with a wealth of microbiological information, that may be difficult to interpret. The field is rapidly evolving with interesting methods of analysing flow cytometric data being published the last few years. Efforts are underway to apply such computational methods to drinking water monitoring, as by Sadler et al. (2020) and Favere et al. (2020). Future research areas can include evaluations of different advanced fingerprinting methods for drinking water monitoring, as well as more intricate computational analysis methods including machine learning. Such advancements may make flow cytometry an even more appropriate method for drinking water analysis.

4.1.5 Comparison between HPC and flow cytometry

It can be argued that HPC is no longer the most appropriate method for analysing bacteria in drinking water. Flow cytometry has numerous advantages over HPC, including speed, reproducibility, automation and sensitivity. Therefore, the method should be able to replace HPC for evaluating drinking water quality after maintenance or new pipe installations. It can be noted that the results from the methods cannot be expected to correlate (Van Nevel et al. 2017b), but this should not serve as an argument against flow cytometry. Rather, the lack of correlations elucidates flaws in the HPC method.

4.2 Microbiological processes in new or repaired pipes

The second aim of this thesis was to examine the natural microbiological process that occurs in new or repaired pipes. An observation is that there appears to be a specific pattern in the HPC, for all cases presented in this thesis. For all the sampling points downstream of pipe breaks or denoted samples rather than reference (fig. 7, fig. 14, fig. 16), the HPC increased after a time to then decrease and reach stable values close to zero. These patterns were however not clearly present in flow cytometry data such as total cell counts, at least not for Teckomatorp (fig. 12) or Vomb (fig. 15). For Torna Hällestad, the pattern was surprisingly similar for HPC and flow cytometry metrics (fig. 16) – a similarity which may

however be coincidental. It can be speculated what the origins and causes of the patterns in HPC are, what the implications are and when exactly the values begin to increase.

4.2.1 Origins and causes of the pattern

The literature on the biofilm formation process in pipes is relatively scarce, making it difficult to make qualified speculations on the origins and causes of the pattern in HPC. Studies on the time-frame of biofilm formation, as by Learbuch et al. (2021), indicate that the biomass and community composition of the biofilm can reach a stable situation after four to eight weeks. Douterelo et al. (2014) also found a three-fold increase in the cell coverage of the biofilm between day 14 and 28. Such studies can be compared with the Torna Hällestad pipe installation, which started to display elevated values in both HPC and flow cytometry cell counts after 14 days. Part of this increase may be related to biofilm formation — for instance, if the secondary colonisers of the biofilm trigger a release of the primary colonisers of the biofilm around this time point. Note that this is very speculative and needs to be investigated further. For instance, 16s rRNA sequencing could be performed to reach insights about this.

Case 1 and 2 in this project are from two pipe breaks and are therefore not straightforward to compare with the biofilm formation studies by Learbuch et al. (2021) or Douterelo et al. (2014). The study by Douterelo et al. (2016), which was made on pipes in operation, offers a more relevant comparison. After flushing the pipes, it was found that the subsequent regrowing biofilm had a different microbiological community structure than the pre-flushing biofilm. This indicates that some of the bacteria in the biofilm was released by the flushing, and also that the uptake of bacteria by the biofilm may change. Thereby, two potential factors affecting the bacterial concentration of the bulk water occur in parallel. The timeframe of the increase was similar for both pipe breaks, with elevated values in the first few days (fig. 7 and fig. 14). The time until stabilisation was then approximately 14 days. Note that part of the elevated HPC may be due to actual contaminations, pathogenic or not, owing to the pipe break and related maintenance work.

4.2.2 Implications of a potential general pattern

Whether the pattern that has appeared for HPC is general for maintenance operations or pipe installations remains to be concluded. The project that this thesis is part of will continue to investigate this question. If the pattern turns out to be generally prevailing, valuable conclusions may be drawn. For instance, new pipe installations are sometimes delayed in their approval due to elevated HPC values without any apparent reason. If there is a generally occurring pattern in HPC after a pipe installation, water utilities can learn to expect elevated values after a certain time (e.g., 14 days). The sampling can be adjusted to not take samples on days which are expected to be elevated due to biofilm formation,

and thereby resources can be saved. Regarding pipe breaks, the potential savings may be even larger since mayor pipe breaks can cause a wide variety of problems (such as reduced water supply, disturbed traffic conditions, increased workload for personnel and a substantial waste of drinking water).

4.3 Conclusions

To conclude, flow cytometry has shown to be a sensitive and feasible method of assessing changes in microbial water quality. Recent computational methods, including calculations of the Bray-Curtis dissimilarity metric from flow cytometric fingerprints, present a more nuanced analysis than merely cell counts. New computational methods can likely improve the usefulness of flow cytometry even more. The natural microbiological process in new or repaired pipes appears to follow a common pattern among the cases studied, at least for HPC. The pattern could possibly be related to biofilm formation.

5 Acknowledgements

I would like to recognise and thank those who have helped me during this project. Thanks to my supervisors Catherine Paul and Markus Fröjd for assisting with all kinds of inquiries in the process. Also, thanks to Sandy Chan, Josefin Barup, and Lena Meyer for providing ideas, data, and valuable feedback. I am also grateful to Niklas Gador for help and discussions about the data analysis.

6 References

Chan, Sandy, Kristjan Pullerits, Alexander Keucken, Kenneth M. Persson, Catherine J. Paul, and Peter Rådström. 2019. "Bacterial Release from Pipe Biofilm in a Full-Scale Drinking Water Distribution System." *Npj Biofilms and Microbiomes* 5 (1): 3–10. https://doi.org/10.1038/s41522-019-0082-9.

Chan, Sandy, Kristjan Pullerits, Janine Riechelmann, Kenneth M. Persson, Peter Rådström, and Catherine J. Paul. 2018. "Monitoring Biofilm Function in New and Matured Full-Scale Slow Sand Filters Using Flow Cytometric Histogram Image Comparison (CHIC)." *Water Research* 138 (July): 27–36. https://doi.org/10.1016/j.watres.2018.03.032.

Cheswick, Ryan, Elise Cartmell, Susan Lee, Andrew Upton, Paul Weir, Graeme Moore, Andreas Nocker, Bruce Jefferson, and Peter Jarvis. 2019. "Comparing Flow Cytometry with Culture-Based Methods for Microbial Monitoring and as a Diagnostic Tool for Assessing Drinking Water Treatment Processes." *Environment International*. https://doi.org/10.1016/j.envint.2019.06.003.

Clement, Lieven, and Olivier Thas. 2014. *flowFda: Flow Cytometry Functional Data Analysis.* Manual.

Cowle, Matthew W., Gordon Webster, Akintunde O. Babatunde, Bettina N. Bockelmann-Evans, and Andrew J. Weightman. 2020. "Impact of Flow Hydrodynamics and Pipe Material Properties on Biofilm Development Within Drinking Water Systems." *Environmental Technology (United Kingdom)* 41 (28): 3732–44. https://doi.org/10.1080/09593330.2019.1619844.

Douterelo, I., S. Husband, V. Loza, and J. Boxall. 2016. "Dynamics of Biofilm Regrowth in Drinking Water Distribution Systems." *Applied and Environmental Microbiology* 82 (14): 4155–68. https://doi.org/10.1128/AEM.00109-16.

Douterelo, I., R. Sharpe, and J. Boxall. 2014. "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* 117 (1): 286–301. https://doi.org/10.1111/jam.12516.

Ellis, B, R. Gentleman, F. Hahne, N. Le Meur, D. Sarkar, and M. Jiang. 2020. *flowViz: Visualization for Flow Cytometry*. Manual.

Ellis, B, Perry Haaland, Florian Hahne, Nolwenn Le Meur, Nishant Gopalakrishnan, Josef Spidlen, Mike Jiang, and Greg Finak. 2021. *flowCore: flowCore: Basic Structures for Flow Cytometry Data*. Manual.

Favere, Jorien, Benjamin Buysschaert, Nico Boon, and Bart De Gusseme. 2020. "Online Microbial Fingerprinting for Quality Management of Drinking Water: Full-Scale Event Detection." *Water Research* 170 (March): 115353. https://doi.org/10.1016/j.watres.2019.115353.

Flemming, H.-C. 2002. "Biofouling in Water Systems–Cases, Causes and Countermeasures." *Applied Microbiology and Biotechnology* 59 (6): 629–40. https://doi.org/10.1007/s00253-002-1066-9.

Gabrielli, Marco, Andrea Turolla, and Manuela Antonelli. 2021. "Bacterial Dynamics in Drinking Water Distribution Systems and Flow Cytometry Monitoring Scheme Optimization." *Journal of Environmental Management* 286 (May): 112151. https://doi.org/10.1016/j.jenvman.2021.112151.

Gatza, E, F Hammes, and E Prest. 2013. "Assessing Water Quality with the BD Accuri C6 Flow Cytometer." *White Paper. BD Biosciences*.

Hammes, Frederik, Michael Berney, Yingying Wang, Marius Vital, Oliver Köster, and Thomas Egli. 2008. "Flow-Cytometric Total Bacterial Cell Counts as a Descriptive Microbiological Parameter for Drinking Water Treatment Processes." *Water Research* 42 (1-2): 269–77. https://doi.org/10.1016/j.watres.2007.07.009.

Hemdan, Bahaa A., Gamila E. El-Taweel, Pranab Goswami, Deepak Pant, and Surajbhan Sevda. 2021. "The Role of Biofilm in the Development and Dissemination of Ubiquitous Pathogens in Drinking Water Distribution Systems: An Overview of Surveillance, Outbreaks, and Prevention." *World Journal of Microbiology and Biotechnology* 37 (2): 36. https://doi.org/10.1007/s11274-021-03008-3.

Koch, Christin, Ingo Fetzer, Hauke Harms, and Susann Müller. 2013a. "CHICan Automated Approach for the Detection of Dynamic Variations in Complex Microbial Communities." *Cytometry Part A* 83A (6): 561–67. https://doi.org/10.1002/cyto.a.22286.

Koch, Christin, Ingo Fetzer, Thomas Schmidt, Hauke Harms, and Susann Müller. 2013b. "Monitoring Functions in Managed Microbial Systems by Cytometric Bar Coding." *Environmental Science & Technology* 47 (3): 1753–60. https://doi.org/10.1021/es3041048.

Learbuch, K. L. G., H. Smidt, and P. W. J. J. van der Wielen. 2021. "Influence of Pipe Materials on the Microbial Community in Unchlorinated Drinking Water and Biofilm." *Water Research* 194 (April): 116922. https://doi.org/10.1016/j.watres.2021.116922.

Liu, Sanly, Cindy Gunawan, Nicolas Barraud, Scott A. Rice, Elizabeth J. Harry, and Rose Amal. 2016. "Understanding, Monitoring, and Controlling Biofilm Growth in Drinking Water Distribution Systems." *Environmental Science and Technology* 50 (17): 8954–76. https://doi.org/10.1021/acs.est.6b00835.

Monaco, Gianni, Hao Chen, Michael Poidinger, Jinmiao Chen, João Pedro de Magalhães, and Anis Larbi. 2016. “flowAI: Automatic and Interactive Anomaly Discerning Tools for Flow Cytometry Data.” *Bioinformatics* 32 (16): 2473–80. https://doi.org/10.1093/bioinformatics/btw191.

Neu, Lisa, and Frederik Hammes. 2020. “Feeding the Building Plumbing Microbiome: The Importance of Synthetic Polymeric Materials for Biofilm Formation and Management.” *Water (Switzerland)* 12 (6). https://doi.org/10.3390/w12061774.

Oksanen, Jari, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, et al. 2020. *Vegan: Community Ecology Package*. Manual.

Props, Ruben, Frederiek-Maarten Kerckhof, Pieter Monsieurs, Mohamed Mysara, Lieven Clement, and Nico Boon. 2018. *Phenoflow*. Manual.

Props, Ruben, Pieter Monsieurs, Mohamed Mysara, Lieven Clement, and Nico Boon. 2016. “Measuring the Biodiversity of Microbial Communities by Flow Cytometry.” *Methods in Ecology and Evolution* 7 (11): 1376–85. https://doi.org/10.1111/2041-210X.12607.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Manual. Vienna, Austria: R Foundation for Statistical Computing.

Rubbens, Peter, and Ruben Props. 2021. “Computational Analysis of Microbial Flow Cytometry Data.” *mSystems* 6 (1). https://doi.org/10.1128/msystems.00895-20.

Rubbens, Peter, Ruben Props, Cristina Garcia-Timermans, Nico Boon, and Willem Waegeman. 2017. “Stripping Flow Cytometry: How Many Detectors Do We Need for Bacterial Identification?” *Cytometry Part A* 91 (12): 1184–91. https://doi.org/10.1002/cyto.a.23284.

Rubbens, Peter, Ruben Props, Frederiek-Maarten Kerckhof, Nico Boon, and Willem Waegeman. 2021a. “Cytometric Fingerprints of Gut Microbiota Predict Crohn’s Disease State.” *The ISME Journal* 15 (1): 354–58. https://doi.org/10.1038/s41396-020-00762-4.

Rubbens, Peter, Ruben Props, Frederiek-Maarten Kerckhof, Nico Boon, and Willem Waegeman. 2021b. “PhenoGMM: Gaussian Mixture Modeling of Cytometry Data Quantifies Changes in Microbial Community Structure.” *mSphere* 6 (1). https://doi.org/10.1128/mSphere.00530-20.

Sadler, Marie C., Jérémy Senouillet, Simon Kuenzi, Luigino Grasso, and Douglas C. Watson. 2020. “Computational Surveillance of Microbial Water Quality With Online Flow Cytometry.” *Frontiers in Water* 2. https://doi.org/10.3389/frwa.2020.586969.

Safford, Hannah R., and Heather N. Bischel. 2019. “Flow Cytometry Applications in Water Treatment, Distribution, and Reuse: A Review.” *Water Research* 151 (March): 110–33. https://doi.org/10.1016/j.watres.2018.12.016.

Schleich, Caroline, Sandy Chan, Kristjan Pullerits, Michael D. Besmer, Catherine J. Paul, Peter Rådström, and Alexander Keucken. 2019. “Mapping Dynamics of Bacterial Communities in a Full-Scale Drinking

Water Distribution System Using Flow Cytometry." *Water* 11 (10): 2137. https://doi.org/10.3390/w11102137.

SLVFS. 2001. "Statens livsmedelsverks föreskrifter om dricksvatten."

Søborg, Ditte Andreasen, Torben Lund Skovhus, Bo Højris, Jørn-Ole Andreasen, and Kurt Brinkman Kristensen. 2020. "Characterizing the Development of Biofilm in PE Pipes Through 1.5 Years in the Non-Chlorinated Danish Drinking Water Distribution System." In *IWA Biofilms 2020 Virtual Conference: Emerging Trends and Developments in Biofilm Processes*.

Van Nevel, Sam, Benjamin Buysschaert, Karen De Roy, Bart De Gusseme, Lieven Clement, and Nico Boon. 2017a. "Flow Cytometry for Immediate Follow-up of Drinking Water Networks After Maintenance." *Water Research* 111: 66–73. https://doi.org/10.1016/j.watres.2016.12.040.

Van Nevel, Sam, S. Koetzsch, C. R. Proctor, M. D. Besmer, E. I. Prest, J. S. Vrouwenvelder, A. Knezev, N. Boon, and F. Hammes. 2017b. "Flow Cytometric Bacterial Cell Counts Challenge Conventional Heterotrophic Plate Counts for Routine Microbiological Drinking Water Monitoring." *Water Research* 113 (April): 191–206. https://doi.org/10.1016/j.watres.2017.01.065.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2013. "Animation: An R Package for Creating Animations and Demonstrating Statistical Methods." *Journal of Statistical Software* 53 (1): 1–27.