Flow cytometric bacterial cell counts challenge conventional heterotrophic plate counts for routine microbiological drinking water monitoring

S. Van Nevel\(^1\), S. Koetzsch\(^2\), C. R. Proctor\(^{2,3}\), M. D. Besmer\(^2\), E. I. Prest\(^4\), J. S. Vrouwenvelder\(^{4,5,6}\), A. Knezev\(^7\), N. Boon\(^8\), F. Hammes\(^2,5\)

\(^1\)Center for Microbial Ecology and Technology (CMET), Ghent University, Coupure Links 653, B-9000 Gent, Belgium
\(^2\)Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland
\(^3\)Department of Environmental Systems Science, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, Switzerland
\(^4\)Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, The Netherlands
\(^5\)Wetsus, Centre of Excellence for Sustainable Water Technology, Oostergoweg 9, 8911 MA Leeuwarden, The Netherlands
\(^6\)King Abdullah University of Science and Technology (KAUST), Water Desalination and Reuse Center (WDRC), Division of Biological and Environmental Science and Engineering (BESE), Thuwal 23955-6900, Saudi Arabia
\(^7\)Het Waterlaboratorium, J.W. Lucasweg 2, 2031 BE Haarlem, The Netherlands

\(^5\) Shared last author

\(^*\) Correspondence to: Frederik Hammes, Eawag, Swiss Federal Institute of Aquatic Science and Technology; Überlandstrasse 133, CH-8600 Dübendorf, Switzerland; phone: +41 (0)58 765 51 52; fax: +41 (0)58 765 55 47; E-mail: frederik.hammes@eawag.ch Webpage: www.eawag.ch.

This document is the accepted manuscript version of the following article:

This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/
Abstract

Drinking water utilities and researchers continue to rely on the century-old heterotrophic plate counts (HPC) method for routine assessment of general microbiological water quality. Bacterial cell counting with flow cytometry (FCM) is one of a number of alternative methods that challenge this status quo and provide an opportunity for improved water quality monitoring. After more than a decade of application in drinking water research, FCM methodology is optimised and established for routine application, supported by a considerable amount of data from multiple full-scale studies. Bacterial cell concentrations obtained by FCM enable quantification of the entire bacterial community instead of the minute fraction of cultivable bacteria detected with HPC (typically < 1% of all bacteria). FCM measurements are reproducible with relative standard deviations below 3% and can be available within 15 minutes of samples arriving in the laboratory. High throughput sample processing and complete automation are feasible and FCM analysis is arguably less expensive than HPC when measuring more than 15 water samples per day, depending on the laboratory and selected staining procedure(s). Moreover, many studies have shown FCM total (TCC) and intact (ICC) cell concentrations to be reliable and robust process variables, responsive to changes in the bacterial abundance and relevant for characterising and monitoring drinking water treatment and distribution systems. The purpose of this critical review is to initiate a constructive discussion on whether FCM could replace HPC in routine water quality monitoring. We argue that FCM provides a faster, more descriptive and more representative quantification of bacterial abundance in drinking water.

Keywords: cultivation; microbiological drinking water quality; flow cytometry (FCM); heterotrophic plate counts (HPC); routine water monitoring;
1. Introduction

Drinking water treatment and distribution systems are designed and operated to safeguard the hygienic quality and ensure the aesthetic quality of the water from source to tap. With this in mind, monitoring is a non-negotiable and legislated requirement worldwide. There is a recognised and accepted need to monitor, characterise and understand the general microbiological performance/response of individual treatment steps, especially under changing environmental and operational conditions (Reasoner, 1990; Lautenschlager et al., 2013; Pinto et al., 2012). There is furthermore the need to monitor the general microbiological behaviour of treated water during distribution, particularly to detect potential contamination or deterioration due to biologically unstable water or distribution systems (Prest et al., 2016; Pinto et al., 2014). From a water utility perspective, microbiological methods used for such general water quality monitoring would ideally meet the criteria of being relevant, simple, reliable, rapid and cost-effective.

Heterotrophic plate counts (HPC) is the descriptive term for a group of similar methods used routinely by water utilities for general microbiological monitoring of drinking water. The method enumerates a variety of heterotrophic bacteria that are cultivable on semi-solid nutrient-rich media under defined incubation conditions (Allen et al., 2004; Rice et al., 2012; Gensberger et al., 2015). The basic HPC method was proposed well over a century ago (Koch, 1881) and was for a considerable time regarded as indicative of the hygienic quality of drinking water (Sartory, 2004). However, during the 1980's and 1990's it was decisively concluded that HPC measurements have no hygienic relevance (WHO, 2003; Sartory, 2004). Increasingly, HPC was regarded as a process variable to monitor a range of events and/or processes relevant to the general microbiological quality of drinking water in treatment and distribution systems (Reasoner, 1990; WHO, 2003; Sartory, 2004). For most of the previous century, HPC was regarded as the best available technology for drinking water process monitoring, and HPC data contributed towards considerable advances in our understanding of drinking water microbiology (Chowdhury, 2012).
In the last two decades, a number of powerful quantitative and molecular methods have emerged for water analysis (e.g., adenosine tri-phosphate (ATP) analysis, flow cytometry (FCM), 16S rRNA gene amplification and sequencing). Application of these new techniques showed that bacterial communities in drinking water were vastly more abundant and complex than what was previously understood from research based on cultivation-dependent methods (Berry et al., 2006; Hoefel et al., 2003). Current evidence suggests that the drinking water microbiome consists of as many as 9’000 distinct taxa, with total numbers ranging between 1’000 – 500’000 bacteria ml\(^{-1}\) (Proctor and Hammes, 2015; Bautista-de los Santos et al., 2016).

FCM is one exciting “new” method capable of rapidly and accurately counting and characterising practically all bacteria in drinking water. FCM has already been used for microbiological characterisation and quantification in natural aquatic habitats for several decades (Legendre and Yentsch, 1989; Trousselier et al., 1993), but was only recently introduced as a method for drinking water analysis (Hoefel et al., 2003, 2005a, 2005b; Hammes et al., 2008). All early drinking water FCM studies immediately confirmed the growing awareness of the considerable numerical divide between the total bacteria and the fraction of cultivable bacteria in drinking water (Hoefel et al., 2003; Hammes et al., 2008). Multiple drinking water studies comparing FCM and HPC data argued that FCM is more meaningful for use as a process variable, and questioned the future relevance of HPC measurements (Hoefel et al., 2005a; Hammes et al., 2008; Ho et al., 2012; Liu et al., 2013a; Gillespie et al., 2014).

Here we evaluate the last 15 years of FCM developments and applications in the field of drinking water analysis, and we argue that routine HPC analysis no longer qualifies as the best available technology for the above-stated criteria of relevance, simplicity, reliability, speed and cost-effectiveness. The purpose of this critical review is to initiate a constructive discussion on whether FCM can and should replace HPC as the primary process variable in routine microbiological water quality monitoring. We approached this by briefly assessing the history, advantages and disadvantages of HPC as a process variable, followed by a consideration of several alternative methods that may be suitable as alternatives. We then argue the case
for FCM as the method of choice, covering both the advantages and disadvantages of the methodology. We also compare FCM to HPC and ATP with extensive data sets collected over the last decade and outline how FCM could be applied as a monitoring method in the future.

2. Enumeration of bacteria by HPC

2.1. 130 years of HPC development and application

Around 1850, John Snow demonstrated the relationship between cholera prevalence and water consumption from a certain well and concluded (without knowing the causative agent) that drinking water was the transmitter of the disease (Sedlak, 2014). At that time, smell, appearance, taste and basic chemical analysis were the only tools available to water utilities for assessing drinking water quality (Payment et al., 2003). This changed considerably after Robert Koch published his gelatine plate method in 1881, for the first time offering the possibility to isolate and cultivate pure bacterial colonies and to enumerate bacteria (Koch, 1881). In the following years, the method was improved by replacing gelatine with agar, and was applied routinely to full-scale treatment systems for assessing particle filtration efficacy and microbiological water quality (Frankland and Frankland, 1894; Payment et al., 2003; Reasoner, 2004). In the same period, Koch proposed a limit of 100 colony forming units per millilitre (cfu mL\(^{-1}\)) for preventing cholera outbreaks (Koch, 1893; Exner et al., 2003). During the 130 years following its first publication (Koch, 1881), the HPC method underwent a range of modifications including new media compositions and different incubation times and temperatures (Sartory, 2004; Reasoner, 2004). In the context of drinking water routine monitoring, these modifications were aimed towards detecting the largest possible fraction of bacteria in a given sample (Frankland and Frankland, 1894; Reasoner and Geldreich, 1985). As a consequence of the numerous method modifications, standardised HPC methods cover a wide range of conditions, including different media formulations like plate count agar (PCA) or R2A-agar (see Table S1), different incubation temperatures ranging from 20 °C to 40 °C and incubation times ranging from hours to weeks (WHO, 2003;
Allen et al., 2004; Rice et al., 2012). These variations are not inconsequential: it is well known that variations in incubation conditions affect the number and composition of bacteria recovered (LeChevallier et al., 1980; Reasoner and Geldreich, 1985; Reasoner, 1990; Gensberger et al., 2015). Nevertheless, even with these modifications, Koch’s original HPC method and proposed limits associated with it are essentially still present in drinking water legislation worldwide (Table 1).

Operational limits for HPC are still regularly incorporated in drinking water legislation (Table 1). Maximum values range from 20 to 500 cfu mL$^{-1}$ depending on the country and the sampling location (Allen et al., 2004). In some countries, maximum values are increasingly replaced by a guideline stating that ‘no abnormal change (NAC)’ should be detected, although guidelines are not clear on how NAC is defined. Some countries have only very recently changed their HPC guidelines. Compared to one decade ago, the European Union, Canada and Australia for example have excluded their HPC upper limit in drinking water legislation (Radcliff, 2003), even though individual EU countries still maintain HPC upper limit guideline values (Table 1).

2.2. Advantages and applications of HPC

One major advantage of HPC data is that a positive result is an undeniable indicator of viability for the cells that formed colonies. Given the well-known lack of silver-bullet methods distinguishing between life and death in bacteria, the ability to identify viable organisms should not be underestimated (Hammes et al., 2011). In addition, changing the incubation conditions enables researchers to isolate different types of organisms as pure cultures, which has through the years facilitated detailed characterisations of numerous drinking water bacteria. Moreover, HPC methods are relatively low cost, simplistic and operators can compare HPC data to more than a century of historical data worldwide to aid interpretation and decision-making (Douterelo et al., 2014).

The application of HPC as an important variable for monitoring a wide range of microbiologically relevant
events and processes in drinking water treatment and distribution systems has been reviewed and
discussed extensively in the works of Reasoner (1990), Sartory (2004), Allen et al. (2004) and Chowdhury
(2012) (Table 2). Some specific examples include: (i) assessment of chlorine disinfection efficacy (e.g.
LeChevallier et al., 1984); (ii) studying the bacteriological activity in biofiltration systems (e.g., Camper et al.,
1986); (iii) tracking microbiological changes as a result of regrowth and biological instability of drinking
water (e.g., Francisque et al., 2009; Uhl and Schaule, 2004; Prest et al., 2016a); (iv) quantifying batch
growth of bacteria during incubation (i.e. stagnation) of nano-filtered drinking water (Liikanen et al., 2003)
and unfiltered drinking water (Uhl and Schaule, 2004), and during overnight stagnation in building plumbing
(e.g., Pepper et al., 2004; Lautenschlager et al., 2010). Based on evidence such as these, the same WHO
expert meeting that found that HPC values have no hygienic relevance (above), concluded that HPC can be
used to monitor a range of relevant microbiological processes in drinking water (WHO, 2003). It is,
however, important to note that the WHO proposes no specific guidelines (i.e., methods, thresholds,
interpretations) on monitoring HPC in drinking water.

2.3. Disadvantages: what is HPC missing?

HPC has basic drawbacks in that it is time and labour consuming and the time-to-result ranges between 2
and 10 days, which is not ideal for fast decision-making and reactions to problems. Moreover, we argue
below that data from routine HPC methods represent neither the abundance nor the composition of
bacteria in drinking water, thus seriously drawing into question the value of this method as a relevant
process variable.

2.3.1. Abundance

Right from the start (Frankland and Frankland, 1885) and throughout the development of HPC methods
(Reasoner and Geldreich, 1985), there was a strong focus on establishing HPC methods capable of detecting
the largest possible fraction of bacteria in a water sample. Reasoner and Geldreich (1985) noted that “it
seems appropriate to use a medium that will provide the highest estimate (of viable bacteria) possible to
follow changes in bacterial quality of water related to treatment or water quality degradation in the distribution system”. However, already in the late 19th century some of the first microbiologists realised that the number of colonies growing on agar plates was not reflecting the real number of bacteria present in the water sample observed with direct microscopy (Winterberg, 1898; Amann, 1911). Although branded ‘non-selective’ and clearly intended to recover a wide range of microorganisms from water, HPC media and methods are highly selective for bacteria growing under the specific incubation conditions (Allen et al., 2004; Gensberger et al., 2015). As Frankland and Frankland (1894) aptly stated: “It might be supposed that it would be easy to find a medium which would suit the requirements of all micro-organisms [...] but, as a matter of fact [...] media which are suitable for the growth of some are utterly unsuitable for the cultivation of others”. Early comparisons showed that microscopic counts were detecting up to 150 times more bacterial cells compared to HPC (Amann, 1911). By the mid-20th century, this was universally recognised (Lewis et al., 2010) and later branded ‘The Great Plate Count Anomaly’ (Staley and Konopka, 1985). More recent estimations of the fraction of the total bacterial community detected by HPC are usually lower than a few percent in drinking water. For example, values of 0.01 % (WHO, 2003), 0.001 – 6.5 % (Hammes et al., 2008) and 0.05 – 8.3 % (Burtscher et al., 2009) have been reported in recent literature, depending on the water sample origin and on the applied HPC and total cell count methods.

Multiple studies have examined the so-called ‘unseen majority’, referring to the major part of the bacteria not detected with conventional HPC methods but observed with microscopy, FCM, next generation sequencing and other cultivation-independent methods. Seminal reviews on the topic are available for more detailed information (Kell et al., 1998; Bogosian and Bourneuf, 2001; Green and Keller 2006; Oliver, 2010; Epstein, 2013), but the overall consensus is that, apart from lethally injured or dead bacteria, this ‘unseen majority’ consists of two basic groups: (i) bacteria belonging to strains that are regarded as cultivable but for some reason enter into a so-called ‘viable but not cultivable’ (VBNC) state (Bogosian and Bourneuf, 2001) and (ii) the so-called ‘uncultivable bacteria’, which are simply not cultivable by conventional HPC methods.
First of all, it is not reasonable to argue that most bacteria in drinking water are lethally injured or dead. Several studies have examined the fraction of intact, active and respiring bacteria in drinking water. In non-chlorinated drinking water, 20 – 70 % of the total cells detected with FCM after SYBR Green I staining were demonstrated to be active (esterase activity measured with CFDA staining and FCM) and 70 – 80 % of the cells had intact membranes (measured with propidium iodide staining and FCM) (Berney et al., 2008; Helmi et al., 2014a). When chlorine residual was lost in a chlorinated system, samples contained 50 – 60 % cells with intact membranes (Kahlisch et al., 2012). Chlorinated tap water samples have also shown an increase in intact cells and ATP, a measure for active biomass, during extended network residence times (Nescerecka et al., 2014), indicating regrowth of living cells. While no single method is viewed as conclusive when assessing the viability and activity of complex microbiological communities (Hammes et al., 2011), these high percentages consistently indicate that the majority of cells are alive.

With respect to the VBNC state, it has been shown that normally cultivable bacteria can remain undetected by HPC due to the presence of injured cells, which may recover and regain cultivability. Another explanation is dormant cells (also referred to as resting cells or persisters), which are physiologically in a VBNC state due to adverse environmental conditions, but can get activated at more favourable conditions or even at random moments and then become cultivable again (Kell et al., 1998; Bogosian and Bourneuf 2001; Oliver, 2005; Epstein, 2013).

Uncultivable bacteria that are not detected with conventional HPC methods have been the subject of considerable discussion. One proposed explanation for so-called uncultivability is the presence of excessive nutrient concentrations in any type of HPC medium compared to drinking water. For example, R2A-agar, often suggested to be low in nutrients and therefore more suitable for cultivating drinking water bacteria (Reasoner and Geldreich, 1985), has an organic carbon concentration up to 800 times in excess of drinking water (Hammes et al., 2008). Efforts for optimizing bacterial growth conditions, including the adaptation of macro- and micronutrients, relieving oxidative stress, the use of synthetic environments mimicking the natural conditions and new cultivation techniques, enabled the growth of some previously uncultivable
bacterial species (Kaeberlein et al., 2002; Rappe et al., 2002; Sangwan et al., 2005; Wang et al., 2009; D’Onofrio et al., 2010). However, millions of different microbiological species are thought to exist and over 99% have never been cultivated (Achtman and Wagner, 2008; D’Onofrio et al., 2010).

2.3.2. Composition

Multiple studies have identified HPC isolates in order to better understand the hygienic and/or practical relevance of HPC data (e.g., LeChevallier et al., 1980; Reasoner, 1990; Allen et al., 2004; Gensberger et al., 2015). For example, LeChevallier et al. (1980) concluded that chlorination selected for gram-positive genera and that distributed drinking water comprised up to 30% opportunistic pathogens. These studies all identified a broad variety of genera and showed that different organisms were isolated based on sample origin and HPC incubation conditions. However, during the last two decades it has become clear that cultivation-isolation based methods completely underestimate the complexity of the drinking water microbiome (Proctor and Hammes, 2015; Bautista-de los Santos et al., 2016). Unfortunately, only a few studies have directly compared the composition of HPC isolates with the total community composition based on 16S rRNA amplification (Farnleitner et al., 2004; Burtscher et al., 2009). Burtscher et al. (2009) concluded that 16S profiles after HPC isolation and directly from water samples differed completely, and showed that the dominant HPC community comprised primarily copiotrophic bacteria while the dominant drinking water community comprised typical oligotrophic aquatic bacteria.

In summary, there is agreement in literature that HPC data vastly underestimate the actual bacterial concentrations in drinking water, and there is some clear evidence that the bacteria detected on HPC plates are not the dominant species in water samples. Moreover, there exists to our knowledge no clear evidence that this small cultivable fraction of bacteria detected by HPC is properly and consistently representative for behaviour of the entire bacterial community in any given water, and for that matter relevant to the challenges in the various application areas of treatment and distribution systems (Table 2). This point is underscored, for example, in the context of bacterial regrowth during distribution. A recent book by leading
drinking water experts, focussing solely on regrowth problems, concluded unequivocally that: “HPC, in combination with the standard of 100 cfu mL⁻¹, is not suited for assessing the level of microbial growth in distribution systems” (van der Kooij and Veenendaal, 2014) and “HPC [...I are not ideal parameters for regrowth assessment” (van der Kooij and van der Wielen, 2014b).

3. Alternative methods for bacterial quantification are available

Given the differences between various HPC regulations and guidelines (Table 1), the very large and inconsistent difference between the number and composition of bacteria detected with HPC and the actual bacterial content of drinking water (above), and the complex challenges for which process variables are needed (Table 2), it is imperative to question whether HPC is still the way to go for routine microbiological water analysis in the 21st century. Drinking water utilities and researchers clearly need accurate and reliable methods to quantify and characterise microbiological changes during treatment and distribution (Table 2; Prest et al., 2016b), so abolishment of HPC methods without a suitable replacement would in itself be a serious step backwards. For bacterial (re)growth during drinking water distribution, van der Kooij and van der Wielen (2014b) stated that “assessment and monitoring of the level of microbial regrowth requires a method for the quantification of the biomass of all active bacteria” and suggested as examples ATP, total cell concentrations and the total DNA concentration in samples. In fact, since the early establishment of HPC, a suite of other methods has been developed to quantify bacteria in order to study/monitor general microbiological quality of drinking water. Table 3 compares six of these methods, but it is acknowledged that several more alternatives exist (Lopez-Roldan et al., 2013). Microscopy cell counting is an established methodology and can be combined with a broad range of fluorescent and non-fluorescent dyes (e.g., DAPI, acridine orange) and probes (e.g., labelled antibodies) to assess total bacteria, viable bacteria or specific bacterial sub-groups. Aside from some automated approaches (Zeder and Pernthaler, 2009), microscopy is far too labour-intensive and operator-subjective for routine application and is therefore used predominantly as a research tool. Molecular assays, such as 16S-rRNA quantitative real-time PCR (qPCR),
can serve as indirect measurements of absolute bacterial counting by measuring the abundance of 16S rRNA gene copies in extracted DNA (Nadkarni et al., 2002; Hoefel et al., 2005b). When the same approach is applied to extracted RNA, the information can serve as viability assay, and with more specific primers, it can target bacterial sub-groups. While continuous and rapid progress is being made in this field, routine application is currently limited by time-intensive nucleic acid extraction steps, PCR amplification bias and difficulties with respect to viability assessment (Nocker et al., 2007). Moreover, recent data suggests that current 16S-based approaches may not accurately represent the full extent of diversity in the system (Hug et al., 2016). To overcome PCR bias, it is also possible to simply quantify the total amount of extracted DNA or RNA, but this approach is less sensitive than PCR methods and still subject to nucleic acid extraction biases (Hwang et al., 2012; Salter et al., 2014). Measurement of ATP is a noteworthy alternative method, which is fast, relatively simple and representative of all viable organisms in a water sample (Hammas et al., 2010a; van der Wielen and van der Kooij, 2010). ATP analysis has been championed for several decades as a cultivation-independent measure of viable biomass in general (Holm-Hansen and Booth, 1966; Karl, 1980), and more recently specifically for drinking water and biofilm quantification (van der Kooij et al., 2003; Vrouwenvelder et al., 2008; Hammas et al., 2010a; van der Wielen and van der Kooij, 2010; Nescerecka et al., 2016b; Magic-Knezev and van der Kooij, 2004). However, routine application of ATP analysis is currently limited by a lack of standardised comparable methods and interference from inorganic compounds in the water (e.g., iron, manganese) as well as non-bacterial and/or extracellular ATP (Hammas et al., 2010a). Finally, FCM for measurement of bacterial cell concentrations has developed tremendously during the last decade (discussed below).

4. FCM cell concentrations as an alternative to HPC

FCM is a fast, accurate, quantitative and reproducible technique for counting the total number of bacteria when a general nucleic acid stain is used (Hammas et al., 2008; Wang et al., 2010; Prest et al., 2013) or the number of viable bacteria when combined with viability stains (Berney et al., 2008; Helmi et al., 2014b).
More recently, the technique was expanded towards creating FCM fingerprints of bacterial communities to allow more detailed characterisation of those bacterial communities (De Roy et al., 2012; Prest et al., 2013; Koch et al., 2014). It is argued here that FCM cell concentrations can replace HPC as a suitable process variable for routine drinking water analysis due to the (i) high level of information, (ii) high accuracy and reproducibility, (iii) speed and automation possibilities and (iv) overall reasonable costs compared to other methods (Figure 1).

4.1. Historical FCM developments with respect to drinking water analysis

FCM characterises and quantifies individual suspended particles by passing them one by one through a light source, typically a laser beam (Figure 1). Fluorescent particles are excited by the light source and emit light at a higher wavelength (Shapiro, 2003) (Figure 1). Particles of interest can either be autofluorescent (e.g. algae that contain chlorophyll), or be made fluorescent, such as bacteria after staining with fluorescent dyes (e.g., SYBR Green I) (Hammes and Egli, 2010). FCM was initially applied for the analysis of mammalian cells, but in 1977 it was introduced by microbiologists to characterise suspended bacteria (Bailey et al., 1977; Paauw et al., 1977). The method was at first not broadly used in microbiology due to expensive instrumentation and technical difficulties stemming from the small size of bacteria compared to mammalian cells (Wang et al., 2010). However, developments in hardware performance, costs and ease of handling and the concomitant emergence of novel stains for specific bacterial features finally made FCM more popular in microbiology from the 1990s onwards (Hammes and Egli, 2010). To our knowledge, Vesey et al. (1991) applied FCM for the first time for drinking water analysis for detecting Cryptosporidium oocysts, while Appenzeller et al. (2002) reported the first FCM-based detection of bacteria in drinking water, studying the sorption of E. coli on FeOOH. One year later, Hoefer et al. (2003) applied FCM for detecting the physiologically active bacterial community in drinking water and highlighted the difference in orders of magnitude between HPC and FCM data. Research applying FCM in drinking water now includes an extensive range of studies. For example, FCM was used for the characterisation of water treatment processes (Hammes et al., 2018; Ho et al., 2012; Van Nevel et al., 2012; Vital et al., 2012; Helmi et al.,
specific disinfection processes (Phe et al., 2005; Ramseier et al., 2011; Wert et al., 2013) and viability assessment (Pianetti et al., 2005; Berney et al., 2008). Regrowth and biological stability were studied in both chlorinated and non-chlorinated drinking water distribution networks (Hoefel et al., 2005a, 2005b; Vital et al., 2012; Lautenschlager et al., 2013; Liu et al., 2013a; Prest et al., 2013; Gillespie et al., 2014; Nescerecka et al., 2014; Wen et al., 2014; Van Nevel et al., 2016) and building/premises plumbing (Lautenschlager et al., 2010; Lipphaus et al., 2014). Measurements of assimilable organic carbon were performed with FCM as the bacterial enumeration method (Hannes and Egli, 2005; Van Nevel et al., 2013a). Specific pathogens were detected and pathogen growth potential was assessed with FCM (Vital et al., 2007, 2010; Yang et al., 2010; Keserue et al., 2012; Van Nevel et al., 2013a). Finally, FCM has been used for studying drinking water bacterial ecology, by detecting changes in the community composition and creating FCM fingerprints (i.e. statistical interpretations of raw FCM data) of specific water samples (Kahlisch et al., 2010; De Roy et al., 2012; Douterelo et al., 2014; Prest et al., 2014).

Monfort and Baleux (1992) compared fluorescence microscopy with FCM total cell concentrations (TCC) and concluded that FCM is reliable to enumerate bacteria in both pure cultures and natural samples. In order to distinguish bacteria from abiotic particles, various general nucleic acid stains have been tested (Wang et al., 2010), and the stains SYBR Green I, SYBR Green II and SYTO 9 were found to deliver high quality results for total bacterial cell counting in fresh water environments (Lebaron et al., 1998; Hammes et al., 2008). While different stains would in theory serve the purpose, a recent extensive standardisation of the FCM-TCC method was based on SYBR Green I, and this standardised method was accepted as a guideline method into Swiss drinking water legislation (SLMB, 2012; Prest et al., 2013). As major advantages, the FCM-TCC method was shown to be fast (10 minutes staining time, < 1 minutes analysis time), accurate (< 3 % relative standard deviation on all measurements), reproducible (<7 % variability between different laboratories) and suitable for high throughput processing (SLMB, 2012; Prest et al., 2013; Van Nevel et al., 2013b) (Figure 1).

4.2. FCM provides relevant quantitative process information
The consensus application value of HPC as a routine drinking water variable is the enumeration of bacteria, either for evaluating/monitoring treatment processes (e.g., filtration or disinfection efficiency) or for assessing the general microbiological quality of raw and treated water (Section 2; Table 2). Here we argue that multiple pilot scale and full-scale studies have demonstrated that for every HPC application-area (Table 2) there are several examples where FCM performed equally or better in providing accurate quantitative and qualitative information that is relevant to the process and/or system under investigation. Table 2 highlights selected FCM studies and specific examples that are presented in the Supplementary Information (SI) and discussed below.

Source water contamination: FCM analysis of temporal fluctuations in source water was reported previously (Besmer et al., 2014, 2016a, 2016b; Kötzsch and Sinreich, 2014; Besmer and Hammes, 2016) and is illustrated in Figure 2 and Example S1. In many of these examples, high frequency online FCM was used over extended time periods to quantify dry weather baseline values and microbiological changes caused by regional precipitation events with respect to their frequency and magnitude. Figure 2 shows a specific example from a karstic spring, which is vulnerable to precipitation-induced contamination. The spring was sampled hourly in the 24 hours immediately following a precipitation event (15.4 mm in 24 h), which increased spring discharge from 1’260 L min\(^{-1}\) to 1’400 L min\(^{-1}\) within hours of the event. The increased spring discharge coincided with a substantial increase in FCM-TCC from 1.4 x 10\(^{5}\) cells mL\(^{-1}\) to a maximum of 3.7 x 10\(^{5}\) cells mL\(^{-1}\). The FCM data gives a clear quantitative and qualitative description of the microbiological response in the spring with respect to the temporal evolution and magnitude. In contrast, the HPC data fluctuated on hourly time scales between 200 and 2’000 cfu mL\(^{-1}\), neither describing the trend, nor the magnitude of the microbiological response in the spring to the precipitation event.

Monitoring treatment processes: FCM was used to characterise drinking water treatment processes in a large number of studies (e.g., Hoefel et al., 2005a, 2005b; Hammes et al., 2008, 2010b; Ho et al., 2012; Vital et al., 2012; Lautenschlager et al., 2014; Helmi et al., 2014b; Example S2). In studies by Hammes et al.
(2008, 2010b) and in Example S2 (a subset of data from Vital et al. (2012)), disinfection through ozonation was not correctly measured with routine HPC analysis, due in part to low cultivability of bacteria in the source water, but could clearly be shown with FCM-TCC. Helmi et al. (2014b) measured increased bacterial abundance after activated carbon filtration with different FCM methods, but these increases went undetected in some cases with routine HPC methods. Similar changes in bacterial numbers with various biofiltration steps were further elucidated in Hammes et al. (2010b), Velten et al. (2011) and Vital et al. (2012). In one of the clearest examples of treatment characterisations, Ho et al. (2012) monitored raw water and the microbiological performance of four different drinking water treatment options on pilot scale during 12 months. In that study, HPC data showed no difference between the different treatment options. In stark contrast, the FCM data clearly characterised the significant differences between the treatment options, which led the authors to conclude that “[...] FCM was shown to be a better monitoring tool than HPCs, which allowed for more definitive comparisons to be made between each of the treatment streams” (Ho et al., 2012).

Growth/instability during distribution: FCM characterisation of distributed drinking water was reported in Hoefel et al. (2005a), Vital et al. (2012), Lautenschlager et al. (2013), Nescerecka et al. (2014), van Nevel et al. (2016) and Prest et al. (2016a) and is illustrated in Examples S3 – S4. In an early study, Hoefel et al. (2005a) described bacterial increases in excess of 1 x 10^5 cells mL^{-1} measured with FCM during distribution of chloraminated drinking water, which went undetected with conventional HPC measurements. These authors concluded that their data “[...] casts doubt on the use of HPC data as an indicator of distribution system integrity following a loss of chloramine residual caused by ammonia-oxidizing bacteria” (Hoefel et al., 2005a). In a similar vein, Nescerecka et al. (2014) demonstrated substantial microbiological growth from about 1 x 10^4 cells mL^{-1} up to 4 x 10^5 cells mL^{-1} during distribution of unstable chlorinated drinking water with FCM after viability staining (measuring intact cells after SYBR Green I and propidium iodide staining). A subset of this data is shown in Example S3, showing that routine HPC measurements failed to recognise the magnitude of instability in that particular system, exceeding the guideline value of 100 cfu mL^{-1} in only one of 39 samples. A recent 2-year study of a full-scale distribution system without residual chlorine (Prest et
al., 2016a; Example S4) revealed a distinct seasonal variation in flow cytometric intact cell concentration (FCM-ICC) data from about 5 x 10^4 cells mL^{-1} up to 2 x 10^5 cells mL^{-1}. Routine HPC measurements corroborated the increased bacterial numbers in summer, but failed to quantify the magnitude of change in the system. In fact, HPC numbers exceeded the Dutch standard of 100 cfu mL^{-1} only twice in the entire measurement period (Example S4). More importantly, the FCM data enabled the researchers to pinpoint the final biological filters in the treatment plant as the primary source of change, and growth during distribution as the secondary source (Prest et al., 2016a). In a final example, Gillespie et al. (2014) studied biological stability in chlorinated and chloraminated drinking water distributions and reported a correlation between free chlorine concentrations and FCM-ICC data. Interestingly, while we do not recommend here that FCM replace faecal indicator monitoring, these authors observed that most samples that tested positive for coliforms also showed elevated FCM-ICC values and concluded that FCM measurements are a potentially complementary tool for detecting quality failures in systems with residual disinfectants.

Water storage and building plumbing: FCM studies of bacteriological changes during drinking water storage and/or stagnation prior to consumption include Siebel et al. (2008), Wang et al. (2008), Lautenschlager et al. (2010), Mimoso et al. (2015), Lipphaus et al. (2014) and Gillespie et al. (2014). Mimoso et al. (2015) used online FCM to characterise bacterial growth during stagnation of gravity driven membrane filtered river water. The high frequency FCM data enabled accurate and detailed quantification of the rate and extent of bacterial growth, mirroring similar FCM studies for bottled water (Wang et al., 2008) and older HPC growth studies (e.g., Uhl and Schaule, 2004; Reasoner, 1990). On the building plumbing level, Lautenschlager et al. (2010) quantified the extent of bacterial increases following overnight stagnation of non-chlorinated drinking water in residential households. This study demonstrated considerable increases in all microbiological variables (i.e. FCM, ATP and HPC), although HPC did not detect the magnitude of change measured with FCM and ATP (Lautenschlager et al., 2010). As a final example, Lipphaus et al. (2014) studied multiple taps in residential and office building plumbing systems receiving chlorinated drinking water, using FCM viability analysis (SYBR green I and propidium iodide staining). The authors showed elevated cell concentrations in taps that were infrequently used, and washout from single taps as a function of water
458 use. Example S5 shows FCM and HPC data for a similar washout study from a drinking water tap during
daily use.

460 In the above section we described several examples where straightforward FCM analysis was in our opinion
461 equal or superior to HPC analysis as a process variable and indicator of general microbiological water
462 quality. This does not diminish in any way the extended history of HPC applications in water quality
463 monitoring. There may well also be cases where specific focus on a sub-group of bacteria provides insights
464 into small microbiological changes that will go undetected with a total cell concentration measurement
465 approach (Vital et al., 2012). With respect to FCM applications, there is an on-going need to systematically
466 document the use of FCM data (e.g., case studies in SI) and particularly to link microbiological changes
467 observed with FCM to specific events and/or problems (e.g., system malfunctioning, consumer complaints).
468 In this regard, it is recognised that the current data sets are dominated by studies from drinking water
469 systems without residual disinfectants, and additional studies on chlorinated systems are needed to fill this
470 gap.

471 4.3. Added qualitative value of FCM: fingerprinting and community interpretations

473 Apart from the combination of FCM with various fluorescent dyes and the determination of TCC and ICC
474 (section 5.1), the fluorescence and scatter detectors deliver information that can be used for a more
475 detailed analysis by creating a microbiological fingerprint of the water. Essentially, these FCM fingerprinting
476 methods are statistical analyses of multivariate FCM data (e.g., size, fluorescent colour, fluorescence
477 intensity) that varies in complexity and which represents the distribution of raw data in FCM plots (De Roy
478 et al., 2012; Prest et al., 2013). Such FCM fingerprinting methods are sensitive for detecting small changes
479 and shifts within the bacterial community, which are overlooked by enumeration alone (De Roy et al., 2012;
Prest et al., 2013; Koch et al., 2014; Prest et al., 2014). For example, Prest et al. (2013) detected the
480 contamination of drinking water by 4 % wastewater effluent bacteria, based on the combination of FCM
481 cell concentrations and a basic FCM fingerprinting method (quantifying high (HNA) and low (LNA) nucleic
acid content bacteria), and showed in a separate study (Prest et al., 2014) that changes in the microbiological community composition by 16S rRNA gene analysis can be detected early through changes in the FCM fingerprint, a method further elucidated by Props et al. (2016). De Roy et al. (2012) likewise used advanced fingerprinting methods to detect the bacterial physiology adaptations within three hours after minor nutrient addition in drinking water. Finally, Van Nevel et al. (2016) examined the biological stability of a drinking water network, where certain water samples showed elevated FCM cell concentrations. The application of fingerprinting methods to these samples suggested that bacterial growth was taking place to a large extent in the household taps rather than in the drinking water network (Van Nevel et al., 2016). With respect to FCM fingerprinting methods, the challenge is to establish methodological toolboxes that are sensitive and robust, standardized, applicable to FCM data generated with different instrument types, broadly available and sufficiently easy to apply by non-specialist users.

High-throughput amplicon sequencing is currently a popular method for microbiological community profiling, with increasing applications in drinking water monitoring and characterisation (Bürtscher et al., 2009; Pinto et al., 2012, 2014; Prest et al., 2014). One inherent limitation to such community profiling is that data are presented as relative abundances, and differences in cell concentrations between samples are not considered (Props et al., 2016). Needless to say, the latter information can completely alter the interpretation of a community profiling dataset. A number of studies have combined sequencing data (relative abundance) with FCM-TCC (total absolute abundance) to derive and compare sample-specific absolute taxon abundances (specific absolute abundance). For example, Lautenschlager et al. (2013) attributed minor biological instability in a non-chlorinated drinking water distribution system to a 20 % increase in the specific absolute abundance of Comamonadaceae. Prest et al. (2014) used the same approach to quantify the absolute abundance increase of some phyla (e.g., Proteobacteria) and decrease of others (e.g., Bacteroidetes) during full-scale distribution of non-chlorinated drinking water. The use of the two methods in concert enhances interpretation of bacterial dynamics in drinking water systems. Combining multiple methods also offers a potential starting point to develop a multi-disciplinary theoretical framework for bacterial growth and other dynamic processes in drinking water treatment and distribution.
4.4. FCM is reproducible

From a statistical perspective, HPC only counts between 0 and 300 colonies in a well-chosen dilution, while FCM analysis usually collects between 50 and 20'000 events for drinking water sample analysis (Hammes et al., 2008). HPC analysis often shows relative standard deviations of 30 % up to 100 % (Hammes et al., 2008; Prest et al., 2013). In stark contrast, the inter-laboratory and instrument variability are < 7 % for FCM (SLMB, 2012), or even < 2.5% for a single operator and instrument (Prest et al., 2013) (see also Figure 3). Finally, FCM results are mutually comparable when measured with a standardised method, while HPC methods already have heavily differentiated over the years, hindering comparison between laboratories (Reasoner, 2004; SLMB, 2012).

The large discrepancy between HPC and FCM reproducibility and accuracy is demonstrated in Figure 3. In this particular example, two raw water samples (groundwater and spring water), a sample from the effluent of a drinking water treatment plant and a drinking water sample from a household tap (all non-chlorinated water samples), were measured independently in triplicate by three different accredited routine laboratories (Koetzsch et al., 2012). All three laboratories used identical methods: FCM-TCC was measured with SYBR Green I staining as described in (SLMB, 2012; Prest et al., 2013) and HPC was measured according to Swiss guidelines (PCA, 30 °C, 72 h). The results showed that only 0.005 % (range = 0.0035 % - 0.0084 %) of the total bacterial cells were detected on average by HPC methods, while the HPC values and reproducibility were so low that no clear differences between different water samples could be detected (relative standard deviation, RSD = 88 %) (Figure 3). In contrast to cultivation, the FCM-TCC results had a RSD as low as 6.9 %, and allowed clear discrimination between all four water samples (Figure 3). This data emphasises the need for standardisation, especially with respect to future routine application of FCM in the water industry (SLMB, 2012; Prest et al., 2013; Nescerecka et al., 2016a).
4.5. FCM speed, automation and online analysis potential

The usefulness of any monitoring variable is heavily influenced by the time from sample collection to when results are available. HPC incubation usually takes several days. For example, the incubation times for HPC agar plates described in ‘Standard methods for the examination of water and wastewater’ (Rice et al., 2012) range from 2 to 7 days. However, by the time these results are available, the drinking water of concern has spread throughout the distribution network and has been consumed widely. In contrast, FCM results can be available within 15 minutes after sampling, enabling immediate action and identification of problematic samples for further investigation or remediation steps. Moreover, automated multi-well plate analysis, which is a feature on nearly all modern FCM instruments, easily allows the measurement of up to 500 samples within a day with only one operator and one instrument (Van Nevel et al., 2013b). This enables researchers and utilities to rapidly screen considerably larger numbers of samples with the same or even less labour input than was feasible with conventional cultivation-based methods.

Many dynamic processes and events in drinking water treatment and distribution systems occur on short time-scales (minutes-to-hours), and conventional grab sampling and cultivation-based analysis approaches fail to correctly detect and characterise such dynamics. Online analysis is the obvious solution to this, and a variety of online microbiological tools and sensors have been developed during the last few years (Lopez-Roldan et al., 2013). Recent developments in fully automated online technology allow continuous FCM measurements for several subsequent weeks (Hammes et al., 2012; Brogniaux et al., 2013, Besmer et al., 2014). For example, Besmer et al. (2016a, 2016b) and Besmer and Hammes (2016) characterised precipitation-induced fluctuations in raw water and operationally induced fluctuations in treated water at high temporal resolution, resulting in total sample numbers in the thousands – far in excess of what is remotely possible with conventional sampling and analysis tools (Example S1). This enabled a detailed characterization of the bacterial baseline concentrations and fluctuations in specific systems on a level of detail not previously possible, and also demonstrated the potential for early warning systems and strategic sampling strategy design (Besmer et al., 2016a, 2016b; Besmer and Hammes, 2016).
4.6. FCM analysis can be cost-beneficial

A final argument for selecting any method for application in routine laboratories is incontrovertibly the cost. A cost comparison for HPC versus FCM was done recently (Helmi et al., 2014b), which showed higher cost of FCM compared to HPC for 100 samples. However, for the purpose of this review the comparison was revised based on our in-house instrumentation and protocols for FCM, as well as information from a routine laboratory for HPC analysis, including all amortization, consumables, quality control and labour costs (Table S2). The FCM cost was estimated at $214 per day for 100 samples processed, compared to $1’030 for the same number of analyses with HPC. The initial FCM hardware investment is considerable, but automation and low consumable use can keep per sample costs minimal. In the case of HPC, the initial investment is lower, but the procedure is labour-intensive, making the cost increase for HPC almost linear to number of samples. Both methods would cost $155 per day for a daily processing of 15 samples, and FCM becomes more profitable for any higher number of water samples. It should be noted that both comparisons (Helmi et al., 2014b) are based entirely on high-income countries, where labour costs are decisive. For low-income countries, the high initial investment for FCM may be more demanding, and labour costs for HPC would be less. However, FCM instrumentation costs have been steadily decreasing during the last decade with the development of simplistic bench-top systems and increased market competition.

5. Arguments against FCM methods

Every analytical method faces some drawbacks. Challenges for FCM discussed below include: (i) the difficulties in distinguishing between viable and non-viable bacteria, (ii) subjective data analysis and (iii) problems in dealing with bacterial aggregates and clusters.
5.1. Detecting disinfection: how dead is dead?

A critical counter-argument against FCM-TCC is the inclusion of dead cells in the enumeration. Disinfection (e.g., chlorination) is common in drinking water treatment worldwide and in such a case the use of FCM-TCC provides limited information and can lead to erroneous conclusions. Viability staining (e.g., for membrane integrity, membrane potential or metabolic activity) combined with FCM can serve as an alternative (Berney et al., 2008; Helmi et al., 2014b). A broad range of viability stains for different bacterial targets exist (e.g., SYTOX Green for damaged membranes and carboxyfluorescein diacetate (CFDA) for esterase enzyme activity), providing information that is regarded as indicators of bacterial viability and/or activity (Hammes et al., 2011).

As a straightforward viability marker for drinking water, we propose the use of membrane integrity, since it measures a cellular property that is targeted by conventional disinfection based on chlorine and ozone (Ramseier et al., 2011). Moreover, severe membrane damage is viewed as a conservative indicator of cell death (Lisle et al., 1999; Berney et al., 2008) and is therefore arguably a safe indicator for disinfection efficiency (Ramseier et al., 2011). The combined use of propidium iodide with SYBR Green I offers the possibility for FCM assessment of bacterial cell membrane integrity, allowing discrimination between cells with intact and damaged membranes (Figure 4) (Berney et al., 2008; Vital et al., 2012). This FCM-ICC is recommended as additional measurements to FCM-TCC whenever disinfection is applied during treatment or when residual disinfectant is maintained in distribution. Recently, the use of FCM-ICC methodology showed instability and growth of bacteria in different chlorinated drinking water distribution networks (Nescerecka et al., 2014; Gillespie et al., 2014) and on the household level (Lipphaus et al., 2014).

One specific challenge to the use of FCM-ICC is the evaluation of UV-C-disinfection. The primary mode of action of UV-C-disinfection is nucleic acid damage; the formation of pyrimidine dimers and other photoproducts of the nucleic acids inhibit replication and transcription and thereby prevents the bacteria
from multiplication (Villarino et al., 2003; Hijnen et al., 2006). However, during low/moderate UV-C
disinfection, the bacterial membranes remain intact (Nocker et al., 2007), making this damage
undetectable by propidium iodide staining. In this specific case, HPC analysis will regard these cells as dead,
and has an advantage over FCM.

A large number of viability dyes are commercially available and have been tested in various research
settings (Berney et al., 2008; Helmi et al., 2014a). The future challenge in this field is to select dyes based
on (i) the mechanism of cellular death that is evaluated (e.g., chlorination permeabilising cells) and (ii) the
mechanistic action of the dye (e.g., penetrating permeabilised cells). Only with these two aspects matched
appropriately, can one establish meaningful standardised protocols (Nescerecka et al., 2016a).

5.2. Is FCM quantification subjective and user-specific?

An often-mentioned critique towards FCM quantification for routine monitoring is the use of variable
instrument settings and the need for manual gating to separate the bacterial signals from the background
(De Roy et al., 2012; Aghaepour et al., 2013; Prest et al., 2013). Until recently, this gating strategy was
subjective and mainly based on personal experience of the operator, which is a serious disadvantage for
standardised analysis. However, several strategies emerged recently to address this issue. Firstly, Prest et
al. (2013) combined an optimised and fixed staining protocol with a fixed gating strategy (i.e. no need for
manual gate adjustments by the operator) to acquire stable and reproducible results. Secondly, an
increasing number of researchers are working on circumventing gating entirely by doing a gating-
independent statistical processing (including cluster recognition) of the data (De Roy et al., 2012;
Aghaepour et al., 2013; Koch et al., 2014). Widespread application will require further work to establish
gating strategies that are sensible, independent of user-bias or instrument-bias and easily applicable to
FCM data generated with different instruments.

5.3. Cell clumps, clusters and aggregates
FCM analysis detects single cells and/or bacterial aggregates, but does not necessarily discriminate between the two, which is a disadvantage that is shared with HPC (Shapiro, 2003). As a consequence, the presence of clumps of sloughed biofilm or colonised suspended particles will potentially lead to under-counting of bacteria (van der Kooij and van der Wielen, 2014a). To address this problem, some authors applied mild sonication to break up bacterial aggregates in wastewater and biofilm samples and validated their procedure using microscopic techniques (Foladori et al., 2010; Ma et al., 2013). However, it is clear that sonication based approaches have to be tested and validated extensively to find the optimal sonication dose for every specific matrix. For example, an activated sludge sample requires a higher sonication dose compared to a settled wastewater sample (Ma et al., 2013), while excess sonication damages bacterial cells (Buesing and Gessner, 2002) and thereby affects viability measurements. Routine application of a sonication step to all samples would also increase the time requirements for the FCM procedure and thereby partly counteract one of the main advantages of FCM. Bulk analysis methods such as ATP-analysis or nucleic-acid-based molecular methods in theory do not face this problem and can be applied even when clumps and clusters are part of the samples (Liu et al., 2013b).

6. FCM data do not correlate with HPC data

The data above highlighted a numerical discrepancy of several orders of magnitude between FCM and HPC values and good correlations between FCM and HPC data should not be expected. However, Hoefel et al. (2003) correctly argued that if a statistical relationship between FCM and HPC data existed, it would facilitate easier incorporation of rapid FCM methods in the routine water analysis sector. Along these lines, a host of studies compared the findings from new/rapid methods with HPC data, usually reporting poor correlations (e.g., Hoefel et al., 2003; Siebel et al., 2008; Burtscher et al., 2009; Nescerecka et al., 2014). These studies were often limited in amount of data and variety of water samples. Thus, for the purpose of this review, the largest dataset (n > 1’800) to date was compiled comparing HPC results with FCM results.
for a variety of water samples spanning a decade of drinking water research. The dataset consists of groundwater and surface water used for the production of drinking water, samples collected during water treatment, finished drinking water and samples of drinking water networks and household installations, from both chlorinated and non-chlorinated waters from four European countries. HPC values were measured using the routine methods prescribed by the different local drinking water legislation, namely The Netherlands (PCA, 22 °C, 72 h), Switzerland (PCA, 30 °C, 72 h), Latvia (PCA, 36 °C, 72 h) and Belgium (YEA, 22 °C, 68h). While not identical, we believe that these HPC methods were sufficiently similar to enable detection of correlations and trends should they exist. FCM-TCC was in all cases based on SYBR Green I staining as described by Hammes et al. (2008) or Prest et al. (2013), FCM-ICC was based on an additional viability staining with propidium iodide for membrane integrity, as described by Nescerecka et al. (2014) or Van Nevel et al. (2013b). The dataset comprises unpublished data from accredited drinking water analysis laboratories in Switzerland and The Netherlands (1’467 data points) as well as from published data (355 data points) (Koetzsch and Sinreich, 2014; Nescerecka et al., 2014; Prest et al., 2014; Sinreich et al., 2014) and on-going research projects (1’138 data points).

The overall dataset shows extremely weak correlation between either FCM-TCC and HPC (Kendall correlation test: $R^2 < 0.1; p < 0.001, \tau = 0.20$) or FCM-ICC and HPC (Kendall correlation test: $R^2 < 0.1; p < 0.001, \tau = 0.25$) (Figure 5). On average, less than 1 out of 3’000 cells (or 0.032 %) detected with FCM was detected by HPC. These correlations did not improve substantially when the data was clustered by country/method, by chlorinated or non-chlorinated water type, or by source water or drinking water (data not shown). While the poor correlations were not particularly surprising considering the large numerical discrepancy between HPC and FCM data, it raises a problematic issue. We discussed above that HPC and FCM methods are used for the same purpose in drinking water monitoring (Table 2; Section 4). Hence, the lack of any meaningful correlation between these two variables suggests that they do not provide the same information and thus should not be used for the same purpose. One may argue that the methods provide different (and potentially complimentary) information and can be used in concert with each other. However, given the practical and financial constraints faced by water utilities and routine laboratories, it is
highly unlikely that the combined use of two microbiological methods for general water quality analysis would find sufficient traction beyond a transitional period.

7. FCM data correlate strongly with intracellular ATP data

Although FCM and HPC did not have a strong correlation due to the constraints and bias of the plating method, cultivation-independent methods should, in theory, be complimentary or in agreement. ATP measurement is one such method that is often promoted for drinking water analysis (Hammes et al., 2010a; van der Wielen and van der Kooij, 2010; Nescrecka et al., 2016b) and intracellular ATP data was previously shown to correlate strongly with FCM-ICC data (Hammes et al., 2010a). This correlation occurs even though these two independent methods differ considerably. ATP is a bulk measurement of the ATP molecule that is unique to all living organisms, while FCM is a single cell method that distinguishes viability states in bacteria based on the reaction of fluorescent dyes with the cells (Figure 4). Some studies argued that ATP measurements are less sensitive than FCM to small changes (Liu et al., 2013a), but more useful when particles, clusters and/or biofilms are measured (Liu et al., 2013b). For the purpose of this review we compared FCM-ICC and intracellular ATP measurements, based on a large dataset (n = 1'441) of samples collected over a decade from different water types in different studies (Figure 6). This dataset is a subset of the data shown in Figure 5. To enable comparison, intracellular ATP analysis of all samples shown here was done according to a similar protocol with reagents from a single supplier (for details, see Hammes et al. (2010a) and Prest et al. (2014)), and all FCM-ICC analysis was done with the same protocol (described above). Statistical analysis confirmed the strong correlation between ATP and FCM-ICC (p < 0.001 according to the Kendall correlation test, τ = 0.60; R² = 0.73 for linear regression). While some correlations between FCM data and intracellular ATP data have been shown in previous studies (e.g., Nescrecka et al., 2014), such a large collection of data from diverse samples and studies has not. Both intracellular ATP and FCM-ICC measurements were previously argued to represent the majority of viable bacteria in drinking water (Hammes et al., 2010a; van der Wielen and van der Kooij, 2010). The overall correlation strongly suggests
that both these cultivation-independent variables may well be used for the same purpose and that conversion factors can be employed to broadly relate intracellular ATP data with bacterial cell concentrations (or vice versa) when only one method is used.

8. Applying FCM for routine microbiological water monitoring

FCM cell concentrations can be used in routine monitoring, similarly to HPC, as a meaningful process variable. For a water utility, this would typically mean characterising spatial variability (e.g., source, treatment steps and various locations in the network) (e.g., Vital et al., 2012; Nescerecka et al., 2014) as well as short and long-term temporal variability (e.g., hours, days, weeks, months per location) (e.g., Ho et al., 2012; Besmer and Hammes, 2016) and detect potential problematic situations. However, due to the range of different factors (source water type, treatment processes, environmental conditions) that affect microorganisms in water, it does not make sense to set a universal absolute upper limit for FCM cell concentrations. We argue that it would make more sense to set the operational goal to “no abnormal change”, as is currently the case for HPC in many countries (Table 1). This requires the establishment of a baseline of bacterial concentrations that can be expected in normal situations and as well as of the natural fluctuations occurring in a given system (Besmer et al., 2014). This baseline value will be dependent on (i) raw water source: surface water has usually considerably higher cell concentrations compared to spring- and groundwater (Leclerc, 2003), (ii) treatment plant design: for example biological filtration is known for elevating cell concentrations (Hammes et al., 2008), (iii) use of final disinfection and a disinfectant residual, which lowers (intact) cell concentrations (e.g., Hoefel et al., 2003), (iv) seasonal variations, whereby groundwater tends to be more stable over time than spring and surface water (Richardson et al., 2009; Pinto et al., 2014; Prest et al., 2016a). Proper baseline establishment requires extensive, long term monitoring and in turn allows accurate detection of deviations (Besmer et al., 2014; Besmer et al., 2016a), which should be evaluated both in relative and absolute changes in bacterial cell numbers. For example, an increase from 5’000 to 55’000 cells mL⁻¹ (i.e. 1’100 % increase) may be considered by some as more
alarming than an increase from 100'000 to 150'000 cells mL⁻¹ (i.e. 50 % increase), although the absolute
increase (50'000 cells mL⁻¹) was in fact identical. On the other hand, bacteria growth in a distribution
network from 5'000 to 10'000 cells mL⁻¹ is the same relative increase (100 % increase) as growth from
100'000 to 200'000 cells mL⁻¹ (100 % increase). However, based on a conversion factor of 10⁷ bacteria
grown per μg of carbon (Van Nevel et al., 2013a), only 0.5 μg·C L⁻¹ is needed for the first situation but 20-
times more assimilable organic carbon (ca. 10 μg·C L⁻¹) would be required for the second example,
suggesting the latter situation is considerably more alarming from a biological stability perspective. While
we expressly do not want to propose upper FCM limit values for drinking water here, the large data set
from multiple studies presented in Figure 5 suggests that drinking water typically does not exceed a value
of 5 x 10⁵ cells mL⁻¹. The latter value is higher than 95 % of the data in Figure 5, excluding all raw water
samples.

Although universal threshold/guideline values may not be sensible for FCM data, we believe that
individualised, utility-specific limits to support decision-making could be established after detailed
monitoring of the particular systems. Implementation of FCM by water utilities would therefore require
several years (e.g. minimum two years) of parallel FCM and HPC measurements, thereby building-up a solid
database and gaining confidence in the FCM data from their own system. This effort is, however,
worthwhile given the considerable advantages of FCM in comparison with HPC as discussed in this review.
9. Conclusions

HPC played an important role in drinking water management and general microbiological quality control over the past century, but this review questions whether HPC is still the best available technology for process and general water quality monitoring. We argue that FCM cell counting is a suitable alternative to replace HPC for routine microbiological drinking water monitoring for the following 8 reasons:

1. *Abundance*: HPC detects considerably less than 1% of the total bacteria in a water sample and often does not detect the dominant species, while FCM detects all bacteria that are present;

2. *Relevance*: Multiple studies showed that FCM is a meaningful process variable, providing relevant information on various drinking water treatment processes and microbiological water quality changes during distribution – the same application areas where HPC is routinely used;

3. *Speed*: FCM results can be available within 15 minutes of samples arriving in the laboratory, while HPC incubation typically requires 2–7 days;

4. *Reproducibility*: Inter-laboratory FCM comparisons have relative standard deviations on cell concentrations below 10%, while single operator measurements have relative standard deviations below 3%;

5. *Flexibility*: In addition to absolute cell numbers, FCM can provide information on bacterial viability and bacterial identity based on different fluorescent dyes and probes;

6. *Added value*: Multivariate FCM data can be used to create a unique flow cytometric fingerprint of the bacterial community, which improves rapid detection of small changes to that community;

7. *Costs*: Depending on labour costs and preferred instruments, FCM costs are equal or lower than those of HPC from about 15 water samples or more per day;

8. *Automation*: FCM offers easy automation options, opening exciting doors on the prospect of fully automated online FCM analysis.
Acknowledgements

Sam van Nevel was supported by the project grant no. G.0808.10N and the travel grant V424114N of the FWO Flanders and the Inter-University Attraction Pole (IUAP) ‘μ-manager’ funded by the Belgian Science Policy (BELSPO, 305 P7/25); Emmanuelle Prest and Hans Vrouwenvelder were supported by funding from King Abdullah University of Science and Technology (KAUST) and Evides Waterbedrijf; the Inter-University Attraction Pole (IUAP) ‘μ-manager’ funded by the Belgian Science Policy (BELSPO, 305 P7/25); Caitlin Proctor was supported by MERMAID, a Marie Skłodowska-Curie Initial Training Network, under grant number 607492. We thank Alina Nescerecka, the Kantonal Laboratory Zürich, Industrielle Werke Basel – Wasserkraft (IW) and Pidpa for the shared data, Lisa Neu for literature support and Synthia Maes for the critical reading of the manuscript.
REFERENCES


Infrastructuur en Milieu (2011). Besluit van 23 mei 2011, houdende bepalingen inzake de productie en distributie van drinkwater en de organisatie van de openbare drinkwatervoorziening (Drinkwaterbesluit).


Lewis, K., S. Epstein, A. D’Onofrio and L. L. Ling (2010). "Uncultured microorganisms as a source of


Table 1: An overview of the variety in drinking water legislation and guidelines with regard to HPC. For agar compositions, see Table S1.

<table>
<thead>
<tr>
<th>Region</th>
<th>Media</th>
<th>Temp.</th>
<th>Time</th>
<th>Upper limit</th>
<th>Comment</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>United States</td>
<td>Plate Count Agar</td>
<td>48 h</td>
<td>35°C</td>
<td>&lt; 500 cfu ml⁻¹</td>
<td>• Concern that values above limit interferes with coliform and E. coli recovery methods</td>
<td>USEPA 2009</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>Yeast extract Agar</td>
<td>22 °C</td>
<td>68 ±4 h</td>
<td>NAC*</td>
<td>• UK guidelines allow for the use of R2A agar when deemed necessary</td>
<td>Anonymous 2012</td>
</tr>
<tr>
<td>France</td>
<td>Plate Count Agar</td>
<td>22 °C</td>
<td>72 h</td>
<td>NAC NAC</td>
<td>• Change should not to exceed 10-fold the “usual” results</td>
<td>Ministère de la santé et des solidarités 2007</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>Plate Count Agar</td>
<td>22 °C</td>
<td>72 h</td>
<td>100 cfu ml⁻¹</td>
<td>• Based on yearly geometrical mean value.</td>
<td>Infrastructuur en Milieu 2011</td>
</tr>
<tr>
<td>Germany A</td>
<td>Low nutrient agar³</td>
<td>22 ±2 °C</td>
<td>68 ±4 h</td>
<td>NAC NAC</td>
<td>• Some German utilities still follow the older TrinkwV1990 guidelines with different media and specified limiting values</td>
<td>TrinkwV 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36 ±2 °C</td>
<td>44 ±4 h</td>
<td></td>
<td>• After treatment</td>
<td></td>
</tr>
<tr>
<td></td>
<td>High nutrient agar</td>
<td>20 ±2 °C</td>
<td>44 ±4 h</td>
<td>20 cfu ml⁻¹</td>
<td>• At the tap</td>
<td>TrinkwV 1990</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36 ±1 °C</td>
<td>44 ±4 h</td>
<td>100 cfu ml⁻¹ 100 cfu ml⁻¹</td>
<td>• After treatment</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• In the network</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Untreated source (spring) water</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>Yeast extract Agar</td>
<td>22 ±2 °C</td>
<td>68 ±4 h</td>
<td>NAC</td>
<td></td>
<td>VMM 2014</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Plate Count Agar</td>
<td>30 ±1 °C</td>
<td>72 ±3 h</td>
<td>20 cfu ml⁻¹</td>
<td>• After treatment</td>
<td>EDI 2014 ISO 4833</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>300 cfu ml⁻¹ 100 cfu ml⁻¹</td>
<td>• In the network</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Untreated source (spring) water</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>Standard Methods Agar</td>
<td>35 °C</td>
<td>48 h</td>
<td>No limit</td>
<td>• If used as an indicator, numbers should be established on a system-specific basis.</td>
<td>Health Canada 2012</td>
</tr>
<tr>
<td>Australia</td>
<td>Yeast extract agar, R2A agar</td>
<td>20–22 °C</td>
<td>72–120 h</td>
<td>No limit</td>
<td>• Guidelines allows flexibility regarding media and incubation conditions</td>
<td>NHMRC and NRMMC 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35–37 °C</td>
<td>24–48 h</td>
<td>No limit</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* NAC: no abnormal change.

³ The German TrinkwV2001 guidelines specify low nutrient agar (6 g L⁻¹ tryptone and 3 g L⁻¹ yeast extract). The older TrinkwV1990 guidelines use high nutrient agar (1 % beef extract and 1 % peptone).
Table 2: Microbiologically relevant processes and/or events in drinking water treatment and distribution systems and that are conventionally monitored with HPC methods\(^1\). A selection of related FCM studies for each category is given, together with links to specific examples presented in the supplementary information section of this manuscript.

<table>
<thead>
<tr>
<th>Relevant process or event</th>
<th>Selection of related FCM studies</th>
<th>Examples (^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
<td>Detecting contamination of drinking water sources</td>
<td>Besmer et al. (2014, 2016a, b); Besmer and Hammes (2016)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Assessment of chemical disinfection efficacy</td>
<td>Hammes et al. (2008, 2010b); Vital et al., (2012)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Assessment of membrane filtration efficacy</td>
<td>Hammes et al. (2010b)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Growth and/or detachment during biofiltration</td>
<td>Hammes et al. (2008, 2010b); Vital et al. (2012)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Operational influences on microbiological water quality</td>
<td>Besmer et al. (2016a, b)</td>
</tr>
<tr>
<td><strong>Distribution</strong></td>
<td>Temporal (in)stability in microbiological water quality</td>
<td>Prest et al. (2016); Besmer et al. (2016a, b); Nescerecka et al. (2014)</td>
</tr>
<tr>
<td><strong>Distribution</strong></td>
<td>Spatial (in)stability in microbiological water quality</td>
<td>Nescerecka et al. (2014); Vital et al. (2012); Lautenschlager et al. (2013)</td>
</tr>
<tr>
<td><strong>Plumbing</strong></td>
<td>Water quality deterioration in building plumbing</td>
<td>Lautenschlager et al. (2010); Lipphaus et al. (2014); Siebel et al. (2008)</td>
</tr>
<tr>
<td><strong>Storage</strong></td>
<td>Bacterial growth in drinking water during storage</td>
<td>Mimoso et al. (2015); Wang et al. (2008)</td>
</tr>
</tbody>
</table>


\(^2\) Examples are summarised in case study format in the supplementary information section.
Table 3: Overview of methods used for general bacterial quantification in drinking water.

<table>
<thead>
<tr>
<th>Method</th>
<th>Measures</th>
<th>Principle</th>
<th>Viability</th>
<th>Labor</th>
<th>Time-to-result</th>
<th>Online</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPC</td>
<td>Cultivable bacteria</td>
<td>Growth</td>
<td>Yes</td>
<td>Medium</td>
<td>Days to weeks</td>
<td>No</td>
<td>Reasoner (1990)</td>
</tr>
<tr>
<td>FCM</td>
<td>Cell concentration</td>
<td>Staining</td>
<td>Yes(^1)</td>
<td>Low</td>
<td>Minutes</td>
<td>Yes</td>
<td>Prest et al. (2013)</td>
</tr>
<tr>
<td>Microscopy</td>
<td>Cell concentration</td>
<td>Staining</td>
<td>Yes(^1)</td>
<td>High</td>
<td>Minutes to hours</td>
<td>No</td>
<td>Burtscher et al. (2009)</td>
</tr>
<tr>
<td>ATP</td>
<td>ATP concentration</td>
<td>Enzymatic</td>
<td>Yes</td>
<td>Low</td>
<td>Minutes</td>
<td>Yes</td>
<td>Nescerecka et al. (2016b)</td>
</tr>
<tr>
<td>qPCR</td>
<td>16S rRNA gene copies</td>
<td>Gene amplification</td>
<td>Yes(^2)</td>
<td>High</td>
<td>Hours to Days</td>
<td>No</td>
<td>Lopez-Roldan et al. (2013)</td>
</tr>
<tr>
<td>Nucleic acid quantification</td>
<td>Total DNA/RNA</td>
<td>Fluorescence /Absorbance</td>
<td>Yes(^2)</td>
<td>High</td>
<td>Hours to Days</td>
<td>No</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^1\) Requires specific viability staining with dyes such as propidium iodide
\(^2\) Quantifies viability when RNA is extracted and analysed
Figure legends

Figure 1: Overview of the main advantages of flow cytometry for drinking water monitoring. Abbreviations used: FCM – flow cytometry; TCC – total cell concentration; ICC – intact cell concentration; HPC – heterotrophic plate counts; RSD – relative standard deviation

Figure 2: Microbiological monitoring of a karstic spring following regional precipitation and subsequent increased spring discharge (A). Samples collected hourly for 24 h were analysed with flow cytometry (FCM) for total cell concentrations (TCC; SYBR Green I staining) and with heterotrophic plate counts (HPC; PCA, 30 °C, 72 h) (B).

Figure 3: Comparison of flow cytometric total cell concentration (FCM-TCC) and heterotrophic plate count (HPC) sensitivity and accuracy on four non-chlorinated water samples. Four water samples were analysed in triplicate by three routine laboratories using both HPC (PCA, 30 °C, 72 h) and FCM-TCC (SYBR Green I staining). Box plots show the median (solid line), the 25 and 75 percentiles (box), as well as the absolute data points (crosses). On average only 0.005 % of the total bacterial cells were detected by HPC (B), while FCM discriminated clearly between the different water samples (A). The cultivability of the samples, expressed as HPC cell count divided by FCM cell concentration, remained below 0.01 % (C). Figure redrafted from Koetzsch et al. (2012).

Figure 4: The principle of flow cytometric total cell concentration (FCM-TCC) and intact cell concentration (FCM-ICC) measurements. FCM-TCC is based on SYBR Green I, which enters and stains all bacterial cells with nucleic acids, independently of membrane integrity. FCM-ICC is based on dual staining with both SYBR Green I and propidium iodide, of which propidium iodide is only able to penetrate and stain bacterial cells with damaged cell membranes. Therefore, FCM-ICC enables the differentiation between bacterial cells with intact membranes and damaged bacterial cells.
Figure 5: Comparison between flow cytometry total cell concentration (FCM-TCC) and heterotrophic plate count (HPC) results (A) and intact cell concentration (FCM-ICC) with HPC results (B) in a broad range of water samples. Samples originate from different (non)chlorinated water treatments, including source, treatment and distribution network samples (FCM-TCC: \( n = 3'675 \), FCM-ICC: \( n = 1'835 \)). It is clear that the correlation between HPC and FCM-TCC or FCM-ICC is very weak (\( R^2 < 0.1 \) for linear regression in both cases).

Figure 6: The comparison of flow cytometry intact cell concentration (FCM-ICC) with intracellular adenosine tri-phosphate (ATP) concentrations in water samples. Samples originate from different chlorinated and non-chlorinated systems, including source, treatment and distribution network samples (\( n = 1'441 \)). There is a clear correlation between both parameters, which is expected since both methods are a direct assessment of potentially viable bacterial cells, either intact or active.
<table>
<thead>
<tr>
<th>Relevant information</th>
<th>Rapid analysis</th>
<th>Reproducible measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM-TCC and ICC data have been used extensively to monitor relevant bacterial changes during treatment and distribution.</td>
<td>Less than 15 minutes is needed from the time when the sample reaches the laboratory until the final result is available.</td>
<td>Single operator/instrument variability is &lt; 3 % RSD; Multi operator/instrument variability is &lt; 10 %.</td>
</tr>
<tr>
<td><strong>Example:</strong> Hammes et al. (2008)</td>
<td><strong>Example:</strong> Prest et al. (2013)</td>
<td><strong>Example:</strong> Prest et al. (2013)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Flexible staining</th>
<th>Low costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM can be combined with a variety of fluorescent dyes and probes targeting bacterial DNA content, different aspects of viability, activity and identity.</td>
<td>FCM analysis can be cost-effective compared to HPC analysis when more than 15 samples per day are analyzed, based on hardware and labor costs.</td>
</tr>
<tr>
<td><strong>Example:</strong> Berney et al. (2007)</td>
<td><strong>Example:</strong> This paper</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multi-variable data</th>
<th>FCM fingerprinting</th>
<th>Automation potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCM measurements collect data on fluorescent color, intensity and light scattering (size) for every single particle.</td>
<td>A range of new statistical tools combines multivariate data to track changes in community composition with FCM.</td>
<td>Online FCM enables detailed measurements of the frequency and magnitude of fluctuations in drinking water systems.</td>
</tr>
<tr>
<td><strong>Example:</strong> Hammes and Egli (2010)</td>
<td><strong>Example:</strong> De Roy et al. (2012)</td>
<td><strong>Example:</strong> Besmer et al. (2016)</td>
</tr>
</tbody>
</table>
A: FCM-TCC
RSD: 6.9%

B: HPC
RSD: 88%

C: Cultivability

groundwater  spring water  drinking water (production)  drinking water (tap)
**FCM total cell count**

- **SYBR Green I**

**Intact cell**

**Permeabilised cell**

---

**FCM intact cell count**

- **SYBR Green I + propidium iodide**

**Intact cell**

**Permeabilised cell**
Highlights

Routine drinking water monitoring still relies on heterotrophic plate counts (HPC)

Flow cytometry (FCM) is proposed as a better method for process monitoring

No good correlation was found between FCM and HPC data (n = 3’675)

Good correlations were found between FCM and ATP data (n = 1’441)

FCM advantages are: relevance, speed, accuracy, costs and automation potential
ROUTINE DRINKING WATER ANALYSIS

heterotrophic plate counts (HPC)

flow cytometric total/intact cell counts (TCC/ICC)

selective variable slow labour intensive

sensitive reproducible fast automated

1880 → 2016