3928

# THE RELATIVE IMPORTANCE OF PARTICLE-ASSOCIATED FECAL INDICATOR BACTERIA FOR MICROBIAL WATER QUALITY ASSESSMENT AND MICROBIAL INDICATOR QUALITY IN KANEOHE STREAM.

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

OCEANOGRAPHY

DECEMBER 2004

By
Yves Plancherel

Thesis Committee:

James P. Cowen, Chairperson
Edward A. Laws
Roger S. Fujioka

# ACKNOWLEDGEMENTS

I am grateful to James P. Cowen for serving as my advisor for the last 3 years. The help and comments from Dr. Edward A. Laws and Dr. Roger S. Fujioka have been very helpful. None of this work would have been possible without the contribution of NOAA Sea Grant Hawaii, which provided me with funding for this study, a traineeship and a stipend, which made life in Hawaii possible.

A special thanks goes to my friends and colleagues from the University of Hawaii and elsewhere, who always knew when it was time for beer or when it was time for another beer. In particular, I would like to express my gratitude to Phyllis Lam, Nicolas Cassar, Terri Rust, Jim Falter and Janie Culp and Adrian Glover, with whom I had memorable adventures on land and in the sea.

Last but not least, I would like to thank my parents, Susan and Gabriel Plancherel, for their support throughout my "Hawaiian adventures" but also for the last 26 years. I do not know which part of these 26 years were the hardest for them, but they made them all easy for me.

This thesis is dedicated to my grandmothers Lucie Plancherel-Joie and Marthe Aebischer-Shaffer. Their wisdom has been a source of inspiration.

# ABSTRACT

The basis for recreational water quality assessment relies on the ability to measure a characteristic pre-determined water quality criterion, which can be related to the risk posed by the water body. In terms of microbial water quality, the criteria are the concentrations of microbial indicators (e.g. enterococcus, coliforms, etc.) and the risk is characterized by the rate of illness of a population exposed to the contaminated water. Two methods, the membrane filtration and the most probable number, are routinely used to monitor indicator density. This approach to water quality assessment, however, faces many challenges, amongst which are the questionable reliability of the indicator system and the inability of the available measurement methods to accurately quantify the density of the microbial indicators.

Hawaii's State Law provides recreational water quality standards for enterococci. Nonetheless, this indicator has been shown to be unreliable in tropical environments because levels of enterococcus in excess of the standards are consistently found in unpolluted locations. Identification of positive colonies isolated on enterococcus-specific agar from a control (conservation land) and a contaminated (urban) station shows that at both stations, the majority of the enterococcus species recovered are common inhabitants of the soil and therefore cannot be used for water quality assessment. A minority of enterococcus species with hygienic importance (e.g. *E. faecalis*) was recovered from the urban site, however. Since the membrane filtration assay for enterococcus is not exclusive to the enterococcus species with hygienic significance, enterococci cannot be used unambiguously as water quality indicators in Kaneohe Stream.

The association of cells with aggregates is a known fact in microbial ecology. The culture-dependant methods for microbial water quality assessment, however, are not able to distinguish between multiple cells associated with aggregates. Instead, the colonies originating from each cell merge into one, such that only one colony forming unit per aggregate is detected. Computer simulations reveal that the problem of colony merging does not only occur because of the presence of aggregates, but that colony merging occurs also by chance alone, from the filtration of a randomly distributed cell population, inducing measurement error of typically <20%, depending on the size and number of the colonies. Efforts to assay the measurement error arising from the presence of aggregates were made both theoretically and empirically. Both methods show that this type of measurement error is typically on the order of a few tens of percent. The error, however, varies with the type of microorganism, being largest for heterotrophic plate count (HPC) bacteria and enterococci, and smallest for *C. perfringens*. Proportionally, enterococci were found to associate more readily with large aggregates (>5 μm) than *C. perfringens*, whose spores are primarily free-living, or than HPC bacteria. Overall, however, small aggregates (<5 μm) are responsible for most of the measurement error due to the inability of the membrane filtration method to separate multiple colonies growing from clusters.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS

| SYMBOL | DESCRIPTION |
|---|---|
| **GENERAL** | |
| MF | Membrane filtration |
| MPN | Most probable number |
| HPC | Heterotrophic plate count |
| ENT | Enterococcus |
| CP | *Clostridium perfringens* |
| CFU | Colony forming unit |
| TSS | Total suspended solids |
| **CHAPTER 1** | |
| D | Dose |
| C | Concentration of pathogenic organisms in the source water |
| R | Decimal recovery of the detection method |
| I | Fraction of the organisms capable of infection |
| DR | Efficiency of the water treatment process expressed as a decimal reduction of the concentration |
| V | Volume of water consumed per unit of time (consumption) |
| **CHAPTER 2** | |
| $N_{ENT}$ | Number of animals needed, assuming only one type of animal present, to account for the observed annual enterococcus discharge |
| $N_{CP}$ | Number of animals needed, assuming only one type of animal present, to account for the observed annual *C. perfringens* discharge |
| **CHAPTER 3** | |
| 3.2 Randomness and the Poisson distribution | |
| $\bar{\mu}$ | True mean cell concentration |
| V | Volume |
| N | True number of cells |
| $N_L$ | Number (lower boundary) |
| $N_U$ | Number (upper boundary) |
| R | Critical merging distance |
| $N_E$ | Total number of intervals (distances) between N points (e.g. bacteria on a filter) |
| $n_c(X)$ | Number of clusters of size X |
| $\gamma$ | Constant (cluster size distribution) |
| X | Cluster size, number of bacteria per cluster |
| $\theta$ | Power (cluster size distribution) |
| $\varepsilon_R$ | Measurement error induced by clustering upon filtration of a randomly distributed population of cells |
| M | Slope (regression of $\varepsilon_R$ as a function of N) |

| | |
|---|---|
| B | Intercept (regression of $\varepsilon_R$ as a function of N) |
| $n_i$ | Number of tubes in replicate set *i* (MPN method) |
| $p_i$ | Number of positive (fertile) tube in replicate set *i* (MPN method) |
| $\pi_i$ | Probability of obtaining a positive (fertile) tube (MPN method) in replicate set *i* |
| L | Likelihood (MPN method) |
| $\mu$ | Measured (estimated) mean cell concentration |

3.3 Deviations from the Poisson assumption

| | |
|---|---|
| pgf | Probability generating function |
| G(z) | pgf of the number of clusters per volume |
| g(z) | pgf of the number of cells per cluster |
| H(z) | pgf of the generalized distribution expressing the number of cells associated to clusters per volume |
| $\pi_j$ | Probability that a cluster contains j cells |
| $\lambda_c$ | Mean number of clusters per volume |
| A | Parameter controlling the logarithmic distribution of the number of cells per cluster, $0<A<1$ |
| K | Parameter of the negative binomial distribution. It is difficult to give k a physical interpretation, but it could indicate the number of samples with no clusters (~number of failures) |
| Q | Parameter of the negative binomial distribution. Q=1-P, represents the probability that a cluster has no associated bacteria. |
| P | Parameter of the negative binomial distribution, represents the probability that a cluster contains associated bacteria. |
| M | Mean of the corresponding pgf |
| $\nu$ | Variance of the corresponding pgf |
| $p_r$ | Coefficient of $z^r$ in H(z) |
| $P_{Tot}$ | Probability distribution describing the total number of cells in a sample |
| $N_{Tot}$ | Total number of cells in a sample |
| $N_1$ | Number of particle-associated bacteria |
| $N_2$ | Number of free-living bacteria |
| $\lambda_f$ | Mean number of free-living bacteria |
| F(z) | pgf of the number of free-living bacteria |
| J(z) | pgf of the total number of cells in a sample |
| $P_{max}$ | Maximum probability of the distribution $P_{Tot}$ |
| $\omega$ | Proportionality constant describing the number of bacteria in a sample as a function of the number of free-living cells |
| $N_{MF}$ | Number of CFUs expected from the routine membrane filtration assay |
| $\varepsilon_A$ | Measurement error induced by the presence of bacteria-hosting aggregates in a water sample assayed by the MF method |
| $N(P_{max})$ | Most probable number of cells in a sample containing bacteria-hosting aggregates |

CHAPTER 4

| | |
|---|---|
| $T_m$ | Melting temperature in °C |
| M | Monovalent cation concentration |
| L | Probe length (base pairs) |
| $C_R$ | Measured concentration of indicator bacteria obtained by the routine and unmodified membrane filtration methods |
| $C_H$ | Measured concentration of indicator bacteria present on > 5 μm aggregates |
| $C_A$ | Measured concentration of indicator bacteria-hosting aggregates > 5 μm |
| $C_F$ | Measured concentration of free-living indicator bacteria (< 5 μm) |
| F | Filter area |
| $A_V$ | Area covered by a field of view |
| *F* | Fraction of the filter area analyzed by microscopy |
| $N_V$ | Number of fields of view analyzed |
| $\chi$ | Number of cells that must be present per filter for detection of a single cell to be possible |
| $\Psi$ | Ratio of the concentration of the total culturable bacteria to the bacterial concentration estimated by the routine membrane filtration analysis |
| $\Delta$ | Number of bacteria not accounted for by the routine membrane filtration method |
| S | Fraction of the excess aggregated bacteria ($\Delta$) originating from aggregates smaller than 5 μm |

# Chapter 1 : INTRODUCTION

The goal of water quality assessment is clearly to protect public health, and as such, water quality assessment is only a part in a bigger whole. It should be noted that water quality assessment is an iterative process. In a first phase, the risk, or a parameter indicative of risk, must be measured; secondly, a decision must be made as to whether the measured risk is acceptable. If this is not the case, action must be taken (risk management), the impact of the remediation process or the outcome on public health must be evaluated and the cycle must start again (Bartram et al., 2001). Each phase in this cycle is subject to error and debate, even if the water quality guidelines are derived from a scientific consensus, integrating the best available evidences from a broad expertise. This thesis focuses on an aspect associated with the measurement of the risk.

The quantification of the risk of illness for swimmers, bathers, or consumers exposed to a polluted water body involves the measurement of microbial indicator organism densities.

> "A health recreational water quality *criterion* developed for use with indicator systems is defined as a quantifiable relationship between the density of the indicator in the water and the potential human health risks involved in the water's recreational use. It is a set of facts or a relationship upon which a judgment can be made. A water quality *guideline* derived from the criterion is a suggested upper limit for the density of the indicator in the water that is associated with health risks that are considered unacceptable. The concept of acceptability implies that there are social, cultural, economic, and political as well as medical inputs to the

derivation and that these may vary in time as well as space. A water quality *standard* obtained from the criterion is a guideline fixed by law. (Cabelli, 1983)"

A range of indicator organisms has been put forward for application in the assessment of the sanitary quality of waters (Ashbolt, Grabow & Snozzi, 2001): coliforms, fecal coliforms, fecal streptococci, enterococci, *Escherichia coli*, sulfite-reducing clostridia, *Clostrdidium perfringens*, Bifidobacteria, bacteriophages, coliphages, *Bacteroides fragilis* bacteriophages, and others. The choice of the most appropriate indicator for water quality assessment of recreational waters, however, remains controversial (Ashbolt, Grabow & Snozzi, 2001), as is the reality of a clear and quantifiable relationship between indicator concentration and the rate of illness among swimmers that is used to estimate the health risk to the population inherent to the tested water body (Fleisher, 1990, Favero, 1985).

Depending on the organisms targeted, there exist two widely used (traditional) techniques for their enumeration: the most probable number (MPN) and the membrane filtration (MF) methods. Both of these methods rely on the selective culturing of the target organisms. These methods have the advantage of being relatively simple and inexpensive. It is known, however, that culture-based methods systematically underestimate the total number of targeted organisms in the sample (Ashbolt, 2001; Amann et al. 1995). In addition, the time lag between sample collection and results is quite long, ranging from one to multiple days, depending on the microorganisms. There exist today alternative technologies for the detection and quantification of organisms in the environment (Ashbolt, 2001; Rompré et al., 2002): immunological and nucleic acid methods, fingerprinting methods, microarrays, etc. These newer methods, however,

cannot yet be used for water quality and risk assessment since there exists presently no relationship between the rate of illness in the exposed population and organism density using these alternate protocols.

Estimation of indicator organism density by the "traditional" methods is subject to large measurement error, which translates into poorly constrained relationships between indicator organism density and illness rate (Fleisher 1990). Unfortunately, epidemiological studies often neglect to empirically determine this source of error or to incorporate it into their models. These studies tend to favor spatial or temporal coverage over precision of the measurement, thus failing to account for 5-60% of the total variance (Fleisher & McFadden, 1980; Fleischer, 1990).

The overall purpose of this study is to describe and estimate errors associated with the measurements using the different culture-based methods to assess the concentrations of the indicator organisms, with emphasis on the membrane filtration technique. Specifically, two types of error intrinsic to the culture-based measurements will be described and their consequences for water quality assessment will be discussed.

The first type of error, which is discussed in Chapter 2, relates to the specificity of the medium used. In "traditional" microbiology, the isolation of the target organisms is achieved by the application of physical (temperature) and biochemical (e.g. antibiotics, pH, salts, substrate choice or availability) stresses. Only organisms capable of tolerating the stresses will grow and form colonies, while the others die, remain dormant or are out-competed. With the development of nucleic acid analysis in microbial systematics, however, it has become increasingly evident that phenotype and genotype do not necessarily correlate. This discovery led to 1) the realization of the vast microbial

diversity (DeLong, Wickham and Pace, 1989), 2) an extensive reshuffling of the phylogeny of microorganisms (Hardie & Whiley, 1997), and 3) an increase in the number of described species (DeVriese et al., 2002).

In terms of microbial water quality assessment, which relies on the phenotypic characteristics of the microbes for their enumeration, the concept of species composition of a group (e.g. enterococcus) and the significance of the positive colonies belonging to that group must be re-evaluated. This problem is of particular concern for the enterococcus group. Not only has the composition of this group been rapidly evolving (Hardie & Whiley, 1997), but enterococci are generally believed to be the best indicators of sanitary water quality for recreational waters, having produced the strongest correlation between organism density and swimming-associated illness to date (Cabelli et al., 1982; Cabelli et al., 1983; Cabelli, 1989). Enterococci are consequently very widely used. Despite their popularity, the reliability of enterococci as sanitary indicators has been questioned in Puerto Rico (Muñiz et al., 1989) and in Hawaii (Hardina & Fujioka, 1991, Roll & Fujioka, 1997; Fujioka & Byappanahalli, 2001). Data from a previous study (Fujioka & Byappanahalli, 2001) suggests that some non-indicator species, that is, species whose presence is not exclusive to the presence of fecal pollution, belonging to the *Enterococcus* genus readily grow on the agar used, such that the overall organism density estimate may grossly overestimate the concentration of those enterococci bearing hygienic significance and thus overestimate the risk.

The second type of measurement error is discussed in Chapters 3 and 4. These chapters describe theoretically (Chapter 3) and empirically (Chapter 4) the error that arises from the limitation of the culture-based methods to account for all the cells

associated with aggregates. Neither the MPN nor the MF method can discriminate between multiple cells aggregated together or with other particles. Both methods assume that the bacteria in a sample are randomly distributed and accept this limitation. Yet, the association of bacteria with aggregates is a real and important ecological phenomenon (see references in Appendix F).

The aggregation of pathogens into clusters may have important repercussions with respect to quantitative microbial risk assessment (QMRA), the modern foundation of water quality assessment. In order to define a tolerable risk for a population, a relation must be obtained that describes the response (occurrence of illness) given a certain exposure to a pathogen (dose)—the dose-response relationship.

In a study assessing the risk of infection by *Cryptosporidium* and *Giardia* in drinking water, Teunis et al. (1997) parameterized the dose ($D$) as follows:

$$D = C \cdot \frac{1}{R} \cdot I \cdot 10^{-DR} \cdot V, \tag{1.1}$$

where $C$ is the concentration of pathogenic organisms in the source water, $R$ is the decimal recovery of the detection method, $I$ is the fraction of the concentration that is capable of infection, $DR$ (decimal reduction) is the efficiency of the water treatment process and $V$ is the volume of water consumed per unit of time (consumption). Eq. (1.1) assumes that the cells are randomly distributed. If multiple cells are forming clumps or are associated with aggregates, then the dose should be expressed by:

$$D = D_{free-living} + D_{particle-associated}. \tag{1.2}$$

That is, to accurately estimate the dose, it is necessary to 1) develop a technique to measure the concentration of particle-associated pathogens, 2) estimate the recovery of

this technique, 3) account for the fact that microorganisms associated with aggregates are generally metabolically more active (Crump & Baross, 1996; Kirchman & Mitchell, 1982) and consequently may have the potential of being more infectious than free-living cells, and 4) correct for the lower efficiency of the treatment process against particle-bound microbes, as is the case with chlorine (LeChevallier, Evans & Seidler, 1981; Ridgeway & Olson, 1982; Stewart, Wolfe & Means, 1990) or UV (Emerick et al., 1999; Parker & Darby, 1995; Qualls, Flynn & Johnson, 1983) disinfection.

Because the concentrations of free-living pathogens are generally low, especially in drinking water, the presence of a few aggregates containing pathogens in a sample could produce a dose that is tens to hundreds of percent larger than estimated without taking into account the numerous pathogens associated with the aggregates, thus underestimating the risk greatly (Figure 1.2).

Chapter 3 investigates via mathematical formulations and computer simulations the effects that bacterial clumps have on the measurement error of the concentrations. The first case considered is that of the formation of clusters induced by the process of filtration from a randomly distributed population of cells. The second case describes a mathematical framework for the modeling of bacteria associated with aggregates.

Chapter 4 summarizes the results 1) of laboratory experiments designed to assess the use of different protocols aimed at enumerating bacteria associated with aggregates using membrane filtration, and 2) of a field study investigating the extent of aggregate-association of different indicator bacteria in a tropical stream.

The concluding chapter (5) lists the major findings of the study.

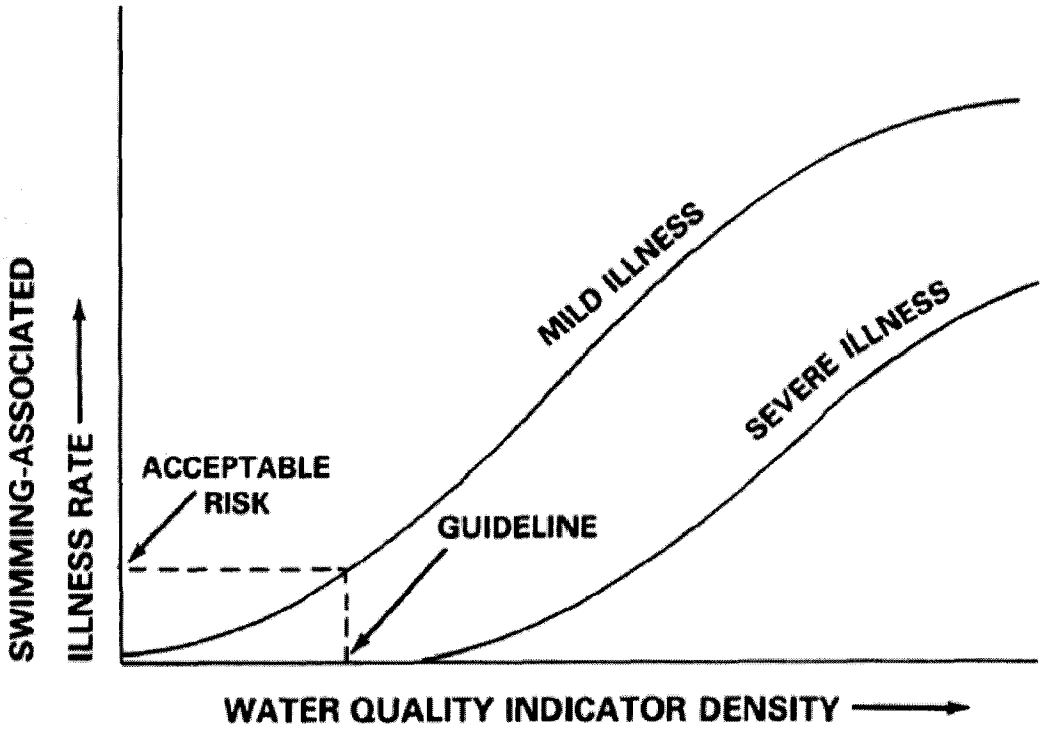


**Figure 1.1: Representation of the relationship of guideline or standard to a recreational water quality criterion (from Cabelli, 1983). An underestimate of the water quality indicator density underestimates the risk. Depending on the characteristics of the "dose-response" relationship, small errors of measurement associated with the indicator density can produce large errors in risk estimates.**

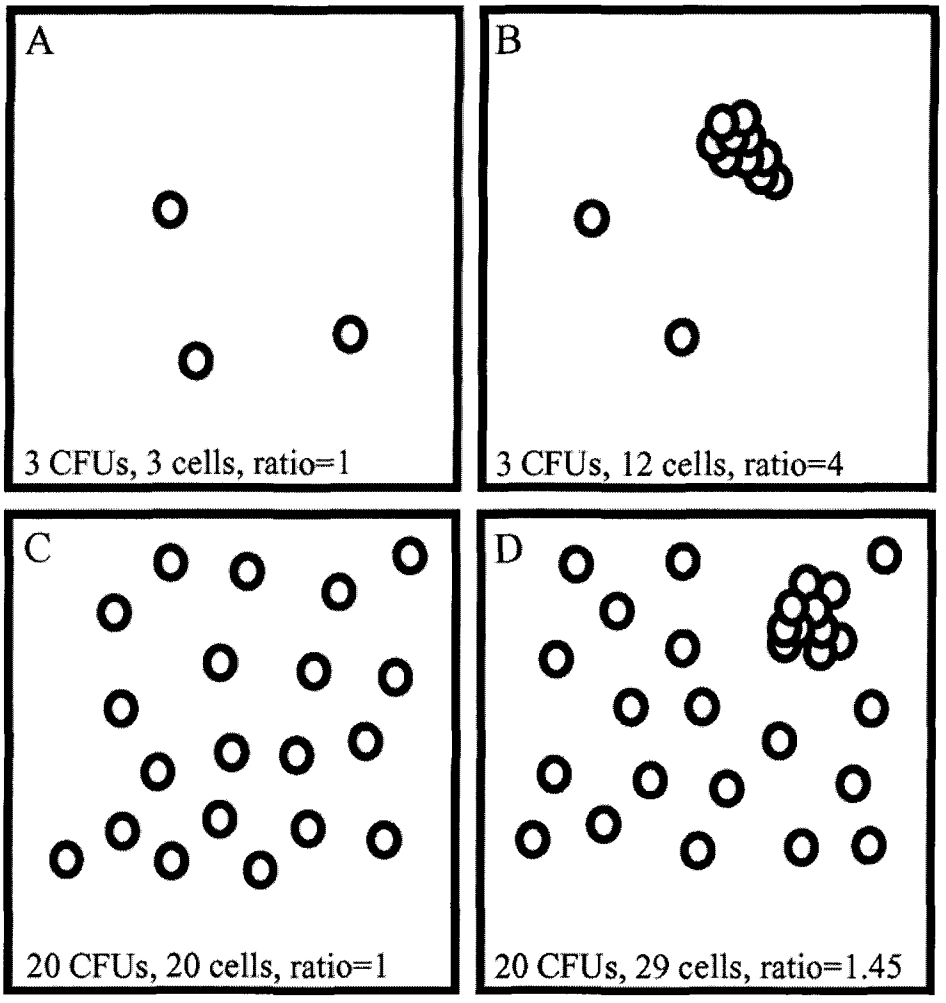


**Figure 1.2: Illustration representing the potential effect of aggregates with many associated bacteria in a situation where the background (free-living) concentration is low (A & B) and high (C & D). The error induced by the presence of an aggregate (10 cells) is greater in low concentration environments than in high concentration environments.**

# Chapter 2 : WATER QUALITY ASSESSMENT OF KANEOHE STREAM

## *2.1 Environmental setting*

### 2.1.1 The characteristics of stream flow

Although streams supply only a few percent of the state's drinking water, streams supply over 50% of the irrigation water (Oki, 2003). Due to Hawaii's variable rainfall intensities, the small drainage basins and the steepness of the valleys, Hawaiian streams are "flashy"; that is, their water level can rise and recede within a few hours, creating a human hazard.

The potential sources of stream flow are: 1) runoff directly from rainfall, both as overland flow or subsurface storm flow that quickly returns infiltrated water to the stream, 2) base flow from groundwater discharge where the stream intersects with the water table, 3) water returned from bank storage, 4) rain falling directly on the stream, and 5) other sources such as returned irrigation water or discharges originating from human activities.

Streams can either gain or lose (due to infiltration, evaporation or diversion) water along their paths. The upper reaches of streams in the state of Hawaii often cross volcanic dikes, thin (a few meters wide), vertical sheets of dense and impermeable rocky intrusion. The arrangement of these dikes allows for the impoundment of the ground water lens at higher altitude, such that in interior mountainous areas, it is possible for streams to gain water from ground water base flow in addition to rainfall. This situation is especially important on windward Oahu, and as a result windward Oahu streams are perennial, with a discharge that is generally less variable than streams of the leeward side (Oki, 2003).

### 2.1.2 The Kaneohe watershed

The islands of Hawaii can be categorized into two physiographic zones: windward and leeward. These terms refer to the exposure of each area to the northeasterly trade winds and associated orographic rainfall. The wet season occurs from October to April, while the dry season lasts generally from May to September (Oki, 2003). The Kaneohe watershed (Figure 2.1) is located on the windward side of Hawaii's most urbanized island, Oahu. It spans elevations from 843m to sea level and encompasses an area of 14.7km$^2$ (De Carlo, Beltran and Tomlinson, 2004).

Hoover (2002) estimated the average daily mean sediment discharge of the Kaneohe watershed from data collected for the years 1999 and 2000 to be 0.471 tons/day, with an average daily minimum and maximum of 0.144 and 52.5 tons/day, respectively. The instantaneous maximum reached 1,130 tons/day in 1999 and 785 tons/day in 2000. The total annual sediment discharge was calculated to be 172 tons. 32% of this discharge occurred during base flow conditions and 68% during storm-runoff events. Correspondingly, the average total annual water discharge from Kaneohe stream was 5.35x10$^6$ m$^3$ over 1999-2000, with 84% of the total during base flow conditions and 16% during storms. A comparison of the water and sediment discharge data readily reveals the dominance of storm events for the mass export of the erosion and weathering products but not for the transport of water. The ratio of the annual sediment discharge to the annual water discharge yields a watershed-averaged mean suspended solid concentration of 12.2 mg/L during base flow and 136.6 mg/L during storms. The median rainfall estimate for the Kaneohe watershed from a 19-year record is 200 cm/y (Hoover, 2002). Forty-one percent of this value is thought to contribute directly to runoff, yielding an

annual sediment discharge, normalized to watershed area and per cm of rainfall contributing to runoff, of 3.48 kg/km$^2$/cm. This value for the year 1999-2000 was significantly lower than the historical USGS-estimate of 8.28 kg/km$^2$/cm.

### 2.1.3 Sampling stations

The uppermost station, LULU, is located on Luluku Stream. It sits about 1.6 km downstream from the watershed divide on conservation land in the Hoomaluhia Botanical Garden (Figure 2.1). In spite of its proximity to the mountains, its elevation is a modest 67 m (De Carlo, Beltran and Tomlinson, 2004). This station represents the control station. It is not absolutely pristine, however, since it drains a small sub-watershed containing a section of a major trans-mountain highway (Likelike Highway) and a banana plantation. Luluku Stream is the first major tributary of Kamo'oali'i Stream upstream of its confluence with Kapunahala Stream. LULU corresponds to the U.S. Geological Survey (USGS) station No. 16270900 and is currently a National Water Quality Assessment Program (NAWQA) station and part of the local integrated sampling effort of the Coastal Intensive Site Network (CISNet) Kaneohe Bay program. LULU is only freely accessible during the open hours of Hoomaluhia Botanical Garden. For security reasons, anybody working at LULU should announce themselves to the garden's manager.

Station KAMO is slightly downstream of the confluence of Luluku Stream with Kamo'oali'i Stream. While Luluku drains mainly the western edge of the Kaneohe watershed, Kamo'oali'i Stream drains the southern sector of the watershed. A multitude of small and mostly intermittent tributaries drain into Kamo'oali'i Stream. In 1980, a

flood control reservoir was built on Kamo'oali'i Stream just upstream of the confluence with Luluku Stream, now the "lake" in Hoomaluhia Botanical Garden. The reservoir has two outlets, one of which discharges the normal amount of water carried by Kamo'oali'i Stream before construction of the dam, while the other outlet is a spillway. Station KAMO matches USGS station No. 16272200. KAMO is located at the onset of residential/suburban Kaneohe and is also being monitored as part of the CISNet program (Cox, 2001). KAMO can best be accessed through a private property located at 45528 Liula Street.

Station KANE is located below the confluence of Kamo'oali'i Stream with Kapuhahala Stream. Kapuhahala Stream drains a small sub watershed on the northern side of the Kaneohe watershed. Both of these streams merge to form Kaneohe Stream. Kaneohe Stream is the only stream that runs through the lower section of the Kaneohe watershed. Because of this, KANE can be regarded as the end-member of the watershed. Large portions of the stream are channelized or hardened (channelized with concrete) in the urban area before the water flows into the Bay (Burr, 2003). Sampling at KANE occurs at the end of the last hardened section of Kaneohe Stream, roughly 50 m downstream from the Kamehameha Highway Bridge. Erosion of the stream banks appears to be an important problem downstream of station KANE. KANE is freely accessible by parking near the Kaneohe public library/police station and cutting across the adjacent soccer field to the stream.

The lowermost station is LKS (Lower Kaneohe Station). It is a suburban site in the lower portion of Kaneohe Stream. It is only slightly (~500 m) upstream from the mouth of the stream, and its elevation is less than 1 m. There are no major tributaries

draining into Kaneohe Stream between station KANE and the stream mouth on Kaneohe Bay. Saltwater intrusions occur during each tidal cycle at LKS. Stations KANE and LKS are also part of the CISNet program. LKS is located on private property at 45189 Wena Street, and authorization should be obtained from the owner before venturing to LKS.

## 2.1 4 Water quality issues

The early focus of the "Clean Water Act" (1977-1987), the federally mandated water pollution control act, was to reduce "point source" discharges of pollutants (Hawaii State Department of Health, 1999). These point sources are typically "end-of-pipe" discharges from industrial sites or treatment plants. Correcting point source emissions has noticeably improved the overall water quality of many sites (Hawaii State Department of Health, 1999). Of course, accidental point source discharges occasionally occur: common and recurring point source problems in Hawaii are the subject of much controversy and include sewer system overflow due to excessive rainfall, leaks and ruptures of sewer mains (Table 2.2). However, in spite of the efforts to remove and control point source pollution, many water bodies still fail to meet water quality standards. The cause for the persistent impairment of many of these water bodies is non-point source pollution. In response, the "Clean Water Act" was revisited and amended in 1987 to accommodate the problem of non-point source pollution (Hawaii Coastal Zone Management Program, 1996).

It is difficult to track non-point source pollution since the sources are diffuse. Remediation efforts are also very difficult. The most common non point source pollutants include soil, nutrients, pesticides, insecticides, oil, litter, lawn clippings, and bacteria

(from cesspools) (Hawaii Coastal Zone Management Program, 1996). The consequences of non point source pollution include waterborne diseases, algal blooms, fish and reef kills, and turbid waters (Hawaii Coastal Zone Management Program, 1996). The U.S. Congress enacted the "Coastal Zone Act Reauthorization Amendments" in 1990 by adding the new Section 6217 entitled "Protecting Coastal Waters" to the older "Coastal Zone Management Act" of 1972. Section 6217 requires that states submit a coastal non-point source pollution control management plan, which is what the state of Hawaii did by creating the "Hawaii's coastal non-point pollution control program" (Hawaii Coastal Zone Management Program, 1996).

There are supposedly no point sources of sewage along Kaneohe Stream or its tributaries, yet, high concentrations of fecal bacteria, along with high levels of nutrients and suspended solids, were recorded in the stream and in water samples collected at Kaneohe Beach park (EPA, 2001). These data justified the blacklisting of Kaneohe Stream by the U. S. Environmental Protection Agency (U.S. EPA) and the Hawaii's Department of Health (DOH). This list identified 31 out of 57 Oahu streams as impaired in 2001, and 28 of 57 streams were classified as impaired in 2002 (Hawaii State Department of Health, 2003). These agencies also listed the Kaneohe Bay waters surrounding the mouth of Kaneohe Stream as a Water Quality Limited Segment (Clean Water Act §303(d) List of Water Quality-Limited Segments).

### 2.1.4.1 The quality of the indicators

Microbial fecal indicators are supposed to mimic the behavior of human pathogens from fecal origins (Sloat & Ziel, 1991). In other words, they should be

consistently and exclusively associated with the human fecal sources of the pathogenic organisms, they should be found in greater number than the pathogens, and they must be equally or more resistant to environmental stresses than the pathogens. Also, they must not proliferate in the environment. Indicators are used instead of pathogens because detecting indicators is simpler, cheaper and less hazardous. Pathogens are also often hard to characterize. Use of intensively studied indicators is intended to mitigate and integrate the risk that the unknown pathogens represent.

The use of coliforms, fecal coliforms and enterococci as indicators for sanitary water quality assessment is controversial in Hawaii and in tropical settings generally. These organisms survive and even grow in soils and waters, which questions their use as fecal pollution indicators (Carillo et al, 1985; Fujioka & Shizumura, 1985; Perez-Rosas & Hazen, 1988; Muñiz et al, 1989; Fujioka & Byappanahalli, 2001; Hardina & Fujioka, 1991). In Hawaii, *Clostridium perfringens* is considered a better indicator (Fujioka & Shizumura, 1985). However, enterococcus is still used as an indicator in Hawaii. The decision to maintain enterococcus on the indicators list for the assessment of Hawaii' streams is based on epidemiological data collected by an EPA study conducted at mainland beaches contaminated with sewage (Cabelli et al. 1979; Cabelli, 1983), In that study enterococci correlated best with the occurrence of gastrointestinal illness in swimmers.

The quality standard set for enterococcus for freshwater is 33 CFU/100ml (Hawaii State Department of Health, 2002), expressed as the geometric mean of no less than 5 samples collected within a 30-day period, and no single sample shall exceed the single sample maximum of 89 CFU/100ml (Hawaii State Department of Health, 2002). In

spite of this, single observations and geometric means of enterococci concentrations are almost systematically one to two orders of magnitude greater than the standards, even at sites with no human activity.

Part of the problem lies in the fact that the culture medium used for the detection of enterococcus (mE agar with transfer on EIA agar) was developped in 1975 (Levin et al. 1975). Since that time, existing non-enterococcus species have been reclassified in the *Enterococcus* genus, and new enterococcus species have been discovered that are not consistently found in human or animal feces (Devriese et al. 1993; Leclerc et al. 1996 ). Their origins are not well known, but some are reported to live in soils or in association with animals or plants (Meier 1998). The strains that are believed to grow exclusively on mE agar (*E. faecalis*, *E. faecium*, *E. gallinarum* and *E. avium*) (APHA, 1998) may not be the only strains that produce positive results with the mE/EIA agar. In order to evaluate the use of the standard membrane filtration technique for the quantification of enterococcus in Hawaii, information about the identity of the strains that grow on the medium used is necessary.

### 2.1.5 Goals

Because of the poor ranking of Kaneohe Stream in terms of water quality, part of this study included the collection of data on the concentration of suspended solids and on a series of microbial indicators: Heterotrophic Plate Count (HPC) bacteria and the fecal indicators Enterococcus (ENT) and *Clostridium perfringens* (CP). The goals of this section are 1) to describe the data collected and interpret the information to assess the

microbial quality of Kaneohe Stream, and 2) to identify the Enterococcus-positive colonies from the mE/EIA agar test to species level.

The specific hypotheses tested under the second goal are: a) a majority of the colonies recovered from mE agar during base flow background conditions are not of major hygienic importance (Table 2.3) and consequently cannot be used as fecal indicators; b) the species composition of the enterococcus reflects the site location within the watershed (botanical garden versus urbanized area); and c) the species composition at each site reflects the type of pollution found within the Kaneohe watershed: fecal pollution (human or animal) and non-fecal sources from plants or soil.

## *2.2 Methodology*

The water samples were collected by grab samples in autoclaved 1L polycarbonate bottles (wide mouth) approximately 10 cm below the air-water interface and as much towards the center of the stream as feasible. The samples were transported to the laboratory in a cooler and on ice. Processing started generally 1-2 hours after collection.

Sampling frequency was irregular and spaced over a >2-year period from April 2002 to June 2004. Stations LULU and KANE were more heavily sampled than KAMO and LKS. Samples were taken either individually or in duplicate or triplicate when possible. When available, replicate samples were averaged—arithmetic mean (Haas, 1996)—to yield one data point per sampling event.

### 2.2.1 Suspended solids

The concentration of suspended solids was measured by filtering a known amount of water (ranging from 800 to 2000 ml) through a pre-weighted GF/F filter. The filter was dried in an oven at 60°C with hygroscopic $CaSO_4$ for at least 3 days before reweighing.

### 2.2.2 Culturing

All indicator organisms were assayed by the membrane filtration technique. Table 2.4 summarizes specifics of the methods used. Briefly, a volume (generally 100 ml) of sample (either directly or from a dilution series) is passed through a membrane filter (nitrocellulose, 0.45 μm pore size, 47mm diameter) under vacuum. The filter is then transferred to a sterile agar plate (47mm) containing the medium. The plate is incubated under the proper conditions with the agar on top and the filter facing downwards to prevent condensation droplets from falling back onto the filter and dispersing colonies.

### 2.2.3 Enterococcus identifications

Identification of enterococcus-positive colonies recovered by the mE/EIA agar tests was performed using the rapid ID 32 Strep system (Biomérieux S.A.). An enterococcus-positive colony was identified, picked from the filter, streaked onto Brain Heart Infusion agar (BHI-agar) and incubated at 41°C for 48-72 hours. Three to four large individual colonies from the BHI plate were resuspended into approximately 2 ml of 0.2 μm syringe-filtered deionized water in autoclaved micro-centrifuge tubes and dispersed by vortexing. Immediately after dispersion, 55 μl of the suspension was pipetted into each of the 32 wells on the test strip, which were incubated at 35°C for 4

hours. The results were read manually according to the table provided by the manufacturer. The results (positive or negative) are translated into a numerical code that is used for identification.

## 2.3 Results

### 2.3.1 Culturing data overview

The data of the microbial concentrations for all stations are presented graphically in Figure 2.2 to Figure 2.4 against total suspended solids. Enterococcus levels are consistently higher than the water quality standard of 33 CFU/100 ml, or the 89 CFU/100 ml for individual samples, at all stations. Levels of *C. perfringens* are low at stations LULU and KAMO but reach or exceed the water quality guideline of 55 CFU/100 ml at station KANE and LKS.

None of the data sets associated with the parameters TSS, HPC, ENT and CP is normally distributed. This was to be expected given the purely positive nature of these measurements (El-Shaarawi, Esterby & Dutka, 1981) and the tremendous differences existing between baseflow and stormy conditions (Hoover, 2002; Jagals, Grabow & DeVillier, 1995; Jagals, 1997). Application of a log-transformation tends to normalize the distributions, however not significantly (Shapiro-Wilk's W tests ,$p<0.01$ for all parameters). Descriptive statistics of the data are shown in Table 2.5.

A non-parametric Kruskal-Wallis ANOVA by ranks test performed on the overall data set for each parameter (TSS, HPC, ENT and CP) suggests that sample location (station) has a significant effect on the measured results for all the parameters ($p<0.001$ for each parameter). The suspended solids concentration data at stations KANE and LKS

are not significantly different (Mann-Whitney U test, $p>0.3$; Table 2.6), but all other comparisons between the station-specific concentrations of suspended solids are significant ($p<0.05$; Table 2.6). HPC levels are not significantly different between stations LULU and KAMO or between stations KAMO and LKS, but the HPC values obtained at station LULU are statistically lower than those from station LKS. Since LKS is tidally influenced, the microbial population contributing to HPC counts at LKS may differ greatly from the HPC-positive microbial population recovered from the other stream stations. In consequence, the comparison between LKS and the other stations in terms of HPC counts bears little physical and microbial significance. LULU and LKS have statistically similar enterococcus populations, as do KANE and LKS. All other comparisons between stations show significantly different concentrations of enterococci. The concentrations of *C. perfringens* are significantly different ($p<0.001$; Table 2.6) between all stations except between stations LULU and KAMO ($p>0.8$; Table 2.6).

### 2.3.2 Enterococcus speciation

Out of a total of 70 colonies picked from station LULU and 71 colonies from station KANE, 41 and 51 tests, respectively, yielded satisfactory results. A test was deemed satisfactory if the identification software returned an overall identification quality between "acceptable" and "excellent" based on the numerical profile provided. The tests that failed this requirement are inconclusive and have not been used in this analysis. The fact that a test is rejected does not necessarily mean that the colony analyzed was not part of the *Enterococcus* genus, rather that the identification was ambiguous, and little confidence can be placed in the proposed identity. Of the 29 inconclusive tests at station

LULU, 19 were still classified as *Enterococcus*, 9 of which were given as *E. casseliflavus*, while the remaining 10 had other non-*Enterococcus* identifications (*Aerococcus* sp., *Leuconostoc* sp.). Twenty tests from station KANE were negative. Seven had non-*Enterococcus* identifications (*Aerococcus* sp., *Streptococcus* sp.) and 13 were classified as *Enterococcus*, 7 of which were *E. casseliflavus*. Only 4% of the total number of tests from the KANE samples gave identifications not compatible with the *Enterococcus* genus, whereas 32% of the LULU tests were incompatible. Satisfactory test results are presented in Table 2.9.

## *2.4 Discussion*

### 2.4.1 Assessment of the culturing data

The data for all the parameters are highly correlated (Spearman R <0.01) with each other. A factor analysis on the log-transformed data was performed to investigate redundancy in the dataset (Table 2.7). If the data for all the parameters behaved exactly the same, the factor analysis would not be able to separate the parameters into different factors. However, if some parameters behave slightly differently, the anomalous parameters should come out of the analysis as distinctive factors. Rigorously, factor analysis requires the data to be normally distributed; even if a log-transformation improves the data's normality status, normality is not statistically achieved. Nonetheless the factor analysis provides interesting information on the relationships of the indicators with each other.

Station LULU and KAMO yield the same results; that is, all the parameters are highly correlated and only one factor can be extracted (eigenvalue>1; Table 2.7). At

station KANE, two factors can be extracted. The variance associated with the first factor, which can be identified as *C. perfringens,* explains almost 60% of the variance of the data set, while a second factor, representing the other 3 parameters (TSS, HPC and CP), explains most of the remaining variance (26%; Table 2.7). At station LKS, the dominating factor isolated (75% of the variance) corresponds to the parameters TSS, HPC and ENT, while a second factor (eigenvalue=0.94), identified as CP, contributes 26% of the variance (Table 2.7). When the factor analysis is performed on the log-transformed data set as a whole with a variable for "station" ranging from 1 to 4, the stations at which the samples were collected, extracted as factor 1, explain up to 61% of the variance (Table 2.7). The second factor encompasses both CP and "station" and explains 20% of the variance. A third and a fourth factor can be extracted, with small associated eigenvalues (<0.5), which contribute another 9% and 6% of the variance, respectively. Factor 3 and 4 are identified as HPC, and HPC, TSS and ENT, respectively (Table 2.7). In summary, the factor analysis shows that in non-urban environments (station LULU and KAMO), all the parameters measured are highly correlated. In the urban environment (stations KANE and LKS), however, CP tends to separate itself from the other parameters, suggesting that a different process controls the concentration of CP and the other variables (TSS, HPC and ENT) in the urban section of the stream. This process may be point or non-point source pollution, whereas the concentrations of TSS, HPC and ENT are primarily controlled by environmental factors.

Because the suspended solid concentration is measured gravimetrically (mg/L), the dense inorganic iron- and aluminum-rich particles produced by the erosion of basalt, which constitute most of the Kaneohe Stream particle population (De Carlo, Beltran &

Tomlinson, 2004), represent most of the suspended solids mass measured in a TSS measurement. Illustrative examples of the types of particles present in Kaneohe stream are shown in Figure 2.5. The sources of the inorganic particles are primarily the soil, the stream banks or the streambed. Consequently, TSS is a parameter that can be used as a tracer of environmental particle input to the stream. If the sample-specific normalization of the microbial data to the TSS data yields consistent ratios between stations not influenced by humans (e.g. LULU) and stations with a human influence (e.g. KANE), this may indicate that the microbial parameter behaves similarly to TSS throughout the watershed. Variation between stations of the ratios obtained by normalizing the microbial concentration to TSS suggests that different station-specific processes control the concentrations of both parameters.

After normalization of the concentrations of the microbial indicators to the sample-specific concentrations of the suspended solids, and further division of this result by the maximum ratio calculated for all stations to cancel the units and obtain values between 0 and 1, the values thus calculated show consistent values for both HPC and ENT for all stations (Figure 2.6, A & B). For CP, however, these values pool in two distinct groups (Figure 2.6, C). One group corresponds to stations LULU and KAMO and the other group to KANE and LKS. This result indicates that the stream contains more CP spores at the urban stations KANE and LKS than the CP levels that would be expected from natural sources.

Hoover (2002) showed that a relationship existed between water discharge and suspended solids concentration in Kaneohe stream. Figure 2.2 and Figure 2.6 (A) showed that a relation exists between HPC and TSS in the Kaneohe watershed. That is, there is a

relationship between HPC and water discharge. If a source discharges a pollutant into a stream at a rate that is unrelated to water discharge, but that stream flow increases, a dilution effect is observed and the measured concentration of the pollutant in the stream, downstream from the source, decreases from its baseflow values. A storm sample was collected on 2/14/2003 and assayed for HPC, ENT, CP and TSS. Normalization of the microbial data to the TSS data shows that the ratio CP/TSS is lower during storm conditions (black symbols with black arrow in Figure 2.6) than during baseflow conditions at stations KANE and LKS. No decrease is observed for the TSS normalized ENT and HPC values, however. This is an indication that the source of TSS, HPC and ENT in Kaneohe stream is the natural environment. The source of CP may be located between stations KAMO and KANE, or on Kapuhahala Stream, which is a tributary to Kaneohe Stream, possibly from leaky cesspools.

Sorensen et al. (1989) concluded, from a study conducted in small streams in Idaho, Utah and California, that *C. perfringens* is a good point-source indicator and could be used to distinguish point-source pollution from non-point source pollution. Samples of cow, horse and sheep feces and farmlot runoff were found to contain low concentrations of *C. perfringens*, whereas the concentrations of other pathogen indicators were high. Municipal wastewater, however, contained elevated *C. perfringens* concentrations. Fujioka & Shizumura (1985) found similarly high *C. perfringens* levels in municipal wastewater effluents.

### 2.4.2 Integration of indicator density with stream flow

Concentration estimates of pathogen indicators in feces of different warm-blooded animals and the excretion rates of these animals are available from Geldreich (1978) and Sorensen et al. (1989) (Table 2.8). Using these values and the stream flow characteristics of Kaneohe Stream given by Hoover (2002) and summarized in an earlier section of this chapter, it is possible to calculate the number (N) of warm-blooded animals required to produce the observed annual discharge of enterococci and *C. perfringens* in Kaneohe Stream. The purpose of this calculation is to illustrate the point that various animals have different impacts on the stream concentrations of the indicators organisms.

Hoover (2002) provides estimates of the annual water discharge and of the fraction of this discharge that is due to baseflow and storm conditions. Pathogen indicator concentration estimates for the different flow conditions are available from this study. The annual discharge of indicator organisms can be calculated using the median indicator concentration for baseflow conditions (Table 2.5) and the concentrations measured during storm flow conditions at station KANE. The annual excretion of indicator organisms for each animal can be obtained from the data presented in Table 2.8. The ratio of the annual discharge of the indicator organisms in Kaneohe stream to the annual excretion of the indicators for each animal provides $N_{ENT}$ and $N_{CP}$. $N_{ENT}$ and $N_{CP}$ represent the number of animals needed, assuming only one type of animal present, to account for the observed annual indicator discharge for enterococcus and *C. perfringens*, respectively. When the N-values are low, the "contamination potential" of the corresponding animal is high and vice-versa.

Large discrepancies exist between the N-values of each indicator organism for a given animal, except for cats and dogs (Table 2.8). The $N_{ENT}$ values are generally lower than the $N_{CP}$ values. That is, animal feces other than human feces may easily influence the concentration of enterococci in the stream, whereas only cat and dog feces may greatly influence the *C. perfringens* levels.

In Hawaii, the populations of wild pigs and feral cats are thriving, suggesting that these animals could be responsible for part of the environmental indicator concentrations in the stream (stations LULU and KAMO). In the urban environments, dogs are a common sight and have the potential for adding abundant enterococci and *C. perfringens* to the stream (stations KANE and LKS). However, since the N-values associated with dogs for both indicators are of similar magnitude, it is unlikely that the presence of dogs at the lower stations could explain the apparent different behavior of *C. perfringens* at these stations compared with enterococci (see discussion associated with Figure 2.6).

### 2.4.3 The quality of enterococci as indicators of fecal pollution

It is clear from the data that the first hypothesis, that most of the enterococci recovered are not of major hygienic importance, is confirmed. A majority of the colonies positively identified at station LULU (58%) and KANE (55%) were of minor hygienic importance (Table 2.10). The second hypothesis, that the species composition of the mE/EIA agar positive colonies would be indicative of the location through the watershed, is also confirmed; the proportion of identifications associated with fecal indicators at station LULU (9%) differs from those at KANE (41%; Table 2.10). In addition, *E. seriolicida* is present often (2 of 5 sampling days, Table 2.11) and in significant numbers

(Table 2.10) at station KANE, but is absent at station LULU. This species was first isolated from diseased specimens of cultured yellowtails and eels in Japan and is known as a fish pathogen (Kusuda et al., 1991). Teixeira et al. (1996) showed that *E. seriolicida* is in fact the same species as *Lactococcus garvieae*. Since the latter is a senior synonym of *E. seriolicida*, *L. garvieae* is currently the preferred name of the species (Teixeira et al. 1996). Fish (e.g. cichlids and gobbies) are abundant at station KANE, while absent at LULU. The presence of *L. garvieae* in the samples likely reflects the presence of fishes.

The enterococcus identification data, coupled with the *C. perfringens* data from the Kaneohe watershed, support the third hypothesis, that the species composition reflects the type of pollution. At station LULU, the level of *C. perfringens* is low and the fraction of enterococci species of minor hygienic importance, especially *E. casseliflavus*, is elevated. This suggests that the soil is the major source of enterococci at this station. *E. casseliflavus* contributes a large fraction of enterococci at station KANE, suggesting that soil is an important source of enterococci at this station as well. However, the fraction of enterococci associated with major hygienic significance (especially *E. faecalis*) is larger at station KANE relative to station LULU. Note, however that *E. faecalis* is also able to grow in soil (R. S. Fujioka, personal communication). In addition, station KANE corresponds to the highest *C. perfringens* concentrations found in the Kaneohe watershed. The latter two lines of evidence suggest the presence of fecal pollution upstream of station KANE but below station KAMO.

Pollution events, however, can be very discrete phenomena. The number of dates sampled for enterococcus identification at each station is low (5 days for each stations). Should a point source event have occurred during any of the total of 10 days sampled, the

percentages calculated above (Table 2.10) could be biased and not represent the long-term situation. The number of positive identification tests analyzed at each sampling date and at each station varies, and the number of satisfactory identifications also differs from day to day. Consequently, it may be preferable to consider the data in terms of frequency of occurrence of the different species per date; that is, any species is counted only once per sampling day (Table 2.11). By comparing Table 2.10 with Table 2.11, the seemingly high percentage attributable to *E. faecalis* for station KANE (Table 2.10) is less pronounced (Table 2.11): *E. faecalis* was only present one of the sampling days (see also Table 2.9). However, the day when *E. faecalis* was recovered was also the day that corresponded to the largest concentration of mE/EIA agar positive colonies (9100 CFU/100ml) at station KANE. *E. casseliflavus* was recovered 4 out of 5 days for both stations, which further highlights the fact that most colonies growing on mE agar from Kaneohe Stream samples cannot be used as fecal indicators.

The data presented here support previous findings (Fujioka & Shizumura, 1985) that enterococcus is not an appropriate water quality indicator in Kaneohe stream, since enterococci are routinely recovered in large quantities at sites with no sources of fecal pollution. Fujioka & Shizumura (1985) measured levels of fecal coliforms, fecal streptococci (a group later renamed enterococci) and *C. perfringens* in different streams on leeward and windward Oahu receiving treated wastewater effluent and concluded that *C. perfringens* was the most suitable of the indicators measured. Both the fecal coliforms and the fecal streptococci were recovered in high quantities at the unaffected upstream stations, making water quality assessment based on these indicators impossible. In a later study, Hardina & Fujioka (1991) identified the soil as an important source of *Escherichia*

*coli* and enterococci in Hawaii's streams, with concentrations of enterococci in surface soil samples ranging from $10^4$ to $10^5$ CFU/100g soil. Assuming that most of the suspended solids observed in the present study represent soil particles, and that a typical suspended solid concentration is ~10 mg/L, the concentration of enterococci in stream waters should be ~0.1-1 CFU/100ml. However, the measured concentration of enterococci (Figure 2.3) ranged from 30 to $10^3$ CFU/100 ml. It is possible that soil is not the only source of enterococci or that cells desorb from soil aggregates when entering the stream because of the increased shear stress or electrolytic dilution (Roper & Marshall, 1974). The streambed has been shown to be a reservoir for bacteria and possibly pathogens (Goyal, Gerba & Melnick, 1977). Surface soil enterococcus concentrations discussed by Hardina and Fujioka (1991) were from the banks of Manoa stream on leeward Oahu and may differ from the local concentration in Kaneohe Stream. Soil samples collected from around the island of Oahu suggest that soil humidity is the factor controlling the concentration of enterococci (R. Fujioka, personal communication). Soils from the humid windward side of the island, where Kaneohe Stream flows, may consequently harbor larger concentrations of enterococcus.

Although the presence of high quantities of enterococci in the soil suggests that this is the most likely source of enterococcus in the stream, the origin of the enterococcus in the soil is still uncertain: are they naturally occurring there, or was the soil subject to pollution? In a previous study, Fujioka and Byappanahalli (2001), identified enterococcus isolates from different soils on Oahu to the species level (Table 2.12). These data contrast with those of station LULU, where less than 10% of the isolates had hygienic importance, compared to values as high as 45% for the soils studied by Fujioka and Byappanahalli

(2001), which is more in line with what is found at station KANE (41%). Unfortunately, Fujioka & Byappanahalli (2001) give no indication regarding the type of the soils that were analyzed. *E. faecalis* was the most prominent fecal indicator in both studies, whereas *E. casseliflavus* was the non-fecal species recovered most often.

The results from Hawaiian environments differ from the data collected by Ferguson et al. (2004) in Orange County, California (Table 2.12). Ferguson et al. (2004) analyzed several hundred suspected enterococcus colonies in different environments. In 6 out of the 9 environments sampled, *Enterococcus* species of major hygienic importance (*E. faecalis* and *E. faecium*) dominated. These environments included seawater, marine or freshwater sediments and seagull stools. *E. gallinarum* was the dominant species in the soil and storm drain water samples, while other species dominated the sewage samples.

*E. gallinarum* and *E. casseliflavus* are very closely related species both phylogenetically and phenotypically (Devriese, Pot & Collins, 1993), Consequently, they are very difficult to differentiate (Devriese et al., 1996). *E. casseliflavus* is commonly isolated from plants, silage and soils, while *E. gallinarum* is mostly isolated from domestic fowls (Collins et al., 1984; Bridge & Sneath, 1982). Because of the low discrimination between these two species, it is possible that different identification methods yield different results. Both Ferguson et al. (2004) and Fujioka and Byappanahalli (2001) used the API 20 STREP system for the identifications, whereas the present study used the rapid ID 32 STREP system. Fujioka and Byappanahalli (2001) tested the API 20 STREP system together with the Biolog System and obtained very different results regarding these two species. The API system identified both *E. casseliflavus* and *E. gallinarum*, whereas the Biolog System assigned the isolates

identified as *E. casseliflavus* from the API system to *E. gallinarum*. It is not known which system is correct, however. The fact that soil samples in both the "Hawaiian" and the "Californian" environment yield high number of *E. casseliflavus* or *E. gallinarum* may be an indication that both *E. gallinarum* and *E. casseliflavus* reflect the same natural soil signature when recovered from water samples.

## 2.5 Conclusions

The levels of enterococcus in the Kaneohe watershed are consistently elevated at all stations and are present in concentrations much greater than the State and EPA water quality standard allows. A majority of the suspected enterococcus colonies recovered on mE agar are of minor or low hygienic importance and should not be used as indicators of fecal pollution. *C. perfringens* is the preferred indicator to assess water quality in Hawaiian streams because it is absent or present in only small quantities at stations with no human influence, but is present at the urban stations where fecal pollution is a plausible source. Based on the data, it is likely that an unknown pollution source, possibly from cesspools, exists between station KAMO and KANE. This source remains to be found, however.

**Table 2.1. Sampling site summary.**

| Station | USGS station No. | Latitude (N) | Longitude (N) | Basin Area ($km^2$) | Altitude (m) | Station active (USGS) |
|---|---|---|---|---|---|---|
| LULU | 16270900 | 21°23'42" | 157°48'44" | 1.14 | 67 | 1967-1998, 2000-2004[b] |
| KAMO | 16272200 | 21°23'47" | 157°48'23" | 9.87 | 36 | 1976-2002 |
| KANE[a] | 16273900 | 21°45'51" | 157°48'13" | 11.34 | 11 | 1959-1980 |
| LKS | --- | --- | --- | 14.7 | <1 | 2000-2004[b] |

[a]Sampling for this study was performed about 200 m downstream from the old USGS station.

[b]Maintained by Dr. Eric H. De Carlo, Department of Oceanography, University of Hawaii.

**Table 2.2. A non-exhaustive list of some of the most recent point source pollution events on the island of Oahu, HI.**

| Date | Event |
|---|---|
| February-04 | Spill of 2,800 gallons of raw sewage due to overflow on the Navy-Marine Golf Course Driving Range |
| February-04 | Overflow of 45,000 gallons of raw sewage on the Kaneohe Preliminary Treatment Facility grounds and 25,000 gallons spill into Kawa Stream, which empties into Kaneohe Bay |
| February-04 | Sand Island Treatment Plant spills 47,250 gallons of sewage on the facility |
| February-04 | Wahiawa Wastewater Treatment Plant spills 40,000 gallons of wastewater due to a 9 minutes power disruption |
| February-04 | Waimanalo wastewater plant spills treated sewage due to heavy rain |
| March-04 | 2 million gallons of raw sewage spill in Sand Island beach park and into Pearl Harbour |
| May-04 | A ruptured line spills un unknown amount of sewage at the Kukanono wastewater pump station in Kailua into Kawai Nui Marsh |
| June-04 | A private contractor accidentally breaks a pressurized sewer line and caused a 6,975 gallons spill of untreated sewage into a storm drains that empties into Ewa Beach |

**Table 2.3. Classification of the 19 species of *Enterococcus* in groups relating to their importance as hygienic indicators (modified from Meier 1998).**

| Indicator role | Species name |
|---|---|
| Major hygienic importance | *E. durans, E. faecalis, E. faecium, E. hirae* |
| Minor hygienic importance | *E. avium, E. cecorum, E. columbae, E. gallinarum* |
| Low hygienic importance, non-fecal species | *E. casseliflavus, E. dispar, E. flavescens, E. malodoratus, E. mundtii, E. pseudoavium, E. raffinosus, E. saccharolyticus, E. sulfureus, E. seriolicida, E. solitarius* |

**Table 2.4. Microorganisms detected, culture media used and incubation details for the different membrane filtrations assay used.**

| Organism | Media | Incubation time (hours) | Incubation temperature (°C) | Reference |
|---|---|---|---|---|
| Enterococcus | mE | 48 | 41.5 | *Standard methods*[a] 9230C |
| *C. perfringens* | mCP | 18-24 | 45[b] | Bisson & Cabelli, 1979 |
| HPC | mHPC | 48 | 35 | *Standard methods*[a] 9215D |

[a] *Standard methods* refers to APHA (1998).

[b] *C. perfringens* must be incubated anaerobically

**Table 2.5. Descriptive statistics of the monitoring data collected at four stations in the Kaneohe watershed (Figure 2.1). TSS=total suspended solids; HPC=heterotrophic plate count bacteria, ENT=enterococcus, CP=*Clostridium perfringens*, see Methodology section and Table 2.4 for details on the quantification protocols.**

| | N | Min | 25th | Mean | Median | 75th | Max |
|---|---|---|---|---|---|---|---|
| | | | | LULU | | | |
| TSS (mg/L) | 28 | 1.13 | 2.30 | 3.35 | 2.92 | 3.40 | 10.67 |
| HPC ($10^5$ CFU/100ml) | 34 | 1.40 | 2.73 | 5.36 | 3.23 | 4.70 | 43.67 |
| ENT (CFU/100ml) | 34 | 38 | 280 | 1920 | 460 | 657 | 47000 |
| CP (CFU/100ml) | 33 | 1 | 1 | 8 | 3 | 6 | 86 |
| | | | | KAMO | | | |
| TSS (mg/L) | 19 | 1.10 | 1.30 | 3.92 | 1.60 | 1.90 | 39.12 |
| HPC ($10^5$ CFU/100ml) | 23 | 0.80 | 2.50 | 6.75 | 4.80 | 7.35 | 44.50 |
| ENT (CFU/100ml) | 23 | 32 | 52 | 654 | 120 | 433 | 10250 |
| CP (CFU/100ml) | 23 | 1 | 1 | 14 | 3 | 6 | 240 |
| | | | | KANE | | | |
| TSS (mg/L) | 28 | 1.65 | 3.20 | 5.27 | 5.10 | 6.00 | 24.77 |
| HPC ($10^5$ CFU/100ml) | 35 | 3.65 | 5.53 | 12.33 | 8.33 | 13.80 | 69.25 |
| ENT (CFU/100ml) | 36 | 217 | 488 | 1594 | 783 | 1333 | 9750 |
| CP (CFU/100ml) | 33 | 34 | 73 | 163 | 109 | 187 | 610 |
| | | | | LKS | | | |
| TSS (mg/L) | 18 | 1.95 | 2.40 | 5.67 | 4.20 | 5.10 | 33.00 |
| HPC ($10^5$ CFU/100ml) | 24 | 2.20 | 3.50 | 7.98 | 5.80 | 7.87 | 52.00 |
| ENT (CFU/100ml) | 24 | 230 | 399 | 1159 | 591 | 923 | 9200 |
| CP (CFU/100ml) | 23 | 17 | 40 | 64 | 63 | 80 | 130 |

**Table 2.6. Mann-Whitney U test analysis of the monitoring data compared station to station; numbers represent the p-values of the tests. The p-values represent the probability that the two data sets compared originate from the same population. TSS=total suspended solids; HPC=heterotrophic plate count bacteria, ENT=enterococcus, CP=*Clostridium perfringens*, see Methodology section and Table 2.4 for details on the quantification protocols.**

| Station | LULU | KAMO | KANE | LKS | |
|---|---|---|---|---|---|
| LULU | | <0.001 | <0.001 | 0.018 | TSS |
| KAMO | 0.362 | | <0.001 | <0.001 | |
| KANE | <0.001 | <0.001 | | 0.339 | |
| LKS | 0.002 | 0.145 | 0.006 | | |
| | **HPC** | | | | |

| Station | LULU | KAMO | KANE | LKS | |
|---|---|---|---|---|---|
| LULU | | <0.001 | 0.004 | 0.158 | ENT |
| KAMO | 0.844 | | <0.001 | <0.001 | |
| KANE | <0.001 | <0.001 | | 0.170 | |
| LKS | <0.001 | <0.001 | <0.001 | | |
| | **CP** | | | | |

**Table 2.7. Factor analysis by the principal component extraction method on different subsets of the log-transformed monitoring data. The factor identifications in bold correspond to those factors whose extracted eigenvalues are greater than 1. The station specific analysis was performed with 4 parameters (TSS, HPC, ENT, and CP), while a parameter indicator of sample location (stations) was added for the analysis performed on the complete dataset (all stations).**

| | Factor | Eigenvalue | % total variance explained | Cumulative % of variance explained | Factor identification |
|---|---|---|---|---|---|
| LULU | 1 | 3.43 | 85.8 | 85.78 | **TSS, HPC, ENT, CP** |
| (N=28) | 2 | 0.37 | 9.4 | 95.15 | TSS, HPC |
| KAMO | 1 | 3.90 | 97.5 | 97.52 | **TSS, HPC, ENT, CP** |
| (N=18) | | | | | |
| KANE | 1 | 2.40 | 59.9 | 59.91 | **CP** |
| (N=27) | 2 | 1.03 | 25.8 | 85.70 | **TSS, HPC, ENT** |
| LKS | 1 | 3.01 | 75.3 | 75.32 | TSS, HPC, ENT |
| (N=18) | 2 | 0.94 | 23.6 | 98.94 | **CP** |
| All | 1 | 3.04 | 60.8 | 60.85 | **STATION** |
| stations | 2 | 1.00 | 20.0 | 80.89 | **CP, STATION** |
| (N=91) | 3 | 0.47 | 9.4 | 90.30 | HPC |
| | 4 | 0.30 | 6.1 | 96.38 | TSS, HPC |

**Table 2.8. Concentrations of fecal indicators excreted in the feces of warm-blooded animals and excretion rates (wet weight) (Geldreich, 1978). The N-values indicate the number of animals required, assuming it is the only animal responsible for the presence of the indicators, to produce the signals observed for ENT and CP in Kaneohe Stream.**

| Animal | ENT[c] ($10^6$ #/g) | CP (#/g) | Excretion (g/day) | $N_{ENT}$ | $N_{CP}$ |
|---|---|---|---|---|---|
| Chicken | 3.4 | 250 | 182 | 525 | $6.1 \times 10^5$ |
| Cow | 1.3 | 200 (80[d]) | 23600 | 11 | 5874 |
| Duck | 54 | - | 336 | 18 | - |
| Horse | 6.3 | 1 (10[d]) | 20000 | 2.6 | $13.8 \times 10^5$ |
| Pig | 84 | 3980 | 2700 | 1.4 | 2580 |
| Sheep | 38 | 199000 (70[d]) | 1130 | 7.6 | 123 ($3.5 \times 10^{5d}$) |
| Turkey | 2.8 | - | 448 | 259 | - |
| Cat | 27 | $25.1 \times 10^6$ | 100[a] | 120 | 11 |
| Dog | 980 | $251 \times 10^6$ | 413 | 0.8 | 0.3 |
| Human | 3 | 1580[b] | 150 | 722 | $1.2 \times 10^5$ |

[a]Assumed value.

[b]Excreted by only 13-35% of humans.

[c]Quantified as fecal streptococci.

[d]Numbers in parenthesis are the data from Sorensen, Eberl & Dicksa (1989), Table 1.

**Table 2.9. Summary of the enterococcus identification data. Values represent the number of colonies identified. The numbers in parenthesis represent the percentage for the given sampling day. N is the date specific number of satisfactory identifications obtained.**

| Hygienic importance | Species | LULU | | | | | KANE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 5/6/04 N=13 | 5/12/04 N=11 | 5/18/04 N=11 | 5/26/04 N=4 | 6/2/04 N=2 | 4/15/04 N=14 | 4/21/04 N=15 | 4/28/04 N=3 | 5/26/04 N=11 | 6/2/04 N=8 |
| Major | *E. faecalis* | | | | | | | 15 (100) | | | |
| | *E. faecium* | 1 (8) | | | | | 2 (14) | | 2 (67) | | |
| | *E. hirae* | | 1 (10) | | | | | | | | 1 (13) |
| | *E. durans* | 1 (8) | | | | | | | | | 1 (13) |
| Minor | *E. gallinarum* | | 1 (10) | | | | | | | | |
| Low | *E. casseliflavus* | 11 (84) | 7 (70) | 4 (36) | 1 (25) | | 12 (86) | | 1 (33) | 2 (18) | 5 (61) |
| | *E. seriolicida* | | | | | | | | | 7 (64) | 1 (13) |
| Unknown | *Lactococcus* sp., *Streptococcus* sp. & *Leuconostoc* sp. | | 1 (10) | 7 (64) | 3 (75) | 2 (100) | | | | 2 (18) | |
| Enterococcus concentration (CFU/100ml) | | 47000 | 653 | 657 | 1080 | 1097 | 2300 | 9100 | 857 | 775 | 4433 |
| *C. perfringens* (CFU/100ml) | | 86 | 6 | 9 | 6 | 3 | --- | 293 | 273 | 35 | 180 |

**Table 2.10. Summary of the enterococcus identifications based on the absolute number of positive tests per station, regardless of the sampling date.**

| Hygienic importance | Species | LULU | | KANE | |
|---|---|---|---|---|---|
| | | N | % | N | % |
| Major | *E. faecalis* | - | - | 15 | 29 |
| | *E. faecium* | 1 | 2 | 4 | 8 |
| | *E. hirae* | 1 | 2 | 1 | 2 |
| | *E. durans* | 2 | 5 | 1 | 2 |
| Minor | *E. gallinarum* | 1 | 2 | - | - |
| Low | *E. casseliflavus* | 23 | 56 | 20 | 39 |
| | *E. seriolicida* | - | - | 8 | 16 |
| Unknown | *Lactococcus* sp., *Streptococcus* sp. & *Leuconostoc* sp. | 13 | 32 | 2 | 4 |

LULU: N=41; KANE: N=51

N is the total number of samples

**Table 2.11. Summary of the enterococcus identifications expressed as the number of days a species was present relative to the total number of days during which samples were taken at either station.**

| Hygienic importance | Species | LULU | | KANE | |
|---|---|---|---|---|---|
| | | D | % | D | % |
| Major | *E. faecalis* | - | - | 1 | 20 |
| | *E. faecium* | 1 | 20 | 2 | 40 |
| | *E. hirae* | 1 | 20 | 1 | 20 |
| | *E. durans* | 1 | 20 | 1 | 20 |
| Minor | *E. gallinarum* | 1 | 20 | - | - |
| Low | *E. casseliflavus* | 4 | 80 | 4 | 80 |
| | *E. seriolicida* | - | - | 2 | 40 |
| Unknown | *Lactococcus* sp., *Streptococcus* sp. & *Leuconostoc* sp. | 4 | 80 | 1 | 20 |

LULU: D=5; KANE: D=5
D is the number of days sampled.

**Table 2.12. A summary of available data on the species composition of the enterococcal community in different environments. The numbers in the table represent % calculated from the number of tests (N) available in each environment. Values highlighted in bold represent the environment-specific maximum percentages.**

| Hygienic importance | Species | California studies | | | | | | | | | Hawaii studies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Marine water[a] | Sediment[a] | Sewage | Soil[a] | Storm drain water[a] | Gull stool[a] | Surfzone, marine water[b] | Santa Ana river sediments[b] | Offshore sediments[b] | Soil[c] | Pristine stream[d] (LULU) | Urban stream[d] (KANE) |
| | | N=246 | N=108 | N=27 | N=23 | N=40 | N=54 | N=338 | N=112 | N=46 | N=36 | N=41 | N=51 |
| Major | *E. faecalis* | **29.3** | 14.8 | 37 | 13 | 22.5 | **53.7** | **29.6** | 20 | **43** | 31 | --- | 29 |
| | *E. faecium* | 22.4 | **35.2** | 37 | 13 | 12.5 | 9.3 | 14.5 | **35** | 20 | 3 | 2 | 8 |
| | *E. hirae* | 9.8 | 11.1 | 11.1 | 4.3 | 5 | 1.9 | 11.5 | 14 | 11 | --- | 2 | 2 |
| | *E. durans* | 0.4 | 0.9 | --- | --- | 2.5 | 1.9 | 0.3 | 2 | --- | 11 | 5 | 2 |
| Minor | *E. avium* | 0.4 | 1.9 | --- | --- | --- | --- | --- | --- | 4 | 3 | --- | --- |
| | *E. gallinarum* | 11.4 | 9.3 | --- | **52.2** | **37.5** | 1.9 | 1.5 | --- | --- | 19 | 2 | --- |
| Low | *E. casseliflavus* | 10.6 | 10.2 | --- | 8.7 | 5 | 3.7 | 7.4 | 11 | --- | **33** | **56** | **39** |
| | *E. seriolicida* | --- | --- | --- | --- | --- | --- | --- | --- | --- | | | 16 |
| | *E. mundtii* | 1.2 | 1.9 | --- | 4.3 | 5 | --- | 2.4 | 4 | --- | --- | --- | --- |
| Unknown | Other | 14.7 | 14.8 | **14.8** | 4.3 | 10 | 27.8 | 16.6 | 14 | 21 | --- | 32 | 4 |

[a] Data from Ferguson et al. 2003, ASLO

[b] Data from Ferguson et al. 2004, ASM

[c] Data from Fujioka & Byappanahalli, 2001

[d] This study

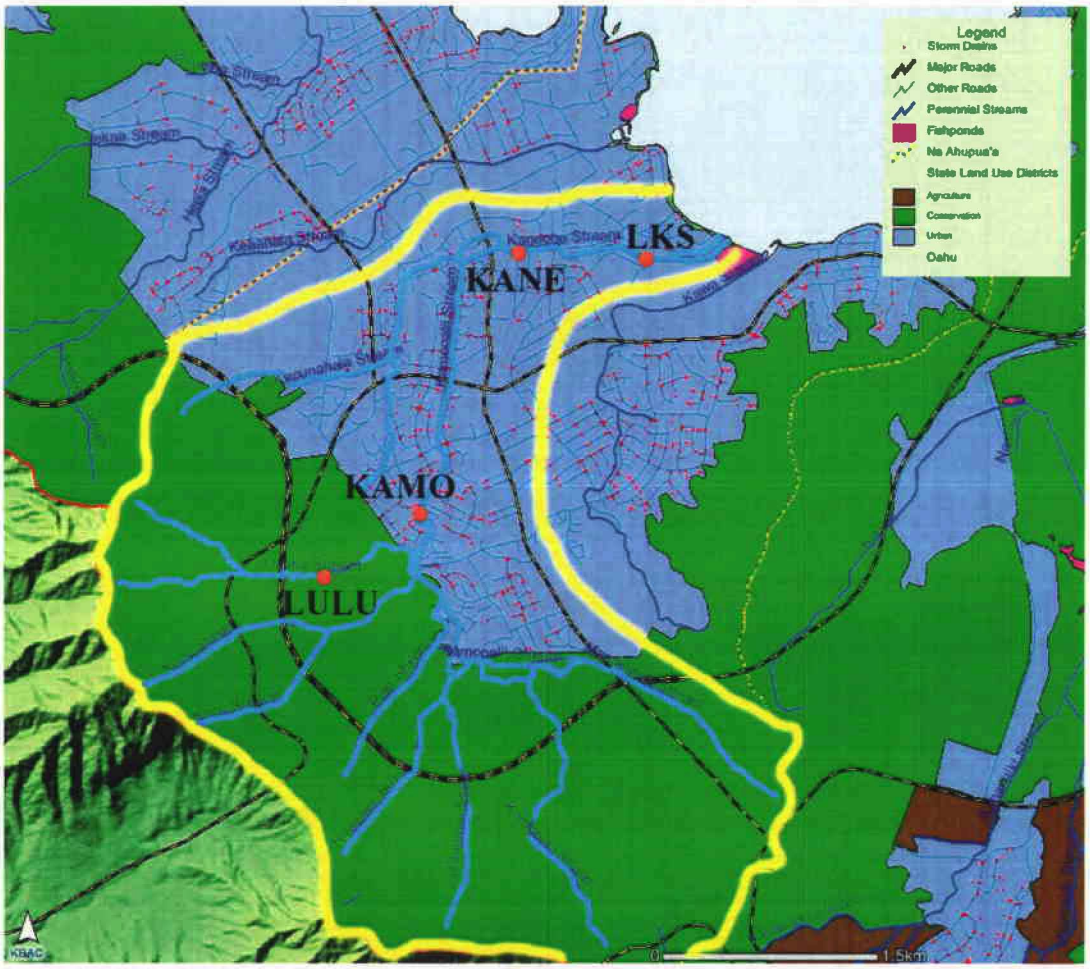


**Figure 2.1. Illustration of the Kaneohe Watershed (highlighted yellow). The sampling stations are shown by the red dots. Conservation land is in green, agricultural land is in brown, urban areas are indigo and streams are highlighted in blue.**

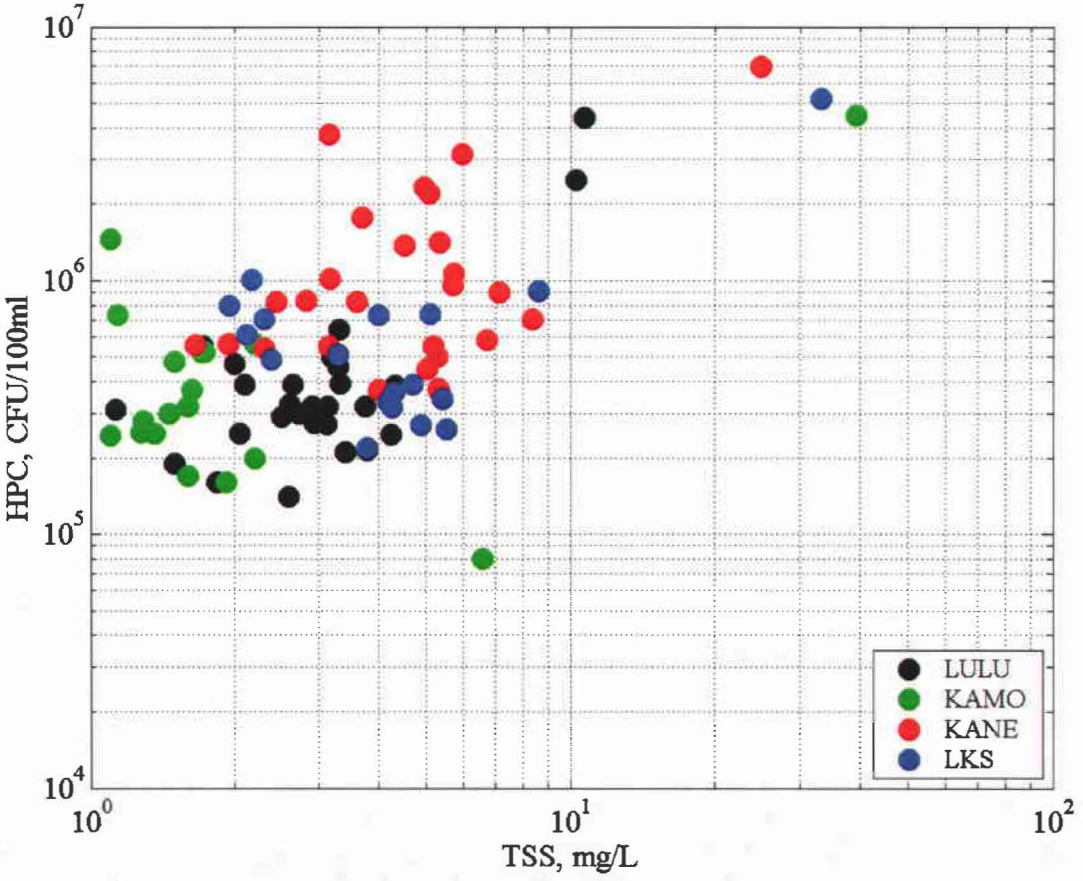


**Figure 2.2. Relationship between TSS and HPC in the Kaneohe watershed. TSS=total suspended solids; HPC=heterotrophic plate count bacteria; see Methodology section and Table 2.4 for details on the quantification protocols.**

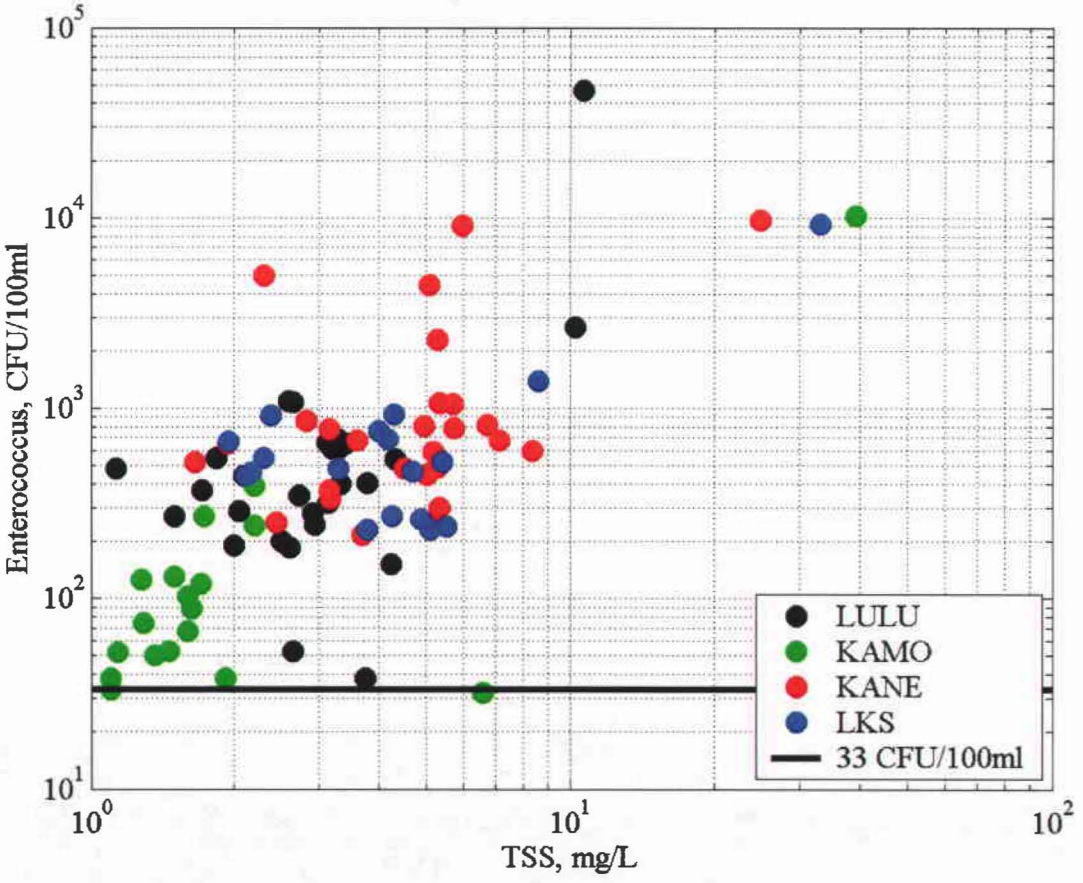


**Figure 2.3. Relationship between TSS and enterococcus. Most enterococcus concentrations exceed the water quality standard of 33 CFU/100ml. TSS=total suspended solids; see Methodology section and Table 2.4 for details on the quantification protocols.**

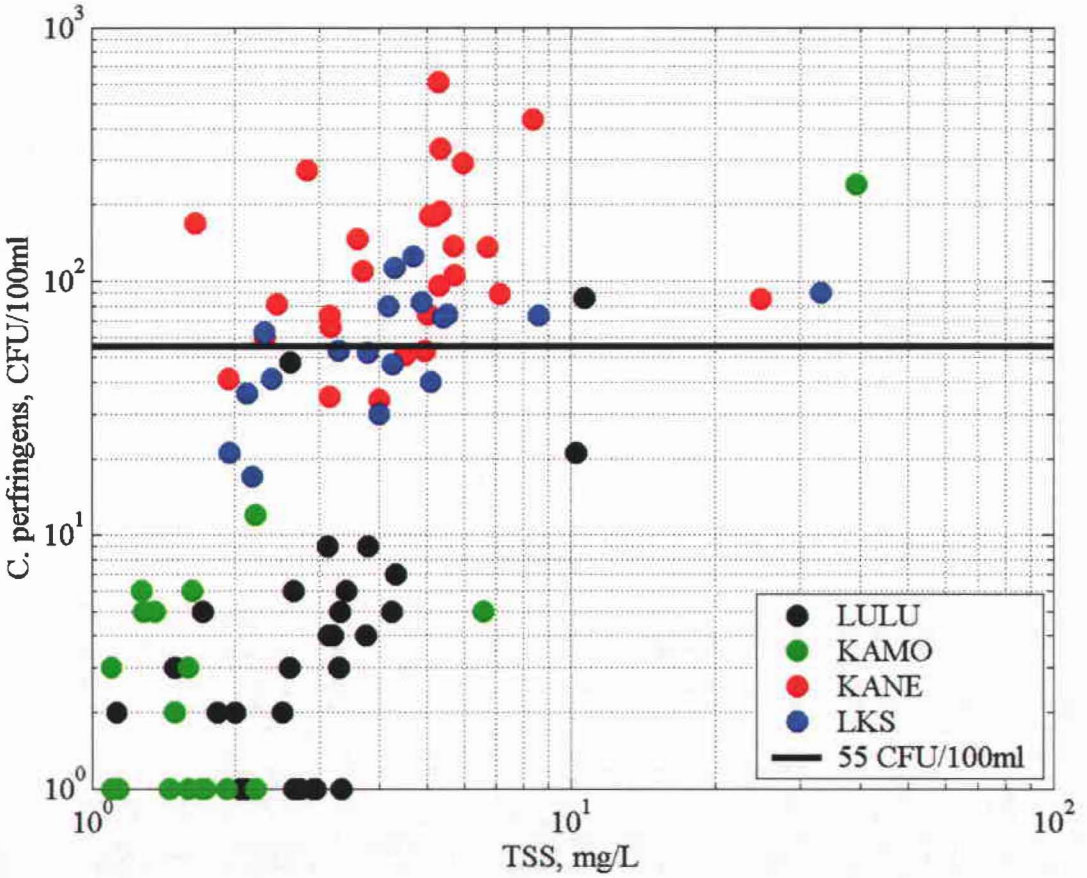


**Figure 2.4. Relationship between TSS and *C. perfringens*. Station KANE and LKS show higher values than expected by extrapolating the LULU an KAMO data to larger TSS values towards their "storm" levels, see Figure 2.6 and text for details. CP levels at stations KANE and LKS often exceed the water quality guideline for this indicator (55 CFU/100 ml).**

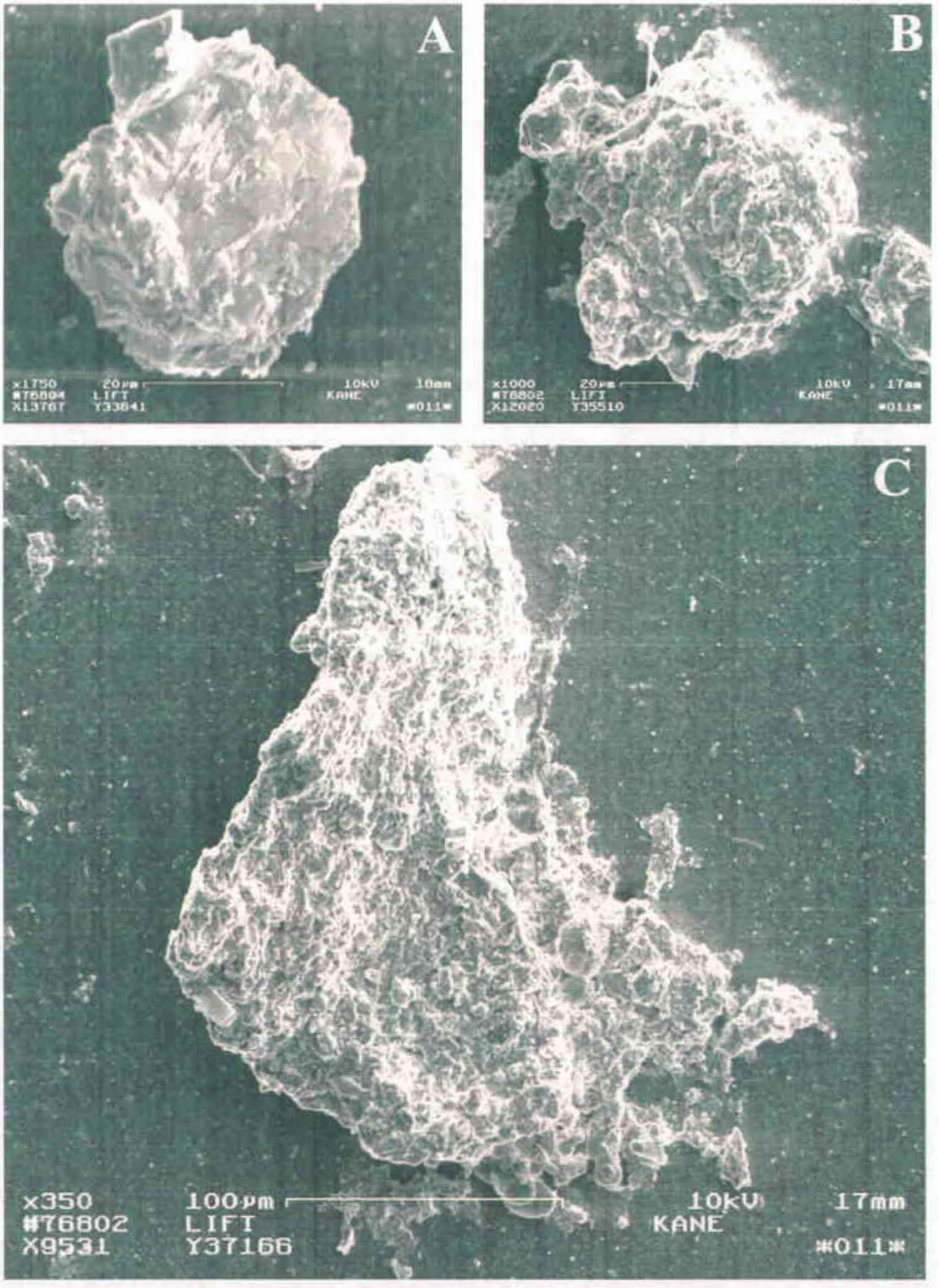


**Figure 2.5. Scanning electron microscope images illustrating that the suspended particles found in Kaneohe Stream are mainly inorganic. Some of them are covered with biofilm (B & C), while other have cleaner surfaces (A). Scale bars are 20 µm (A & B) and 100 µm (C).**

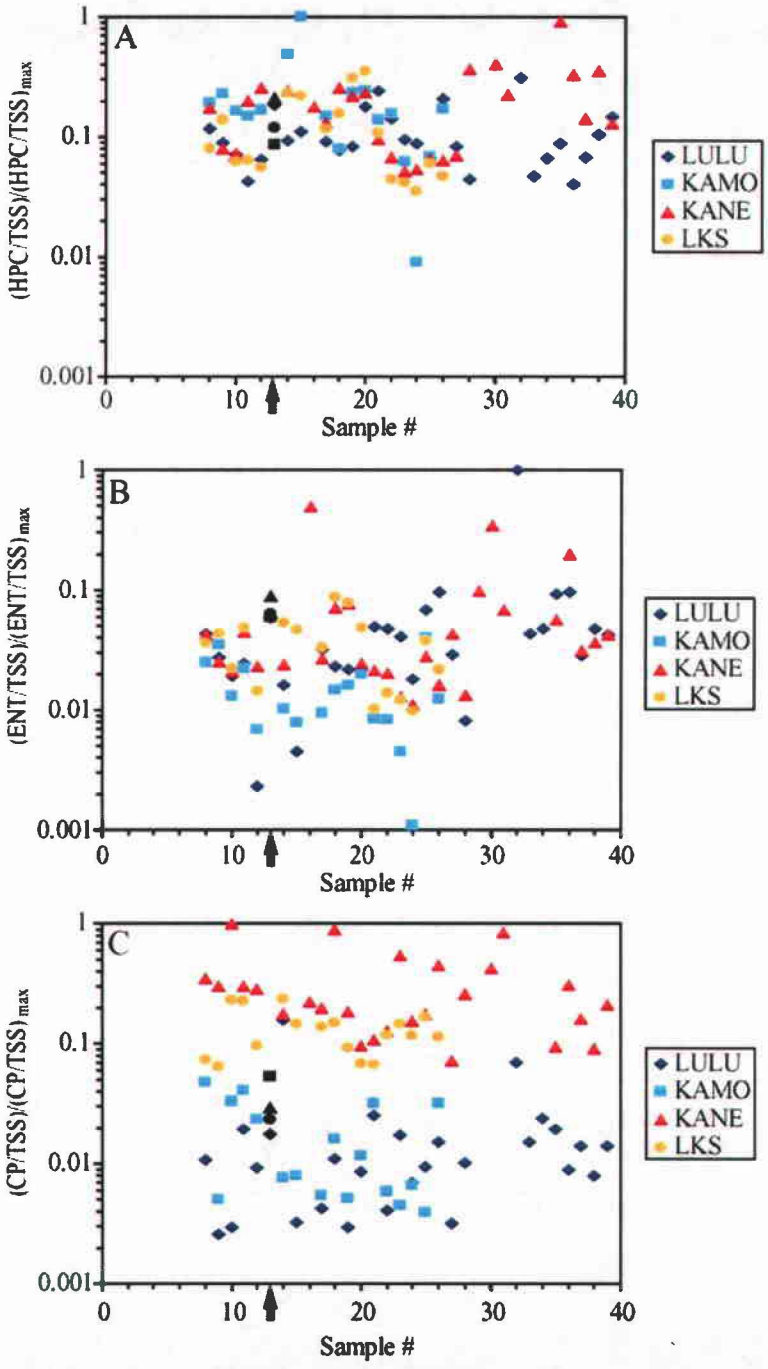


**Figure 2.6. Scaled and TSS-normalized microbial indicator levels at 4 monitoring stations in the Kaneohe watershed for A) HPC, B) ENT and C) CP. The black arrows and symbols represent samples collected during a storm event. Cold color symbols (red/orange) correspond to stations with no or little human disturbance. Warm color symbols (light and dark blue) represent stations impacted by human activity.**

# Chapter 3 : THE EFFECT OF BACTERIA CLUSTERING ON THE PRECISION OF THE CURRENT CULTURE METHODS: A THEORETICAL APPROACH.

## 3.1 Introduction

*"Much of our knowledge must always remain uncertain. The most we can know is in terms of probabilities"(Richard Feynman, 1963, p 6-11)*

The concentration of indicator organisms is generally assessed using culturing techniques: membrane filtration (MF) or most probable number (MPN). In addition to variability in the measurements due to the quality of the growth medium, the strain of organisms to be cultured or the metabolic state of the cells in the sample, variability can also occur because of the specific distribution of cells in the sample. In this chapter, this latter type of variability is discussed. For simplicity, the following discussion assumes that all cells grow into colonies.

If the distribution of bacteria present in a water sample were homogenous, any sub-sample would yield the same concentration. Since homogeneity is not achieved for natural samples, a random distribution is assumed. In a randomly distributed sample, there is an equal probability for any point in space to be occupied by a bacterium at any one time, such that sub-samples can have different concentrations. In theory, the mean of many replicate sub-samples will eventually yield the same concentration that would be measured if the distribution were homogenous, that is the true concentration of the sample.

Now consider what would happen if motile and randomly distributed bacteria collide and stick to each other. The resulting clusters are clearly anomalies in the random

distribution of the sample since the clustered cells are not independent of one another anymore. Although it is possible that the distributions of the clusters and of the single cells are random in the sample, the overall distribution of cells is not random. The following discussion addresses the question of how this scenario affects the cell density measurements and the implications for risk assessment. To measure the concentration of bacteria in a sample, it is necessary to have analytical techniques with detection limits that allow for the differentiation of all cells. Do culture methods provide this level of detection limit?

The goal of this chapter is to theoretically constrain the measurement errors associated with the culture methods commonly used for water quality assessment (membrane filtration and most probable number) induced by the two scenarios depicted above. The first section (Randomness and the Poisson distribution) discusses the error due to the inherent imperfections of the culture techniques in use; that is, an error occurring even if cells are randomly distributed. The second section (Deviations from the Poisson assumption), addresses the impact of groups of bacteria on the accuracy of the culture-based measurements.

## *3.2 Randomness and the Poisson distribution*

Commonly used culture-based methods for quantitative microbial risk assessment assume that the microbes present in the sample are distributed randomly. Randomness in general can be expressed mathematically with the Poisson distribution. The probability (P) that the number of organisms in a volume V(x) will equal N when the mean cell density is $\overline{\mu}$ is given by (Haas et al. 1999):

$$P(x=N)=\frac{(\bar{\mu}V)^N}{N!}\exp(-\bar{\mu}V). \quad (3.1)$$

The cumulative probability that a sample will have between $N_L$ and $N_U$ organisms is obtained by summing all the probabilities given in Eq. (3.1) from $N=N_L$ to $N=N_U$. If

$$N_U \to \infty$$

$$P(N_L \le x \le \infty)=1-P(0 \le x \le (N_L-1))=1-\sum_{N=0}^{N=(N_L-1)}\frac{(\bar{\mu}V)}{N!}\exp(-\bar{\mu}V). \quad (3.2)$$

For randomly distributed bacteria, at some points in space the local concentration may be large compared to the bulk concentration. Assuming Poisson conditions, how do these heterogeneities affect the culture-based measurements for the purpose of water quality?

### 3.2.1 The membrane filtration approach

It is generally assumed that the membrane filtration technique (MF), as opposed to the MPN method, provides a direct count of the target organisms present in the sample. However, the method results in data formulated in terms of colony forming units (CFU) per volume. The error associated with the potential discrepancy between CFU density and organism density is unknown.

The concept of CFU as a countable unit is imprecise. If all cells on a membrane filter initiate growth upon incubation, the number of CFUs corresponds to the minimum number of separable cells on the agar that give rise to visible colonies. These colonies generally range in sizes from 1 to 3 mm in diameter, whereas cells are only microns in size. Consequently, on a microscopic level, CFUs may consist of pairs, chains and clusters as well as single cells. The unit CFU is a variable quantity, and each CFU does

not have the same weight, a fact that is rarely considered in the field of quantitative microbial risk assessment. It is only possible to assume that at least one bacterium must have produced one CFU, but the possibility that multiple bacteria grew into one CFU after the merging of their respective micro-colonies cannot be excluded.

### 3.2.1.1 The role of chance alone

Assume that a sample is composed totally of single cells; there are no pairs, chains or clusters of bacteria. Further assume that the cells are distributed randomly in a volume, V. Consequently, upon filtration, there is a uniform probability that the cells will fall anywhere on the filter, such that the position of the cells on the filter will be random, too. It is possible that by chance alone, multiple cells may fall in close proximity to each other. If the distance that separates two cells is smaller than a critical value R, their respective colonies will merge and a single CFU will be produced. What is the critical distance R? What is the frequency of colony merging? What effect does colony merging have on the relationship between the number of CFUs and bacterial density?

This situation can be simulated on a computer. The algorithm Mfchance_2.m (Appendix B) was written for Matlab. Its function is to randomly assign coordinates to a number, N, of virtual bacteria, calculate the distances between each pair of bacteria and count how many CFUs are produced from clusters and how many are produced from single cells.

#### *3.2.1.1.1 Computer simulation*

The critical distance R, below which colonies merge, is unknown and probably varies depending on the physiological state of the cells. An active and healthy cell might

produce a larger colony than an inactive or damaged cell given a constant incubation time. In this study, R is understood as the radius of a circle that has, at its center, the bacterium (Figure 3.1). Given that colonies typically range from 1 to 3 mm in diameter, typical R-values range from 0.5 to 1.5 mm. The radius is believed to be a better parameter to define merging than the diameter because when only the edges of 2 colonies touch, it is possible to differentiate these colonies as two entities, even if, theoretically, they have merged (Figure 3.1).

The total number of intervals (distances) between N bacteria is given by the following equation:

$$N_E = (N-1)\frac{N}{2}. \tag{3.3}$$

Even if only less than 1 % of the total number of intervals between any two bacteria on the filters are below 2 mm (Figure 3.2, d), $N_E$ is large relative to N, such that the number of potential mergers is elevated. For N=70, $N_E$ is 2415. 1 % of this value corresponds to 24 distances, involving a maximum of 48 bacteria if only clusters of 2 cells are considered, a very large number relative to a total of 70 cells. Illustrative examples are shown in Figure 3.3 using different conditions for N (50 and 100) and R (0.75 and 1.5 mm). In these examples, between 81% and 96% of the true number of cells on the filter could be recovered as CFU. The difference between CFU counts and the number of cells increases as N and R increase. To assess whether these results are typical, this exercise was repeated a large number of times, with different initial conditions for N and R.

#### *3.2.1.1.2 Results*

Simulation results for 1000 hypothetical filters for different combination of bacterial concentrations (N) and R-values are shown in Figure 3.4 to Figure 3.10. At low N and with small colonies (small R), the number of single-cell CFUs closely approximates N; however, with increasing colony size (large R), the difference between N and the number of CFU produced becomes very large (Figure 3.4). The number of colonies originating from more than one bacterium increases non-linearly as N increases (Figure 3.5). The total number of cells trapped in clusters corresponds to the total number of cells in the sample minus the concentration of single-cell CFUs and increases with increasing N and R (Figure 3.6).

Most clusters are small (pairs), but larger clusters exist (Figure 3.7). The number of clusters of a given size ($n_c(X)$) decreases with increasing cluster size (X) following approximately a power law

$$n_c(X) = \gamma X^{\theta}. \tag{3.4}$$

The exponent ($\theta$) increases as the critical colony size is increased (Figure 3.8, A), meaning that it becomes more and more probable to find a larger cluster when the colonies become larger. $\theta$ was determined five-times for R-values ranging from 0.125 to 2 mm, in 0.0625 mm intervals. The regression model through the different values of $\theta$, $\theta=5.05\ln(R)-4.95$, yields an $R^2$ of 0.98 (Figure 3.8, A). A plot of the constant ($\gamma$) against R (Figure 3.8, B) shows a clear break in the neighborhood of $\gamma=1.3$. For $\gamma<1.3$, the constants were satisfactorily fit with a power law of the form $\gamma=1.31R^{-0.69}$ ($R^2=0.96$), while for $\gamma>1.3$, a linear fit ($\gamma=-3.29R+5.02$) fit the data well ($R^2=0.97$).

The mean number of "measurable" CFUs resulting from N initial bacteria consistently underestimates the true number of bacteria (Figure 3.9). The difference between the true number and the number of CFUs increases as the size of the colonies (R) and the number of the bacteria present (N). The relative error, expressed as the difference between the two parameters normalized by the true bacterial concentration, increases proportionally with N (Figure 3.10). For a given R, the error ($\varepsilon_R$) can be adequately described by a simple linear model,

$$\varepsilon_R = MN + B . \tag{3.5}$$

For a range of R (0<R<~1.3), the slopes (M) fit a simple quadratic function of the form $M=0.0012R^2$, but deviations from this relationship for greater R (Figure 3.8, C) are obvious. The y-intercept, B, is nearly 0 for smaller R, but drops significantly when R increases beyond ~1.3 (Figure 3.8, D). The dependency of the error on the square of the colony radius is evidence that for smaller R values, the error is a direct function of colony area. When R increases, however, clusters of clusters rather than primary clusters of individual colonies start to form at a significant rate, such that the relative error increases dramatically and can no longer be adequately described by Eq.(3.5) for all N.

#### *3.2.1.1.3 Conclusions*

The results presented above show a very strong dependency on R. This emphasizes the importance of keeping the incubation time as short as possible to keep colonies small. If R becomes too large, the number of CFUs produced that originate from more than 1 microbe reaches a peak. At this point, the total number of bacteria per CFU

increases sharply. Under these circumstances, the relative error can only be approximated by a linear fit for small N.

The recommended CFU density on a membrane filter (47 mm diameter) ranges from 20-60 for enterococcus (APHA, 1998), 30-300 for HPC (APHA, 1998), or generally 10-100 (Anon, 1983). The relative error at these densities is typically less than 25% when colonies are large (R=1.5 mm) and is much less for smaller colonies (less than 5% for R=0.5 mm) (Figure 3.10). It is important to note that the error described above does not disappear with replication. Replication in this case only makes the error predictable for a given value of R.

Increasing the diameter of the membrane filters would minimize the counting error due to colony merging, but is impractical. An increase in filter diameter would increase the range of bacteria concentrations that could be measured on a filter, decreasing the need for dilution, and providing more accurate numbers. A doubling of the filter diameter would decrease the bacteria density by a factor of 4, which in turn would also reduce the relative error due to clustering on the filter by a factor 4. If the error involved currently is judged acceptable, then doubling the filter size mimics the effect of a 4x dilution. Practically, however, a doubling of the filter diameter necessitates an appropriate 4 times increase in agar, which can be very costly to produce (e.g. mCP agar), further boosting the cost per sample.

### 3.2.2 The most probable number approach

The most probable number technique (MPN) is an alternative method used to quantify the concentration of microbes in a sample. The MPN method was developed

well before the MF method. Unlike MF, the MPN approach is a maximum likelihood technique in which no direct count of organism is made, but where the observable is the occurrence of growth. Several sterile tubes containing growth medium are inoculated with aliquots from a serial dilution series of from an original sample. The seeded tubes are incubated and the number of fertile (positive) and sterile (negative) tubes is recorded. Bacterial density estimates are calculated from the number of fertile tubes as follows.

For any one set (*i*) of replicates, the probability (P) of observing $p_i$ positive (fertile) tubes out of $n_i$ trials can be expressed by (Haas et al. 1999):

$$P(p_i) = \frac{n_i!}{p_i!(n_i - p_i)!} \pi_i^{p_i} (1 - \pi_i)^{n_i - p_i}. \tag{3.6}$$

The binomial coefficients $\binom{n_i}{p_i} = \frac{n_i!}{p!(n_i - p_i)!}$ represent the total number of ways to pull $p_i$ positives out of $n_i$ trials (CRM, 1992). $\pi_i^{p_i}$ represents the probability of having a fertile tube $p_i$ times. This has to be further multiplied by the probability of having sterile tubes for the remaining times: $(1 - \pi_i)^{n_i - p_i}$.

If the organisms are randomly distributed without aggregation of any kind and if all the organisms have an equal 100% chance to produce growth (i.e. the only explanation for a negative result is the absence of an organism), it is possible to define the probability ($\pi_i$) that a tube will have one or more organisms in it and be fertile if one knows the true mean cell concentration of the sample ($\bar{\mu}$) and the volume of the sample in the tubes (*V*), by:

$$\pi_i = 1 - \exp(-\bar{\mu}V). \tag{3.7}$$

Notice from Eq.(3.2) that $\pi_i = P(1 \le x \le \infty)$. After substituting Eq. (3.7) into Eq. (3.6) and multiplying the terms for all volumes ($V_i$), which are practically equivalent to different dilutions, the likelihood function (L) is (Haas et al. 1999):

$$L = \prod_{i=1}^{k} \frac{n_i!}{p_i!(n_i - p_i)} [1 - \exp(-\bar{\mu} V_i)]^{p_i} [\exp(-\bar{\mu} V_i)]^{n_i - p_i} . \quad (3.8)$$

The most probable number is the value of $\bar{\mu}$, that maximizes L in Eq. (3.8). This optimization must be done numerically, except when k=1, at which point an analytical solution can be found (Appendix A):

$$\bar{\mu} = -\frac{1}{V} \ln \frac{n - p}{n} . \quad (3.9)$$

In practice, the case of k=1 is rarely used for water quality assessment. Commonly, k=3; that is, the multiple tube most probable number assay is composed of 3 different volumes (e.g. 3 decimal dilutions), $V_{10ml}$, $V_{1ml}$, $V_{0.1ml}$, each with usually 5 tubes ($n_{10ml}$=$n_{1ml}$=$n_{0.1ml}$=5). After incubations, the number of positive tubes corresponding to each volume are recorded (e. g. $p_{10ml}$, $p_{1ml}$, $p_{0.1ml}$). These numbers can be plugged in Eq. (3.8), which can be numerically solved to yield the value of $\bar{\mu}$ that maximizes L.

The small-scale heterogeneities in bacterial concentration arising from the random distribution of mono-dispersed bacteria do impact each individual tube in an MPN assay. Nevertheless, the effect of these fluctuations is built into the derivation of the MPN theory in the form of $\pi_i$. In consequence, if the cells in the samples are Poisson (randomly) distributed, the precision of the MPN estimate is a function of the experimental design (number of dilutions, number of aliquots and number of volumes used). Maximum likelihood measurements show increasing bias as the number of

replicates is reduced. Methods have been developed to mathematically correct for the bias and constrain the error bounds due to the low number of replicates practically achievable (Norden 1973, Salama et al. 1978, Mehrabi & Matthews 1995). Unlike the MF method, for which there will always be underestimation of the true bacterial population because of colony merging, independently of the number of replicates, the maximum likelihood methods become asymptotically more accurate as the number of replicates is increased, assuming the cells are randomly distributed.

## *3.3 Deviations from the Poisson assumption*

### 3.3.1 Statistics of clusters and "heavy-weight" CFUs

In the previous exercise, chance was determined to be a possible factor contributing CFUs with more than one bacterium. The number of cells per CFU was generally low, however. Let us now consider the possibility that not all the microbes in a sample are "free-living" and randomly distributed but that some of these microbes live aggregated. Under these circumstances, the error of the MF measurement due to the presence of aggregated bacteria depends on the number of cells present in these aggregates.

Suppose that clusters of individuals, rather than single individuals are Poisson distributed, and the number of individuals per group is a random variable with its own probability distribution. In this particular case, the distribution that arises is termed a "generalized distribution" (Pielou, 1977). If the probability generating function (pgf) (Appendix C) of the number of clusters per volume is G(z), and the probability

generating function of the number of individuals per cluster is g(z), then the pgf of the generalized function (H(z)) can be expressed by:

$$H(z) = G(g(z)). \tag{3.10}$$

By defining $P_i$ the probability that a unit contains $i$ clusters ($i \in \mathbb{N}$) and $\pi_j$ the probability that a cluster contains j individuals ($j \in \mathbb{N}$), we can write (Pielou, 1977):

$$G(z) = \sum_i P_i z^i \tag{3.11}$$

$$g(z) = \sum_j \pi_j z^j \tag{3.12}$$

**such that equation (3.10) becomes**

$$H(z) = \sum_i P_i \left[ \sum_j \pi_j z^j \right]^i. \tag{3.13}$$

If we assume that the number of clusters per unit volume (G(z)), independently of cluster size, is a Poisson variable with a mean number of clusters $\lambda_c$, G(z) can be expressed by

$$G(z) = \exp(\lambda_c (z-1)). \tag{3.14}$$

This last result can be obtained by applying the definition of pgf found in Appendix C to equation (3.1).

We must now define a probability distribution describing the number of bacteria per cluster. Pielou (1977) proposed a solution for a similar problem, taking as example the number of caterpillars per shoot. Pielou's solution, presented in Eq. (3.15) to Eq. (3.28), can be adapted for the purpose of this study. It is probable to find a large number of clusters with a low number of bacteria and very few clusters with a large number of

bacteria. That is, we could assume that P(X) is adequately described by a logarithmic distribution such that (Pielou, 1977)

$$P(X) \propto \frac{A^X}{X}, \tag{3.15}$$

where A is a constant. Consequently

$$\sum_{X=1}^{\infty} P(X) \propto \frac{A^1}{1} + \frac{A^2}{2} + \frac{A^3}{3} + \ldots = -\ln(1-A). \tag{3.16}$$

It is necessary to define more precisely what is intended by the term "cluster". Here, cluster is taken to mean "aggregate", or the grouping of target bacteria and other suspended particles. A cluster is any entity containing more than just one cell of the target organism: be it one target organism stuck to one grain of sand or multiple target bacteria stuck together.

Since, by the above definition, a cluster is only a cluster when at least one bacterium is associated with a suspended particle, X=0 is excluded, and

$$\sum_{X=1}^{\infty} P(X) = 1, \tag{3.17}$$

Relation (3.15) formally becomes

$$P(X) = \frac{-1}{\ln(1-A)} \frac{A^X}{X}, \text{ for X=1,2,... and 0<A<1.} \tag{3.18}$$

The proportionality constant can be found by noticing that Eq. (3.16) will only equal Eq. (3.17) if Eq. (3.16) is multiplied by $\frac{-1}{\ln(1-A)}$.

Simple application of the definition of pgf (Appendix C) yields:

$$g(z)=\frac{-1}{\ln(1-A)}\left[Az+\frac{A^2}{2}z^2+\frac{A^3}{2}z^3+\frac{A^3}{2}z^3+\ldots\right] \tag{3.19}$$

Using the relationship in equation (3.16), the result can be simplified to:

$$g(z)=\frac{\ln(1-Az)}{\ln(1-A)}. \tag{3.20}$$

Therefore, the generalized distribution is obtained by:

$$H(z)=G(g(z))=\exp\left\{\lambda_c\left[\frac{\ln(1-Az)}{\ln(1-A)}-1\right]\right\}. \tag{3.21}$$

Equation (3.21) can be greatly simplified if we let

1) $\lambda_c = k\ln Q$

2) $A=\frac{P}{Q}$

3) $Q=1+P$.

Using relation 2) and 3) we can write

$$\ln(1-A)=\ln\left(1-\frac{P}{Q}\right)=\ln\left(\frac{1}{Q}\right)=\ln(1)-\ln(Q)=-\ln(Q). \tag{3.22}$$

such that Eq. (3.21) becomes

$$
\begin{aligned}
H(z) &= \exp\left\{k \ln Q \left[\frac{\ln\left(1-\frac{P}{Q}z\right)}{\ln\frac{1}{Q}} - 1\right]\right\} \\
&= \exp\left\{k \ln Q \left[\frac{\ln\left(1-\frac{P}{Q}z\right) - \ln\frac{1}{Q}}{\ln\frac{1}{Q}}\right]\right\} \\
&= \exp\left\{-k \ln\left(1-\frac{P}{Q}z\right)\right\} \cdot \exp\left\{k \ln\frac{1}{Q}\right\} \\
&= \exp\left\{-k \ln\left(1-\frac{P}{Q}z\right)\right\} \cdot \exp\{-k \ln Q\} \\
&= \left(1-\frac{Pz}{Q}\right)^{-k} Q^{-k} \\
&= (Q-Pz)^{-k}
\end{aligned}
\tag{3.23}
$$

With Q-P=1, the mean of this pgf is

$$m = H'(1) = kP \tag{3.24}$$

and the variance will be given by

$$H''(1) + H'(1) - (H'(1))^2 = kP(1+P) = kPQ. \tag{3.25}$$

The variance (v) can be expressed by a function of the mean (Pielou, 1977) as

$$v = m + \frac{m^2}{k}. \tag{3.26}$$

From this last relation, we conclude that when k is small, the variance will be large, which in terms of clusters of bacteria means that the range of probable number of bacteria per volume will be large and the results from individual samples will vary greatly (Figure 3.11).

By the theory of the pgf (Appendix C), the probability of finding exactly r clustered bacteria in a volume of sample is given by the coefficients of $z^r$ in the expansion of H(z). As an example, we can take k=-4 and expand H(z):

$$\begin{aligned} H(z) &= Q^4\left(1-\frac{P}{Q}z\right)^4 \\ &= Q^4\left[1^4-4\frac{P}{Q}z+6\left(\frac{P}{Q}\right)^2 z^2-4\left(\frac{P}{Q}\right)^3 z^3+\left(\frac{P}{Q}\right)^4 z^4\right] \end{aligned} \tag{3.27}$$

We quickly recognize the coefficients [1,4,6,4,1] as they are the binomial coefficients. The result of Eq. (3.27) can be generalized to yield the coefficients of $z^r$:

$$p_r = \frac{k(k+1)...(k+r-1)}{r!}Q^{-k}\frac{P^r}{Q^r} = \frac{\Gamma(k+r)}{r!\Gamma(k)}\frac{P^r}{Q^{k+r}} = \frac{\Gamma(k+r)}{r!\Gamma(k)}P^rQ^{-r-k} \tag{3.28}$$

This result, indeed, is the negative binomial distribution (Pielou, 1977), whose mean and variance we have already calculated in Eq. (3.24) and Eq. (3.26).

### 3.2.2 Measurement bias associated with particle-bound microbes: Theory

The total number of organisms in a sample is given by the sum of the free-living and the clustered bacteria. Knowing the probability distribution of both of these populations, one should be able to determine the probability distribution for $P_{Tot}$, the probability distribution of the total number of microbes ($N_{Tot}$) in the sample. This is not simple, however, since there are many ways to produce $N_{Tot}$ bacteria out of two populations. For example, the number 5 can be obtained by 5+0, 4+1, 3+2, 2+3, 1+4 and 0+5; that is, 6 (=$N_{Tot}$+1) different ways. In other words, $P_{Tot}$ is a function of multiple variables, such that a graph of $P_{Tot}$ is not simply a line, but a surface. The table of

isotopes in chemistry is a good analogy to this situation, where the atomic mass is given by the sum of the number of protons and neutrons, such that different elements can have the same atomic masses (isobars). In the present case, the isobars are the samples with the same total number of bacteria but with different number of clustered bacteria ($N_1$) and free-living microbes ($N_2$).

The product of the probability distributions of the two variables gives the resulting probability of finding $N_{Tot}$ bacteria in a sample. Mathematically,

$$P_{Tot} = P^{N_{Tot}=N_1+N_2}_{(X=N_2;Y=N_1)} = P(X = N_2)\cdot P(Y = N_1). \tag{3.29}$$

From equation (3.1) and (3.28),

$$P_{Tot} = \frac{\lambda_f{}^{N_2}}{N_2!}\exp(-\lambda_f)\cdot\frac{\Gamma(k+N_1)}{N_1!\Gamma(k)}P^{N_1}Q^{-N_1-k}, \tag{3.30}$$

where $\lambda_f$ is the mean concentration of the free-living cells. By substituting

$$Q = 1+P, \tag{3.31}$$

$$P = \frac{A}{1-A} \tag{3.32}$$

$$\text{and } k = \frac{\lambda_c}{\ln\left(1+\frac{A}{1-A}\right)}, \tag{3.33}$$

with $\lambda_c$ representing the mean number of clusters in the sample, equation (3.30) becomes

$$P_{Tot} = \frac{\lambda_f{}^{N_2}}{N_2!}\exp(-\lambda_f)\cdot\frac{\Gamma\left(\frac{\lambda_c}{\ln\left(1+\frac{A}{1-A}\right)}+N_1\right)}{N_1!\Gamma\left(\frac{\lambda_c}{\ln\left(1+\frac{A}{1-A}\right)}\right)}\left(\frac{A}{1-A}\right)^{N_1}\left(1+\frac{A}{1-A}\right)^{-N_1-\frac{\lambda_c}{\ln\left(1+\frac{A}{1-A}\right)}} \tag{3.34}$$

with the condition that $N_{Tot} = N_1 + N_2$.

We already know from Figure 3.11, that the variance will be large when k is small. Since k plays a big role in controlling the spread of the data and since it is a function of A, it is important to understand what control A has on k (Figure 3.12). Recall that A controls the distribution of bacteria per cluster: the cluster size distribution. Larger values of A indicate more large clusters. Consequently, as A approaches 1, the variance around the mean of $P_{Tot}$ will be large.

The pgf of equation (3.34) (J(z)) can be derived by multiplying the pgfs of the Poisson distribution describing the distribution free-living bacteria (F(z))

$$F(z) = \exp(\lambda_f (z-1)) \tag{3.35}$$

and the pgf describing the number of bacteria in clusters (H(z), Eq. (3.23)): $J(z) = F(z) \cdot H(z)$. This yields:

$$J(z) = \exp\left(\lambda_f (z-1)\right) \cdot \left(Q - Pz\right)^{-k} \tag{3.36}$$

The first derivative is given by J'(z), using the fact that $(f \cdot g)' = f'g + fg'$:

$$\begin{aligned} J'(z) &= \left(\lambda_f \exp\left(\lambda_f (z-1)\right)\right) \cdot \left(Q - Pz\right)^{-k} + \left(\exp\left(\lambda_f (z-1)\right)\right)\left[(-k)(-P)(Q-Pz)^{-k-1}\right] \\ &= \exp\left(\lambda_f (z-1)\right) \cdot \left(Q - Pz\right)^{-k} \cdot \left(\lambda_f + \frac{kP}{(Q-Pz)}\right) \end{aligned} \tag{3.37}$$

such that the mean (J'(1)) is

$$m = J'(1) = \lambda_f + kP. \tag{3.38}$$

The second derivative J''(z) is

$$
\begin{aligned}
J''(z) &= \lambda_f{}^2 \exp\left(\lambda_f (z-1)\right)(Q-Pz)^{-k} + \lambda_f \exp\left(\lambda_f (z-1)\right) kP(Q-Pz)^{-k-1} \\
&\quad + \lambda_f kP \exp\left(\lambda_f (z-1)\right) kP(Q-Pz)^{-k-1} + kP(kP+P)\exp\left(\lambda_f (z-1)\right)(Q-Pz)^{-k-2} \\
&= \exp\left(\lambda_f (z-1)\right)(Q-Pz)^{-k}\left[\lambda_f{}^2 + kP(Q-Pz)^{-1} + \lambda_f kP + \left(k^2P^2 + kP^2\right)(Q-Pz)^{-2}\right]
\end{aligned}
\tag{3.39}
$$

such that

$$
\begin{aligned}
J''(1) &= \lambda_f{}^2 + \lambda_f kP + \lambda_f kP + k^2P^2 + kP^2 \\
&= \lambda_f{}^2 + 2\lambda_f kP + kP^2(k+1)
\end{aligned}
\tag{3.40}
$$

The variance becomes $J''(1) + J'(1)\left[1 - J'(1)\right]$:

$$
\begin{aligned}
v &= \lambda_f{}^2 + 2\lambda_f kP + k^2P^2 + kP^2 + \lambda_f + kP - \lambda_f{}^2 - 2\lambda_f kP - k^2P^2 \\
&= kP^2 + kP + \lambda_f = \lambda_f + kP(P+1) = \lambda_f + kPQ
\end{aligned}
\tag{3.41}
$$

which is simply the variance of the Poisson distribution of free-living cells added to the variance of the negative binomial distribution of the particle-associated cells.

### 3.2.2.1 Model results

To investigate the behavior of Eq. (3.34), different values for the coefficients A, $\lambda_f$ and $\lambda_c$ were assumed. N was assigned and Eq. (3.34) was rewritten in terms of $N_2$, which varied between 1 and N. It was further assumed that $\lambda_c$ was proportional to $\lambda_f$ with a proportionality constant $\omega$: $\lambda_c = \omega\lambda_f$. An argument in favor of this latter assumption is that bacterial concentration and turbidity often correlate (see Chapter 2; Bidle & Fletcher, 1995; Loge et al., 2002). Hypothetically, as the concentration of bacteria increases, the number of clusters also rises.

$P_{Tot}$ was normalized to the maximum probability ($P_{max}$) obtained for all N for the given conditions of A, $\lambda_f$ and $\lambda_c$. In other words, the color scale in Figure 3.13 to Figure

3.15 represents the relative magnitude of the probability under specific conditions, not an absolute number. The maximum probabilities used to normalize each graph are explicitly plotted in Figure 3.16.

Figure 3.13 to Figure 3.15 illustrate how the probability surface ($P_{Tot}$) varies when different numbers of clusters are present in a sample. At low ω-values (left panels, ω=0.1), the colored patch of highest probabilities is almost symmetric and close to the intersection of the pink lines. The vertical pink line is simply $\lambda_f$. The diagonal pink line is given by $N_2 + \lambda_c = N_2 + \omega \cdot \lambda_f$, where $N_2$ is the hypothetical number of free-living microbes. The intersection point represents the expected number of CFUs that could be recorded, ignoring, as for the membrane filtration technique, 1) that colony merging does occur by chance alone during filtration (see section 3.2) and 2) that multiple cells can be present in a cluster. The value of N corresponding to this intersection point is given by summing the mean number of free-living cells and the mean number of clusters ($N_{MF}=\lambda_f+\lambda_c$). At constant A, but with a larger ω (right panels, ω=0.25) the intersection point shifts upwards because of the larger number of clusters present. Notice, that the point of maximum probability increasingly overshoots the intersection point with increasing cluster number. Increasing A, from 0.3 to 0.9 mostly affects the dispersion of the probability patch (Figure 3.13 to Figure 3.15), as discussed earlier and illustrated in Figure 3.12: the variance scales with A.

Assuming that the number of replicates is large, the point of maximum probability $P_{max}$ could be used to calculate an average error committed by not accounting for the bacteria present in clusters in numbers equal or greater than two. The relative error can be expressed by

$$\varepsilon_A = \frac{N(P_{\max}) - (\lambda_f + \lambda_c)}{N(P_{\max})} \tag{3.42}$$

and represents the difference between the projection of the intersection point of the two pink lines onto the vertical N axis and the N-value corresponding to the point of maximum probability ($N(P_{max})$), normalized to that value. Under the conditions considered, the error ranges from 0 to 35% of the total number of bacteria present (Figure 3.17).

### 3.2.2.2 Discussion

The average error presented here is comparable in magnitude to the error term ($\varepsilon_R$) calculated in a previous section (3.2.1) arising from colony merging during the membrane filtration procedure (Figure 3.10) but has the potential to be much greater depending on the relationship used to describe the number of cells in clusters. The variance associated with the presence of clusters can be substantial when the number of clusters is high (high $\omega$) or when the number of bacteria per cluster is large (high A). For individual samples, the error involved by neglecting clustered bacteria can be unpredictably small (~0%) or large (>100%). This uncertainty is reflected in the low $P_{max}$ values obtained in the above analysis (Figure 3.16). The maximum probabilities range from 0.001 to 0.015. In other words, any given combination of $N_1$ and $N_2$ has a low probability of occurrence, even the most probable combinations. These low values, however, are due to the wide range of possible values of $N_1$ and $N_2$ considered. In absolute terms, there is a large probability that a sample will contain a combination of $N_1$ and $N_2$ values that falls within the colored patch (Figure 3.13 to Figure 3.15).

The measurement bias arising from the presence of clustered bacteria in a sample depends on the number of clusters. The number of clusters decreases drastically upon dilution (Figure 3.18). When membrane filtration is used to detect heterotrophic plate count bacteria (HPC), dilutions ranging from $10^{-4}$ to $10^{-6}$ are commonly used. When fecal indicators are targeted, however, such as *Clostridium perfringens* or enterococci, the dilutions used are typically $10^{0}$ to $10^{-2}$. This practice effectively eliminates clusters at high dilutions (low sample volume) such that the HPC measurement *per se* should be little affected by clustered bacteria. On the other hand, at low dilution (large sample volume), the number of clusters may be high, such that clustered bacteria can be abundant. Although measurements requiring multiple dilutions are less accurate, they are more precise. Highly diluted samples almost certainly miss the population of the aggregated bacteria ($\lambda_c \rightarrow 0$ in the measured volume). For these diluted samples, the mean CFU density almost equals $\lambda_f$, and the variance is small ($v \rightarrow \lambda_f$, see equation (3.41)). The error becomes maximal:

$$\varepsilon_A \approx \frac{N(P_{\max}) - \lambda_f}{N(P_{\max})}. \tag{3.43}$$

The simple routine procedure of dilution is responsible for a significant amount of variability between replicates. Even if cells are randomly distributed in a sample, large uncertainties are associated with the estimated counts, a fact that is often ignored but that may explain a large fraction of the variability. For example, Tillett & Farrington (1991) estimated that the 95% confidence interval for a sample diluted 10 times (e.g. $10^{0}$ to $10^{-1}$) from an original volume of 100 ml and producing 100 CFUs is 823-1205, or roughly ± 20% of the estimated count of 1000 in the pre-dilution sample.

Like all models, the one described here relies on a list of assumptions. These include assumptions on the distribution of single cells, the distribution of clusters and the distribution of cells on clusters. The random distributions of single cells and of the number of clusters in a sample may be intuitive, but it is more difficult to justify the assumed logarithmic distribution of the number of cells per cluster.

The results presented here depend heavily on the functionality of the cluster size distribution (Eq. (3.18)). For example, the assumption that a distribution of this form accurately describes the natural phenomenon seems acceptable if the clustering mechanism is of a stochastic nature. However, if a bacterium becomes attached to a particle that is rich in nutrients, it could initiate growth and replication (Grossart et al, 2003; Kiorboe et al., 2003) resulting in a hot-spot of bacterial activity and elevated concentration. Should bacteria attaching to particles systematically form colonies onto these particles, the error calculated above could greatly underestimate the true impact of clustered cells on water quality assessment. Under these circumstances, the errors could easily be larger than an order of magnitude.

## 3.4 Conclusions

The inability of membrane filtration to effectively enumerate all potential CFU when these are in close proximity generates large measurement errors. These errors vary as a function of the number of cells in the sample and the size of the filter and of the colonies, but are also and mostly influenced by the characteristics of the distribution of aggregated bacteria in the sample. Empirical data constraining this distribution and the

variability of this distribution in different environments are necessary to ground truth the model but are unfortunately unavailable.

In summary, when a measurement is made using any culturing technique, both types of error discussed in this chapter will affect that measurement

$$\varepsilon = \varepsilon_R + \varepsilon_A \tag{3.44}$$

For example, imagine that membrane filtration is used and 70 CFU/100ml are reported as the mean of a large number of replicates. If the colonies averaged 2 mm in diameter, then the average number of CFUs that actually were on the filter (estimated from Figure 3.9) is roughly 75 CFU/100ml. Provided equation (3.18) is accurate, depending on the values of A and R, the mean of the bacterial concentration could in practice range from 75 to ~110 CFU/100ml (Figure 3.17). In other words, the "true" mean could have been underestimated by over 50%. This value could be much larger when individual samples are considered.

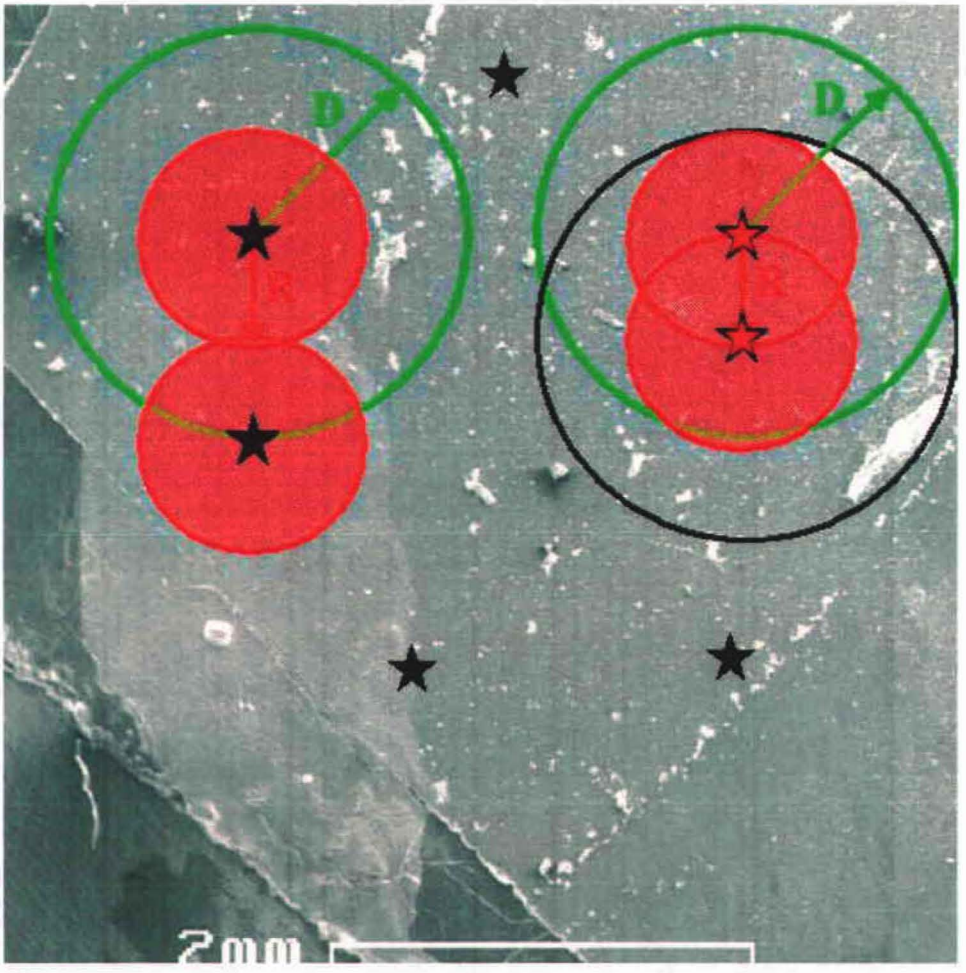


**Figure 3.1. R, the critical distance under which colonies merge, can be understood as the radius of the colonies that form. Stars represent individual bacteria. The red circles are colonies with a radius R and a diameter D=2R. When merging is defined in term of the diameter, represented by the green circles (left), two colonies would be considered merged, when in fact they are clearly distinguishable. Choosing the radius as critical distance (right) represents a more realistic problem.**

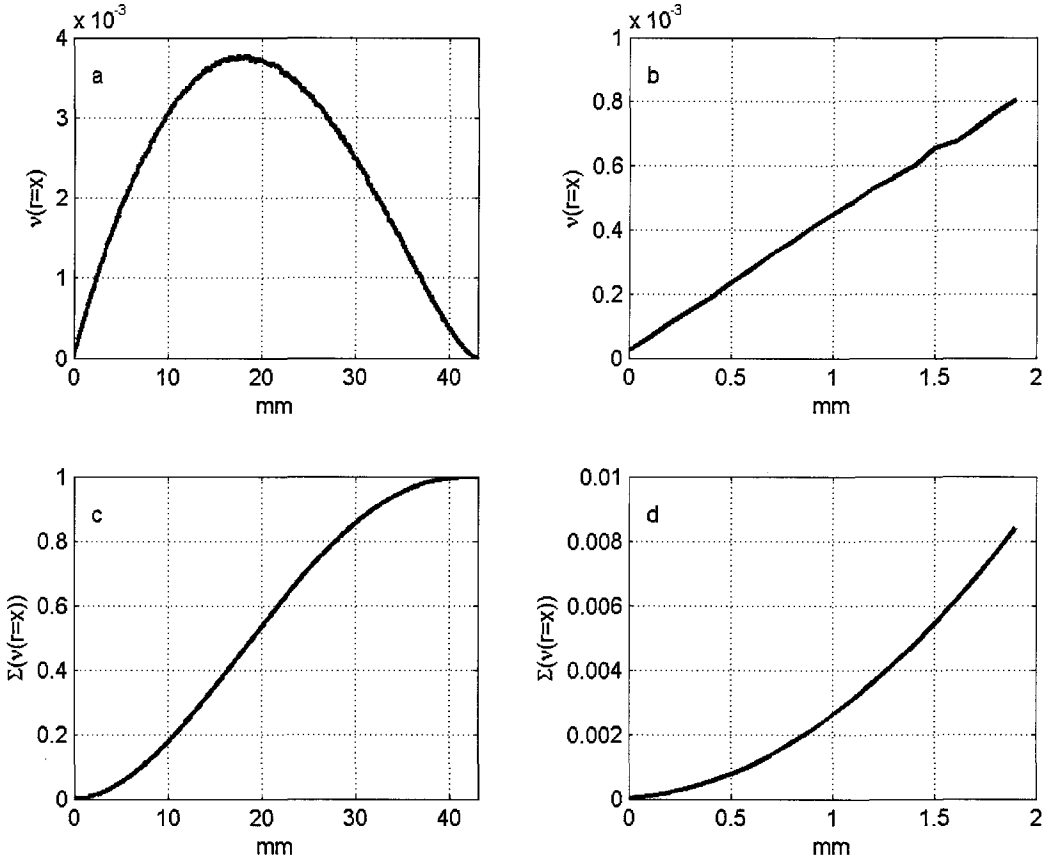


**Figure 3.2. Frequency (ν) distribution of the distances between any two points on a 43 mm diameter disk (a & b) and the corresponding cumulative probability plots (c & d). Interestingly, the distribution is not symetric and skewed towards smaller sizes. The plots on the right (b & d) represent zoomed-in portions of the left plots (a & c) for the distances of interest in the current discussion.**

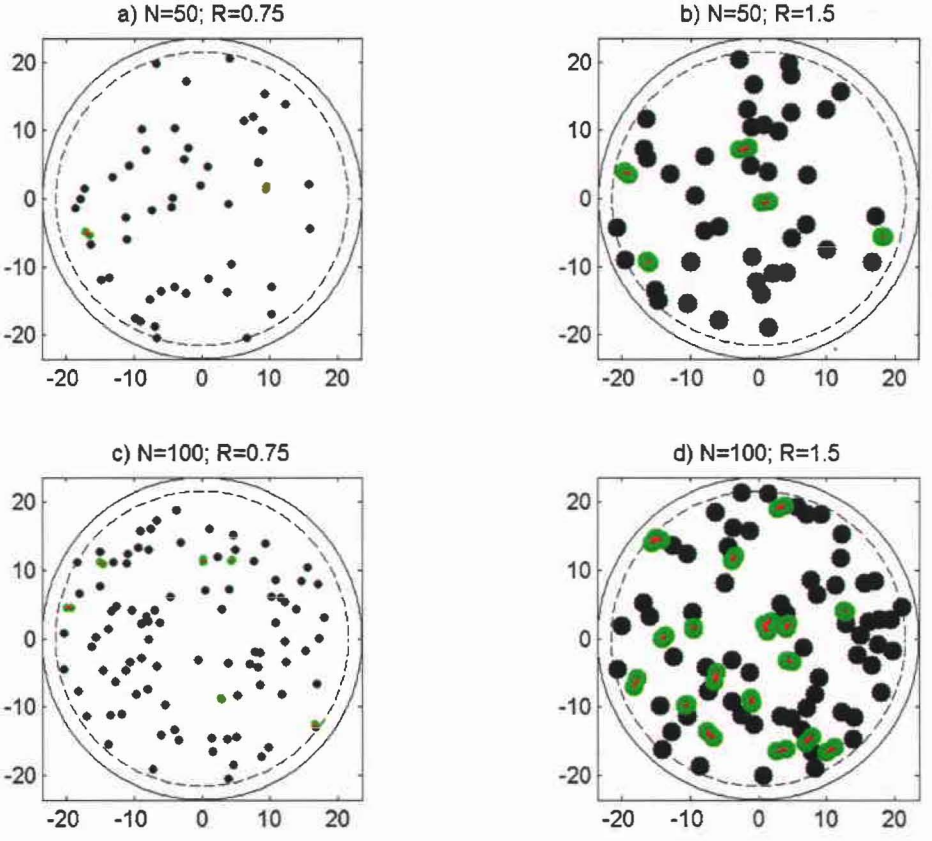


**Figure 3.3. The problem of colony merging becomes increasingly large when the number and the size of the colonies increase. The points in the graphs were scaled to represent the relative size of the colonies compared to the 47 mm diameter filter (circle). The dashed circle represents the effective filtration area (43 mm diameter). The green dots represent the colonies that have merged, or that are closer than R units apart. The red lines illustrate the "links" between the merged colonies. Actual number of CFU measurable by the membrane filtration assay are a) 48, b) 45, c) 93 and d) 81.**

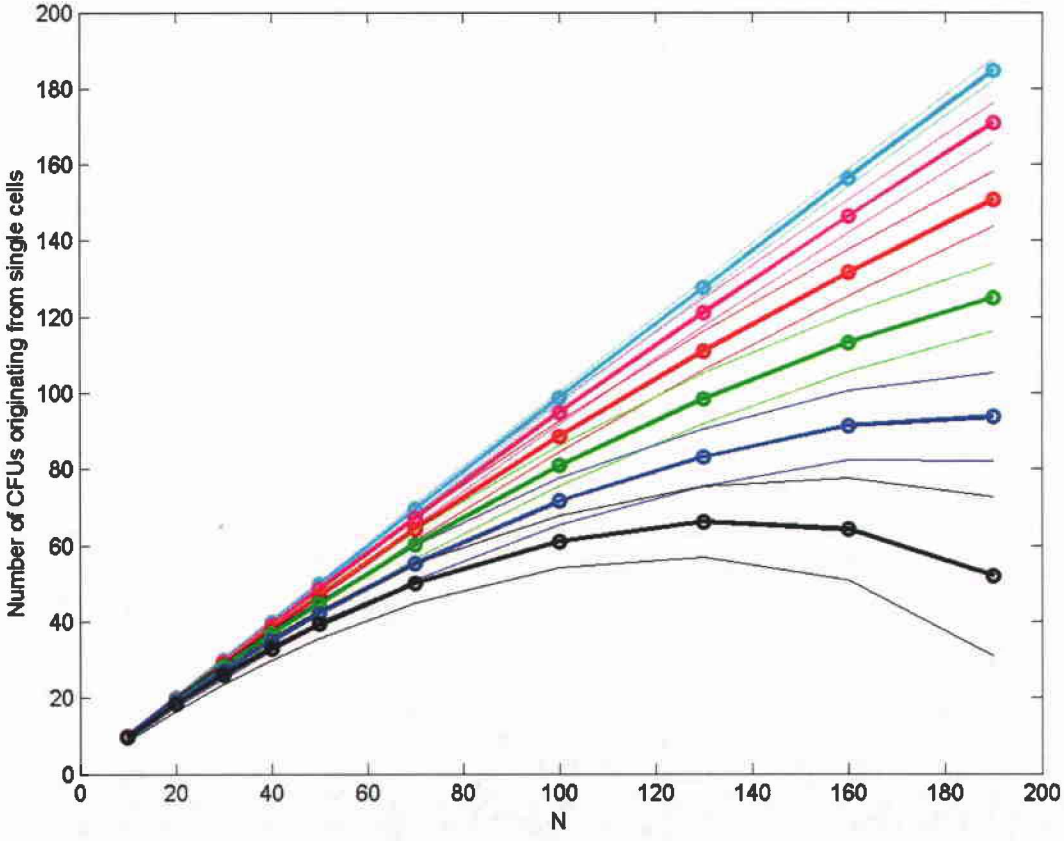


**Figure 3.4. Number of CFUs originating from single cells assuming different bacteria concentrations (N). The colors of the lines represent different R-values: cyan=0.25 mm, magenta=0.5 mm, red=0.75 mm, green=1.0 mm, blue=1.25 mm and black=1.5 mm. The thick lines illustrate the mean of the data and the thin lines are ± 1 standard deviation.**

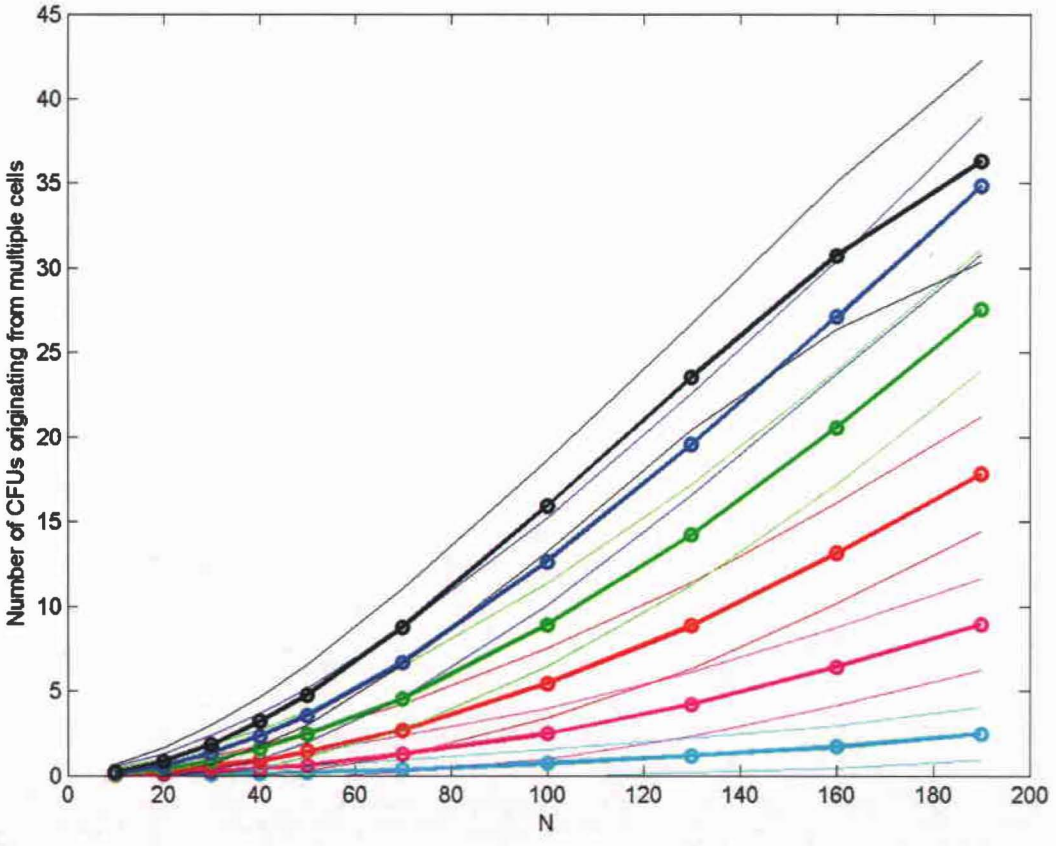


**Figure 3.5. Number of CFUs originating from cell clusters (more than one cell) as function of N and R. The colors of the lines represent different R-values: cyan=0.25 mm, magenta=0.5 mm, red=0.75 mm, green=1.0 mm, blue=1.25 mm and black=1.5 mm. The thick lines illustrate the mean of the data and the thin lines are ± 1 standard deviation.**

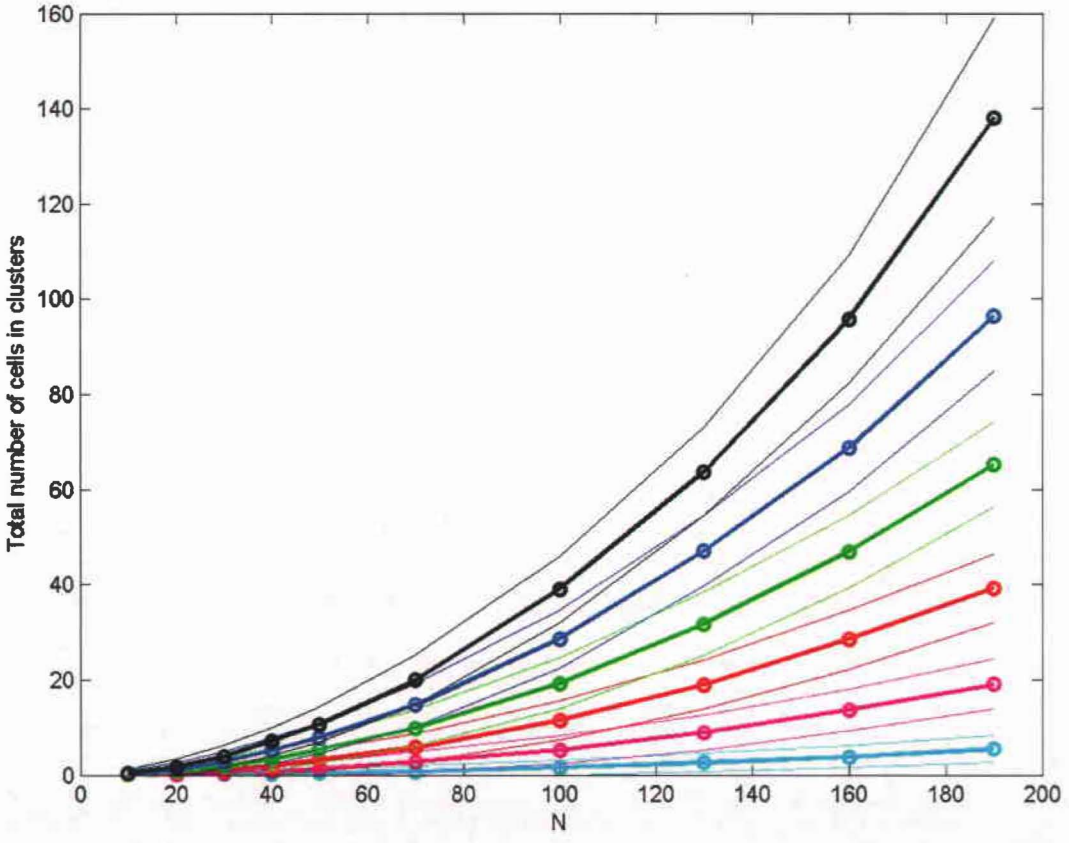


**Figure 3.6. Total number of cells in clusters for a realistic range of N and R conditions. The colors of the lines represent different R-values: cyan=0.25 mm, magenta=0.5 mm, red=0.75 mm, green=1.0 mm, blue=1.25 mm and black=1.5 mm. The thick lines illustrate the mean of the data and the thin lines are ± 1 standard deviation.**

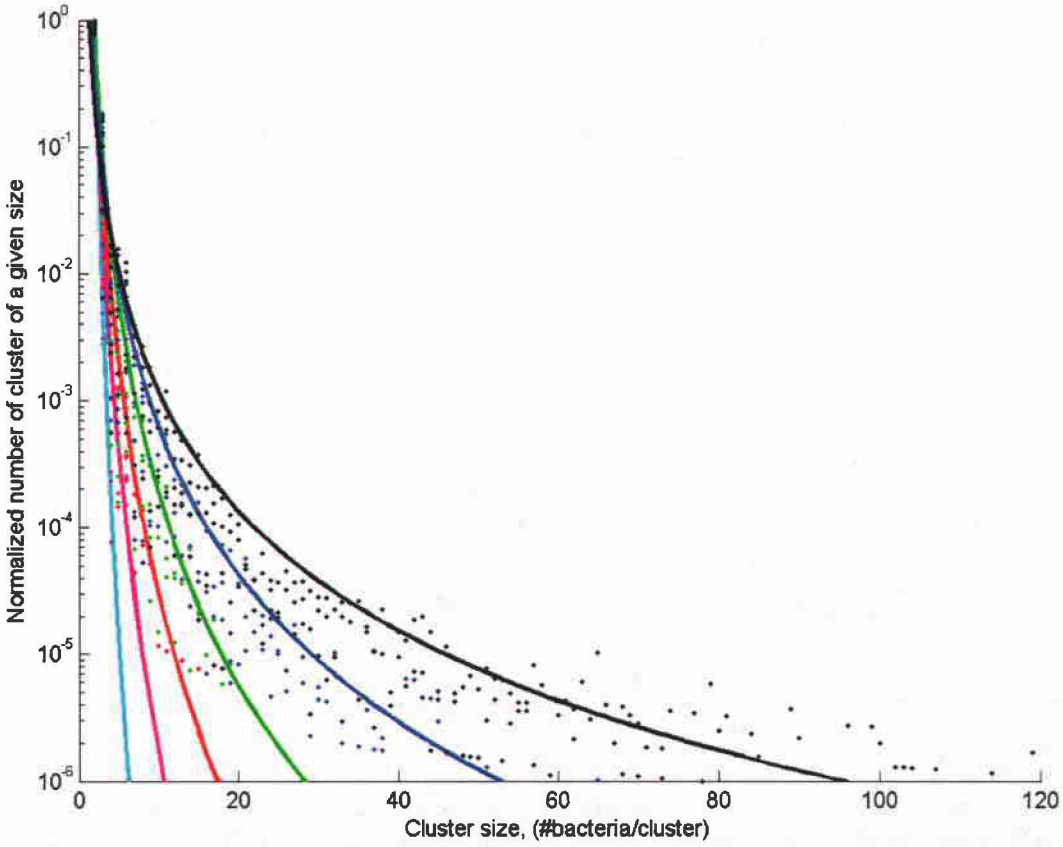


**Figure 3.7. Cluster size distribution: the number of clusters of a given size normalized to the total number of clusters plotted against cluster size, the number of bacteria per cluster. The data are represented by dots. The lines are the corresponding power law fits. The greater R, the more and larger the clusters formed. The colors of the points and lines represent different R-values: cyan=0.25 mm, magenta=0.5 mm, red=0.75 mm, green=1.0 mm, blue=1.25 mm and black=1.5 mm.**

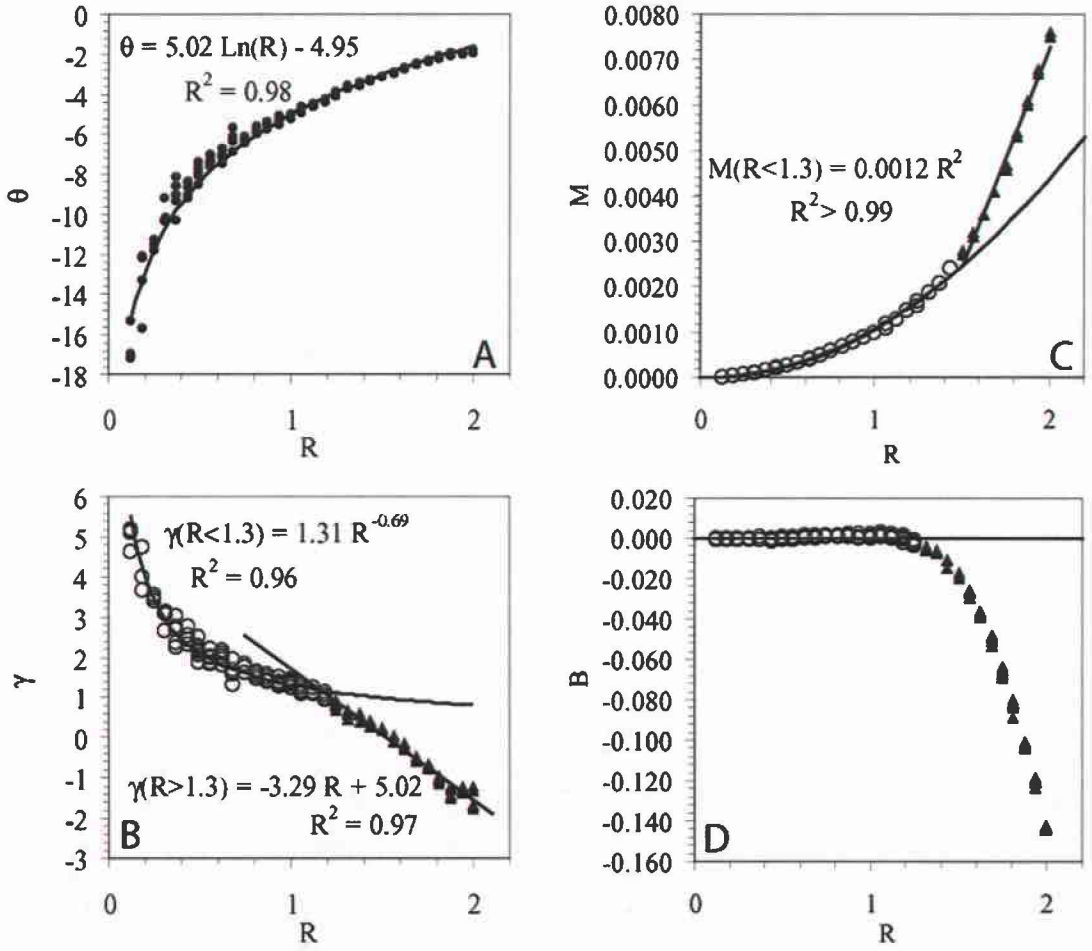


**Figure 3.8. Plots illustrating the variability of the constants for Eq. (3.4) and Eq. (3.5) as a function of R, the critical merging distance. A) θ (power) and B) γ (constant) correspond to the power law fits to the cluster size distributions in Eq. (3.4), and C) M (slope) and D) B (y-intercept) are linear fit constants to the relative error as a function of N in Eq. (3.5). The mode of aggregation under low R conditions (empty circles) is believed to be the aggregation of single cells into clusters, while with increasing values of R (dark triangles), the rate of formation of clusters of clusters is larger than the rate of formation of aggregates from the accretion of single cells. Black lines represent regression fits of the constants for all R (A) and under low and high R conditions (B & C). The horizontal black line in (D) represents B=0.**

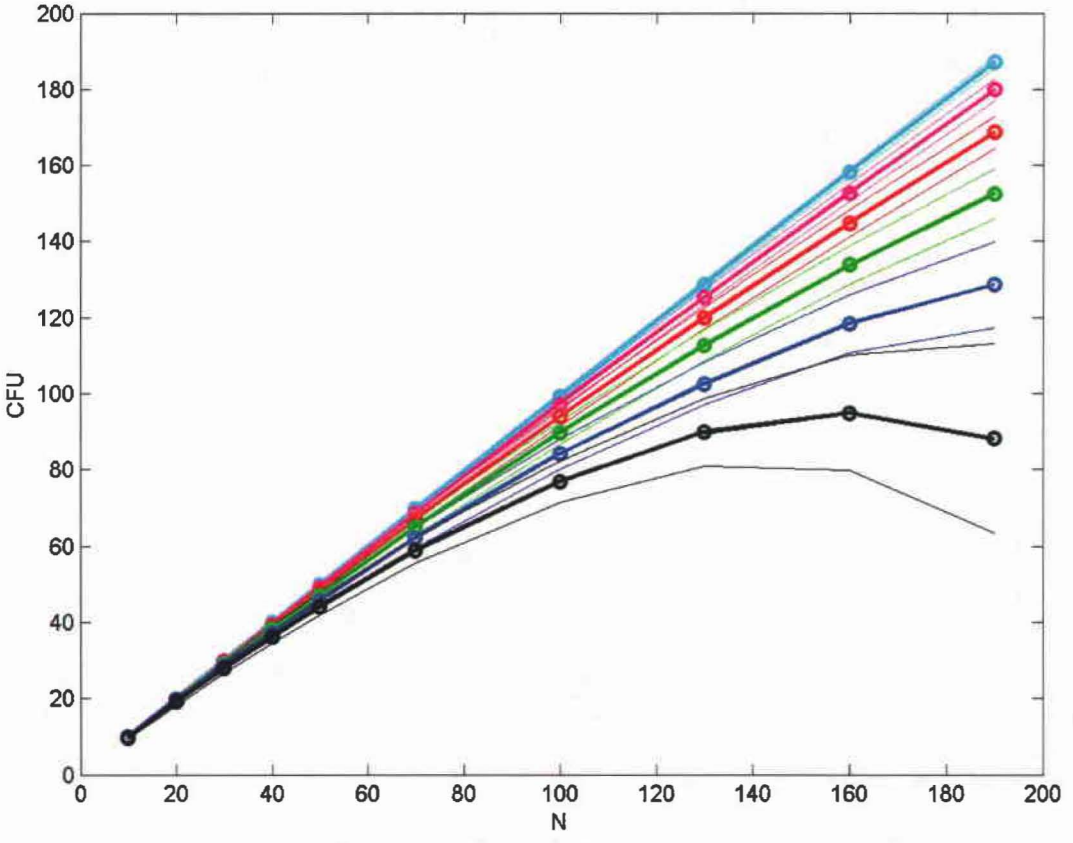


**Figure 3.9. Mean number of CFUs counted under different conditions of N and R. As expected, the larger the colonies, the greater the difference between the number of bacteria (N) and the number of CFUs measured. The colors of the lines represent different R-values: cyan=0.25 mm, magenta=0.5 mm, red=0.75 mm, green=1.0 mm, blue=1.25 mm and black=1.5 mm. The thick lines illustrate the mean of the data and the thin lines are ± 1 standard deviation.**

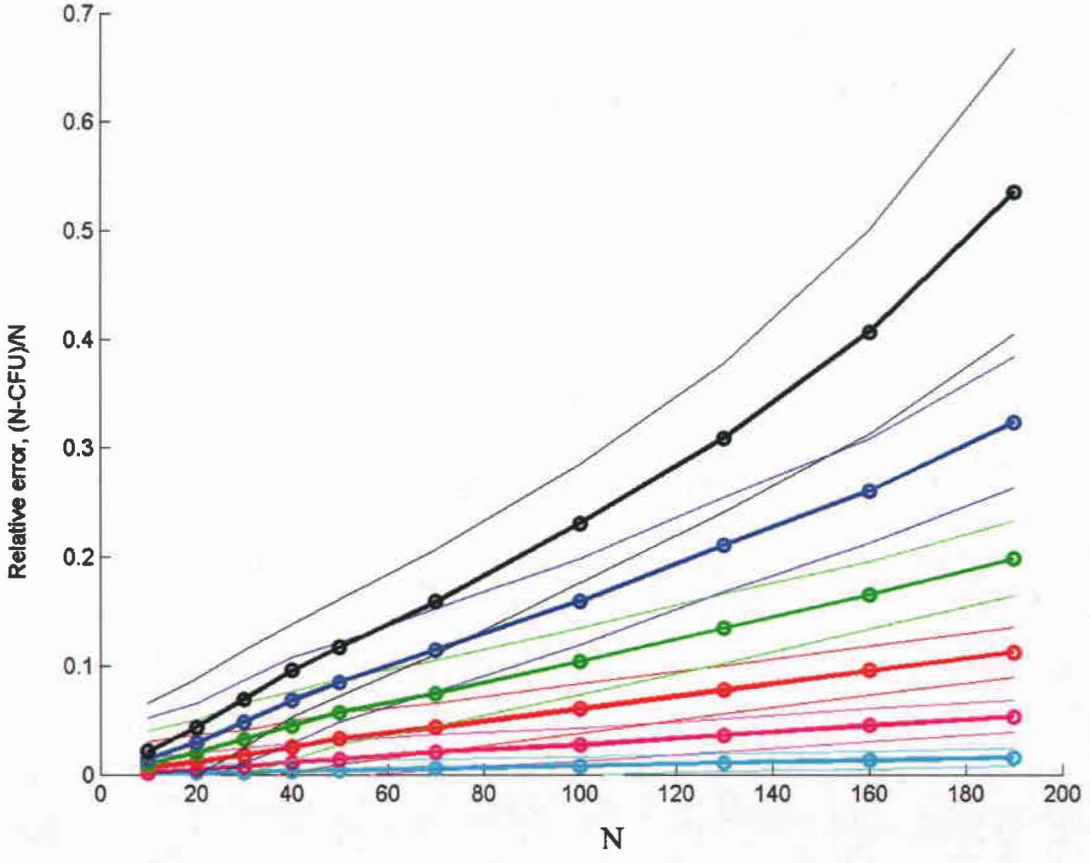


**Figure 3.10. The relative error ($\varepsilon_R$) between the number of CFU and the true number of bacteria (N) increases with increasing N and R. The error ranges from 0 to 25% when N is smaller than 100, but can reach 70% for larger N. The colors of the lines represent different R-values: cyan=0.25 mm, magenta=0.5 mm, red=0.75 mm, green=1.0 mm, blue=1.25 mm and black=1.5 mm. The thick lines illustrate the mean of the data and the thin lines are ± 1 standard deviation.**

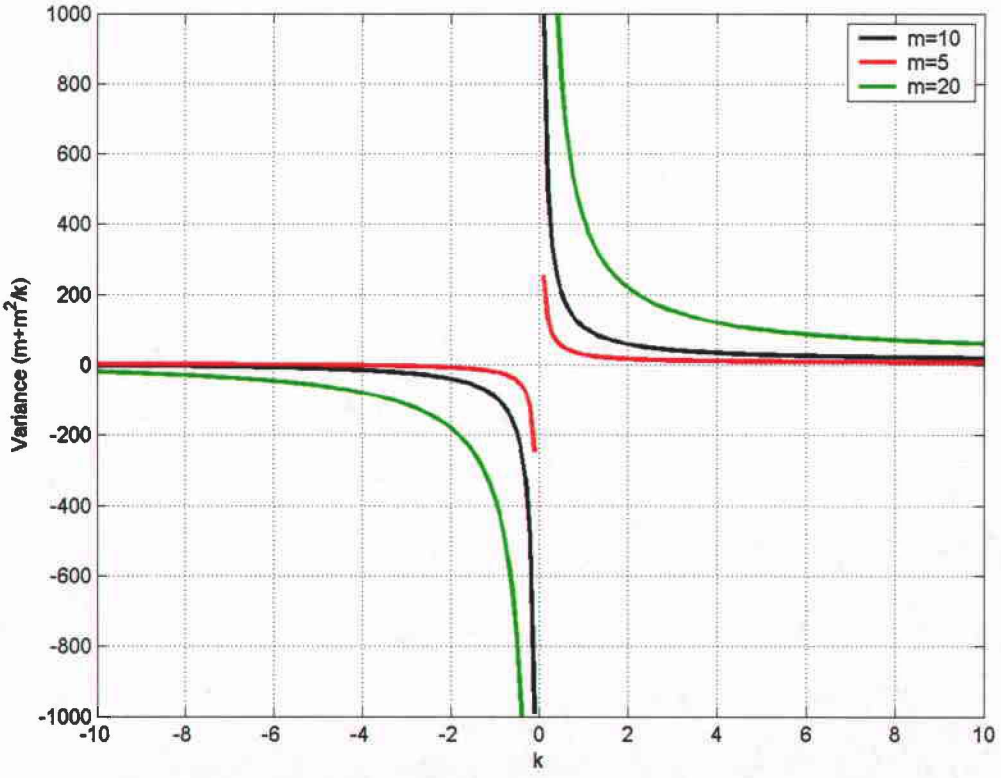


**Figure 3.11. Variance of $(Q\text{-}Pz)^{-k}$ as a function of k. Due to the asymptotic nature of the variance, the uncertainty in the data increases drastically for small k. The variance is larger for larger mean densities of clusters (m).**

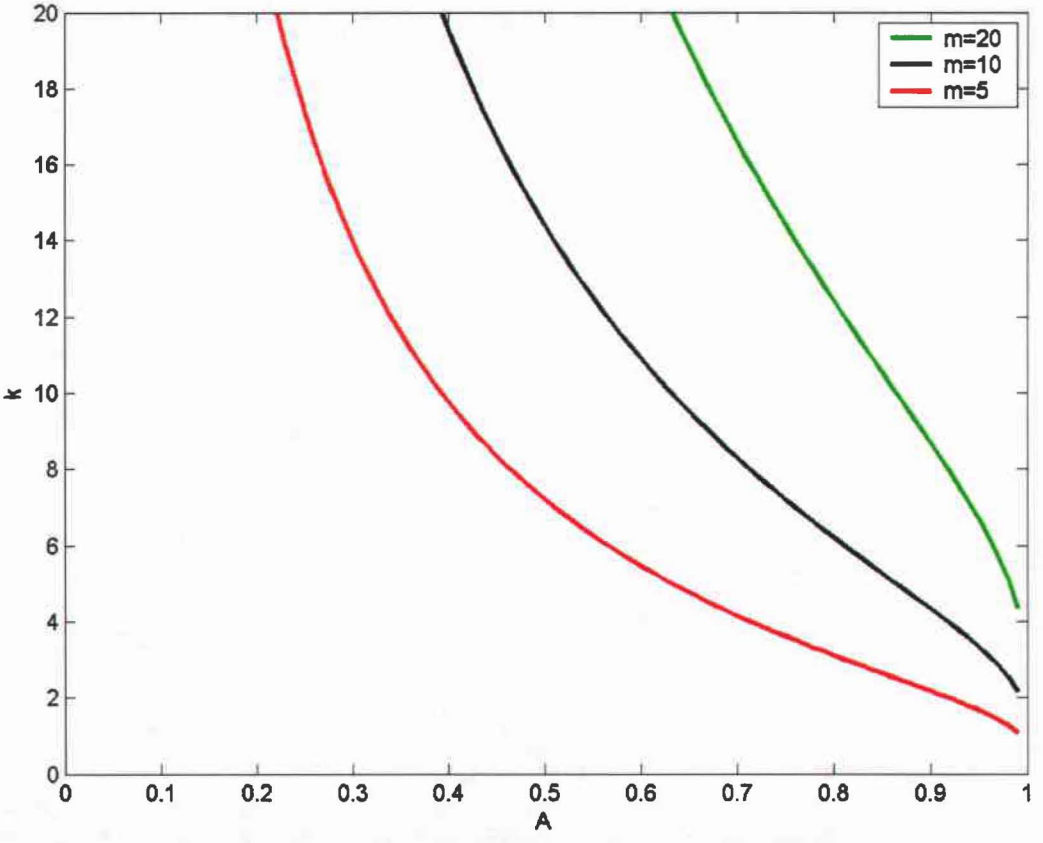


**Figure 3.12. Effect of A on k. The parameter A controls the cluster size distribution; that is the number of bacteria per cluster and the relative quantity of these clusters. Large A values mean that the chances are large for clusters with large number of bacteria associated with them to be present in the sample. When A is large, the variance will be large.**

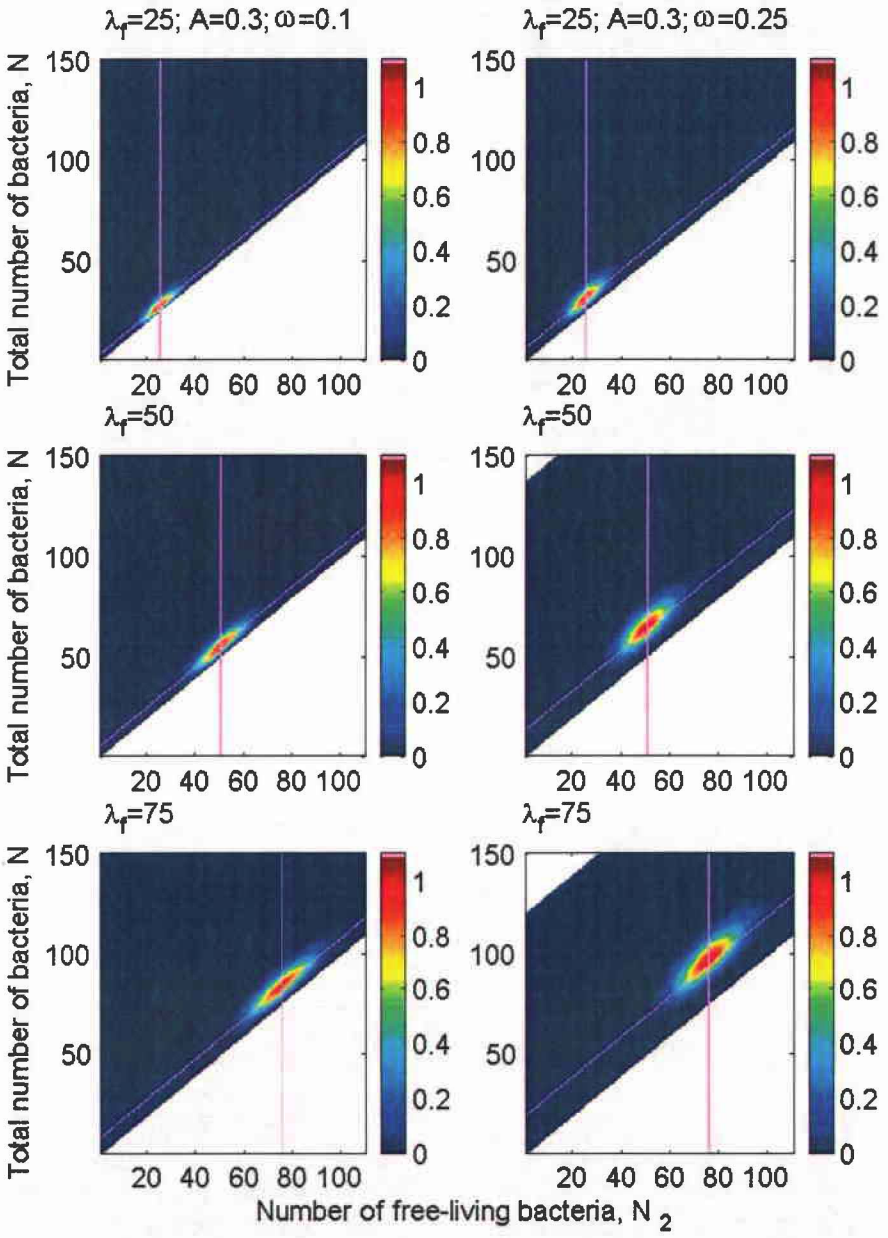


**Figure 3.13. Relative normalized probability ($P_{Tot}/P_{max}$) of finding N bacteria in a sample with A=0.3 and ω=0.1 (left panels), or ω=0.25 (right panels). The vertical pink line represents $\lambda_f$, whereas the diagonal pink line is $N_2+\lambda_c$. The intersection of these two lines corresponds to the expected number of CFUs that would form on an agar plate without considering clustering ($\lambda_f+\lambda_c$).**

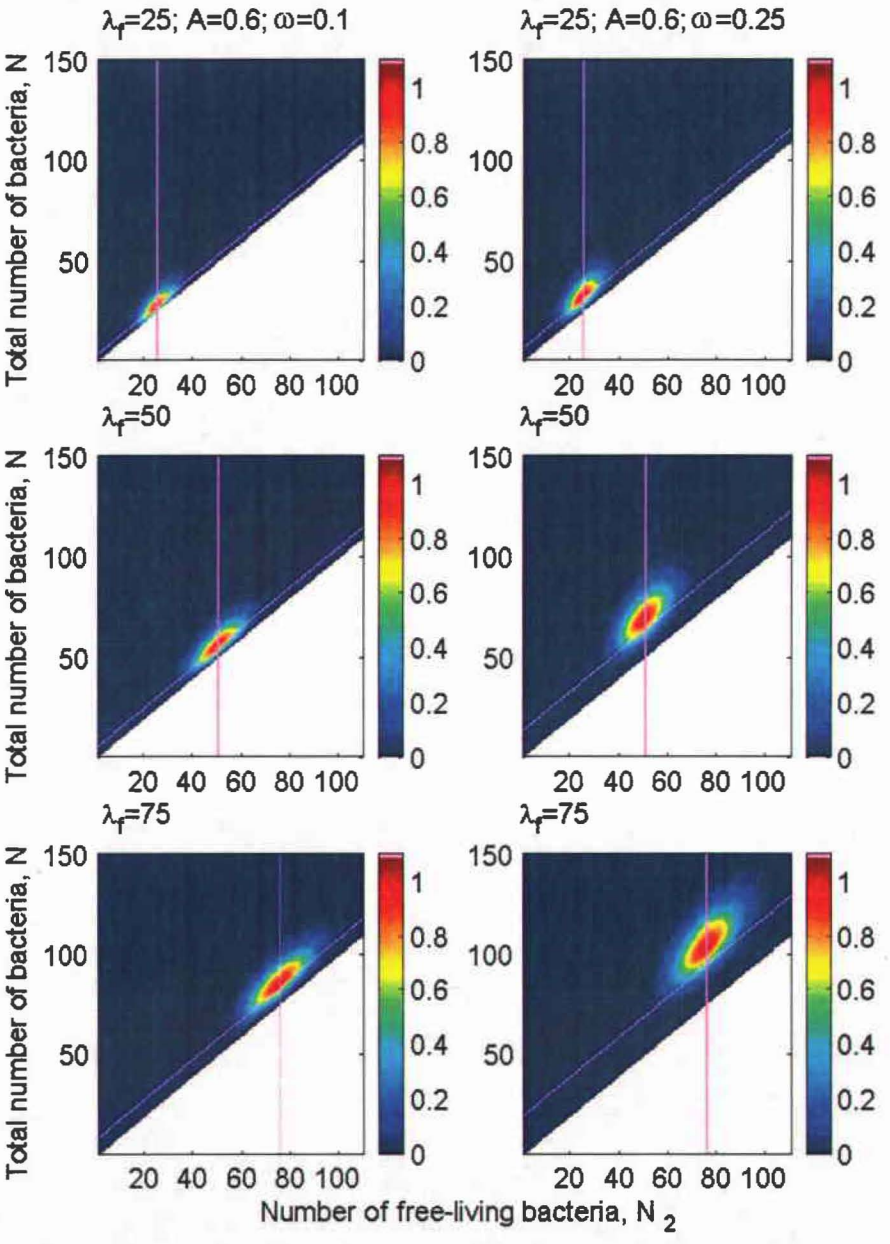


**Figure 3.14. Relative normalized probability ($P_{Tot}/P_{max}$) of finding N bacteria in a sample with A=0.6 and ω=0.1 (left panels), or ω=0.25 (right panels). The vertical pink line represents $\lambda_f$, whereas the diagonal pink line is $N_2+\lambda_c$. The intersection of these two lines corresponds to the expected number of CFUs that would form on an agar plate without considering clustering ($\lambda_f+\lambda_c$).**

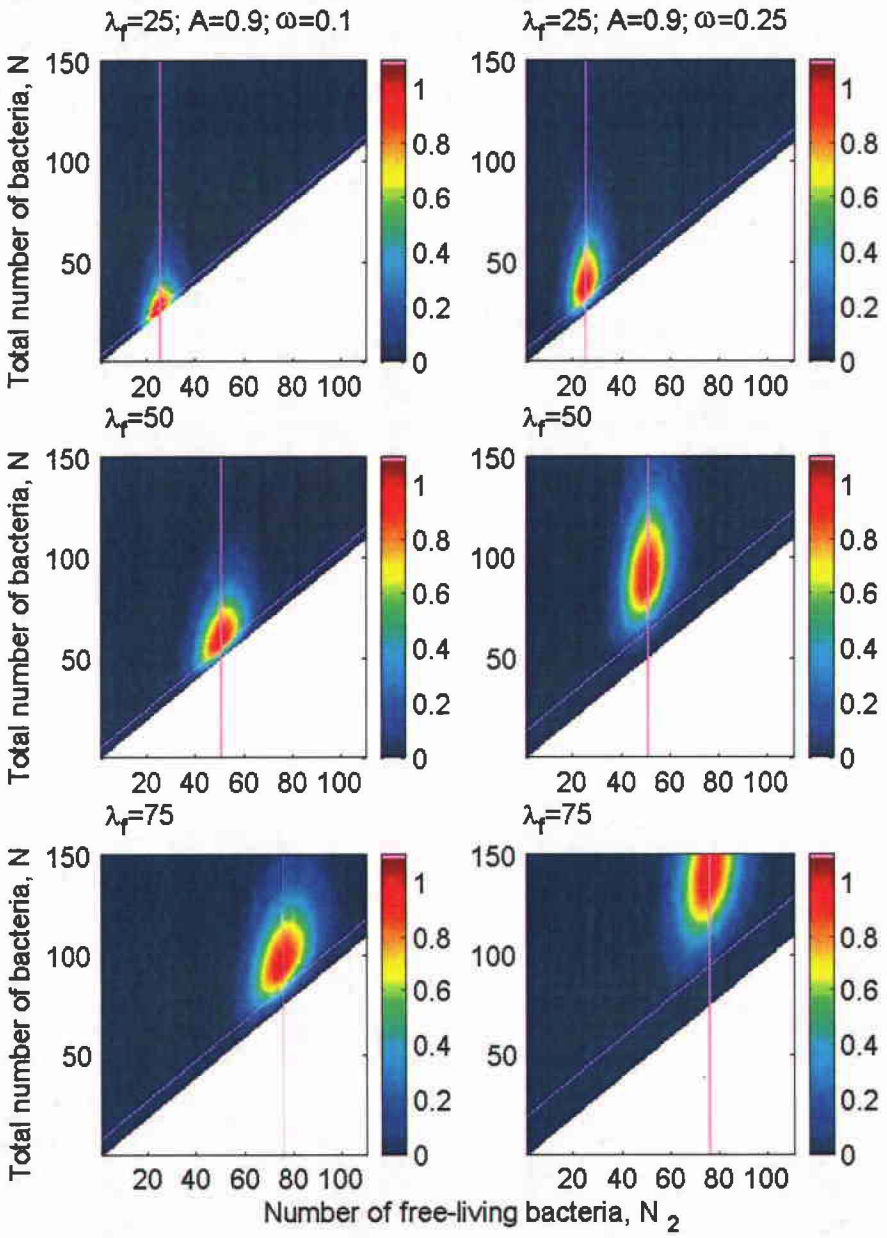


**Figure 3.15. Relative normalized probability ($P_{Tot}/P_{max}$) of finding N bacteria in a sample with A=0.9 and ω=0.1 (left panels), or ω=0.25 (right panels). The vertical pink line represents $\lambda_f$, whereas the diagonal pink line is $N_2+\lambda_c$. The intersection of these two lines corresponds to the expected number of CFUs that would form on an agar plate without considering clustering ($\lambda_f+\lambda_c$).**

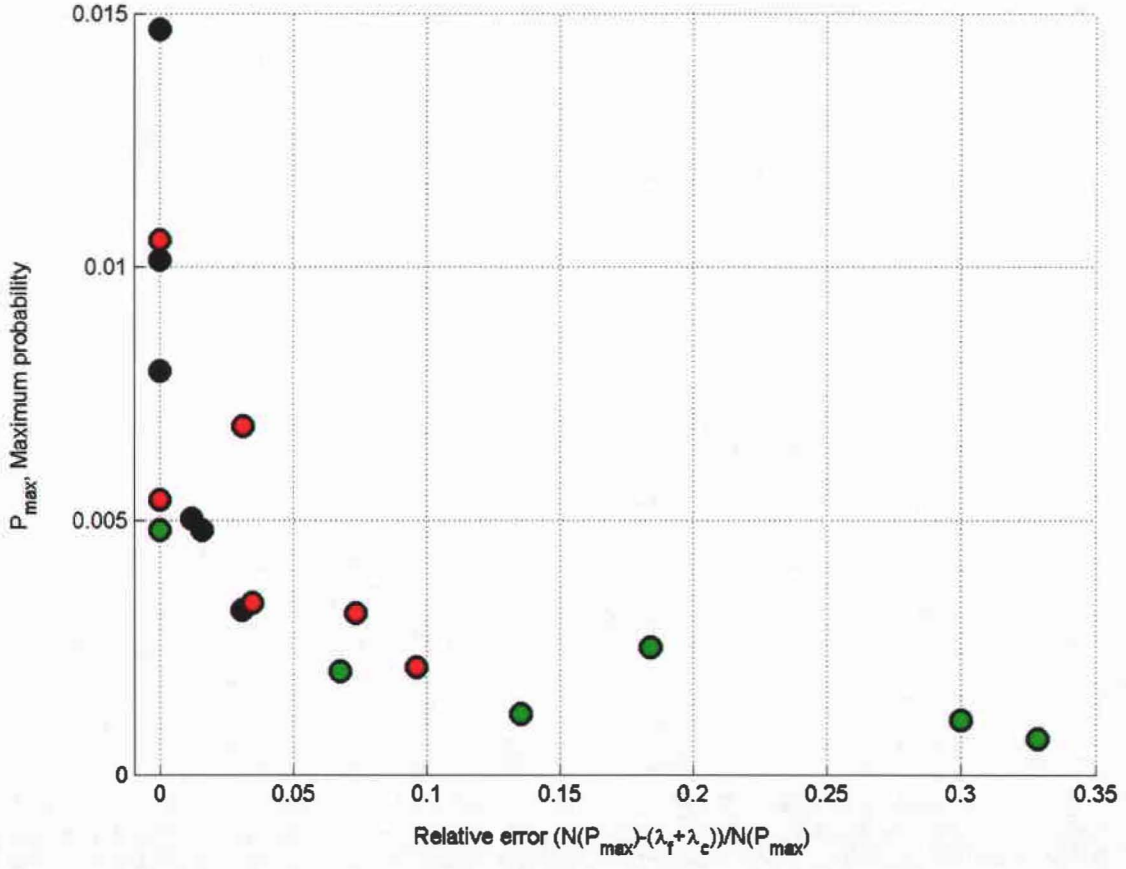


**Figure 3.16. Absolute magnitude of the maximum probability ($P_{max}$) peak plotted against the relative error. $P_{max}$ represent the normalization factors used in the previous plots. The color-coding corresponds to different values of the coefficients, A: A=0.3 (black), A=0.6 (red), A=0.9 (green). The multiple points for each A correspond to different conditions of $\lambda_f$ and $\omega$.**

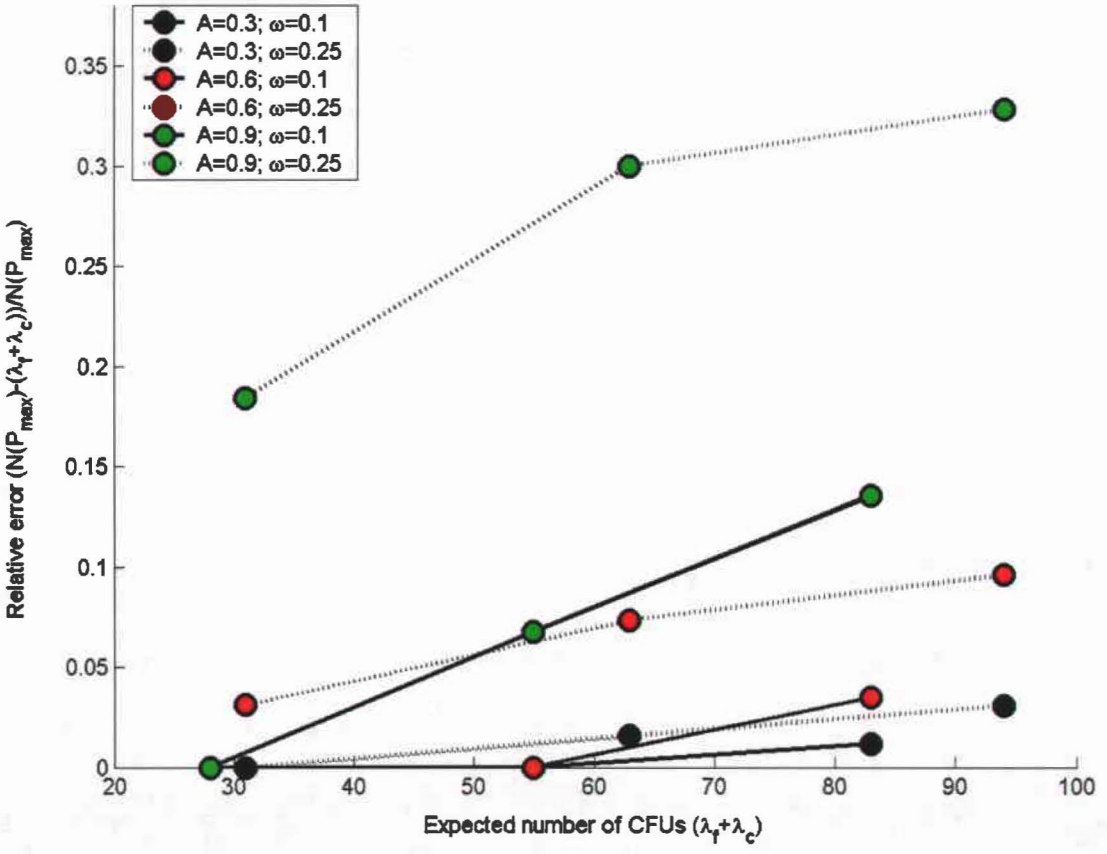


**Figure 3.17. Relative error plotted against the mean number of CFUs that would be expected to grow on an agar plate ($\lambda_f+\lambda_c$) provided a large number of replicates. The model assumes that the total number of clusters (regardless of cluster size) increases proportionally (by the factor $\omega$) to the number of free-living cells explaining the increasing trends with increasing expected CFUs. Larger A values produce larger error.**

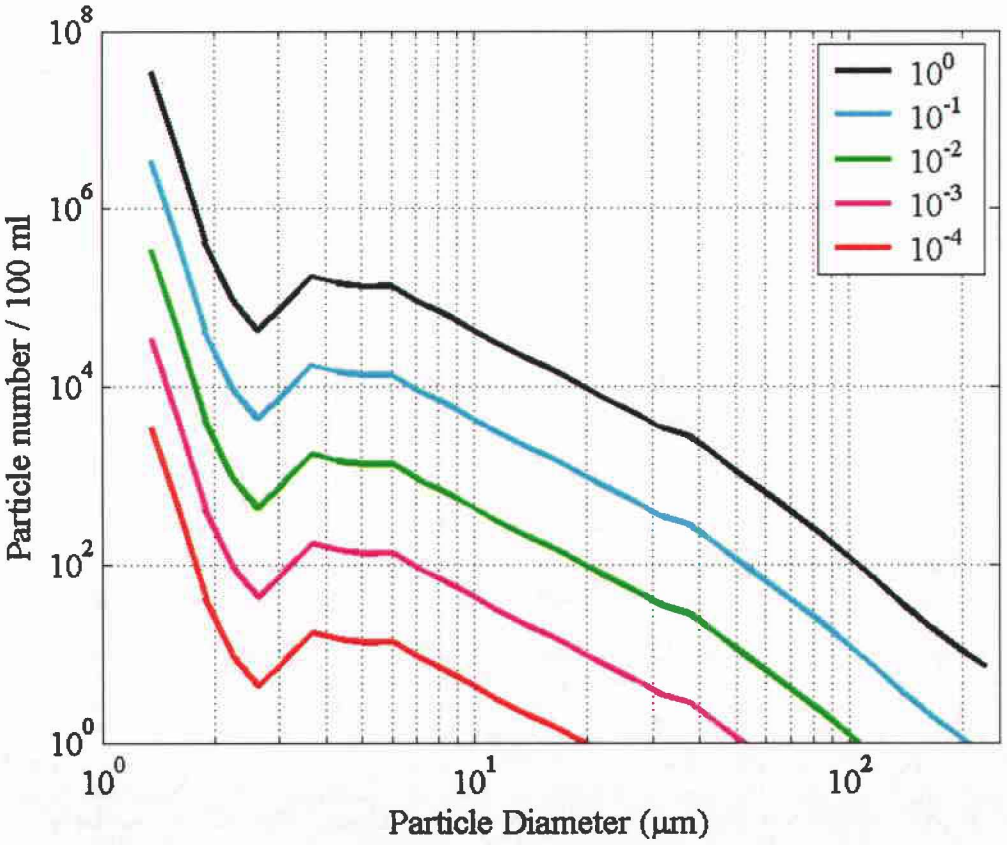


**Figure 3.18. Particle number size distribution measured during baseflow conditions at station LULU (Figure 2.1) with a LISST-100 instrument. Upon dilution, the concentration of the larger particles is greatly reduced, effectively putting a cap to the maximum particle size expected in a sample. At high dilutions, only small particles remain in the sample. (Data courtesy of Dr. Eric De Carlo)**

# Chapter 4 : MEASURING PARTICLE-BOUND FECAL INDICATOR BACTERIA

## *4.1 Introduction*

The standard culture-based methods (most-probable number, membrane filtration, spread plate) used to measure the levels of fecal indicator bacteria in water do not allow for the differentiation between free-living and aggregated microbes. These techniques ignore (or accept) the fact that a portion of the microbial community is associated with particles (Geesey & Costerton, 1979; Griffith et al., 1994). In the membrane filtration or the spread plate methods, aggregates each result in only a single colony-forming unit (CFU) regardless of how many positive bacteria are attached to them. This issue is a well-known limitation of the technique (Fleisher, 1990; Borst & Selvakumar, 2003), but the bias involved in the culture-based concentration measurements due to the presence of aggregates, and the potential for the increased contamination risk associated with these aggregates, remains poorly documented in recreational waters (Borst & Selvakumar, 2003).

In this study, measurements on three groups of microbes have been performed. Both enterococci (ENT) and *Clostridium. perfringens* (CP) are Gram-positive microbes, that is, they possess peptidoglycan-rich cell walls. *C. perfringens* is an obligate anaerobe and survives as spores in the aerobic environment, whereas some studies report that enterococci, which are facultative anaerobes, remain active and even grow in tropical freshwater environments (Muñiz et al., 1989; Fujioka & Byappanahalli, 2001; Hardina & Fujioka, 1991). Enterococci can produce extensive extracellular polymers, which act as "aggregating substances" (Vanek et al. 1999), suggesting that the bias due to particle-

association may be much larger for enterococci than for *C. perfringens*. Heterotrophic plate count (HPC) bacteria, however is not a well defined group as HPC bacteria simply are the most abundant representative of a mixed microbial population culturable aerobically on non-selective nutrient-rich agar medium.

### 4.1.1 Attempts to quantify bacteria associated with aggregates

There is no consensus in the literature on the best methodology to quantify bacteria associated with particles (Appendix F). The definitions of "free-living" and "particle-associated" vary greatly among studies; however, most definitions are operationally defined based on the pore sizes of the filters used. Table 4.1 summarizes the pore sizes of the filters used in the studies reviewed here (Appendix F). Pore sizes ranging from 3 to 10 μm are the most commonly used cutoff.

Of the 52 studies listed (Appendix F), 24 attempted to separate the particle-associated microorganisms from the aggregates. This was achieved, to varying degrees of success, by three main approaches: sonication (6 studies), homogenization (17 studies) or chemical treatment (19 studies). Most researchers used a combination of chemicals with sonication or homogenization. The chemicals used involved mostly detergents or surfactants (Zwittergent 3-12, Tween 20 and 80, pyrophosphate), but enzymes (galactosidases, lipases) and chelating agents (EGTA, EDTA) are also common.

Papers discussing the use of sonication or homogenization for live cells present conflicting results (Boeckelmann, Szewzyk & Grohmann, 2003; Ramsay, 1984; Velji & Albright, 1986; Yoon & Rosson, 1990, Camper et al, 1986). Overall, however, sonication appears to be the preferred tool for desorbing bacteria from aggregates when cells have

been fixed, whereas homogenization is preferred when the detection method relies on culturing. Sonication may interfere with cell viability (Ramsay, 1984, Boeckelmann, Szewzyk & Grohmann, 2003), whereas this does not seem to be the case for homogenization (Camper et al. 1985 a; Parker & Draby, 1995).

### 4.1.2 Goals

The overall purpose of this study is to quantify the abundance of different microbial indicators associated with aggregates. Specifically, the study will focus on heterotrophic plate count (HPC) bacteria and on the fecal indicator organisms enterococcus and *C. perfringens*. Specific goals include 1) to test a protocol that enumerates bacteria associated with aggregates using membrane filtration but does not affect cell viability, 2) to measure the level of association with aggregates of the indicators in Kaneohe Stream and 3) to compare the results of this field study with the data obtained by others (Appendix F), and 4) to assess the bias induced by the presence of multiple cells attached to aggregates on the concentration measurements by routine membrane filtration methods (see section 2.2.2).

## *4.2 Methodology*

### 4.2.1 Culturing

Enterococcus, *C. perfringens* and HPC bacteria were assayed by membrane filtration. The membrane filtration protocols are described in detail in the methodology section of Chapter 2. In addition, pure cultures of *Enterococcus faecalis* (ATCC# 29212, Kwik-Stik™, MicroBioLogics, Inc.) were prepared. The cells were streaked onto tryptic

soy (TS) agar and incubated at 35°C for 48 to 72 hours. A well-isolated colony was picked, transferred into autoclaved TS broth (10ml) and incubated on a shaker (200 rpm) at 35°C for 24 hours. The concentration of cells achieved after growth for 24 hours in the liquid medium was on the order of $10^{10\text{-}11}$ CFU/100ml.

### 4.2.2 Coulter counter

Natural stream particles smaller than 60 μm were isolated by filtering water from Kaneohe Stream station KANE sequentially through a 60 μm Nitex screen and a 5 μm Nuclepore membrane filter (diameter 47 mm) until clogging of the filters. The particles trapped on the 5 μm filter were resuspended with 0.2 μm filtered saline solution (0.9% NaCl). The Coulter counter measures variations in electrical conductivity through a fine jet of solution. When particles pass by the detector through the jet, the resistance increases. The magnitude of the resistance correlates with particle size. Consequently, it is important to use a diluent that conducts electricity; saline solution was used as the diluent throughout this experiment. The aperture of the Coulter counter used was 100 μm. Particles larger than 60 μm were removed to prevent clogging of the aperture. 30 ml aliquots of the particle solution were homogenized at 16,000 rpm, the maximum speed available (DuPont/Sorvall model#17150), for different periods (0, 15, 30, 60, 180 and 360 seconds). Particle concentrations in four different size classes were monitored: 1.5 to 5.5 μm, >5.5 μm, 4.3 μm to 15.6 μm and >15.6 μm.

### 4.2.3 Fluorescence *in situ* hybridization (FISH)

Volumes (1 to 10 ml) of stream water were filtered through white 0.2 μm membrane filters (25 mm diameter). The cells caught on the filters were fixed using a dehydration series. The filters were immediately covered with 3 ml of a 0.2 μm filtered 50% mixture of ethanol and phosphate buffer saline (PBS) solution (130 mM NaCl, 10 mM sodium phosphate buffer, pH 7.4) and incubated at room temperature for 30 minutes. The mixture was drained under vacuum and replaced with 3 ml of an 80% ethanol-PBS mixture for 3 minutes. Again, the solution was drained and the filter covered with ethanol (200 proof) for an additional 3 minutes, before being drained and air-dried.

The cell wall of Gram-positive bacteria (e.g. enterococci and *C. perfringens*) is quite impermeable to oligonucleotide probes relative to Gram negative bacteria (Beihmfor et al., 1993). Fixation with 4% paraformaldehyde was shown to be ineffective for Gram-positives (Roller et al., 1994; Jurtshuk et al., 1992), and instead, an ethanol dehydration series is preferred (Beihmfor et al., 1993; Roller et al., 1994).

The overall FISH procedure used here follows the protocol described in Glöckner et al. (1996) and Lam & Cowen (2004). The membrane filters were cut into four sections and each section was hybridized with a different set of probes. The sections were individually placed into a sterile prewarmed 24-well microtiter plate and covered with 20 μl of temperature equilibrated (46°C) hybridization solution containing 0.9 M NaCl, 20 mM Tris-HCl (pH 7.4), 0.01% sodium dodecyl sulfate (SDS) buffer, 2.5 ng of probe and the appropriate amount of formamide (Table 4.2). The samples were hybridized at 46°C for 2 hours by placing the microtiter plate onto paper towels soaked in hybridization solution lacking the probes within a temperature-equilibrated container with lid to

produce a saturated atmosphere to prevent evaporation of the hybridization solution from the filter. Following hybridization, each filter section was transferred into temperature-equilibrated (48°C) autoclaved 20 ml glass vials filled with the probe-specific washing solution and incubated in a water bath incubator for 15 minutes without agitation. The washing solutions contain the appropriate concentration of NaCl (Table 4.2), plus 20 mM Tris-HCl, 5 mM EDTA and 0.01% SDS (sodium-dodecyl-sulfate). After 15 minutes, the washing solution was drained, replaced with fresh prewarmed washing solution and placed in the water bath for another 15 minutes. After the second washing step, the filters were air dried on KimWipes™ in the dark and either mounted onto microscope slides with FluoroGuard reagent (Bio-Rad Laboratories, Inc.) or returned to the hybridization chamber for another hybridization reaction with a different probe. The slides were stored at –20°C until analyzed under an Eclipse E400 epifluorescence microscope at a magnification of 1000x.

The specificities of the probes used in this study are summarized in Table 4.2. The probe Cp2 had been used for dot blot hybridization, but not for whole cell hybridization, such that no empirical information about the specific hybridization and washing conditions was available. Formamide is used in the hybridization solution to maintain a reasonably high hybridization reaction rate at a practical temperature, which is lower than the melting temperature of the probe. When the probe-specific concentration of formamide to use has not been experimentally optimized, it is possible to approximate the formamide concentration required in the hybridization solution with the following equation for DNA-RNA hybrids (Wahl et al. 1987, Hames & Higgins, 1995):

$$T_m = 79.8 + 18.5 \cdot \log M + 58.4(molefractionGC) + 11.8(molefractionGC)^2 - \frac{820}{L} - 0.5(\%\, formamide) \quad (4.1)$$

where M represents the monovalent cation concentration (0.9 M NaCl in the present case), and L is the length of the probe in base pairs. Since the goal is to set a hybridization temperature that is constant for all probes (46°C), this value can be used instead of $T_m$ (the melting temperature) and the *% formamide* (%FA) can be adjusted accordingly. The GC mole fraction of the Cp2 probe is 0.45, which yields a formamide concentration of 41%.

Probe Cp2 was shown to be species specific (Roenner & Stackebrandt 1994), and it should consequently be 100% complementary to its exclusive target, *C. perfringens*. Under these circumstances, the use of a low NaCl concentration in the washing solution is recommended (Whal et al., 1987; Hames & Higgins, 1995). Meier et al. (1997, 1998) suggested the use of 80 mM NaCl for the probes Enc131 and DB8. Since both of these probes are also highly specific (Table 4.2), 80 mM was used for the probe Cp2 as well.

### 4.2.4 Particle settling experiments

Water samples were collected from stations LULU and KANE (Figure 2.1) in autoclaved 1L polycarbonate bottles. Once in the laboratory, three 1L samples from a given station were combined into a 3L autoclaved glass Erlenmeyer flask and hand-shaken vigorously for 10 seconds. Two hundred fifty ml of the batch was directly assayed by membrane filtration for HPC, ENT, and CP without additional treatment, while a parallel 250 ml aliquot was treated with the proposed desorption protocol (homogenization and Camper solution, see section 4.3.2.3). The untreated and treated

unsettled subsamples will be referred to as the CONTROL and the WHOLE samples, respectively. In addition, three 500 ml and three 250 ml autoclaved separation funnels were filled with sample and allowed to settle for 4 hours at room temperature. A similar settling time was used by Schillinger & Gannon (1985). After 4 hours, the top 83 ml of each funnel of equivalent volume (250 or 500 ml) was pipetted off and combined in an autoclaved polypropylene bottle, yielding two subsamples constituting the fraction of the samples gravity-stripped of the largest particles (Figure 4.1). In a similar manner, the bottom 83 ml of each funnel were drawn using the valve located at the bottom of the funnels, producing two samples corresponding to fractions enriched with the larger settled particles (Figure 4.1). An additional 83 ml were drained from the bottom valve and discarded. Finally, 83 ml were drawn from the valve from each 500 ml funnel. This last subsample essentially originates from the middle of the funnel (Figure 4.1). Two and one-half ml of autoclaved concentrated (100 x) Camper desorption solution were added to each subsample. The subsamples were incubated for 30 minutes on ice in a cooler before homogenization. The concentration of microorganisms in the treated samples was measured using the routine membrane filtration technique (see methodology section in Chapter 2). Homogenization (16,000 rpm) of the subsamples was performed as 5 x 50 ml aliquots (practical volume of the homogenization chamber). The shaft and the homogenization chamber were rinsed with 70% ethanol, flamed on a Bunsen burner and exposed to UV for at least 10 minutes between each subsample.

The subsamples resulting from the pooled top 83 ml layers from either type of funnel (250 ml or 500 ml) will be referred to as the $TOP_{250}$ or the $TOP_{500}$ samples. Similarly, the pooled bottom 83 ml subsamples are defined as the $BOTTOM_{250}$ and

$BOTTOM_{500}$ samples. The subsamples collected from the center of the funnels, after collection of the BOTTOM samples and after discarding the next 83 ml layer, are exclusive to the 500 ml funnels and will be referred to as the $MIDDLE_{500}$ samples.

### 4.2.5 Particle filtration experiments

Two 250 ml aliquots were taken in parallel from a 1L polycarbonate sample bottle. One aliquot was assayed by the traditional membrane filtration methods to produce $C_R$, the CFU concentration measured with no additional treatment (Figure 4.2). The second aliquot was first passed through a 5 μm Nuclepore membrane filter (47mm in diameter). Every filter was UV-sterilized, and the glass filter tower was rinsed with ethanol, flamed on a Bunsen burner and exposed to UV between each filtration. The filtrate (<5 μm) was assayed similarly to the control ($C_R$) aliquot for HPC, ENT and CP, yielding CFU counts originating from "free-living" cells, $C_F$, operationally defined by the filter pore size (Figure 4.2). Blank samples consisting of 50 ml of autoclaved phosphate buffer solution run occasionally between samples indicated that these sterilization efforts were satisfactory since growth was never observed on any of the type of agar used.

### 4.2.6 The number of positive colonies per aggregate

In order to measure the number of positive colonies originating from aggregates containing indicator bacteria, two measurements are required: the number of aggregates with associated bacteria ($C_A$) and the number of microbes associated with these aggregates ($C_H$); see Figure 4.2 for an illustration. The particles were separated from the bulk sample by filtration through a 5 μm Nuclepore membrane.

Immediately after filtration, the filter was transferred into a sterile Petri dish whose edge had been heated on a Bunsen burner and shaped into a spout to funnel liquid out of the dish. A few drops of Camper solution were placed in the middle of the filter using a sterile syringe equipped with a 0.2 μm filter and a needle. Using flamed forceps to hold the filter in place in the dish, the particles were gently scraped off the filter with a rubber spatula (Fisher Scientific, Cat. # 14-105A). Particles remaining on the rubber spatula were rinsed off using the syringe assembly with Camper solution and the rinsate was collected. The particles retained on the filter were washed off by holding the Petri dish vertically with the spout at the bottom above the receiving container and squirting 10 ml of Camper solution using the syringe assembly.

To measure the number of colony-producing aggregates ($C_A$), the collected particles were resuspended in phosphate buffer to achieve a 4 x concentration factor relative to the original volume filtered and assayed by membrane filtration for HPC, ENT and CP. The use of a concentration factor was necessary to achieve counts between 20 and 80 CFU/100ml (APHA, 1998), especially for CP and ENT, which are both present in smaller concentrations than HPC. In order to determine the number of CFUs associated with >5 μm particles ($C_H$) the particles scraped off the filters were incubated for 30 minutes in 1 x Camper solution in a cooler, subjected to homogenization (16000 rpm) for 3 minutes on ice as 5 x 50 ml, resuspended in phosphate buffer to achieve a 4 x concentration factor (as for $C_A$) and assayed by the routine membrane filtration methods.

## *4.3 Results and discussion*

### 4.3.1 Sonication

The use of sonication to desorb bacteria for the assessment of aggregated microbes using the membrane filtration method was investigated by subjecting stream samples to different sonication times. Samples were collected on 4 occasions from Manoa Stream on the University of Hawaii at Manoa Campus (leeward Oahu) between January and February 2002. The sampling bottles (autoclaved 1L polycarbonate bottles) were hand-shaken vigorously for 15 seconds before 100 ml aliquots were transferred into autoclaved glass 250 ml Erlenmeyer flasks. The flasks were placed in a 130 W ultrasonic bath (Branson Ultrasonics Corp.) with a water depth of 3 cm. The sonication time varied from 0 to 60 seconds.

The results from this experiment show that sonication does not consistently increase the counts of either HPC bacteria or enterococci for sonication times between 0 and 60 seconds (Figure 4.3). The data have been normalized to the maximum value obtained for each date (normalized index). Consequently, the maximum value for each experiment is 1. The normalized index values for HPC range from 0.45 to 1, but no trend is observed with respect to sonication time (Figure 4.3, A). The index values for enterococcus range from 0.73 to 1 (Figure 4.3, B). No trend is observed relating the index values and sonication time. The sonication time corresponding to the maximum index value (1) varies for each date sampled. The sonication times matching the maxima for HPC differ from those of enterococcus, except for the January 14$^{th}$, 2002 sample, where the maxima for both organisms occurred with the same sonication time (5 seconds).

These results can be interpreted in different ways: a) the number of particle-associated bacteria was too small for a signal to be detected, b) sonication, under the conditions of this experiment, is ineffective at separating bacteria from particles or c) the negative effects of sonication (possible reduction of cell viability) is balanced by the number of cells that become separated from particles, such that the index remains more-or-less constant within the sonication times tested.

Since sonication did not produce an increase of the counts for either HPC bacteria or enterococcus, the next step was to assess the effect of sonication on cell viability for a pure culture of *E. faecalis.* A pure culture of *E. faecalis* was prepared as described in the methodology section above and diluted 1000 fold with autoclaved phosphate buffer. Fifty ml aliquots of the diluted culture were distributed into 100 ml autoclaved glass beakers and subjected to different sonication times ranging from 0 to 360 seconds. After treatment, the sonicate was processed normally through the dilution series and assayed by the traditional membrane filtration method on mE agar (see Chapter 2 for details). Three sets of experiments with a range of sonication times (0, 15, 30, 60, 180 and 360 seconds), were performed sequentially from the same dilution batch. The concentration of *E. faecalis* decreased with increasing sonication time in all three sets (Figure 4.4). Most of the decrease occurred during the first 60 seconds of sonication.

Cultured cells behave very differently than cells in the environment (Costerton, Irvin & Cheng, 1981; Costerton et al., 1987), and it is possible that the effects of sonication could differ for environmental cells. If environmental cells are more resistant to stresses than cultures, then the results presented in Figure 4.4 could over-dramatize the effect that sonication has on environmental cells. However, the present experiment

indicates that sonication showed either no effect (Figure 4.3) or a negative effect (Figure 4.4) on HPC bacteria or enterococci.

### 4.3.2 Homogenization

#### 4.3.2.1 Viability experiment

The effect of homogenization was tested on a pure culture of *E. faecalis*. A 1000-times diluted batch of the pure culture was prepared as described above. Fifty ml aliquots of the batch were transferred into the stainless steel homogenization chamber provided with the homogenizer, and the solution was subjected to homogenization at 16,000 rpm (Dupont/Sorvall model17150) for various periods (0, 15, 30, 60, 180 and 360 seconds) on ice. The homogenization chamber, the homogenizer shaft and blades were sterilized with 70% ethanol, flamed on a Bunsen burner and exposed to a germicidal UV lamp between each homogenization. Three runs were performed sequentially from the same diluted batch of culture. The concentration of *E. faecalis* did not vary with homogenization time (Figure 4.5), indicating that homogenization does not interfere with the viability of the cells. However, this experiment provides no indication of whether homogenization is effective at separating particle-associated bacteria.

Since cultures of *E. faecalis* are rich in large cell clusters (Figure 4.6, A & B) compared to other species (Figure 4.6, C & F), an increase of CFU would be intuitively expected in the homogenization experiment if homogenization were effective at breaking apart cells from aggregates. The homogenization procedures were performed on a 1000 x dilution of the original culture (Figure 4.6, D and E). Comparison of panels A and B, which represent a 100 x dilution, with panels D and E, corresponding to a 1000 x

dilution, indicates that the process of dilution (in this example 10 x) is an effective filter that discriminates against large aggregates (see also Figure 3.18). Dilution is essentially a statistical filter; it removes specimens that are rare. In this case, dilution removes the larger aggregates such that, relatively speaking, the diluted sample becomes relatively enriched in smaller aggregates and free-living cells. Consequently, the homogenization and sonication procedures performed on the diluted cultures were performed on samples depleted in large aggregates compared to the original culture. It is possible that the concentration of cell clusters was low enough in the diluted batch culture such that treatment did not produce a detectable increase down the dilution series: the optimal dilution for the membrane filtration was $10^{-8}$, which is five decimal dilutions beyond the homogenized dilution. Consequently, the absence of an increase in colony counts after either sonication or homogenization in the pure culture experiments does not necessarily mean that the treatment was ineffective. However, observing a decrease does mean that the treatment had a negative effect on the survival potential of the cells towards the treatment in question. Consequently, the experimental results suggest that sonication is not appropriate. Homogenization may be appropriate but requires further verification of its effectiveness at breaking apart bacteria from aggregates.

### 4.3.2.2 Fragmentation experiment

In order to verify the impact of homogenization on stream water particles, the number of stream particles in different size classes was measured with a Coulter counter, after varying homogenization time at constant speed (Figure 4.7). The data were normalized to the initial particle concentration of each size class ($C_0$) to yield numbers

between 0 and 1 and facilitate the comparison of the results obtained for the different size classes. The largest portion of the decrease in concentration occurs during the first 60 seconds of homogenization for all the size classes monitored. The largest effect of the homogenization procedure is on the largest particles. The number of particles greater than 15.6 μm decreased by about 75% after 360 seconds of homogenization, while the number of particles in the range 1.5 to 5.5 μm was reduced by 20% only. The number of particles between 4.3 and 15.6 μm diminished roughly 50%, a value similar to the one observed for the bin comprised of all particles greater than 5.5 μm. These results indicate that bacterial cells, which are small (~1 μm), are little affected by the homogenization process. This statement is confirmed by the previous experiment, which showed that the viability of *E. faecalis* cells was not impaired by homogenization. The larger aggregates, however, are effectively fragmented by the treatment.

The fact that a decrease in particle counts was observed for the smallest size bin (1.5-5.5 μm) upon homogenization suggests that a majority of the fragments produced by homogenization are smaller than 1.5 μm. Because of the greater number of the small particles relative to the large ones (Bader 1970), great numbers of small fragments must be produced from the less abundant large particles for an increase of the small particles to be detectable. For example, if homogenization produces $10^5$ two-micron fragments but the number of two-micron particles naturally occurring in the sample is already $10^7$, the number of two-micron particles will increase by only 1%. Thus, the lack of a measurable increase of small particles does not necessarily indicate that the fragments produced are

smaller than 1.5 μm, but a decrease of the 1.5-5.5 μm particles is evidence that homogenization effectively produces fragments smaller than 1.5 μm.

The data presented here contrast with the results of Borst & Selvakumar (2003), who did not see a significant difference between the mean particle size of storm runoff samples that were or were not homogenized (22,000 rpm). The mean particle size is not a practical descriptor of particle size distributions. Particle size distributions in the aquatic environment can generally be well approximated by power laws with negative exponents (Bader, 1970). In other words, the number of small particles is always greater than the number of large particles. Under such conditions, the mean particle size is quite insensitive to the large end of the distribution and is mostly controlled by the number of smaller particles. Homogenization modifies particle size distributions in two ways. It breaks apart larger particles, reducing the number of the larger particles. The greater effect of homogenization on the larger particles practically cuts off the particle size distributions at a maximum size. The second effect derives from the combined effect of the disappearing larger particles and the fact that the fragments produced by homogenization increase the number of particles in the smaller size classes, which decreases the power of the particle size distributions; the exponent becomes more negative (Figure 4.8). The measurement of particle size distribution of aquatic particles is prone to large uncertainties (Agrawal & Pottsmith, 2000). Therefore, it is extremely difficult to measure particle size distributions accurately enough to detect other than very large changes in mean particle sizes before and after homogenization (Figure 4.8). In fact, differences in mean particle sizes large enough to be detected only occur when the power of the particle size distribution is high (less negative), corresponding to situations of

elevated larger particles concentration. The difference between the qualitative nature of particles explains why Borst & Selvakumar (2003), who studied storm runoff particles, did not see a significant change in mean particle size, while Perdek & Borst (2000) observed a decrease in the the mean particle size of combined sewer overflow (CSO) samples (3-8 μm to 2-3 μm).

### 4.3.2.3 The use of chemicals

Camper et al. (1985 a) developed a protocol to desorb bacteria from granular activated carbon (GAC) particles. The authors concluded that homogenization of the GAC particles in a solution of Tris buffer (0.01 M, pH 7.0), Zwittergent 3-12 ($10^{-6}$ M), ethyleneglycol-bis-(β amino-ethyl ether)-N,$N^1$-tetra acetic acid (EGTA, $10^{-3}$ M) and peptone (0.01%) produced the highest removal efficiency of heterotrophic plate count organisms measured by the spread plate method. This desorption solution (Camper solution), whose efficiency with GAC particles was estimated to be 80-90% from synthetically produced aggregates (Camper et al. 1985 a), has been used in a number of studies investigating the association of bacteria with particles in different environments: GAC particles (Camper et al. 1985 a & b, 1986, 1987; LeChevallier et al., 1984; Pernitsky, Finch & Huck, 1997, Stewart, Wolfe & Means, 1990; Stringfellow, Mallon & DiGiano, 1993), secondary effluent waste water (Parker & Darby, 1995), CSO (Perdek & Borst, 2000) and storm water runoff (Borst & Selvakumar, 2003).

The effect of exposure to Camper solution on the viability of a pure culture of *E. faecalis* was tested by preparing two diluted (1000 x) batches of cells, one batch using phosphate buffer as a diluent, and the other using Camper solution. The batches were

kept in a cooler with ice until analysis and treated as if they were natural samples. The concentration of *E. faecalis* in both batches, measured by the membrane filtration method, was monitored as a function of time since inoculation (Figure 4.9). Camper solution prepared in phosphate buffer is a better diluent than phosphate buffer alone. The relative concentration (measured concentration normalized by the maximum concentration) of *E. faecalis* in the modified Camper solution remains constant up to at least 12 hours after inoculation, while the relative concentration of cells in phosphate buffer decreases, with only 20% of the initial concentration remaining after 12 hours of incubation (Figure 4.9). Phosphate buffer is only used to perform the serial dilution. During a routine membrane filtration measurement the sample aliquots do not remain in the buffer for more than a few minutes. Consequently, the poor survival of *E. faecalis* in the buffer should not affect the final estimate of the membrane filtration protocol as long as the dilution is performed shortly before the assay.

### 4.3.3 Fluorescence *in situ* hybridization

Fluorescence *in situ* hybridization (FISH) relies on the hybridization of oligonucleotide probes to specific DNA or RNA targets within the cells (Sayler & Layton, 1990). Since this method theoretically allows for the detection of single cells, it was identified as a potential candidate to a) ground truth the data obtained by culturing and b) quantify the degree of particle association of the target indicator bacteria. However, implementation of a FISH protocol for the enumeration of enterococci and *C. perfringens* faced two major difficulties that could not be overcome in the present study: 1) a strong interfering autofluorescence background signal was emitted by the stream

water particles, even without the addition of the probes, that was visible with both filter sets available, and 2) the low natural abundance of either enterococci or *C. perfringens* challenge the effective detection limit of the FISH stained cells with epifluorescence microscopy.

Theoretically, *in situ* hybridization allows for the specific detection of a single microbe, since it is based on the hybridization of synthetic oligonucleotide probes with the original in-cell matching sequences of nucleotides (either DNA or RNA). A molecular beacon is attached to the probe that can be detected though enzymatic reactions, radioactive photography or in the case of FISH, by epifluorescence microscopy. Practically, a few barriers challenge a successful hybridization, generally involving signal-to-noise ratio. Although satisfactory hybridization signals were obtained on tests performed on pure cultures of both *E. faecalis* and *C. perfringens*, hybridizations performed on environmental samples did not produce useful and unambiguous signals.

To obtain a clear signal, it is first necessary for sufficient copies of the probe to penetrate the cell. This is more easily done on Gram-negative cells than on Gram-positive cells (enterococci, *C. perfringens*) because of the interference of the Gram-positive cell wall (Roller et al., 1994; Beihmfohr et al., 1993; Jurtshuk et al., 1992). In addition, spores (e.g. *C. perfringens*) are highly impermeable. Secondly, the signal (chemical, radioactive or fluorescent) must be strong enough to be detectable. This problem is generally solved by targeting ribosomal RNA, present in greater abundance than DNA, or by designing probes with multiple fluorochromes that increase the strength of the signal but also increase the size of the probes and consequently decrease their potential to penetrate the cells. In the case of rRNA probes, signal intensity depends on the activity of the cells: the

more active a cell, the greater the concentration of ribosomes. This is again a problem for the detection of dormant spores (*C. perfringens*).

The concentrations of the fecal indicators (enterococcus and *C. perfringens*) are low, in terms of absolute number of cells, in Kaneohe stream (Table 2.5); the concentrations are nonetheless higher than the water quality standards or guidelines allow. The largest median concentration for any station was 783 CFU/100ml for enterococcus and 109 CFU/100ml for *C. perfringens*. At a magnification of 1000 x, the area of the field of view through the microscope optics is $10^4$ μm$^2$, whereas the area of the 25 mm diameter filter used for FISH is 4.9x$10^8$ μm$^2$; there are >$10^4$ fields per filter. If one defines $F$ as the area of the filter, $A_V$ the area of a field of view, and $N_V$ the number of fields analyzed, it is possible to compute the ratio ($f$) of total filter area sampled by the microscope to the total filter area: $f = \frac{N_V \cdot A_V}{F} = N \cdot 2.04 \cdot 10^{-5}$. In other words, the number of cells ($\chi$) that must be present per filter for detection of a single cell to be possible is given by $\chi = \frac{1}{f} = \frac{1}{N_V} \cdot 4.9 \cdot 10^4$. It is important to note that $\chi$, experimentally, corresponds to a number per volume. If 100 ml of stream water were filtered, and assuming the largest median concentrations for enterococci or *C. perfringens* for $\chi$, the number of fields that must be analyzed for each of these indicators in order for one bacterium to be detected in at least one of the fields analyzed is 62 for enterococcus and 450 for *C. perfringens*. It is of course unrealistic to believe that 100 ml of stream water can be routinely filtered through a 25 mm diameter 0.2 μm filter. Instead, realistic practical volumes are <10 ml, such that the number of fields to be analyzed to permit

detection of only one bacterium really become >620 or >4500 for enterococcus and *C. perfringens*, respectively.

The second problem comes from the high background fluorescence arising from small particles and colloids present in the stream water. This background makes it very challenging or impossible to accurately distinguish between cellular and other material of bacterial size. Under these conditions, and at the cell concentrations in question, which translate into very low counts on the filter, the effects of false positive counts (the assimilation of non-cellular particles as positive cells) becomes so large that the use of FISH for water quality assessment is not possible. For example, assume the true concentration of enterococci in a stream sample is 783 cells/100ml, a 10 ml aliquot of that stream water is filtered and 620 fields analyzed, then if 2 fluorescent points are recorded as positive, but only 1 signal originated from a cell, the measurement error of the concentration estimate becomes 100%.

Because of the low probability of detection, the high background fluorescence signal and the large number of non-cellular particles present in the stream water, it seems unlidely that the FISH protocol described here could be used efficiently to quantify the number of bacteria associated with aggregates or as a tool for water quality assessment in general.

### 4.3.4 Particle settling experiments

Homogenization and Camper solution were used on environmental samples from stations LULU and KANE (Figure 2.1). The results from the samples subjected to homogenization were compared to the result obtained from the corresponding untreated

control samples (CONTROL). If the ratio is larger than 1, treatment effectively increased the colony counts. An increase is consistently observed for HPC bacteria, with average ratios for the different funnel subsamples (BOTTOMs, MIDDLE, TOPs and WHOLE) ranging from 1.20 to 1.70 (Table 4.3). Mean ratios for enterococci are smaller, ranging from 0.85 to 1.28. The average ratios are smaller still for *C. perfringens*, 0.48 to 1.09, depending on the funnel subsample. The maximum increase in colony counts (ratio=3.16) was recorded for HPC bacteria from the WHOLE water sample. The minimum ratio for HPC (0.42) was produced by one of the $BOTTOM_{500}$ samples. The maximum ratio measured for enterococcus was 1.57 ($BOTTOM_{500}$), and the minimum was 0.47 ($TOP_{500}$). Correspondingly, values for *C. perfringens* were 1.96 (WHOLE) and 0.12 ($BOTTOM_{250}$, $TOP_{500}$ and $TOP_{250}$). Sixty-eight percent of the samples (23 of 34 samples) taken from the bottom of the funnels showed an increase in detected colonies (ratio > 1), regardless of organism type (Table 4.3). This was only the case 38% of the time (13 of 34 samples) for samples pipetted from the top of the funnels.

The Shapiro-Wilk W test for normality shows that, for most of the treatments, the hypothesis that the ratios are normally distributed cannot be rejected ($p>0.05$), except for the HPC ratios corresponding to the WHOLE samples and the *C. perfringens* ratios for the $TOP_{250}$ samples (Table 4.4).

A t-test was used to compare the ratios obtained for each funnel sub-sample for each type of organism. Of the 45 possible comparisons only 6 were significant ($p<0.05$) (Table 4.5). The sample size is small, however, such that the Shapriro-Wilk test for normality is not very powerful and the parametric test is not robust. In consequence, a Bayesian non-parametric binomial test relying on the assumption that any difference

observed between counts of paired samples is due to random error has also been used. A more detailed description of this test can be found in Appendix G. The test results are shown in Table 4.6. Results from the t-test and from the Bayesian binomial test agree well, however.

### *4.3.4.1 C. perfringens*

The ratios obtained from the WHOLE samples for *C. perfringens* were significantly larger than the ratios of the $TOP_{250}$ samples (Table 4.5 and Table 4.6 A). The $TOP_{250}$ ratios are generally lower than both the $BOTTOM_{250}$ and the $BOTTOM_{500}$ samples, however not significantly. In addition, the ratios of the WHOLE samples tend to be higher than either the $MIDDLE_{500}$ or the $TOP_{500}$ samples. Other comparisons did not yield statistically significant differences (Table 4.5 and Table 4.6). The statistical differences and trends observed between treatments suggest that the TOP samples are depleted in *C. perfringens* compared to the BOTTOM and WHOLE samples.

However, 22 out of the 34 (65%) *C. perfringens* ratios are smaller than 1 (Table 4.3). The mean ratio of the $TOP_{250}$ samples is significantly lower than 1 (t-test against mean of 1, Table 4.7). Recall, however, that the Shapiro-Wilk test for normality suggested that the $TOP_{250}$ ratios for *C. perfringens* are not normally distributed (Table 4.4), such that the result of this test for the $TOP_{250}$ ratios is not robust. Nonetheless, all of the $TOP_{250}$ ratios for *C. perfringens* were <1 (Table 4.3) and results from a non-parametric binomial test (binomial test for ratio <1, Table 4.7) indicate that the $TOP_{250}$ ratios are indeed smaller than 1. In addition, the p-values corresponding to the ratios of the $TOP_{500}$ and the $MIDDLE_{500}$ samples are low, 0.103 and 0.074, respectively,

suggesting that the mean of these ratios may be lower than one. This statement is confirmed statistically using the binomial test (Table 4.7). Other mean ratios for CP do not differ from 1 (Table 4.7).

Homogenization, the Camper solution or the exposure of the sample to light at room temperature in a glass funnel may have a negative effect on the recovery of *C. perfringens* compared to untreated samples and may explain the low (<1) ratios.

In summary, the analysis of the ratios between samples homogenized and treated with Camper solution and control samples for *C. perfringens* shows that the bottom layer of the settled sample tends to have a larger concentration of cells compared to the top layer, indicating that *C. perfringens* sinks. Different mechanisms could explain the sinking. Either *C. perfringens* is associated with denser particles or the inherent density of the spores is sufficiently greater than that of water such that many single cells are able to settle to the bottom of the funnel within 4 hours. Based on Stokes Law for particle settling, it is possible to calculate the settling distance that particles of a given size and density travel within 4 hours (Figure 4.10), and compare these distances with the vertical distances between the sub-sample layers in the separatory funnel (Figure 4.1) to assess whether bacterial spores could settle out of the top layers. Because the time required for small particles to reach the terminal settling velocity is much smaller than 4 hours, it is possible for some or all spores with diameters between 1-3 μm and an assumed density of 1.5 g/cm$^3$, to settle out of the surface layer in both the 250 ml and 500 ml funnels. It is possible for spores attached to denser inorganic particles to settle out of the funnel completely.

The fact that the mean ratios of both the $BOTTOM_{250}$ and the $BOTTOM_{500}$ samples and the WHOLE samples do not significantly differ from 1, however, may suggest that *C. perfringens* is not abundantly present on aggregates. The results suggest that association of *C. perfringens* with aggregates is likely not an important problem for the measurement of *C. perfringens* concentrations by culturing.

### 4.3.4.2 Enterococcus

The mean $BOTTOM_{500}$ ratio for enterococcus is significantly greater than the means for both the $TOP_{250}$ and the $TOP_{500}$ samples (Table 4.5 and Table 4.7). The mean ratio corresponding to the WHOLE samples is significantly greater than the mean ratios of either the $TOP_{250}$ or the $TOP_{500}$ samples. These differences are consistent with the hypothesis that some enterococci are found on aggregates.

Some of the data presented in Table 4.5 and Table 4.7 are contradictory, however. The p-value for the test comparing the mean ratios of the $BOTTOM_{500}$ and the $BOTTOM_{250}$ samples is low ($p=0.13$ in Table 4.5 and $p=0.04$ in Table 4.7), and the average ratio of the $BOTTOM_{250}$ samples is significantly smaller than the ratios obtained for the WHOLE samples (Table 4.5 and Table 4.7). The number of replicates is small, however, and the low average ratio observed for the $BOTTOM_{250}$ samples can be attributed to 2 points: KANE 2 and 4, whose ratios are both unusually low (Table 4.3). The low ratios for KANE 4 are due to an unusually large CONTROL value relative to the treated samples. Removing KANE 4 from the analysis increases the mean ratios for all treatments.

Even if differences do exist between treatments, only the ratios of the WHOLE samples differ significantly from 1 (p=0.01, Table 4.7). The p-value corresponding to the $BOTTOM_{500}$ ratios, with a mean ratio of 1.23, is low when the t-test is used (p=0.106) and significant with the binomial test (p=0.01), suggesting that these ratios are consistently larger than 1. Omitting the ratios corresponding to KANE 4, the mean ratio of the $BOTTOM_{500}$ samples increases to 1.34 and becomes significantly larger than 1 with both statistical tests (t-test, p=0.004), and the mean ratio for the WHOLE samples becomes 1.32. Overall, the data presented in Table 4.3 suggest that 1-34% of the total number of enterococci present in a sample is found associated with aggregates.

### 4.3.4.3 HPC

No significant differences were found among the mean ratios of the diverse funnel sub-samples for HPC using the t-test (Table 4.5) and the t-test against a mean of 1 revealed that none of the mean ratios was significantly different from 1 at the 0.05 level (t-test against mean of 1, Table 4.7). The p-values of the latter t-test are generally low, however, with two of them ($BOTTOM_{250}$ samples and $MIDDLE_{500}$ samples) between 0.05 and 0.1, and three other values are smaller than 0.2 (Table 4.7). When the binomial test is used, however, all ratios but the $TOP_{500}$ ratios become significantly greater than 1 (Table 4.7). The HPC ratios in Table 4.3 indicate that HPC bacteria are found on aggregates in sufficiently large proportions for the membrane filtration assay to be affected.

### 4.3.4.4 Differences between organisms

Test results in Table 4.7 indicate that all except the WHOLE ratios are statistically greater for HPC than they are for CP. Yet, the p-value associated with the WHOLE ratios' comparison is low (p=0.12, binomial test A>B, Table 4.7). These results suggests that the ratios between treated samples and the CONTROL obtained for HPC are larger than the ratios obtained for *C. perfringens*. This indicates that HPC bacteria are relatively more abundant on aggregates than *C. perfringens*.

For both statistical tests (t-test and binomial), the ratios obtained from the $BOTTOM_{250}$ and $TOP_{250}$ sub-samples for HPC are significantly greater than the mean ratio obtained for ENT (Table 4.7). The mean ratio of the $MIDDLE_{500}$ is only marginally significantly greater for HPC than the mean ratio of this treatment for enterococcus (Table 4.7). Overall, the HPC ratios tend to be larger than the enterococcus ratios for all the treatments except the $BOTTOM_{500}$, the $TOP_{500}$ and the WHOLE samples.

The comparisons between enterococcus and *C. perfringens* ratios for the $MIDDLE_{500}$ and the $TOP_{250}$ samples, present significant p-values for the binomial test, but only marginally significant p-values for the t-test (Table 4.7). The magnitudes of the ratios obtained for enterococcus tend to be larger than those for *C. perfringens* (Table 4.3). Sixty percent (24 out of 40 values) of the enterococcus ratios are larger than 1, whereas only 35% (12 of 34 values) of the ratios of *C. perfringens* are greater than 1 (Table 4.3). Based on this analysis, it is likely that the association of cells with particles is more important for enterococcus than it is for *C. perfringens*.

### 4.3.4.5 A comparison of the results with the literature

The magnitudes of the ratios shown in Table 4.3 are small compared with ratios estimated from the literature (Table 4.8). The values in Table 4.8 were estimated by dividing measurements of the total bacterial population by measurements that do not account for the fact that multiple cells may be associated with aggregates. For most of the values in Table 4.8, this effectively corresponds to taking the ratio of the concentrations obtained from samples treated to account for aggregated cells to the concentrations of untreated (control) samples, as was done to produce the numbers in Table 4.3.

The data presented in Table 4.3 and Table 4.8 are not easily comparable. The methods used for determining the concentrations differ, the types of organisms are different and the environments sampled do not resemble a tropical stream. The magnitude of the ratios obtained by Perdek & Borst (2000) for CSO and by Borst & Selvakumar (2003) for storm water runoff agree best with the data presented in Table 4.3. In addition, the environments sampled by these two studies, of the studies presented in Table 4.8, are the most closely related to the environment considered in the present study (stream water). The numbers given for enterococcus (fecal streptococcus) are fairly close in all three studies, ranging from 1.15 to 6.6.

The ratios obtained from studies performed on samples from drinking water systems (GAC, sand, anthracite) are larger than those obtained from environmental samples (stormwater runoff, estuary) or from wastewater treatment (primary & secondary effluent, CSO). The particles present in drinking water are either detached fragments of biofilms or particles escaped from the filters used to clean the water (Camper et al., 1985 a & b, 1986, 1987; Pernitsky, Finch & Huck, 1997). These particles are, by nature, rich in

cells while the background concentrations of the drinking water are small, resulting in elevated ratios. In the case of the wastewater samples, both the number of cells associated with aggregates and the number of background (free-living) cells are elevated (Wagner et al. 1998, Emerick et al. 1999), such that the ratios will be lower than those measured in drinking water. Environmental samples are usually dominated by free-living cells (Palumbo, Ferguson & Rublee, 1990; Bidle & Fletcher, 1995; Kirchman & Mitchell, 1982, Clarke & Joint, 1986, Ramsay, 1984, Yoon & Rosson, 1990, Murrell et al., 1999), such that the ratios obtained from DAPI or acridine orange counts are expected to be close to 1, which is indeed observed (Table 4.8).

### 4.3.5 Particle filtration experiments

If the total number of free-living bacteria is large compared with the particle-associated population it may not be possible to detect a positive signal from the particle-associated population even after treatment with homogenization and Camper solution. This is an artifact of the culturing methods, which rely on the decimal serial dilution of the original sample. That is, one colony-forming unit recovered at a dilution of $10^{-5}$, essentially carries the weight of $10^5$ cells. If particles-associated cells account for only less than 10% the total number of free living cells, then the particle-associated cells will be lost in the dilution process. To circumvent this problem, it becomes necessary to physically separate the population associated with aggregates from the free-living population. This can be achieved by filtration (Table 4.1, see references in Appendix F).

The filtration process yields two products: the particles retained by the filter (retentate) and the solution that passes through it (the filtrate). It is theoretically possible

to quantify the number of aggregates with associated bacteria ($C_A$) by the difference between the concentration obtained for the filtrate ($C_F$) and that obtained without the pre-filtration step ($C_R$) (Figure 4.2).The results obtained by this approach are summarized in Table 4.9. Both $C_R$ and $C_F$ derive from 250 ml samples. The $C_F$ samples were pre-filtered though a UV-sterilized 5 μm Nuclepore membrane filter before analysis by the membrane filtration methods (see Methodology section). Given that none of the data presented in Table 4.9 is normally distributed (Shapiro-Wilk test, $p<0.05$), the non-parametric Wilcoxon matched pair test was used to test for differences between $C_R$ and $C_F$ for each type of organisms. The p-values are all greater than 0.1, indicating that the data from the two treatments are *a priori* not statistically different (Table 4.10).

Given the lack of a significant difference between the data, the ratios of $C_F$ to $C_R$ (Table 4.11) should be statistically indistinguishable from 1. The mean and the median of the ratios between $C_F$ and $C_R$ for all the organisms are smaller but not statistically different from 1 (Table 4.11). However, because of the non-normality of the ENT and CP ratios (Shapiro-Wilk test, $p<0.001$), the parametric t-test is not appropriate, and its results are not robust for these organisms. The t-test is used here because an equivalent non-parametric test does not exist. The difference between the mean and 1 is significant ($p=0.027$) for *C. perfringens* if the maximum ratio (5.44, Table 4.11) is omitted from the analysis; in which case, the mean ratio decreases from 0.98 to 0.72. This unusually high value is difficult to explain. Contamination of the filtrate sample with high levels of *C. perfringens* is possible, however very unlikely for there were no sources of *C. perfringens* in the laboratory at the time of the analysis and, if contamination had occurred, it is likely that the levels of HPC bacteria or enterococcus would have increased

as well. An alternative explanation for the abnormally elevated datum could be the presence of one or more aggregates containing enormous amounts of *C. perfringens* that became stuck on the filter and were stripped of their aggregated cells under the force of the vacuum.

The Wicoxon matched pair test shows that the HPC ratios are not different from the ENT ratios ($p>0.8$; Table 4.12). The p-values of the Wilcoxon test are smaller when the ratios for *C. perfringens* are compared to the ratios of HPC and ENT, $p=0.170$ and $p=0.199$, respectively (Table 4.12).

In conclusion, the concentration measurements collected from the filtrate ($C_F$) are not statistically different than the concentration measurements performed on the whole sample ($C_R$), suggesting that the number of particles greater than 5 μm harboring the types of bacteria measured in this study is relatively small (<10% of the measured colonies originate from aggregates >5 μm). On the other hand, the ratios from Table 4.3 indicate that the concentration of the bacteria of interest increased on average between 10-50% after treatment of the samples with the desorption protocol. These two data sets can be reconciled in two ways (Table 4.3 and Table 4.11): 1) on average, >2-5 bacteria are present per bacteria-hosting aggregate larger than 5 μm, or 2) there exists a relatively large number of small aggregates harboring bacteria, smaller than 5 μm, that are not effectively removed by the 5 μm filter. For the second alternative to be the case, 10 to 50% of the colonies detected on an agar plate must originate from small (<5 μm) bacterial clusters (e.g. pairs of bacteria). Statistically, from aggregation theory alone, the odds are greater to find a large number of aggregates with few bacteria than a few

aggregates with many associated microbes, such that the second proposition is more likely than the first.

### 4.3.6 Particles with associated bacteria

None of the data sets, except $C_A$ and $C_H$ for HPC (Shapiro-Wilk test, $p>0.2$), is normally distributed ($p<0.05$; Table 4.13). The non-parametric Wilcoxon matched pair tests suggest that the $C_H$ values for all the organisms are significantly greater than the corresponding $C_A$ numbers (Table 4.13), reinforcing the idea that the desorption protocol separates bacteria from aggregates (Table 4.3).

The ratio $C_A/C_R$ represents the fraction of the colonies produced by the routine membrane filtration ($C_R$) without homogenization nor Camper solution that are attributable to aggregates containing bacteria larger than 5 μm ($C_A$). The mean (median) fraction for HPC is 16% (15%), but only 7% (5%) for ENT and as high as 21% (22%) for CP (Table 4.14). The $C_A/C_R$ ratios for all types of organisms are the same order of magnitude, or slightly larger, than the values obtainable from Table 4.11 using the expression $1-C_F/C_R$, indicating consistency between the data sets obtained from the filtrate and the retentate. The fact that the fractions $C_A/C_R$ are largest for CP suggests that many colonies grown from the membrane filtration protocol on mCP agar originate from particles larger than 5 μm. This is consistent with the results obtained from the settling experiments (Table 4.3), which showed that the top samples were CP depleted and with the results from the filtration experiment (Table 4.11), omitting the 5.44 outlier.

### 4.3.6.1 An estimate of the number of bacteria per aggregate

The ratio $C_H/C_A$ gives an indication of the magnitude of the increase achieved by the desorption protocol. This ratio could be thought of as an "aggregation factor" expressing the sample's average number of bacteria per bacteria-hosting aggregate. The median $C_H/C_A$ ratio for HPC is 1.18, with a maximum of 3.45 and a minimum of 0.53; 5 of 17 HPC ratios are less than 1 (Table 4.14). The ratios for enterococcus are the largest, with a median ratio of 2.02, a minimum of 0.75 and a maximum of 5.4. Only 2 $C_H/C_A$ ratios are smaller than 1 for enterococcus (Table 4.14). The $C_H/C_A$ ratios for *C. perfringens* are the lowest observed. The median ratio is 1.14, the minimum is 0.36 and the maximum is 1.68. A many as 7 of 17 *C. perfringens* ratios are smaller than 1 (Table 4.14). The potential negative effect of the treatment on the recovery of CP has already been suggested from previous results (Table 4.3).

A few studies have attempted to enumerate the number of bacteria associated with aggregates (Table 4.15). As was the case earlier (Table 4.8), previously published numbers are not directly comparable with the data generated in the present study. Interestingly, however, the ratios in Table 4.14 range from 0.36 to 5.40, which compares well with the magnitude of the numbers obtained from previously published work (Table 4.15).

A problem with the method used to determine $C_H/C_A$ ratios is the potential for "by-catch" on the membrane filter of free-living bacteria due to non-ideal filtration; that is, free-living bacteria get trapped on the membrane or onto large particles that clog the pores. By-catch essentially pushes the ratio between $C_H$ and $C_A$ towards 1 by increasing the magnitude of both numbers. For example, if the true $C_A$ is 10 and the true $C_H$ is 15,

but if by-catch of free-living bacteria results in measured $C_A$ of 15 and $C_H$ of 20, then the measured $C_T/C_A$ ratio (20/15=1.33) will underestimate the true ratio of 15/10=1.5.

To investigate the error caused by the by-catch problem, constant known amounts (cell concentration was adjusted after DAPI determination of the concentration in the batch culture) of cells (*E. faecalis*) were mixed in constant volumes (500 ml) containing different amounts of autoclaved stream water (with particles) and sterile phosphate buffer. The goal of these experiments was to assess how the suspended solid concentration in a sample affects the extent of by-catch. The stream water samples used for the first set of experiments (Figure 4.11, A) originated from Manoa Stream on a stormy day (TSS=10.6 mg/L). The second set of experiments (Figure 4.11, B & C) was performed on samples from Kaneohe Stream (station KANE, TSS=5.1 mg/L). Experiments performed with both *E. faecalis* cells and a Gram-negative environmental isolate show similar results (Figure 4.11): the number of cells that are caught drastically increases when the particle concentration per filter (5 μm pore size) reaches 1-1.5 mg/filter.

For each determination of $C_A$ and $C_H$, the particles (>5 μm) from 1L of stream water were collected and resuspended in phosphate buffer ($C_A$) or Camper solution ($C_H$) to achieve a final particle concentration 4 x greater than in the original sample. Initially, this process was accomplished by filtering 500 ml of stream water through each of two 5 μm pore size membranes. In an effort to reduce the "by-catch" problem by decreasing the particle load per filter, this protocol was modified to use multiple filters (up to 6 x 166 ml). A plot of the ratios $C_H/C_A$ versus the suspended solid load per filter suggests that, as the particle load per filter increases, the ratio approaches 1(Figure 4.12). However, there

exist no significant correlations (Spearman R, $p>0.05$) between the ratios for any organism and the suspended solids load per filter.

If the data set for the $C_H/C_A$ ratios (Table 4.14) is separated into those ratios derived from filters with a particle load larger (N=10) than 1.6 mg/filter, which is approximately the critical filter load at which by-catch becomes problematic (Figure 4.12), and those coming from filters with particle loads smaller than 1.6 mg/filter (N=7), including the 3 data points for ENT with a high ratio, then 1) the assumption of normality for both pools of data cannot be rejected, except for the HPC ratios with large amounts of solids per filter, and 2) both sets of data reveal that a weak difference (t-test for independent samples, $p=0.1$) exists between the mean ENT ratios (2.91) with low particle loads and the mean ENT ratios obtained from filters with a large solids concentration (1.78). Similar analysis of the HPC and CP data yield high p-values and suggest that no difference exists between the two subsets for these organisms. Overall, these data may indicate that the by-catch of cells may be a problem for ENT, but the magnitude of the problem is comparatively small for HPC and CP. This could also be evidence for the greater stickiness of ENT (Vanek, 1999) compared with CP spores or Gram-negative dominated HPC (of 20 HPC colonies isolated from Kaneohe Stream on TS agar and resuspended in TS broth, 17 (85%) were Gram negative).

### 4.3.7 Assessing the measurement error due to particle-associated bacteria.

Using the data presented in Table 4.13 and the following relation,

$$\Psi = \frac{C_R + C_H - C_A}{C_R} \tag{4.2}$$

it is possible to compute the ratio ($\Psi$) of the concentration of the total culturable bacteria to the bacterial concentration estimated by the routine membrane filtration analysis without homogenization or Camper solution. The numbers produced by this formula should theoretically be comparable to the ratios of treated WHOLE samples to the corresponding untreated (CONTROL) samples shown in Table 4.3. It is clear that the $\Psi$ values (Table 4.16) are smaller than the corresponding WHOLE/CONTROL ratios (Table 4.3).These two data sets also contrast in the fact that the largest mean WHOLE/CONTROL ratios are observed for HPC (Table 4.3), whereas the greatest mean $\Psi$ ratios correspond to ENT (Table 4.16).

In spite of being relatively small, the mean ENT $\Psi$ ratio (Table 4.16) is significantly larger than 1 (t-test, $p<0.001$), and that for HPC is marginally larger than 1 ($p=0.083$; Table 4.16). The mean ratio for CP is not statistically different from 1 ($p>0.1$). Equivalent observations (Table 4.7) were made for the WHOLE/CONTROL ratios (Table 4.3).

Unlike the $C_H/C_A$ ratios, the $\Psi$ ratios (Table 4.16) are not influenced by the particle load on the filters (Figure 4.13). This is expected since the ratio $\Psi$ relies on the difference between $C_H$ and $C_A$ and not the ratio $C_H/C_A$, such that any effect of by-catch of free-living bacteria automatically cancels out.

The ratios (WHOLE/CONTROL and $\Psi$) in Table 4.3 and Table 4.16 are calculated by generic equations of the form:

$$\psi = \frac{C_R + \Delta}{C_R} = 1 + \frac{\Delta}{C_R} \tag{4.3}$$

where Δ represents the number of bacteria that have not been accounted for by the routine membrane filtration method. The fact that the ratios in Table 4.16 are smaller than those in Table 4.3 means that the Δs associated with the data set in Table 4.16 are smaller.

The issue was raised in a previous experiment comparing the concentrations of $C_R$ and $C_F$, that either a large number of bacteria are associated with rare aggregates, or that a large number of small clusters of bacteria must exist that are not effectively retained on the 5 μm filter. The discrepancy present between the data in Table 4.3 and Table 4.16 suggests that there exist a large number of small (<5 μm) bacterial clusters that are not included in the parameters $C_A$ or $C_H$. Based on the results from the experiments presented here, one can rewrite

$$\frac{\Delta}{C_R} = \frac{\Delta_{<5\mu m}}{C_R} + \frac{\Delta_{>5\mu m}}{C_R} \tag{4.4}$$

The data in Table 4.3 and Table 4.16 can be used to calculate the fraction $\frac{\Delta_{<5\mu m}}{C_R}$:

$$\frac{\Delta_{<5\mu m}}{C_R} = \frac{\Delta}{C_R} - \frac{\Delta_{>5\mu m}}{C_R} = \left(\frac{WHOLE}{CONTROL} - 1\right) - (\Psi - 1) = \left(\frac{WHOLE}{CONTROL} - \Psi\right) = S \tag{4.5}$$

S can be understood as the fraction of the excess (aggregated) bacteria not accounted for by the routine membrane filtration alone that originate from aggregates smaller than 5 μm (Table 4.17). The values of S are all elevated for all the organisms, highlighting the general importance of small bacterial clusters, as opposed to rare large ones. It was suggested earlier that *C. perfringens* was not abundantly found on aggregates (Table 4.3 and Table 4.14). This statement is confirmed in this analysis, which suggests that most *C. perfringens* cells originate from the sample fraction smaller than 5 μm

(median S=0.99, Table 4.17). The median S-value associated with HPC is also high (0.98), while the median S corresponding to enterococcus is lower (0.85), supporting the contention that relatively more enterococci are associated with aggregates greater than 5 μm than are HPC.

In conclusion, the error inherent in not accounting for the fact that multiple bacteria can be associated with aggregates is typically on the order of 10-50% of the routine membrane filtration measurement (Table 4.3). The overall error is largest for HPC and ENT, but small for CP. The largest fraction of the error comes from small bacterial clusters (<5 μm). This is most important for HPC. While the fraction of the error attributable to small clusters is also high for ENT, these organisms showed a greater affinity for larger aggregates relatively to HPC or CP.

## 4.4 Conclusions

The protocol used in this study to desorb bacteria from aggregates does produce greater or similar colony counts after treatment for all types of organisms relative to the membrane filtration assay alone. In addition, homogenization was demonstrated to be a better alternative than sonication since homogenization did not impair cell viability. It is unclear if the use of Camper solution greatly improves the separation of cells from aggregates. Camper solution, however, was shown to be a superior buffer to phosphate buffer in which to incubate microbial samples for extended periods prior homogenization. Unfortunately, a rigorous independent assessment of the efficiency of the protocol on natural stream water samples was not possible because of the limitations inherent to the FISH technique.

The data presented throughout this chapter verify the original hypothesis that enterococci are proportionally more abundant on aggregates than HPC or *C. perfringens*. The relative fractions ($C_H/C_A$) of *C. perfringens* and HPC bacteria on aggregates are similar. A comparison of the present results with other published data is difficult due to differences in the methodologies used, in the model organisms, or in the environment sampled.

It is concluded that the concentration measurements of *C. perfringens* are not significantly biased by the presence of bacteria associated with aggregates, because these organisms are not abundantly found on aggregates. However, large and variable fractions of the HPC and ENT populations are found associated with aggregates. Small (<5 μm) bacterial clusters are responsible for a majority of the bias associated with these organisms, but some larger aggregates are found with a large number of ENT, as shown by the largest aggregation factors recorded for this group.

These results do not indicate that *C. perfringens* is a better pathogen indicator than enterococcus, however, since the extent to which pathogens are or not found on aggregates in unknown.

**Table 4.1. A review of the literature summarizing the pore sizes of the filters used to separate between free-living and aggregated bacteria. Some studies used sequential filtration and consequently contribute multiple sizes to this table (see references in Appendix F).**

| Pore size (μm) | Number of studies | Pore size (μm) | Number of studies |
|---|---|---|---|
| 1 | 2 | 11 | 1 |
| 2 | 1 | 20 | 1 |
| 3 | 4 | 30 | 1 |
| 5 | 4 | 50 | 1 |
| 7 | 1 | 52 | 1 |
| 8 | 3 | 70 | 1 |
| 10 | 5 | 80 | 2 |

**Table 4.2. Properties of the oligonucleotide probes used in this study. All the probes target 16s rRNA and were manufactured by IDT-DNA, Inc. %FA refers to the formamide concentration of the hybridization solution and the NaCl concentration relates to the washing solution.**

| Probe name | Dye+Sequence | Specificity | %FA | NaCl (mM) | Reference |
|---|---|---|---|---|---|
| Cp2 | 5'-/5Cy3/GCT CCT TTG GTT GAA TGA TG -3' | *C.perfringens* | 40 | 80 | Roenner & Stackebrandt, 1994 |
| Enc131 | 5'-/5Cy3/CCC CTT CTG ATG GGC AGG -3' | *E. avium; E. casseliflavus; E. durans; E. faecium; E. flavescens; E. gallinarum; E. hirae; E. malodoratus; E. mundtii; E. pseudoavium; E.* | 35 | 80 | Meier et al. 1997 |
| DB8 | 5'-/56-FAM/TAG GTG TTA GCA TTT CG -3' | *E. faecalis* | 25 | 80 | Meier et al. 1997, Betzl et al. 1990 |
| EUB338 | 5'-/5Cy3/GCT GCC TCC CGT AGG AGT-3' | *Eubacteria* | 20 | 225 | Amman et al., 1990, Schramm et al., 1998 |
| NON338 | 5'-/5Cy3 or 6-FAM/ACT CCT ACG GGA GGC AGC-3' | Negative control | 20 | 225 | Stahl & Amman, 1991, Schramm et al. 1998 |

**Table 4.3. Data table of the ratios obtained by normalizing the CFU concentrations obtained after treatment of the different subsamples with the desorption protocol by the concentrations of the corresponding untreated (CONTROL) samples. Number greater than 1 indicate that the colony count increased after treatment relative to the control.**

| Sample | Date | Separation funnel subsamples | | | | | WHOLE |
|---|---|---|---|---|---|---|---|
| | | $BOTTOM_{500}$ | $BOTTOM_{250}$ | $MIDDLE_{500}$ | $TOP_{500}$ | $TOP_{250}$ | |
| | | | | Heterotrophic plate count | | | |
| LULU 1 | 5/6/2004 | 1.92 | 2.02 | 2.24 | 1.44 | 1.56 | 1.33 |
| LULU 2 | 5/12/2004 | 1.33 | --- | 1.81 | 1.33 | 1.05 | 1.29 |
| LULU 3 | 5/18/2004 | 1.28 | 1.02 | 1.06 | 0.85 | --- | 1.32 |
| KANE 1 | 2/12/2004 | 0.42 | --- | 0.68 | 0.93 | --- | 3.16 |
| KANE 2 | 4/15/2004 | --- | --- | --- | --- | --- | --- |
| KANE 3 | 4/21/2004 | 2.46 | 1.74 | 2.27 | 2.08 | 2.84 | 0.84 |
| KANE 4 | 4/28/2004 | 1.56 | 1.67 | 1.45 | 0.54 | 1.34 | 1.16 |
| Mean | | 1.50 | 1.61 | 1.59 | 1.20 | 1.70 | 1.52 |
| Median | | 1.45 | 1.71 | 1.63 | 1.13 | 1.45 | 1.31 |
| Standard deviation | | 0.68 | 0.42 | 0.64 | 0.54 | 0.79 | 0.83 |
| | | | | Enterococcus | | | |
| LULU 1 | 5/6/2004 | 1.36 | 1.17 | 1.55 | 0.87 | 0.87 | 1.28 |
| LULU 2 | 5/12/2004 | 1.09 | 0.98 | 1.00 | 0.98 | 0.83 | 1.23 |
| LULU 3 | 5/18/2004 | 1.45 | 1.32 | 1.52 | 1.13 | 1.19 | 1.46 |
| KANE 1 | 2/12/2004 | 1.57 | --- | 0.97 | 0.55 | --- | 1.51 |
| KANE 2 | 4/15/2004 | 1.35 | 0.65 | 0.83 | 0.87 | 0.70 | 1.45 |
| KANE 3 | 4/21/2004 | 1.23 | 1.03 | 1.07 | 1.10 | 1.07 | 1.01 |
| KANE 4 | 4/28/2004 | 0.58 | 0.53 | 0.54 | 0.47 | 0.49 | 1.04 |
| Mean | | 1.23 | 0.95 | 1.07 | 0.85 | 0.86 | 1.28 |
| Median | | 1.35 | 1.01 | 1.00 | 0.87 | 0.85 | 1.28 |
| Standard deviation | | 0.32 | 0.30 | 0.36 | 0.26 | 0.25 | 0.20 |
| | | | | *Clostridium perfringens* | | | |
| LULU 1 | 5/6/2004 | 1.80 | 1.52 | 1.26 | 1.16 | 0.69 | 1.96 |
| LULU 2 | 5/12/2004 | 0.33 | 0.67 | 0.17 | 0.50 | 0.17 | 0.50 |
| LULU 3 | 5/18/2004 | 0.23 | 0.12 | 0.46 | 0.12 | 0.12 | 0.65 |
| KANE 1 | 2/12/2004 | 1.16 | --- | 0.96 | 1.10 | --- | 1.04 |
| KANE 2 | 4/15/2004 | --- | --- | --- | --- | --- | --- |
| KANE 3 | 4/21/2004 | 1.19 | 1.33 | 0.37 | 0.75 | 0.72 | 1.30 |
| KANE 4 | 4/28/2004 | 0.62 | 0.81 | 0.55 | 0.29 | 0.72 | 1.07 |
| Mean | | 0.89 | 0.89 | 0.63 | 0.65 | 0.48 | 1.09 |
| Median | | 0.89 | 0.81 | 0.51 | 0.63 | 0.69 | 1.06 |
| Standard deviation | | 0.60 | 0.56 | 0.41 | 0.43 | 0.31 | 0.52 |

**Table 4.4. Results of the Shapiro-Wilk W test for normality performed on the data in Table 4.3. The p-values represent the probability that the distributions are normally distributed. Underlined values have $p<0.05$.**

| Variable | Separation funnel subsamples | | | | | WHOLE |
|---|---|---|---|---|---|---|
| | $BOTTOM_{500}$ | $BOTTOM_{250}$ | $MIDDLE_{500}$ | $TOP_{500}$ | $TOP_{250}$ | |
| HPC | 0.92 | 0.48 | 0.59 | 0.81 | 0.25 | <u><0.01</u> |
| ENT | 0.19 | 0.70 | 0.50 | 0.32 | 0.96 | 0.28 |
| CP | 0.58 | 0.80 | 0.61 | 0.60 | <u>0.02</u> | 0.65 |

**Table 4.5. p-values of the t-test for the comparisons between treatments for the data from Table 4.3. Underlined and bold values are significant at the 0.05 level. Low p-values suggest that the means of the two treatments compared are different.**

| HPC ENT CP | Separation funnel subsamples | | | | WHOLE |
|---|---|---|---|---|---|
| | $BOTTOM_{250}$ | $MIDDLE_{500}$ | $TOP_{500}$ | $TOP_{250}$ | |
| | 0.77 | 0.82 | 0.42 | 0.68 | 0.96 |
| $BOTTOM_{500}$ | 0.13 | 0.38 | **0.03** | **0.04** | 0.73 |
| | 1.00 | 0.40 | 0.45 | 0.21 | 0.55 |
| | | 0.94 | 0.23 | 0.86 | 0.84 |
| | $BOTTOM_{250}$ | 0.53 | 0.56 | 0.59 | **0.04** |
| | | 0.39 | 0.45 | 0.19 | 0.55 |
| | | | 0.28 | 0.81 | 0.87 |
| | | $MIDDLE_{500}$ | 0.23 | 0.26 | 0.19 |
| | | | 0.92 | 0.53 | 0.12 |
| | | | | 0.26 | 0.44 |
| | | | $TOP_{500}$ | 0.98 | **<0.01** |
| | | | | 0.48 | 0.14 |
| | | | | | 0.74 |
| | | | | $TOP_{250}$ | **<0.01** |
| | | | | | **0.05** |

**Table 4.6: Results of a Bayesian binomial test on the data obtained from the settling experiment. A definition of the test is available in Appendix G. The values given in these tables are the p-values of the test.**

A) One-tailed test (X>Y)

| HPC ENT CP | Separation funnel subsamples | | | | WHOLE |
|---|---|---|---|---|---|
| | $BOTTOM_{250}$ | $MIDDLE_{500}$ | $TOP_{500}$ | $TOP_{250}$ | |
| $BOTTOM_{500}$ | 0.38 | 0.47 | 0.77 | 0.40 | 0.72 |
| | 0.97 | 0.87 | >0.99 | 0.99 | 0.48 |
| | 0.47 | 0.76 | 0.81 | 0.89 | 0.29 |
| | | 0.46 | 0.91 | 0.66 | 0.82 |
| | $BOTTOM_{250}$ | 0.37 | 0.80 | 0.71 | **0.03** |
| | | 0.84 | 0.80 | 0.87 | 0.47 |
| | | | 0.91 | 0.54 | 0.71 |
| | | $MIDDLE_{500}$ | 0.82 | 0.84 | 0.16 |
| | | | 0.53 | 0.71 | **0.05** |
| | | | | 0.13 | 0.43 |
| | | | $TOP_{500}$ | 0.57 | **<0.01** |
| | | | | 0.82 | 0.13 |
| | | | | | 0.82 |
| | | | | $TOP_{250}$ | **<0.01** |
| | | | | | 0.07 |

B) One-tailed test (Y>X)

| HPC ENT CP | Separation funnel subsamples | | | | WHOLE |
|---|---|---|---|---|---|
| | $BOTTOM_{250}$ | $MIDDLE_{500}$ | $TOP_{500}$ | $TOP_{250}$ | |
| $BOTTOM_{500}$ | 0.70 | 0.59 | 0.26 | 0.64 | 0.31 |
| | **0.04** | 0.16 | **0.01** | **0.02** | 0.53 |
| | 0.60 | 0.29 | 0.21 | 0.18 | 0.76 |
| | | 0.62 | 0.13 | 0.44 | 0.24 |
| | $BOTTOM_{250}$ | 0.69 | 0.25 | 0.35 | 0.98 |
| | | 0.21 | 0.23 | 0.19 | 0.60 |
| | | | 0.12 | 0.54 | 0.35 |
| | | $MIDDLE_{500}$ | 0.24 | 0.17 | 0.87 |
| | | | 0.53 | 0.37 | 0.98 |
| | | | | 0.91 | 0.61 |
| | | | $TOP_{500}$ | 0.46 | >0.99 |
| | | | | 0.25 | 0.89 |
| | | | | | 0.24 |
| | | | | $TOP_{250}$ | >0.99 |
| | | | | | 0.96 |

**Table 4.7. Statistical comparisons of the ratios given in Table 4.3 using a t-test, a binomial test and a Bayesian binomial test (Appendix G). The numbers given are the p-values associated with the specific tests. Underlined and bold values indicate significant differences ($p<0.05$), values underlined only are marginally significant ($0.05<p<0.1$).**

| Parameters | Separation funnel subsamples | | | | | WHOLE |
|---|---|---|---|---|---|---|
| | $BOTTOM_{500}$ | $BOTTOM_{250}$ | $MIDDLE_{500}$ | $TOP_{500}$ | $TOP_{250}$ | |
| t-test against mean of 1 | | | | | | |
| HPC | 0.137 | 0.063 | 0.076 | 0.420 | 0.176 | 0.186 |
| ENT | 0.106 | 0.681 | 0.644 | 0.179 | 0.224 | **0.010** |
| CP | 0.670 | 0.676 | 0.074 | 0.103 | **0.021** | 0.696 |
| binomial test for ratio > 1 | | | | | | |
| HPC | **0.020** | **<0.001** | **0.020** | 0.340 | **<0.001** | **0.020** |
| ENT | **0.010** | 0.340 | 0.230 | 0.770 | 0.110 | **<0.001** |
| CP | 0.340 | 0.500 | 0.890 | 0.660 | 0.970 | 0.110 |
| binomial test for ratio < 1 | | | | | | |
| HPC | 0.890 | 0.940 | 0.890 | 0.340 | 0.940 | 0.890 |
| ENT | 0.940 | 0.340 | 0.500 | 0.060 | 0.110 | 0.990 |
| CP | 0.340 | 0.190 | **0.020** | 0.110 | **<0.001** | 0.660 |
| t-test on ratios | | | | | | |
| HPC-ENT | 0.386 | **0.019** | 0.095 | 0.164 | **0.038** | 0.484 |
| HPC-CP | 0.134 | 0.070 | **0.011** | 0.084 | **0.015** | 0.307 |
| ENT-CP | 0.215 | 0.830 | 0.064 | 0.318 | 0.055 | 0.372 |
| Bayesian binomial test A>B | | | | | | |
| HPC-ENT | 0.27 | **0.03** | 0.09 | 0.19 | **0.02** | 0.53 |
| HPC-CP | **0.05** | **0.03** | **0.007** | **0.05** | **0.008** | 0.12 |
| ENT-CP | 0.12 | 0.60 | **0.04** | 0.24 | **0.05** | 0.19 |
| Bayesian binomial test A<B | | | | | | |
| HPC-ENT | 0.78 | 0.98 | 0.93 | 0.86 | 0.99 | 0.53 |
| HPC-CP | 0.97 | 0.98 | 0.99 | 0.97 | 1.00 | 0.91 |
| ENT-CP | 0.91 | 0.47 | 0.97 | 0.80 | 0.98 | 0.83 |

**Table 4.8. Estimate ratios derived from the published literature calculated by dividing the concentration measurements of the total bacteria populations to the concentration measurements that do not account for cells present on aggregates. CSO=combined sewer overflow, GAC=globular activated carbon; GAC, Sand and Anthracite relate to the type of filter used to process drinking water and consequently the type of particles found in the samples.**

| References | Organisms | Ratios | Environment | Comments |
|---|---|---|---|---|
| Perdek & Borst (2000) | Fecal coliforms | 5.9 (+/-3.6)[a] | CSO | Average of data in Fig. 2 |
| | Enterococcus | 1.9 (+/-0.7)[a] | CSO | Average of data in Fig. 3 |
| Camper et al. (1985) | *E. coli* | 2.6 | GAC | Table 3, ratio treatment #4/#12 |
| | HPC | 20.1 | GAC | Table 2, ratio treatment #4/#1 |
| Camper et al. (1986) | HPC | 8.6 (50)[b] | GAC | Table 1 |
| | Coliform | 124.3 (1194)[b] | GAC | Table 1, membrane filtration |
| | Coliform | 24.5 (122.2)[b] | GAC | Table 1, MPN |
| Camper et al. (1987) | Coliform | 5 (9)[b] | Sand | Table 3, membrane filtration |
| | Coliform | 43 | Sand | Table 3, MPN |
| | Coliform | 7.7 (20)[b] | Anthracite | Table 3, membrane filtration |
| | Coliform | 7.6 (21)[b] | Anthracite | Table 3, MPN |
| Parker & Darby (1995) | Total coliform | 4.3 (2.8)[a] | Secondary effluent | Table 1, sample 1 & 2 |
| | Fecal coliforms | 1.3 (0.8)[a] | Secondary effluent | Table 1, sample 1 & 3 |
| Pertinsky, Finch & Huck (1997) | HPC | 5.05[c] (151)[b] | GAC | Table 3 |
| Borst & Selvakumar (2003) | Total coliform | 4.8 (1.8)[d] | Stormwater runoff | Table 1 |
| | Fecal coliforms | 27 (7.3)[d] | Stormwater runoff | Table 3 |
| | Fecal streptococcus | 1.15 (6.6)[d] | Stormwater runoff | Table 4 |
| | *E. coli* | 8.5 (0.005)[d] | Stormwater runoff | Table 5 |
| Yoon & Rosson (1990) | DAPI | 1.64 (0.14)[a] | Seawater | Table 2 |
| Bidle & Fletcher (1995) | Acridine orange | 1.5 | Estuary | Table 2 |
| Berman, Rice & Hoff (1988) | Coliforms | 2.8 (5.8)[b] | Primary effluent | Table 4 |
| Palumbo, Ferguson & Rublee (1984) | Acridine orange | 1.03 | Estuary | Table 4 |

[a]+/- standard deviation.

[b]Maximum.

[c]Median.

[d]The two numbers have been measrued with and without (parenthesis) Camper solution.

**Table 4.9. Data table contrasting the CFU concentrations obtained from samples directly measured by the traditional membrane filtration ($C_R$) and those obtained from samples that were pre-filtered though a 5 μm membrane filter to remove the larger particles ($C_F$, filtrate).**

| Station | Date | $C_R$ (CFU/100ml) | | | $C_F$ (CFU/100ml) | | |
|---|---|---|---|---|---|---|---|
| | | HPC x$10^5$ | ENT | CP | HPC x$10^5$ | ENT | CP |
| N | | 22 | 22 | 20 | 22 | 22 | 20 |
| LULU | 07/02/02 | 4.23 | 660 | 1 | 4.15 | 866 | 1 |
| LULU | 07/23/02 | 6.00 | 1260 | 5 | 2.60 | 1440 | 1 |
| LULU | 09/10/02 | 4.17 | 567 | 1 | 3.50 | 300 | 1 |
| LULU | 10/16/02 | 4.97 | 610 | 4 | 4.50 | 595 | 1 |
| LULU | 10/18/02 | 3.93 | 400 | 1 | 3.36 | 410 | 1 |
| LULU | 02/11/03 | 3.20 | 38 | 4 | 1.40 | 30 | 8 |
| KAMO | 06/25/02 | 9.47 | 433 | 6 | 14.40 | 166 | 3 |
| KAMO | 07/23/02 | 6.10 | 477 | 5 | 3.80 | 333 | 1 |
| KANE | 06/12/02 | 15.25 | 4467 | --- | 14.65 | 5800 | 450 |
| KANE | 06/18/02 | 13.30 | 1600 | 200 | 11.80 | 2100 | 266 |
| KANE | 06/25/02 | 9.40 | 933 | 109 | 10.70 | 2230 | 100 |
| KANE | 07/23/02 | 3.83 | 1043 | 90 | 5.70 | 903 | 50 |
| KANE | 09/10/02 | 9.00 | 920 | 210 | 6.60 | 780 | 183 |
| KANE | 02/11/03 | 8.30 | 250 | 81 | 12.75 | 220 | 60 |
| LKS | 06/12/02 | 16.60 | 5200 | --- | 13.40 | 4000 | 260 |
| LKS | 06/18/02 | 8.87 | 1000 | 27 | 2.60 | 1100 | 147 |
| LKS | 06/25/02 | 7.73 | 1133 | 130 | 6.50 | 700 | --- |
| LKS | 07/02/02 | 6.43 | 810 | 54 | 6.43 | 600 | --- |
| LKS | 07/23/02 | 4.47 | 637 | 77 | 4.60 | 590 | 20 |
| LKS | 09/10/02 | 5.50 | 357 | 67 | 4.90 | 285 | 15 |
| LKS | 10/16/02 | 9.07 | 1383 | 73 | 8.40 | 1060 | 35 |
| LKS | 10/18/02 | 7.33 | 760 | 30 | 7.90 | 440 | 20 |

**Table 4.10. Wilcoxon matched pairs test results performed on the data presented in Table 4.9. The lower the p-values, the more likely it is that the parameters originate from different distributions. $C_R$ and $C_F$ cannot be statistically separated.**

| Comparisons ($C_{MF}$ vs. $C_F$) | p-value |
|---|---|
| HPC | 0.105 |
| ENT | 0.291 |
| CP | 0.125 |

**Table 4.11. The ratios ($C_F/C_R$) are obtained from the data in Table 4.9. Ratios lower than 1 indicate that aggregates > 5 μm contributing colonies have been removed from the sample.**

| Stations | Date | Ratio ($C_F/C_R$) | | |
|---|---|---|---|---|
| | | HPC | ENT | CP |
| LULU | 07/02/02 | 0.98 | 1.31 | 1.00 |
| LULU | 07/23/02 | 0.43 | 1.14 | 0.20 |
| LULU | 09/10/02 | 0.84 | 0.53 | 1.00 |
| LULU | 10/16/02 | 0.91 | 0.98 | 0.25 |
| LULU | 10/18/02 | 0.85 | 1.03 | 1.00 |
| LULU | 02/11/03 | 0.44 | 0.79 | 2.00 |
| KAMO | 06/25/02 | 1.52 | 0.38 | 0.50 |
| KAMO | 07/23/02 | 0.62 | 0.70 | 0.20 |
| KANE | 06/12/02 | 0.96 | 1.30 | --- |
| KANE | 06/18/02 | 0.89 | 1.31 | 1.33 |
| KANE | 06/25/02 | 1.14 | 2.39 | 0.92 |
| KANE | 07/23/02 | 1.49 | 0.87 | 0.56 |
| KANE | 09/10/02 | 0.73 | 0.85 | 0.87 |
| KANE | 02/11/03 | 1.54 | 0.88 | 0.74 |
| LKS | 06/12/02 | 0.81 | 0.77 | --- |
| LKS | 06/18/02 | 0.29 | 1.10 | 5.44[c] |
| LKS | 06/25/02 | 0.84 | 0.62 | --- |
| LKS | 07/02/02 | 1.00 | 0.74 | --- |
| LKS | 07/23/02 | 1.03 | 0.93 | 0.26 |
| LKS | 09/10/02 | 0.89 | 0.80 | 0.22 |
| LKS | 10/16/02 | 0.93 | 0.77 | 0.48 |
| LKS | 10/18/02 | 1.08 | 0.58 | 0.67 |
| Normality[a] | | Yes (0.146) | No (<0.001) | No[c] (<0.001) |
| Mean | | 0.92 | 0.94 | 0.98[c] |
| Min | | 0.29 | 0.38 | 0.20 |
| Max | | 1.54 | 2.39 | 5.44[c] |
| Mean = 1[b] | | 0.247 | 0.521 | 0.945[c] |
| Median | | 0.90 | 0.86 | 0.70[c] |

[a]Tested with the Shapiro-Wilk W test.

[b]Using the parametric t-test against a mean of 1, displayed are the p-values.

[c]By removing the 5.44 outlier, the data approach normality (p=0.046), the mean decreases to 0.72 and becomes significantly smaller than 1 (p=0.027), and the median is 0.67.

**Table 4.12. Wilcoxon matched pairs test comparing the ratios (Table 4.11) obtained for the different organisms. Small p-values support the hypothesis that the two distributions compared are different. HPC and ENT yields similar ratios, but the ratios corresponding to CP are likely smaller than those for either HPC or ENT.**

| Comparisons | N | p-value |
|---|---|---|
| HPC - ENT | 22 | 0.808 |
| HPC - CP | 18[a] | 0.170[a] |
| ENT - CP | 18[a] | 0.199[a] |

[a]Without the outlier (CP ratio = 5.44), N=17, the p-value for the HPC-CP comparison is 0.055 and that for the ENT-CP pair is 0.068.

**Table 4.13. Data table of the measurements of $C_R$, $C_A$ and $C_H$ (Figure 4.2) performed on samples collected from station KANE.**

| Date | HPC x $10^5$ (CFU/100ml) | | | ENT (CFU/100ml) | | | CP (CFU/100ml) | | |
|---|---|---|---|---|---|---|---|---|---|
| | $C_R$[c] | $C_A$[d] | $C_H$[e] | $C_R$ | $C_A$ | $C_H$ | $C_R$ | $C_A$ | $C_H$ |
| 03/18/03 | 5.40 | 0.74 | 0.39 | 5000 | 101 | 80 | 60 | 10 | 10 |
| 03/25/03 | 5.50 | 0.75 | 0.96 | 370 | 5 | 27 | 73 | 9 | 14 |
| 04/01/03 | 5.55 | 0.33 | 0.35 | 520 | 9 | 37 | 170 | 5 | 6 |
| 04/08/03 | 5.60 | 0.36 | 0.52 | 650 | 12 | 59 | 41 | 3 | 2 |
| 04/24/03 | 13.75 | 2.02 | 2.74 | 485 | 15 | 40 | 51 | 8 | 7 |
| 05/08/03 | 9.00 | 1.94 | 2.85 | 675 | 28 | 63 | 89 | 21 | 20 |
| 05/14/03 | 4.50 | 1.10 | 1.08 | 445 | 36 | 49 | 74 | 17 | 21 |
| 05/21/03 | 3.65 | 0.81 | 1.23 | 300 | 29 | 47 | 335 | 20 | 21 |
| 05/29/03 | 3.75 | 0.89 | 0.65 | 255 | 12 | 30 | 96 | 28 | 25 |
| 06/06/03 | 5.85 | 1.61 | 1.67 | 820 | 93 | 136 | 136 | 55 | 76 |
| 06/13/03 | 7.05 | 1.89 | 1.88 | 595 | 120 | 130 | 435 | 110 | 136 |
| 01/27/04 | 3.70 | 0.15 | 0.50 | 760 | 11 | 15 | 34 | 2 | 1 |
| 05/26/04 | 37.80 | 1.84 | 2.01 | 775 | 100 | 75 | 35 | 8 | 13 |
| 06/02/04 | 22.10 | 2.62 | 3.18 | 4433 | 152 | 208 | 180 | 21 | 34 |
| 06/16/04 | 10.70 | 1.84 | 2.17 | 790 | 107 | 221 | 106 | 29 | 39 |
| 06/18/04 | 23.33 | 1.10 | 1.68 | 803 | 90 | 192 | 53 | 28 | 37 |
| 06/23/04 | 9.63 | 1.99 | 1.84 | 1057 | 58 | 116 | 138 | 39 | 14 |
| Normality[a] | <0.001 | 0.278 | 0.285 | <0.001 | 0.026 | 0.022 | <0.001 | <0.001 | <0.001 |
| Mean | 10.40 | 1.29 | 1.51 | 1,102 | 57 | 90 | 124 | 24 | 28 |
| Median | 5.85 | 1.10 | 1.67 | 675 | 36 | 63 | 89 | 20 | 20 |
| Min | 3.65 | 0.15 | 0.35 | 255 | 5 | 15 | 34 | 2 | 1 |
| Max | 37.80 | 2.62 | 3.18 | 5000 | 152 | 221 | 435 | 110 | 136 |
| Wicoxon[b] | | 0.022 | | | 0.003 | | | 0.130 | |

[a]Normality was assessed with the Shapiro-Wilk test. Expressed are the p-values of the tests. Small p-value suggest that the distribution is not normal.

[b]The results from the Wilcoxon matched pair test suggest that the distributions for $C_A$ and $C_H$ for all organisms are different (low p-values).

[c]$C_R$ are the CFU concentrations measured by the traditional membrane filtration method.

[d]$C_A$ represents the concentration of colonies that form on agar plate that originate from aggregates.

[e]$C_H$ is the total CFU concentration present on the aggregates.

**Table 4.14. Ratios of the data presented in Table 4.13. The ratio $C_A/C_R$ represent the fraction of the colonies growing subsequent to a traditional membrane filtration measurement that originate from aggregates. The fraction $C_H/C_A$ indicates the average number of CFU recoverable from aggregates.**

| Date | $C_A/C_R$ | | | $C_H/C_A$ | | |
|---|---|---|---|---|---|---|
| | HPC | ENT | CP | HPC | ENT | CP |
| 03/18/03 | 0.14 | 0.02 | 0.17 | 0.53 | 0.79 | 0.98 |
| 03/25/03 | 0.14 | 0.01 | 0.12 | 1.28 | 5.40 | 1.58 |
| 04/01/03 | 0.06 | 0.02 | 0.03 | 1.08 | 4.23 | 1.14 |
| 04/08/03 | 0.06 | 0.02 | 0.07 | 1.43 | 5.02 | 0.63 |
| 04/24/03 | 0.15 | 0.03 | 0.16 | 1.36 | 2.67 | 0.88 |
| 05/08/03 | 0.22 | 0.04 | 0.23 | 1.47 | 2.24 | 0.99 |
| 05/14/03 | 0.24 | 0.08 | 0.23 | 0.98 | 1.37 | 1.27 |
| 05/21/03 | 0.22 | 0.10 | 0.06 | 1.53 | 1.60 | 1.02 |
| 05/29/03 | 0.24 | 0.05 | 0.29 | 0.73 | 2.50 | 0.92 |
| 06/06/03 | 0.27 | 0.11 | 0.41 | 1.04 | 1.46 | 1.37 |
| 06/13/03 | 0.27 | 0.20 | 0.25 | 0.99 | 1.08 | 1.24 |
| 01/27/04 | 0.04 | 0.01 | 0.05 | 3.45 | 1.43 | 0.57 |
| 05/26/04 | 0.05 | 0.13 | 0.22 | 1.10 | 0.75 | 1.67 |
| 06/02/04 | 0.12 | 0.03 | 0.11 | 1.22 | 1.37 | 1.68 |
| 06/16/04 | 0.17 | 0.14 | 0.27 | 1.18 | 2.07 | 1.34 |
| 06/18/04 | 0.05 | 0.11 | 0.53 | 1.52 | 2.13 | 1.31 |
| 06/23/04 | 0.21 | 0.05 | 0.28 | 0.92 | 2.02 | 0.36 |
| Normality[a] | 0.123 | 0.030 | 0.237 | 0.001 | 0.006 | 0.757 |
| Mean | 0.16 | 0.07 | 0.21 | 1.28 | 2.24 | 1.11 |
| Median | 0.15 | 0.05 | 0.22 | 1.18 | 2.02 | 1.14 |
| Min | 0.04 | 0.01 | 0.03 | 0.53 | 0.75 | 0.36 |
| Max | 0.27 | 0.20 | 0.53 | 3.45 | 5.40 | 1.68 |

[a]Normality was assessed with the Shapiro-Wilk test. Expressed are the p-values of the tests. Small p-value suggest that the distribution is not normal.

**Table 4.15. A literature review summarizing measurements of the number of bacteria per particle and the fraction of particles with associated bacteria.**

| Reference | Organisms | Number of bacteria per particle | Fraction of particles with bacteria | Environment | Comments |
|---|---|---|---|---|---|
| Kirchman & Mitchell (1982) | Acridine orange | 7.7 | 24-62% | Brackish marsh | Table 4 |
| | Acridine orange | 4.9 | 35-87% | Salt marsh | Table 4 |
| Berman, Rice & Hoff (1988) | HPC | 1-24.5 | | | Table 4 |
| Ridgeway & Olson (1981) | SEM | 10-100 | <17% | Drinking water | Most particles with attched bacteria size between 10-50 μm |
| Ridgeway & Olson (1982) | SEM | 5-10 to>100 | 1% | Drinking water | Particles with attached bacteria are >10 μm |
| Stewart, Wolfe & Means (1990) | HPC | 0-7 | | GAC | |
| | SEM | 8% of the particles have 0 cells<br>77% of the particles have 1-50 cells<br>7% of the particles have 51-100 cells<br>8% of the particles have >100 cells | | | |
| Stringfellow, Mallon & DiGiano (1993) | HPC | 15-24 | 7% | GAC | Calculated from particles count and total attached bacteria |
| Emerick et al. (1999) | Coliforms | | 0-32% | Various wastewater treatment process | Table 2 |

**Table 4.16. Ratios (Ψ) of the total number of CFU in a sample to the number of CFU measured by the routine membrane filtration method. These data can be compared with the data presented in Table 4.3, which show ratios greater than the ones presented here, presumably due to the presence of a large number of small (<5 μm) clusters of bacteria.**

| Date | $\Psi=(C_R+C_H-C_A)/C_R$ | | |
|---|---|---|---|
| | HPC | ENT | CP |
| 03/18/03 | 0.935 | 0.996 | 0.996 |
| 03/25/03 | 1.038 | 1.059 | 1.071 |
| 04/01/03 | 1.005 | 1.054 | 1.004 |
| 04/08/03 | 1.028 | 1.073 | 0.973 |
| 04/24/03 | 1.053 | 1.052 | 0.980 |
| 05/08/03 | 1.101 | 1.052 | 0.997 |
| 05/14/03 | 0.996 | 1.030 | 1.061 |
| 05/21/03 | 1.116 | 1.058 | 1.001 |
| 05/29/03 | 0.936 | 1.071 | 0.976 |
| 06/06/03 | 1.011 | 1.052 | 1.152 |
| 06/13/03 | 0.998 | 1.017 | 1.060 |
| 01/27/04 | 1.096 | 1.006 | 0.978 |
| 05/26/04 | 1.005 | 0.968 | 1.145 |
| 06/02/04 | 1.026 | 1.013 | 1.077 |
| 06/16/04 | 1.031 | 1.144 | 1.094 |
| 06/18/04 | 1.025 | 1.127 | 1.166 |
| 06/23/04 | 0.984 | 1.056 | 0.821 |
| Normality[a] | 0.347 | 0.308 | 0.162 |
| Mean | 1.023 | 1.049 | 1.033 |
| Median | 1.025 | 1.052 | 1.004 |
| Min | 0.935 | 0.968 | 0.821 |
| Max | 1.116 | 1.144 | 1.166 |
| t-test vs. 1[b] | 0.083 | <0.001 | 0.133 |

[a]Shapiro-Wilk test. Significant p-values (<0.05) suggest a non-normal distribution.

[b]t-test of means against a constant of 1. Low p-values suggest a difference bewteen the mean and 1.

**Table 4.17. S, the fraction of the excess (aggregated) bacteria not accounted for by the routine membrane filtration that originate from aggregates smaller than 5 μm, calculated with a Monte Carlo method using the data means and standard deviation calculated from Table 4.3 and Table 4.16.**

| Organism | S median |
|---|---|
| HPC | 0.98 |
| ENT | 0.85 |
| CP | 0.99 |

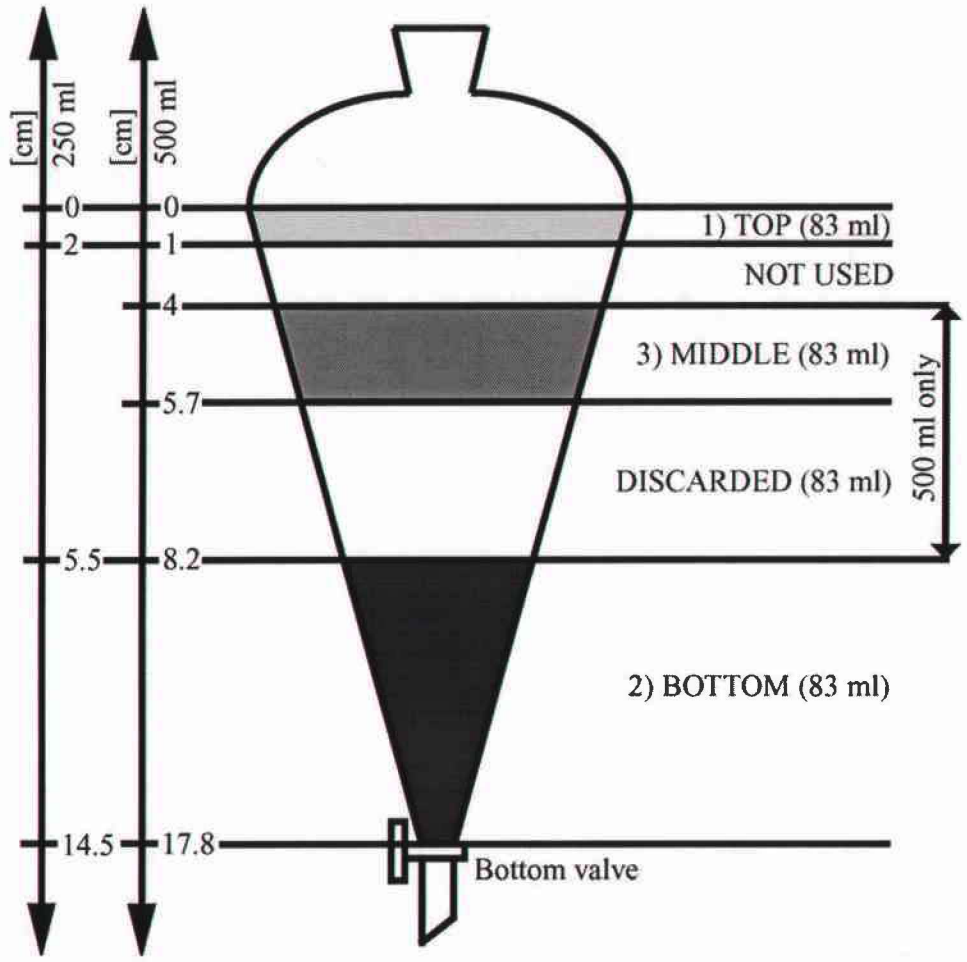


**Figure 4.1. Illustration of the different subsamples taken from the separation funnel and used in the particle settling experiment. After 4 hours undisturbed, the top 83 ml were pipetted off from the 3 funnels of identical volumes (TOP). The bottom 83 ml of each funnels (BOTTOM) were collected using the bottom valve. An additional 83 ml was drained through the valve and the next 83 ml were collected (MIDDLE). The values on the left scale bars show the vertical distances spanned by each subsample in the 250 ml and the 500 ml funnels.**

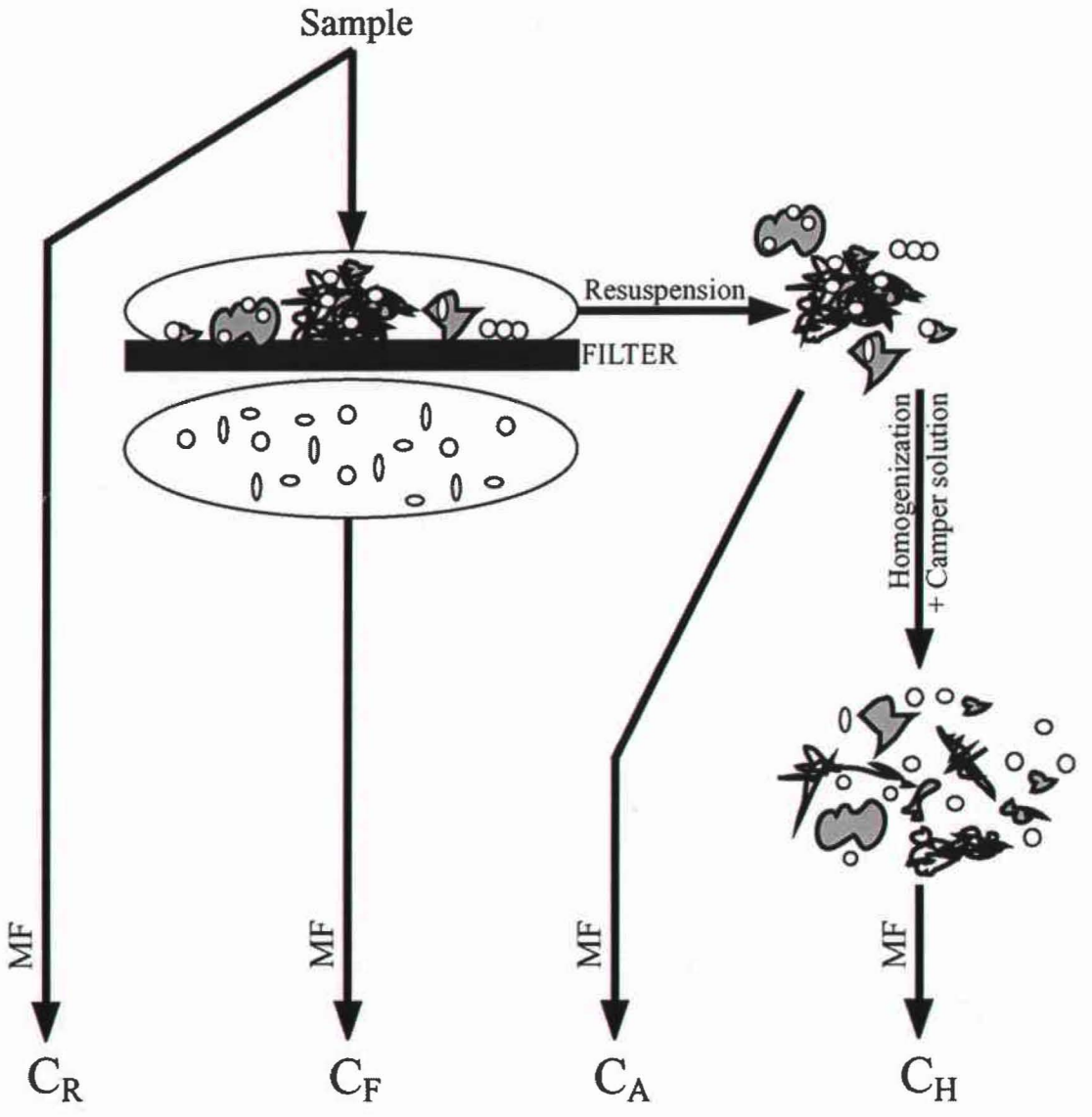


**Figure 4.2. The concentration of cells in the sample is directly assayed by the routine membrane filtration (MF) techniques to yield $C_R$. The remaining volume of the sample is passed through a membrane filter (pore size 5 μm) to separate (operationally) free-living cells (filtrate) from the cells associated to aggregates (retentate). The filtrate is assayed by the MF methods to produce $C_F$, the CFU concentration in the filtrate. The retentate is either resuspended from the filter into phosphate buffer and assayed directly with the MF approach to produce a measurement of the number of aggregates producing colonies ($C_A$), or it is resuspended into Camper solution and homogenized to separate the bacteria from the aggregates. After homogenization, the solution is assayed by the MF methods. This treatment yields an estimate of the sample average number of bacteria associated with aggregates larger than 5 μm ($C_H$). Because the measurement of $C_A$ and $C_H$ required the filtration of large volumes of water through the 5 μm filter, multiple filters were used and the resuspended retentate for each filter was pooled before assay or treatment.**

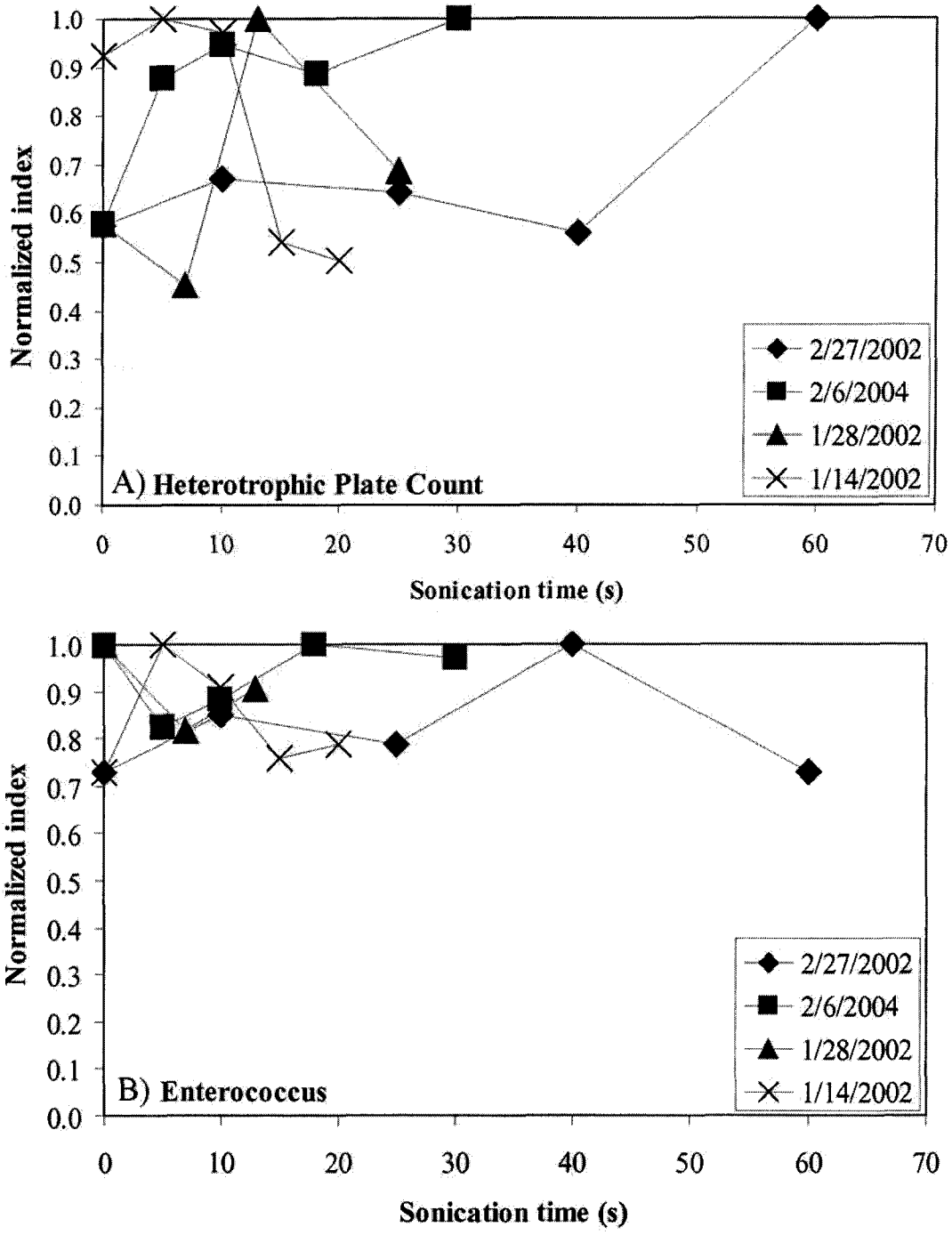


**Figure 4.3. Sonication does not significantly increase the number of HPC bacteria (A) or enterococci (B) from environmental samples. The normalized index refers to the value obtained by dividing the concentrations obtained by the maximum concentration on the given sampling date.**

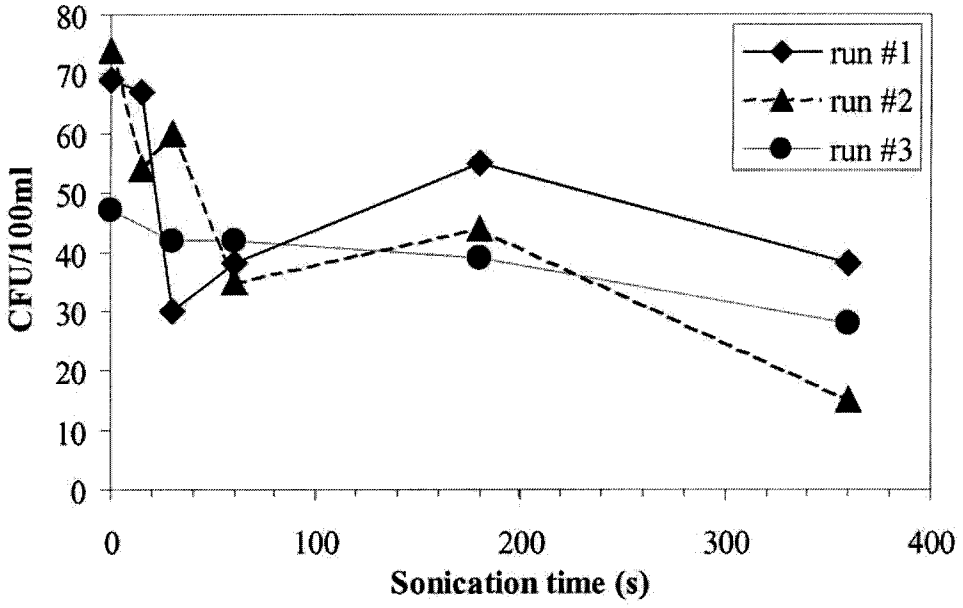


**Figure 4.4. The number of positive colonies recovered on mE agar from a diluted pure culture of *E. faecalis* decreases with increasing sonication time.**

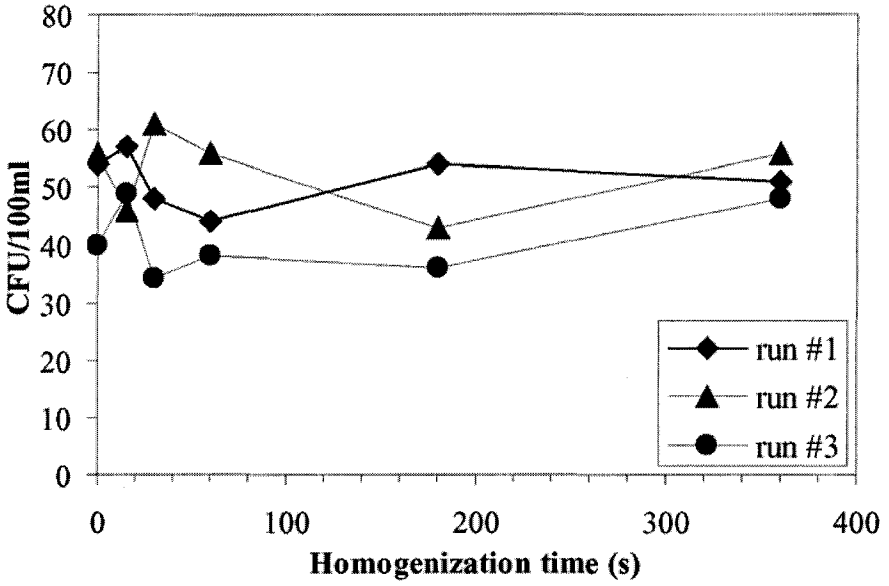


**Figure 4.5. The use of high-speed homogenization (16000rpm) has no measurable negative effects on a diluted pure culture of *E. faecalis*.**

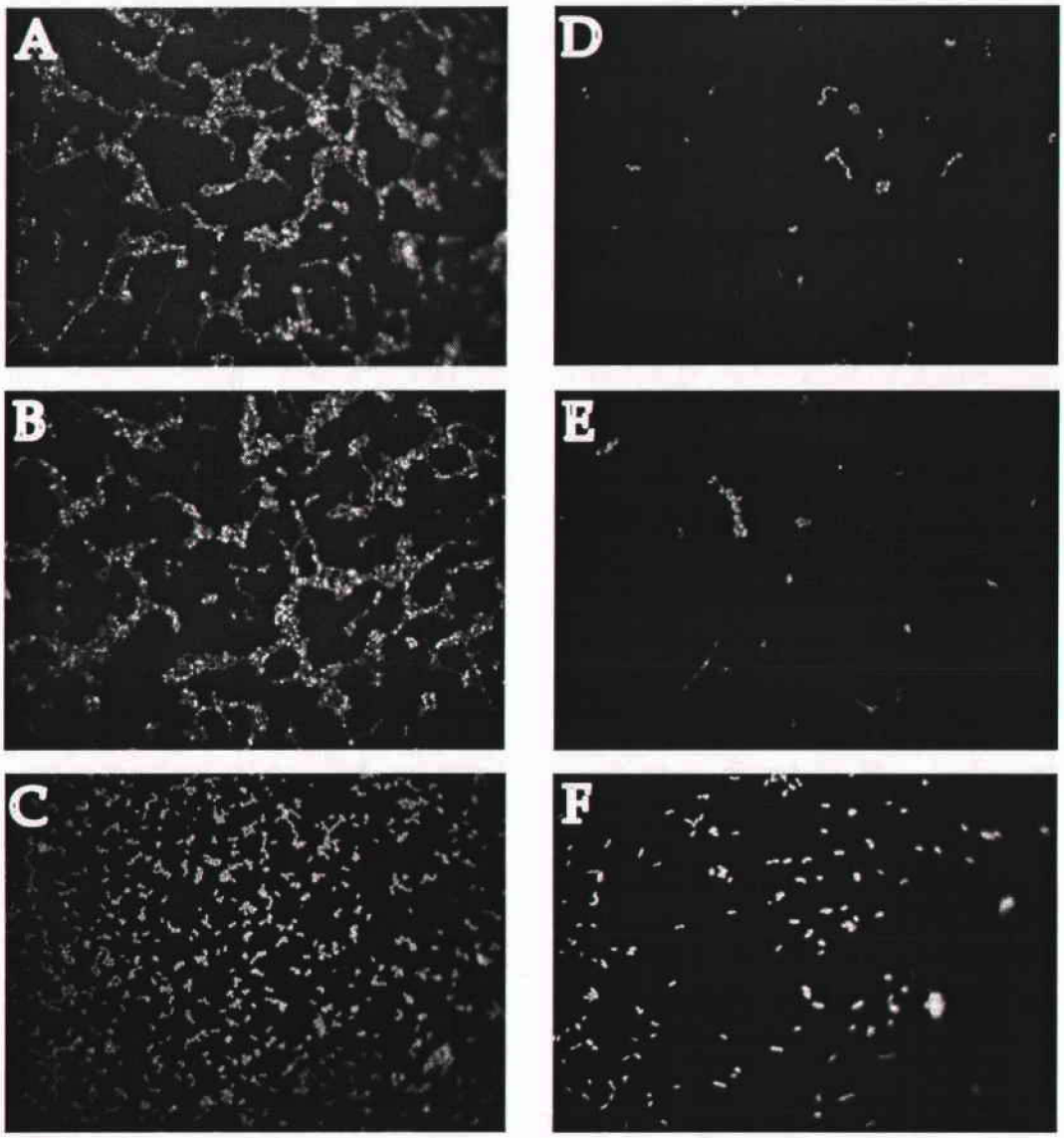


**Figure 4.6. Epifluorescence microscope images of different pure cultures stained with DAPI, magnification 1000x. A & B) *E. faecalis*, diluted 100x from stock culture; C) unknown environmental isolate (#1), diluted 100x from stock culture; D & E) *E. faecalis*, diluted 1000x from stock culture, F) unknown environmental isolate (#3) diluted 100x from stock culture.**

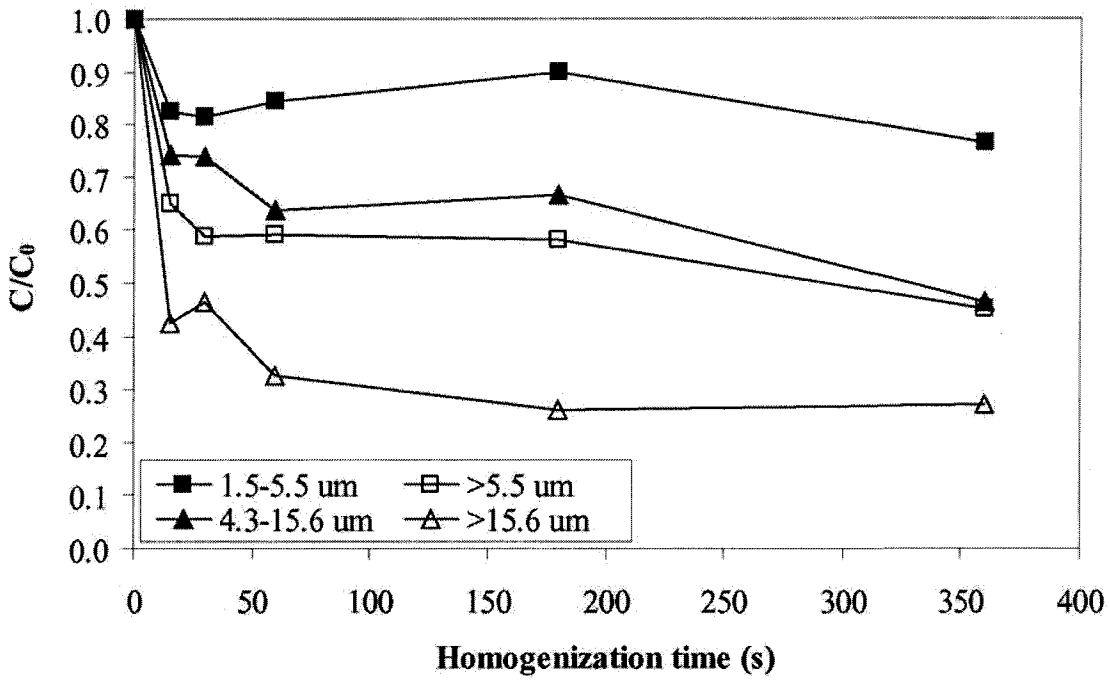


**Figure 4.7. Effect of homogenization time on the populations of particles of different sizes measured with a Coulter counter in a 0.9% NaCl solution. C represents the concentration of particles in the sample and $C_0$ is the non-homogenized concentration of particles of the given size class.**

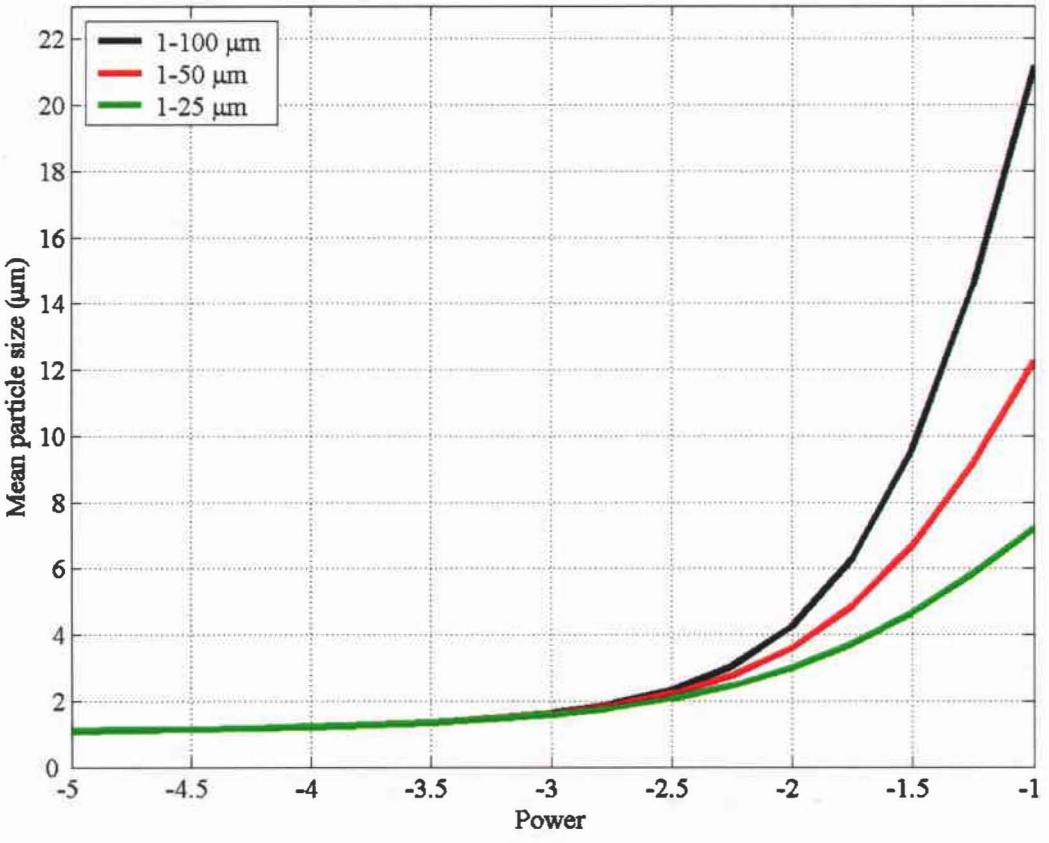


**Figure 4.8. Evolution of the mean particle size defined by the weighted average of different particle size distributions given by power laws with powers ranging from –1 to -5 and different size ranges. Homogenization reduces the number of large particles, which limits the maximum particle size of the distributions and decreases its power. The mean particle size of the different distributions varies little.**

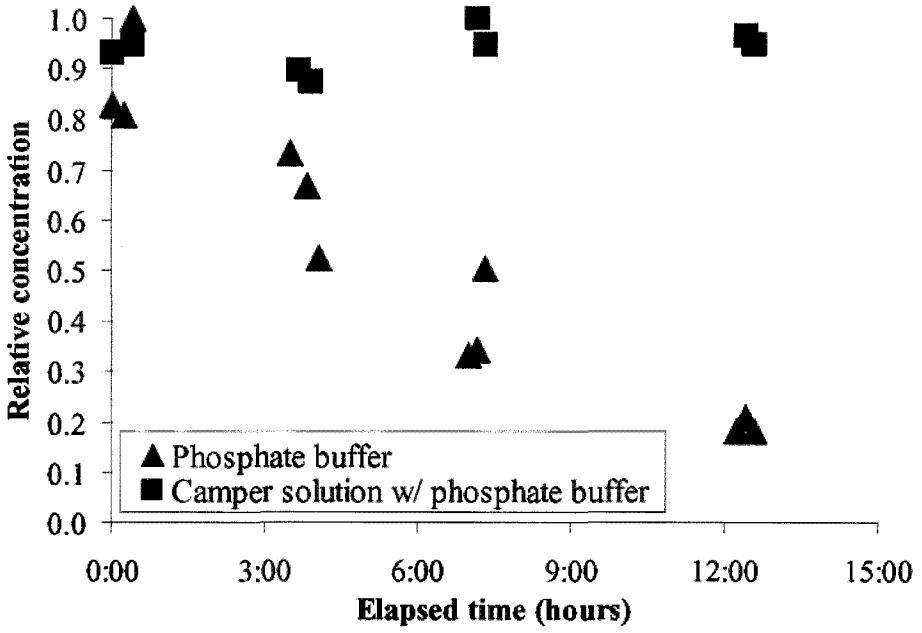


**Figure 4.9. In the long term, a mixture of Camper solution and phosphate buffer is a better diluent than the recommended phosphate buffer (APHA, 1998). The relative concentration refers to the concentration measured at a given time normalized to the concentration measured at the time of inoculation.**

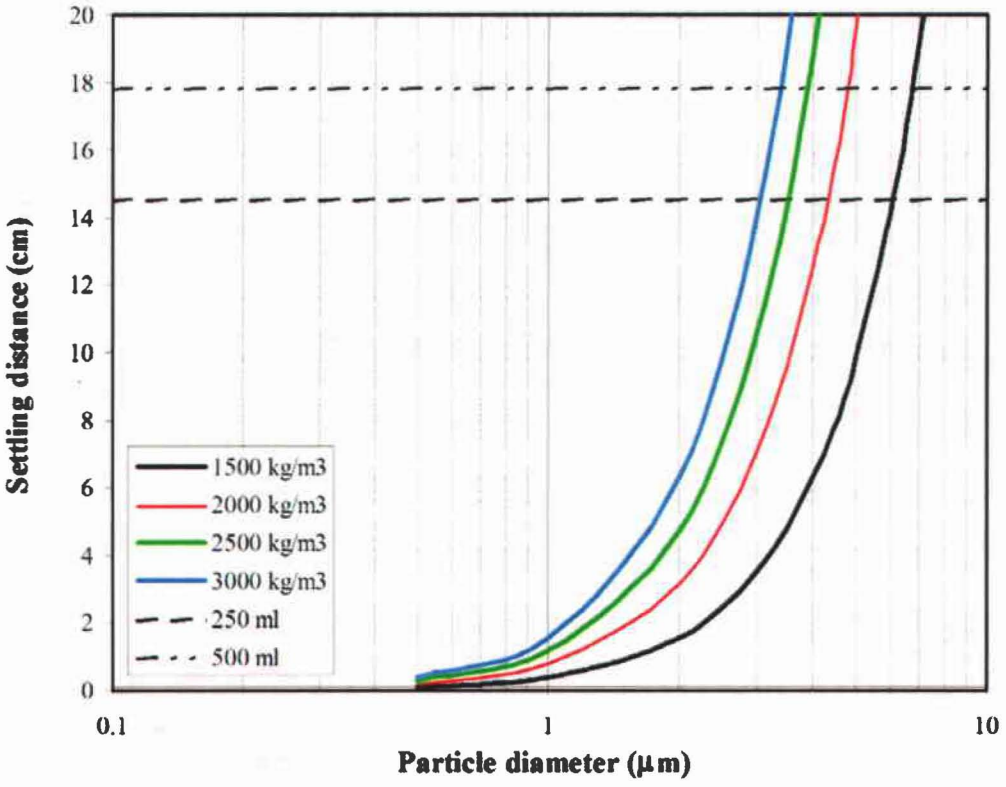


**Figure 4.10. Settling distances that particles in water can achieve after 4 hours assuming the particles settle with a terminal velocity predictable with Stokes Law. The terminal velocities were calculated using an online tool at http://www.filtration-and-separation.com/settling/settling.htm. Even small particles with densities close to that of water may settle significant distances in the separation funnel (Figure 4.1). The 250 ml and 500 ml lines correspond to the height of the water in the specific separation funnels.**

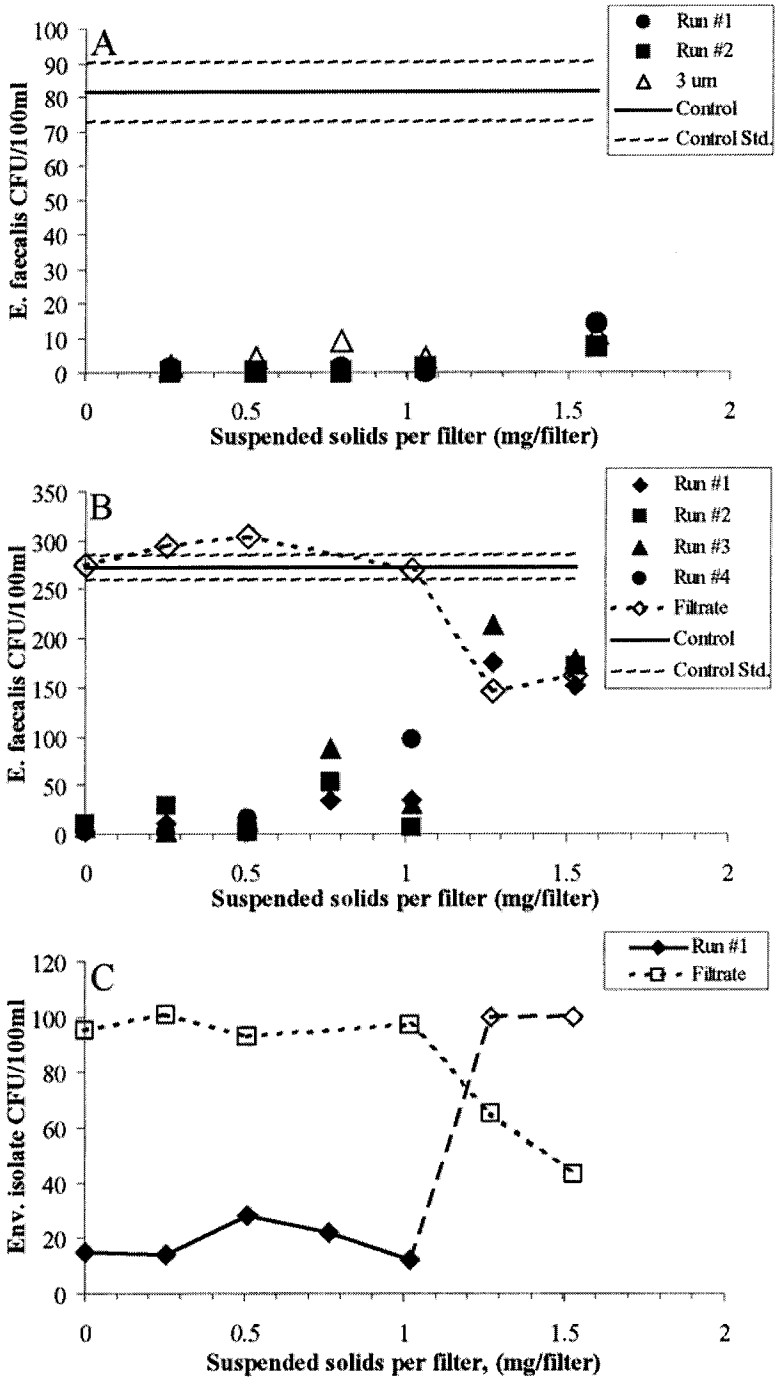


**Figure 4.11. Experiments performed to measure the extent of "by-catch" of free-living bacteria depending on the particle loading on the filter (5 μm pore size). A) Low concentrations (~80 CFU/100ml) of *E. faecalis* added to mixtures of Manoa Stream waters and sterile phosphate buffer. B) Seeding of Kaneohe Stream samples (station KANE) with larger concentration (~275 CFU/100ml) of *E. faecalis* and C) with a Gram negative environmental isolate (~100 CFU/100ml). In panel C, the two plates with the largest load of suspended solids (empty diamonds with dashed line) were recorded as too numerous to count; consequently, these data points were assigned the same value as the maximum concentration of the filtrate.**

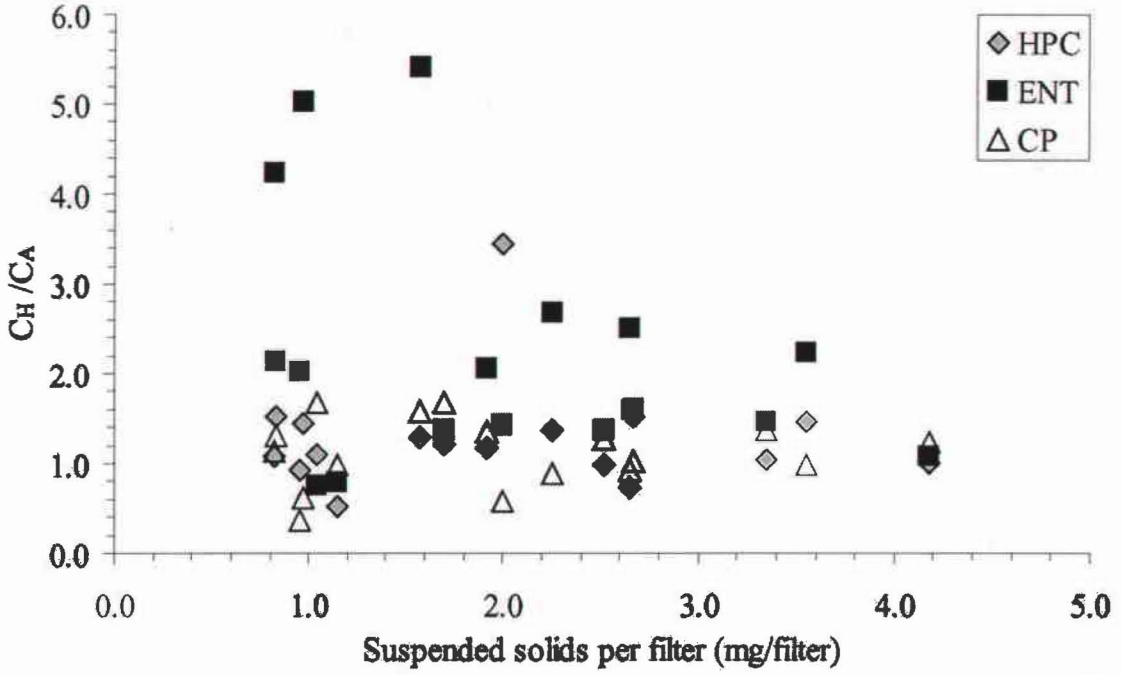


**Figure 4.12. The ratio $C_H/C_A$ approaches 1 as the suspended solids load on the filter used to separate free-living bacteria from the particle-associated population increases. The ratios of HPC and CP are indistinguishable from 1. This is also the case for the ENT ratios corresponding to filter loads >1.6 mg/filter. Ratios associated with filter loads <1.6 mg/filter are significantly greater than 1.**

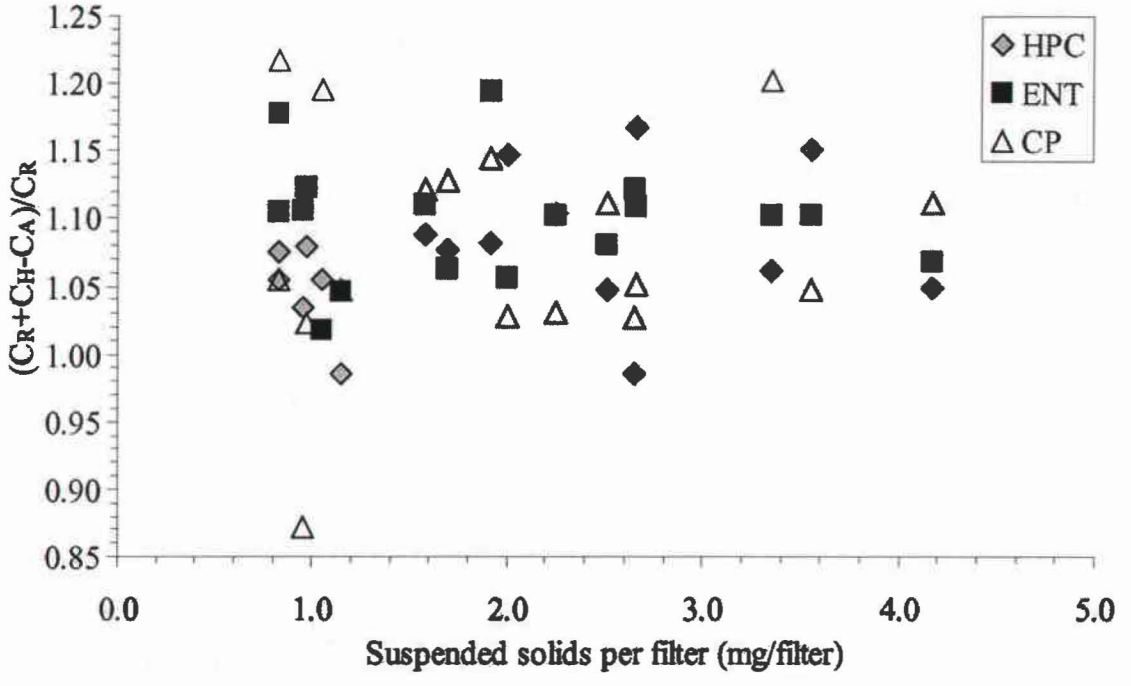


**Figure 4.13. The ratio (Ψ) of the total concentration of cells estimated by membrane filtration using pre-filtration through a 5 μm filter and homogenization to the routine membrane filtration measurements do not vary with the suspended solids load on the filters.**

# Chapter 5 : CONCLUSIONS

*"Science itself is recognizing that its own methods are ultimately no more objective that those of the arts." (Howard Felperin, from Wilber (1998, p31))*

1) Data on the concentrations of heterotrophic plate count (HPC) bacteria, enterococcus, *C. perfringens* and total suspended solids obtained from 4 stations located within the Kaneohe watershed suggest the presence of a source of *C. perfringens* between station KAMO and KANE. The type of source (point versus non-point) cannot be unambiguously determined. Circumstantial evidence and support from the literature suggest the presence of a point source, however.
2) The species composition of the positive colonies growing on enterococcus specific mE agar shows that a majority of the species recovered from stream water samples have low hygienic significance and are likely part of the natural soil microbial community. Variations in the species composition can be related to the location of the sampling station in the watershed, and allows for inferences on the type of microbial pollution present at the station. These data do not support reliance on enterococcus concentrations, measured by the mE/EIA membrane filtration assay, for water quality assessment of Kaneohe stream, and possibly Hawaiian streams generally.
3) The measurement errors associated with the membrane filtration measurements due to the presence of multiple cells per aggregate can be substantial. The results of the theoretical analysis shows that the formation of clusters upon filtration of a randomly distributed monodisperse population can, depending on the sizes of the

colonies, generate counting errors between 0 and 20%, but this error can increase drastically when the number of large colonies increases.

4) The presence of aggregates with multiple associated cells generates errors (underestimates) that are typically between 0-35% of the "true" cell concentration. The error is highly dependant on the function assumed to describe the cluster size distribution.

5) The errors induced by the clumping of bacteria upon filtration or by the presence in the sample of aggregates with multiple associated cells do not disappear with replication. Replication simply allows for these errors to fall within a certain (predictable) range.

6) Empirical data suggest that the error induced by the inability of the membrane filtration methods to account for all but a single cell present on aggregates agrees well with the theoretical prediction. Indeed, when whole sample volumes are subjected to homogenization with Camper solution, a protocol that has been shown to separate cells from aggregates, the mean colony counts increase on average by 9%, 28% and 52% for *C. perfringens*, enterococci, and HPC, respectively (Table 4.3) relative to untreated control samples. The sample specific increase, however, is very variable; the maximum increase observed was 440%.

7) When the particles greater than 5 μm are isolated (by filtration and resuspension), and the bacteria content of these larger aggregates is analyzed, the relative increase in colony counts achieved (2 to 5%) relative to the untreated controls is smaller than the values obtained when treated whole samples are assayed. The dichotomy between the two data sets (treated whole samples and particulate

fraction) suggests that on average 84-99% of the increase observed in the whole samples subjected to homogenization and Camper solution is attributable to small aggregates (<5 μm).

8) Because the counts of *C. perfringens* did not increase significantly upon homogenization and because very few *C. perfringens* cells were found associated with aggregates, determinations of *C. perfringens* concentrations by membrane filtration are not influenced by the inability of the culture methods to account for more than one cell per aggregate. HPC and enterococcus measurements, however, are both underestimating the potential colony forming units of the samples because both types of organisms can be found associated with aggregates.
9) Even if in rare instances the problem of particle-association of fecal indicator bacteria may be substantial, water quality assessment relies on the correlation of the number of colony-forming units, not cell number, to a measured rate of illness. The fact that the number of colony-forming units does not correspond to the number of cells in a sample has no effect assuming that the error (underestimation) is constant and systematic. In other words, the potential for error in the risk assessment phase induced by the presence of particle-associated cells does not lay in the magnitude of this error but in the variability of this error in time and space. Further study should be dedicated to investigate this variability.
10) The reader should keep in mind that other factors besides particle-association are able to induce measurement error, for example the physiological state of the cells, the variability in the quality of the media and, as discussed in Chapter 2, the composition of the microbial community.

# APPENDICES

## *Appendix A: Derivation of the most probable number estimate μ for the case k=1.*

It is difficult, starting from equation (3.8), to calculate $\frac{dL}{d\bar{\mu}} = 0$. Instead, it is easier to compute $\frac{d(-\ln(L))}{d\bar{\mu}} = 0$, thus minimizing –ln(L) instead of L. We can start by noticing that $-\ln(n_1 \cdot n_2 \cdot \ldots \cdot n_k) = -\sum_{i=1}^{k} \ln(n_i)$. Using this relationship on (3.8), we obtain:

$$
\begin{aligned}
-\ln(L) &= \sum_{i=1}^{k} \ln\left[\frac{n_i}{p_i!(p_i - n_i)}[1-\exp(-\bar{\mu}V_i)]^{p_i}[\exp(-\bar{\mu}V_i)]^{n_i-p_i}\right] \\
&= \sum_{i=1}^{k} \ln\frac{n_i}{p_i!(p_i-n_i)} + \sum_{i=1}^{k} \ln\left[1-\exp(-\bar{\mu}V_i)\right]^{p_i} + \sum_{i=1}^{k} \ln\left[\exp(-\bar{\mu}V_i)\right]^{n_i-p_i}. \\
&= C + \sum_{i=1}^{k} p_i \ln\left[1-\exp(-\bar{\mu}V_i)\right] + \sum_{i=1}^{k} (n_i - p_i)\ln\left[\exp(-\bar{\mu}V_i)\right]
\end{aligned}
$$

Now, with k=1, $-\ln(L') = p\ln\left[1-\exp(-\bar{\mu}V)\right] + (n-p)(-\bar{\mu}V)$.

The differentiation is as follows:

$$
\begin{aligned}
\frac{d(-\ln(L'))}{d\bar{\mu}} &= p\frac{d}{d\bar{\mu}}\left(\ln[1-\exp(-\bar{\mu}V)]\right) + \frac{d}{d\bar{\mu}}\left(p\bar{\mu}V - n\bar{\mu}V\right) \\
&= p\frac{1}{1-\exp(-\bar{\mu}V)}\frac{d}{d\bar{\mu}}(1-\exp(-\bar{\mu}V) + (p-n)V \\
&= \frac{pV\exp(-\bar{\mu}V)}{1-\exp(-\bar{\mu}V)} + (p-n)V = 0
\end{aligned}
$$

$$
\begin{aligned}
p\exp(-\bar{\mu}V) &= n - n\exp(-\bar{\mu}V) - p + p(\exp(-\bar{\mu}V) \\
&= (n-p) + (p-n)\exp(-\bar{\mu}V)
\end{aligned}
$$

$$\exp(-\bar{\mu}V)(p-p+n)=n-p$$

$$\bar{\mu}=-\frac{1}{V}\ln\left(\frac{n-p}{n}\right)$$

## *Appendix B: Mfchance_2.m for Matlab*

```
%MFchance_2.m
%++++++++++++++++++++++++++++++++++++++++++++++++++++++++
%Written by Yves Plancherel (3/2004) as MFchance.m
%Corrected and modified to be MFchance.m V.2 (8/04)
%to account for circular filters
%++++++++++++++++++++++++++++++++++++++++++++++++++++++++
%The purpose of MFchance.m is to simulate the filtration of bacteria to
%look at the role of chance in creating clusters on the filter.
%It monitors the number of clusters, their size as well as other
%parameters such as number of single cells, total cells and the total
%number of colony forming units that form.

%The adjustable parameters are:
%  D, the filter diameter
%  r, the critical radius of a colony
%  nb, the concentration of microbes
%  nmax, the number of runs of the program
%++++++++++++++++++++++++++++++++++++++++++++++++++++++++
clear all
%set initial conditions:
D=43;  %47mm is the diameter of the membrane filters used -2mm each side
     %hidden by filter tower
r=1  % r is the critical colony radius
nb=[10:10:50,70:30:190];   %the concentration of microbes
     %There is no limit as to the number of concentrations in nb.
nmax=1000;  %The number of runs that the program will go through.
       %This is equivalent to the number of filters analyzed.

%Begining of the analysis
  for lnb=1:length(nb);    %Determines the number of concentration
                  %that the program will try
    N=nb(lnb);  %Number of bacteria in the virtual sample

    %Build a symetric matrix with 0 and 1.
    in=ones(N,N);
    for line=1:N;
      for col=1:N;
        if in(line,col)==in(col,line)
          in(line,col)=1;
          in(col,line)=0;
        end
      end
    end %of making "in" matrix
```

```
%Begining of the filter analysis
for n=1:nmax;

  clear V %Initialize V, used to count
  V=[];  %the number of bacteria/cluster

  %Instead of set the r value for all runs, it is
  %possible to chose r randomly between some set
  %values:
  %to generate a uniform r distrubution
  %ra=0.1;rb=1.5;
  %r=ra+(rb-ra)*rand(1);
  %to generate a normal r distrubution with only positive values
  %r=1+sqrt(0.5)*randn(1);
  %while r<0
  %r=1+sqrt(0.5)*randn(1);
  %end

  %produes randomly generated points in a circle of diameter D
  a=-D/2;
  b=D/2;
  for i=1:N
    x(i)=a+(b-a)*rand(1);
    y(i)=a+(b-a)*rand(1);
    t=x(i)^2+y(i)^2;
    while t>(D/2)^2;
      x(i)=a+(b-a)*rand(1);
      y(i)=a+(b-a)*rand(1);
      t=x(i)^2+y(i)^2;
    end
  end

  X=[x' y'];
  Y=pdist(X);    %caluclates the distances
  sY=squareform(Y);
  md=sY.*in;     %removes half of matrix
              %and all diagonal elements
  %Finds all the distances smaller than r:
  [sil sic]=find(0<md&md<r);

  %Prepare matrix of points found smaller than r
  k=[sil,sic];
  [sl sli]=sort(sil);
  rk=[sl,sic(sli)];
```

```
i=1; %initalize cluster counting variable

%Begining of loop that measures cluster size
while size(rk,1)>0;
  rs=rk(1,:);    %compares first line to all
                 %other lines
  rk(1,:)=[];    %sets first line = []

  %finds all the data that fit first line
  [a1 a2]=find(rk==rs(1));
  [b1 b2]=find(rk==rs(2));
  c=[];  %initialize c
  if isempty(a1)==1;
    c=cat(1,b1);
  elseif isempty(b1)==1
    c=cat(1,a1);
  else
    c=cat(1,a1,b1);
  end

  %if data were found that fit first line,
  %the cluster is greater than 2, so find
  %the other points connected to it
  if isempty(c)==0
    %assures that there are not multiples in c
    for irk=1:size(c,1);
      xrk=c(irk);
      fxrk=find(c==xrk);
      if length(fxrk)>1;
        c(fxrk(2:length(fxrk)))=0;
      end
    end
    fc=find(c~=0);
    nc=c(fc);

    v=rk(nc,:); %Define new matrix of point
                %that fit first line
    V=[rs;v];  %put all the data back together
    V=cat(1,V(:,1),V(:,2));

    %removes multiples
    for irv=1:size(V,1);
      xrv=V(irv);
      fxrv=find(V==xrv);
```

```
    if length(fxrv)>1;
      V(fxrv(2:length(fxrv)))=0;
    end
  end
  fv=find(V~=0);
  V=V(fv);

  rk(c,:)=[]; %removes all the points
          %that were connected

  sizeVo=size(V,1); %measures size of cluster

  sizeR=0; %initialize sizeR
  %Begining of new loop for "secondary..." points
  while sizeR<1;  %condition
    %for all the points connected to first line
    for iv1=1:size(V,1)
      [v11 v21]=find(rk==V(iv1));
      if isempty(v11)==0
        %removes multiples
        for irk1=1:size(v11,1);
          xrk1=v11(irk1);
          fxrk1=find(v11==xrk1);
          if length(fxrk1)>1;
            v11(fxrk1(2:length(fxrk1)))=0;
          end
        end
        fv11=find(v11~=0);
        nv11=v11(fv11);

        v1=rk(nv11,:);
        v1=cat(1,v11(:,1),v1(:,2));
        V=[V;v1]; %define V again with all
              %points that fit first line
        rk(v11,:)=[]; %remove all connected
                %points
      end
    end
    %check condition of while loop
    sizeV1=size(V,1);
    sizeR=sizeVo/sizeV1;
    sizeVo=sizeV1;
  end
else %alternative if cluster was only a pair
  V=rs';
```

```
        end %of secondary while loop

      tv(n,i)=size(V,1); %matrix of cluster size
      i=i+1; %cluster counting variable

      end %of while rk

      if size(rk,1)==0;
        tv(n,i)=0; %failsafe for tv variable
      end

   NC(lnb,n)=i-1; %number of clumps per run

   end %of n loop

clear sumtv %prevents "spillover"
sumtv=sum(tv,2)';
TCc(lnb,1:nmax)=sumtv; %number of bacteria involved in clumps per run
FC(lnb,1:nmax)=N-sumtv; %number of single cell CFU per run
colony(lnb,1:nmax)=FC(lnb,:)+NC(lnb,:); %number of colonies per run

%creates matrices of bacteria number in clusters
%of different sizes that were detected
ftv=find(tv~=0);
%need a failsafe if tv=0 (occurs when nmax is low)
if isempty(ftv)==1
   display('No clusters were detected')
else
   fftv=tv(ftv);
   sfftv=sort(fftv);
   %removes multiples
   for iftv=1:length(sfftv);
      xftv=sfftv(iftv);
      fxftv=find(sfftv==xftv);
      if length(fxftv)>1;
        sfftv(fxftv(2:length(fxftv)))=0;
      end
   end
   fsfftv=find(sfftv~=0);
   nsfftv=(sfftv(fsfftv))';
   for yuh=1:length(nsfftv);
      fyuh=find(tv==nsfftv(yuh));
      syuh(yuh)=size(fyuh,1);
      if isempty(fyuh)==1;
        syuh(yuh)=0;
```

```
        end
      end
      TVQ(lnb,1:length(nsfftv))=nsfftv;
      TVN(lnb,1:length(syuh))=syuh;
      %Massage matrices for calculations and plotting
      ftvq=find(TVQ==0);
      TVQ(ftvq)=NaN;
      ftvn=find(TVN==0);
      TVN(ftvn)=0;
      T=(TVN-1)./TVQ;
      fT=find(isnan(T)==1);
      T(fT)=0;

   end %of ftv failsafe started line 205
end %of nb loop

mNC=mean(NC,2); %mean number of clumps
stdNC=std(NC,0,2);  %STD number of clumps
mFC=mean(FC,2); %mean number of single cell CFU
stdFC=std(FC,0,2);  %STD no. of single cell CFU
mcolony=mean(colony,2); %mean no. of CFU
stdcolony=std(colony,0,2); %STD no. of CFU
mTCc=mean(TCc,2);   %mean total no. of bacteria in clusters
stdTCc=std(TCc,0,2); %STD tot. no. of bact. in clusters

%fits power laws through cluster size distributions
for i=1:size(TVN,1);
   clear lx ly tx ty fx fy ptvno
   xtvn(i,1:size(TVN,2))=TVQ(i,:);%/nb(i);
   ytvn(i,1:size(T,2))=T(i,:)/sum(T(i,:));
   fytvn=find(ytvn(i,:)==0);
   ytvn(i,fytvn)=NaN;
   lx=log10(xtvn(i,:));
   ly=log10(ytvn(i,:));
   fx=find(lx~=-Inf);
   tx=lx(fx);
   ty=ly(fx);
   fy=find(isnan(ty)==0);
   tx=tx(fy);
   ty=ty(fy);
   ptvno=polyfit(tx,ty,1);
   ptvn(i,1)=ptvno(1);
   ptvn(i,2)=ptvno(2);
end
%large N go further in distribution and provide more data.
```

```
%I arbitrarily choose the fits of the larger N
%for the mean and standard deviation of the estimate.
%Small N bias fit to larger power exponents because of
%lack of formation of larger clusters.
mptvn=mean(ptvn(length(nb):size(ptvn,1),:),1)
%std of ptvn is meaningless if only one N is fit.
%stdptvn=std(ptvn(length(nb):size(ptvn,1),:),0,1);

errorfit=polyfit(nb',(nb'-mcolony)./nb',1)

figure(1);
plot(nb,mFC,'ko-','LineWidth',2)
hold on
plot(nb,mFC+stdFC,'k-')
plot(nb,mFC-stdFC,'k-')
xlabel('N')
ylabel('Number of CFUs originating from single cells')

figure(2)
plot(nb,mNC,'ko-','LineWidth',2)
hold on
plot(nb,mNC+stdNC,'k-')
plot(nb,mNC-stdNC,'k-')
xlabel('N')
ylabel('Number of CFUs originating from multiple cells')

figure(3)
plot(nb,mTCc,'ko-','LineWidth',2)
hold on
plot(nb,mTCc+stdTCc,'k-')
plot(nb,mTCc-stdTCc,'k-')
xlabel('N')
ylabel('Total number of cells in clusters')

figure(4)
plot(nb,mcolony,'ko-','LineWidth',2)
hold on
plot(nb,mcolony+stdcolony,'k-')
plot(nb,mcolony-stdcolony,'k-')
xlabel('N')
ylabel('CFU')

figure(5)
hold on
plot(nb,(nb'-mcolony)./nb','k-o','LineWidth',2)
```

```
plot(nb,(nb'-mcolony+stdcolony)./nb','k-')
plot(nb,(nb'-mcolony-stdcolony)./nb','k-')
xlabel('N')
ylabel('Relative error, (N-CFU)/N')

X=0:1:N;
figure(6)
for i=1:size(xtvn,1);
   hold on
   plot(xtvn(i,:),ytvn(i,:),'k.','MarkerSize',8)
end
plot(X,10^mptvn(2)*X.^(mptvn(1)),'k-','LineWidth',2)
hold on
set(gca,'YScale','log')
xlabel('Cluster size, (#bacteria/cluster)')
ylabel('Normalized number of cluster of a given size')
```

## *Appendix C: Probability generating functions (PGF)*

Probability generating functions are useful tools to add distributions of independent variables (e.g. section 3.3). This appendix is intended as an introduction to pgf.

Lets imagine one is throwing a fair die and one wishes to know the probability (P) that each number (X) will roll. These are given by:

| X | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| P(X=r) | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 | 1/6 |

**Here, X is the random variable and P(X=r) is the probability of throwing value r. A probability generating function is a polynomial (power series) whose coefficients are the probabilities associated with the different outcomes. In the case of the die, the probability generating function is:**

$$G(z) = 0z^0 + \frac{1}{6}z^1 + \frac{1}{6}z^2 + \frac{1}{6}z^3 + \frac{1}{6}z^4 + \frac{1}{6}z^5 + \frac{1}{6}z^6.$$

There is nothing special about a die, however, and a probability generating function can be written for almost any discrete distribution. The general form of G(z) is:

$$G(z) = P(X=0)z^0 + P(X=1)z^1 + P(X=2)z^2 + \ldots$$

Probability generating functions have interesting properties in that they can greatly reduce the amount of work necessary to analyze distributions. For example, if one wants to know what are the probabilities of obtaining Y, the sum of two fair dice, one simply has to square G(z):

$$G(z)\cdot G(z) = 0z^0 + 0z^1 + \frac{1}{36}z^2 + \frac{2}{36}z^3 + \frac{3}{36}z^4 + \frac{4}{36}z^5 + \frac{5}{36}z^6 + \frac{6}{36}z^7 + \frac{5}{36}z^8 + \frac{4}{36}z^9 + \frac{3}{36}z^{10} + \frac{2}{36}z^{11} + \frac{1}{36}z^{12}$$

and the coefficients yield the probabilities of obtaining the sum of the two dice, indicated by the exponents of z.

Some important properties of the pgf are

1) $G(0) = P(X = 0)$

2) $G(1) = \sum_r P(X = r) = 1$

3) $G'(1) = \frac{dG(z)}{dz} = \sum_r r \cdot P(X = r) = E(X)$, where E(X) is the expectation value of X, which can be associated to the mean of the variable X.

4) $G''(1) = \frac{d^2G(z)}{dz^2} = \sum_r r \cdot (r-1) \cdot P(X = r) = E(X(X-1))$. From this last result, it is possible to compute the variance of the variable X, since:

$$\begin{aligned} G''(1) + G'(1) - (G(1))^2 &= E(X(X-1)) + E(X) - (E(X))^2 \\ &= E(X^2) - E(X) + E(X) - (E(X))^2 \\ &= E(X^2) - (E(X))^2 \\ &= V(X) \end{aligned}$$

## *Appendix D: PGF of the binomial distribution*

Given the definition of pgf in Appendix C and the definition of the binomial distribution, the pgf for the binomial distribution is

$$\begin{aligned} G(z) &= \binom{n}{0}(p)^0 q^n z^0 + \binom{n}{1} p^1 q^{n-1} z^1 + \binom{n}{2} p^2 q^{n-2} z^2 + \binom{n}{3} p^3 q^{n-3} z^3 + \ldots \\ &= \binom{n}{0}(pz)^0 q^n + \binom{n}{1}(pz)^1 q^{n-1} + \binom{n}{2}(pz)^2 q^{n-2} + \binom{n}{2}(pz)^3 q^{n-3} + \ldots \\ &= (q+pz)^n \end{aligned}$$

Accordingly, with p+q=1, p,q<1, and positive n

$$G'(z) = n(q+pz)^{n-1} p$$

which will yield the expectation value

$$E(X) = G'(1) = np$$

and

$$G''(z) = n(n-1)(q+pz)^{n-2} p^2$$

with the corresponding variance

$$\begin{aligned} V(X) &= G''(1) + G'(1) - (G'(1))^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= n^2 p^2 - np^2 + np - n^2 p^2 \\ &= np(1-p) \\ &= npq \end{aligned}$$

For the negative binomial, the index is negative and q-p=1.

## *Appendix E: the Gamma function (Γ(n))*

The Gamma function is a generalization of the factorial n! from integers to any positive n. For a real number n>0, Γ(n) is defined by (Stein & Barcellos 1992; Greenberg 1998):

$$\Gamma(n) = \int_0^\infty \exp(-x)x^{n-1}dx\,.$$

This equation can be integrated by parts with $u=x^{n-1}$ and dv=exp(-x)dx to yield

$$\Gamma(n) = -x^{n-1}\exp(-n)\Big|_0^\infty + (n-1)\int_0^\infty \exp(-x)x^{n-2}dx\,.$$

When evaluated, the first term of this equation goes to 0. Integrating the second term produces

$$\Gamma(n) = (n-1)\left[(n-2)\int_0^\infty \exp(-x)x^{n-3}dx\right]$$

and we immediately notice that the integral was indeed $\Gamma(n-1)$, such that

$$\Gamma(n) = (n-1)\Gamma(n-1)$$

for x>1, which is a recursion formula. If n is a positive integer

$$\begin{aligned}\Gamma(n+1) &= n\Gamma(n) = n(n-1)\Gamma(n-1)\\ &= \ldots = n(n-1)(n-2)\ldots(1)\Gamma(1)\end{aligned}$$

Because

$$\Gamma(1) = \int_0^\infty \exp(-x)dx = 1$$

the last result become

$$\Gamma(n+1) = n!$$

## *Appendix F: Literature review of studies focusing on the association of cells with particles.*

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Acinas, Anton & Rodriguez-Valera | 1999 | Sea | Sequencing | 8 | N/A | < 8 μm community is more diverse. |
| Bakken | 1985 | Soil | Microscopy, culturing | - | Homogenization | 2-5% of cells enumerated by microscopy produce colonies. Provides cell-size distribution. |
| Berman, Rice & Hoff | 1988 | Wastewater | Culturing | 7 | Homogenization | > 7 μm fraction is more resistent to disinfection. Homgenization produces greater concentration measurments of coliforms. |
| Bidle & Fletcher | 1995 | Sea, estuary | Microscopy, LMW-RNA | 3 | N/A | <3 μm and >3 μm communities are different. 1-18% of the cells are particle-associated. |
| Boecklemann et al. | 2000 | River | Microscopy, culturing, sequencing | - | N/A | β-Proteobacteria and Clostridia are the most abundant members of the 'river-snow' community. The community structure on particles changes with season. |
| Boecklemann, Szewzyk & Grohmann | 2003 | Soil | Microscopy, culturing | - | Sonication + chemicals | Development of a desorption protocol that achieves up to 22 fold higher counts. 3 of 7 strains tested showed decreased survival after sonication. |
| Borst & Selvakumar | 2003 | Stormwater runoff | Culturing | - | Homogenization + chemicals | Fecal indicators counts generally increase after treatment. |
| Camper et al. (a) | 1985 | Driking water | Culturing | - | Homogenization + chemicals | Globular activated carbon particles can sustain populations of $10^5$ to $10^7$ CFU/g. |
| Camper et al. (b) | 1985 | Driking water | Culturing | - | Homogenization + chemicals | Pathogens can produce populations (>$10^5$ CFU/g) on globular activated carbon particles, but they are outcompeted by river microorganisms. |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Camper et al. | 1986 | Driking water | Culturing, Microscopy | - | Homogenization + chemicals | Original description of the Camper protocol for the desorption of cells from globular activated particles. |
| Camper et al. | 1987 | Driking water | Culturing | - | Homogenization + chemicals | Globular activated carbon particles can support larger bacterial population than sand or anthracite filters |
| Carlson et al. | 1968 | Streamwater | Culturing | - | N/A | Increase in di-cation (Ca) concentration promotes viruses adsorption onto kaolinite clay. |
| Clarke & Joint | 1986 | Estuary | Microscopy | - | N/A | A larger fraction of the bacteria are attached to particles in freshwater than in seawater. |
| Crump & Baross | 1996 | Estuary | Tracer | 20 | N/A | Particle-associated bacteria are more active. Rotifers could be the main grazers or particle-associated bacteria. |
| Crump, Armbrust & Baross | 1999 | Estuary | Microscopy, tracer, sequencing | 3, 10 | N/A | Communities of particle-attached bacteria are different in the river, in the estuary and in the ocean, and develop into unique communities, whereas 48% of the free-living clones were similar in all three settings. |
| Emerick et al. | 1999 | Wastewater | Culturing, microscopy | 10, 80 | N/A | After disinfection, >80 μm particles bear more bacteria than the smaller particles (10-80 μm). Bewteen 1-30% of the particles contained coliforms. Particles >10 μm are important for shielding bacteria from disinfection |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Gale, Pitchers & Gray | 2002 | Drinking water | Culturing | - | Chemicals | Water treatment produces spatial heterogeneity of anaerobic spores samples, such that spot samples will underestimate the true population for these organisms. The problem is concentration/volume dependant and can be explained statistically. |
| Geesey & Costerton | 1979 | River | Microscopy | - | N/A | > 75% of the bacterial population exist as free-living cells in spite of the large suspended solids concentration (>220 mg/L). The bulk of the suspended sediments range between 5-50 μm. |
| Gerba & Schaiberger | 1975 | Sea | Culturing | - | N/A | Viruses entering coastal waters from land readily adsorb onto particles, what prolongs their survival and promotes their accumulation in sediments that are subject to resuspension. |
| Gerba & McLeod | 1976 | Marine sediments | Culturing | - | N/A | Coliforms 1-2 order of magnitude larger exist in sediments relative to the water above. Sediments allow for longer survival and growth. Autoclave sediments are even more efficient; autoclaving is believed to liberate nutrients. |
| Gough & Stahl | 2003 | Freshwater sediments | Microscopy | - | N/A | The authors recommend the use of 8.5-10 cells per field (higher dilution) instead of 30 cells/field to reduce the problem due to masking from particles. |
| Grabow, de Villiers & Prinsloo | 1991 | Shellfish | Culturing | 5 | Homogenization + chemicals | Prefiltration of a sample through a 5 μm filter considerably reduces the counts. |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Grossart et al. | 2003 | Laboratory experiments | Microscopy | - | N/A | Growth dominates over attachment after a few hours of exposure of bacteria to synthetic agar particles. |
| Kiorboe et al. | 2003 | Marine snow | Microscopy | - | N/A | Bacteria are the first colonizers of marine snow, followed by flagellates and ciliates. Colonization is initially a diffusion process, until growth occurs. |
| Kirchman & Mitchell | 1982 | Ponds, marshes | Tracer, microscopy | 1, 3 | Homogenization | Particle-associated bacteria are less abundant by more active than free-living cells. Most particles (>50%) are colonized by a few bacteria. Only 1-2 bacteria could be seen on mineral-like particles. <5% of the particles were heavily colonized (20-50 cells). |
| Lam & Cowen | 2004 | Deep sea hydro-thermal plumes | Microscopy | 5, 10 | Sonication | Sonication and size fractionation produce total counts 148-326% greater. Optimization of sonication time is recommended for different types of samples. |
| LeChevallier, Evans & Seidler | 1981 | Drinking water | Culturing, microscopy | - | Homogenization + chemicals | Disinfection efficiency is negatively correlated with turbidity. MPN method is recommended in water with turbidity >5NTU or when presence/absence test are sufficient. |
| LeChevallier et al. | 1984 | Drinking water | Culturing, microscopy | - | Homogenization + chemicals | Globular activated carbon shield bacteria from chlorine disinfection. |
| Loge et al. | 2002 | Wastewater | Microscopy | 11, 80 | N/A | Formation of particles with associated coliforms is influenced by the particle concentration, the coliform concentration and the mean residence time. Between 0-25% of the particles between 11-80 μm have associated coliforms. |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Manz et al. | 1993 | Drinking water | Microscopy | - | N/A | Greater physiological activity of surface-associated cells. Only Bacteria were sampled, no Archaea or Eukarya were recovered. |
| Murrel et al. | 1999 | Estuary | Tracer, microscopy | 1 | N/A | Most clumps were very small (<5 μm). There are many more ($10^{5-6}$ times) mineral-like particles than organic aggregates such that even if there is/are less carbon, cells, … per minearl-like aggregates, these are still important. |
| Palumbo, Ferguson & Rublee | 1984 | Estuary | Tracer, microscopy | 0.6, 3, 8 | N/A | 93-99% of the cells are found in the <3 mm fraction. Cells in the >3 mm fraction are more active, but their low abundance dwarfs their ecological role. |
| Parker & Darby | 1995 | Wastewater | Culturing | - | Homogenization + chemicals | Faster homogenization (19,000 rpm) is more effective at disrupting aggregates and produce larger bacterial counts than the homogenization speed used by Camper et al. (1986). |
| Perdek & Borst | 2000 | Combined sewer overflow | Culturing | - | Homogenization + chemicals | Variability between replicates is large but highest homogenization speed gives the greatest recovery of bacteria (22,000 rpm). |
| Pernitsky, Finch & Huck | 1997 | Drinking water | Culturing | - | Homogenization + chemicals | Particle-associated bacteria in finished driking water were found to have little impact on public health because of the low number of particles released and the low number of bacteria per particles. |
| Phillips et al. | 1999 | Sea | Sequencing, culturing | - | N/A | *Nitrosomonas* are found mostly on particles whereas *Nitrosospira* are mostly free-living. |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Plummer, Owens & Herbert | 1987 | Estuary | Tracer | - | N/A | Bacteria associated to permanently suspended particles are responsible for most of the high activity associated with particle-associated bacteria. |
| Qualls, Flynn & Johnson | 1983 | Wastewater | Culturing | 8, 70 | N/A | Samples filtered through a 8 μm filter were less resistent to UV disinfection. The survival of the 70 μm filtered sampled were only slightly lower than unfiltered samples. |
| Ramsay | 1984 | Soil | Culturing, microscopy, tracer | - | Homogenization, sonication | Sonication is more efficient for the extraction of bacteria from silt loam, but shaking, blending or sonication are equally good for sandy soils. Sonication, however, was responsible for cell lysis. Naturally particle-bound cells are more tighly held onto the particles than Bacillus added to particles in the larboratory. |
| Ridgeway & Olson | 1981 | Drinking water | Microscopy | - | N/A | 17% of the particles analyzed had between 10-100 associated bacteria. Most particles ranged between 10-50 μm. |
| Ridgeway & Olson | 1982 | Drinking water | Microscopy, culturing | 2 | N/A | 1% of the particles bore attached cells, most of which were rod shaped and ofter enrobed in extracellular material. Particles with attached bacteria were usulally >10 μm. Sample fraction >2 μm was more resistent to chloring disinfection. |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Roper & Marshall | 1974 | Marine sediments | Microscopy, culturing | - | Electrolytic dilution | There is a critical point in ionic strength where colloidal montmorillonoidal clay, phage and E. coli desorb from each other. Below this critical point, E. coli is protected from the phage by a coating of colloid on the cell, above the critical ionic strength, protection comes from the entrapment of E. coli and phage onto particles. |
| Roper & Marshall | 1979 | Estuary | Culturing | - | N/A | Longer survivability of coliforms in sediments. Both the adsorption of the cells onto particles and the adsorption of colloids onto the cells provide protection. |
| Schallenberg, Kalff & Rasmussen | 1989 | Sediments | Microscopy | - | N/A | Sediments affect the staining efficiency of DAPI. Larger concentration (5 μg/ml) of DAPI is recommended. |
| Schillinger & Gannon | 1985 | Urban stromwater | Culturing | 5, 10, 30, 52 | N/A | Generally, >50% of the bacteria do not settle after 4 hours and are not filtered (>5 μm). Variation exist between species regarding adsorption, with fecal coliforms adsorbing less than *Klebsiella* spp., *Pseudomonas aeruginosa* or Gram-negative cells. |
| Stewart, Wolfe & Means | 1990 | Drinking water | Culturing | 10 | Homogenization + chemicals | Attached bacteria are highly resistent to chlorine. |
| Stoodley et al. | 2001 | Drinking water, chemostat | Culturing, microscopy | - | Sonication + vortexing + chemicals | Focus on the detachment of biofilm fragments. Most fragments are small with a few cells, but rarely, large fragments with 100 to >1000 bacteria are released that account for 10-20% of the total cell count. |

Appendix F. (continued) Literature review of studies focusing on the association of cells with particles.

| Author | Year | Environment | Methodology | Filter (μm) | Cell desorption | Findings/comments |
|---|---|---|---|---|---|---|
| Stringfellow, Mallon & DiGiano | 1993 | Drinking water | Culturing, microscopy, energy dispersive X-ray analysis | 5 | Vortex / homogenization + chemicals | The majority of the cells collected from the drinking water have no attached bacteria, only 13 of 181 particles were colonized, representing only 0.01% of the total bacteria population. |
| Velji & Albright | 1986 | Marine sediments, fecal matter, kelp blade | Microscopy | - | Sonication + chemicals | Use of formaldehyde as a fixative and treatment with pyrophosphate and ultrasound yieded greater counts. |
| Wallner, Erhart & Amann | 1995 | Wastewater | Microscopy, flow cytometry | - | N/A | 0-8% of the Hoechst stained cells could be identified by hybridization with specific probes and flow-cytometry. The presence of particles is a big problem. |
| Weiss et al. | 1996 | Lake | Microscopy | - | N/A | $>10^6$ cells/aggregate, >55% of them *Bacteria*. Comparison of lake aggregates and laboratory made aggregates. |
| Yoon & Rosson | 1990 | Sea | Microscopy | - | Sonication / homogenization + chemicals | Comaparison between the use of sonication and homogenization. Correlation between turbidity and the number of particle-associated bacteria. |

## *Appendix G: resamplebinbayesian.m, a non-parametric alternative to a t-test for small populations*

```
function [S,S_lower,S_upper,P,sig,pvalue]=resamplebinbayesian(data_1,data_2,tail)
%=========================================================================
% resamplebinbayesian.m is a binomial test with a twist of bayesian and
% montecarlo.
% [S,S_lower,S_upper,P,sig,pvalue]=resamplebinbayesian(data_1,data_2,tail)
%=========================================================================
% The INPUTS are:
%  data_1 and data_2: two column vectors containing the data to be
%     compared
%  tail: must be equal to either 1 or 2
%  If tail=1, a one-tailed test is used, in this case, the underlying
%     assumption is that data_1 is greater than data_2. In this case, the
%     function assumes the 95% confidence interval to be asymetric, with
%     the accepted 5% error concentrated at either end of the
%     distribution.
%  If tail=2, a two-tailed test is used, in which case, no assumptions is
%     used as of which data vector is greater than the other. In such a case,
%     however, the test is "less sensitive" as the 95% confidence interval is
%     split to both ends of the distribution (2.5% at each end).
%
% The OUTPUTS are:
%  S: the general probability distribution
%  S_lower: the lower 95% confidence interval
%  S_upper: the upper 95% condidence interval
%  sig:   'different' means that the case considered is different relative
%     to the general distribution
%        'not different' means that any difference between the case
%        studied and the general distribution could be due to chance
%  pvalue: pvalue of the test. It is computed by computing the fraction of
%     the number of pvalues of the general test falling above (or below) the
%     case P-value, to the total number of p-values obtained by the bayesian
%     combinations of the data.
%=========================================================================

ndata1=length(data_1);
ndata2=length(data_2);
data=[data_1;data_2];
ndata=length(data);
A=combnk(1:ndata,ndata1);
[r,c]=size(A);
```

```
D=([1:ndata]'*ones(1,r))';
M=D;
for i=1:r
   for j=1:c
      M(i,A(i,j))=NaN;
   end
end

for m=1:r
   fm=find(isnan(M(m,:))==0);
   DM(m,1:size(fm,2))=M(m,fm);
end

for k=1:r
   [N(k),NP(k),p(k)]=binbayesian(data(A(k,:)),data(DM(k,:)),0.5,tail);
end

S=sort(p)';
if tail==2
   S_lower=S(round(0.025*size(S,1)));
   S_upper=S(round(0.975*size(S,1)));
else
   S_lower=S(round(0.05*size(S,1)));
   S_upper=S(round(0.95*size(S,1)));
end


[n,np,P]=binbayesian(data_1,data_2,0.5,1);

if P<=S_lower
   sig='different';
   P<=S_lower
   fp=find(S<=P);
   pvalue=length(fp)/r;
else
   sig='not different';
   fp=find(S<=P);
   pvalue=length(fp)/r;
end
%=================================================================
%=================================================================

function [N,NP,p]=binbayesian(data_1,data_2,P,tail)
%=================================================================
%binbayesian.m
```

```
%[N,NP,p]=binbayesian(data_1,data_2,P)
%=========================================================================
% binbayesian.m calculates the binomial probability given the probability P
% and the data in data_1 and data_2.
% This function may be useful to test for differences between two small
% vectors by setting P=0.5, which assumes that any difference seen between
% data_1 and data_2 is due to chance. If the p-values spit out by
% binbayesian.m is lower than 0.05, then the probability that the
% difference between data_1 and data_2 is due to chance (that is P=0.5)
% is small and one should consider that the values in data_1 are greater
% than those in data_2 systematically.
%  INPUTS:
% tail is a number, either 1 or 2. Choose 1 for a one-tailed test and 2 for
% a two-tailed test.
% A one-tailed test can be used if the investigator suspects that data_1 is
% greater than data_2, if there are no reasons to believe that data_1 is
% greater than data_2, then tail should be set to 2.
%
% The OUTPUTS are
%   N: the number of paired cases considered
%   NP: the number of pairs yielding a ratio data_1/data_2 greater than 1
%   p: the p-value of the test
%=========================================================================
   ndata1=length(data_1);
   ndata2=length(data_2);
for i=1:length(data_1)
   for j=1:length(data_2)
      D(i,j)=data_1(i)/data_2(j);
   end
end
[r,c]=size(D);
ratio=[];
for k=1:c;
   ratio=cat(1,ratio,D(:,k));
end
N=length(ratio);
np1=find(ratio>=1);
NP=[length(np1)];
if tail==1;
   p1=1-binocdf(NP(1),N,P);
elseif tail==2;
   p1=2*(1-binocdf(NP(1),N,P));
end
p=[p1];
```

# REFERENCES


Acinas SG, A. J, F. Rodriguez-Valera (1999). Diversity of free-living and attached bacteria in offshore Western Mediterranean waters as depicted by analysis of genes encoding 16S rRNA. Appl Environ Microbiol. 65: 514-22.

Agrawal, Y. C. and H. C. Pottsmith (2000). "Instrument for particle size and settling velocity observations in sediment transport." Marine Geology 168: 89-114.

Amann, R., L. Krumholz and D. A. Stahl (1990). "Fluorescent-oligonucleotide probing of whole cell for determinative, phylogenetic, and environmental studies in microbiology. Journal of Bacteriology 172(2): 762-770.

Amann, R., W. Ludwig and K.-H. Schleifer (1995). "Phylogenetic idenfication and in situ detection of individual microbial cells without cultivation." Microbiological Reviews 59(1): 143-169.

Anon. (1983). Department of the Environment, Department of health and social security. Public Health Laboratory Services. The Bacteriological Examination of Drinking Water Supplies. London: HMSO."

APHA (American Public Health Association) (1998). "Standard Methods for the examination of water and wastewater." 20th Edition, Eds: Clesceri, L.S., A. E. Greenberg and A. D. Eaton.

Ashbolt, N. J. (2001). "Chapter 9: Methods to identify and enumerate frank and opportunistic bacterial pathogens in water and biofilms" In: L. Fewtrell & J Bartram (Eds.),World Health Organization (WHO), "Water Quality: Guidelines, Standards and Health".

Ashbolt, N. J., W. O. K. Grabow and M. Snozzi (2001). "Chapter 10: Indicators of microbial water quality." In: L. Fewtrell & J Bartram (Eds.),World Health Organization (WHO), "Water Quality: Guidelines, Standards and Health".

Bader, H. (1970). "The hyperbolic distribution of particle sizes." Journal of Geophysical Research 75(15):2822-2830.

Bakken, L. R. (1985). "Separation and purification of bacteria from soil." Applied and Environmental Microbiology 49(6):1482-1487.

Bartram, J., L. Fewtrell and T.-A. Stenström (2001). "Harmonised assessment of risk and risk management for water-related infectious disease: an overview." In: L. Fewtrell & J Bartram (Eds.),World Health Organization (WHO), "Water Quality: Guidelines, Standards and Health".

Beimfohr, C., A. Krause, R. Amann, W. Ludwig and K.-H. Schleifer (1993). "In situ identification of Lactococci, Enterococci and Streptococci." System. Appl. Microbiol. 16: 450-456.

Berman, D., E. W. Rice and J.C. Hoff (1988). "Inactivation of particle-associated coliforms by chlorine and monochloramine." Applied and environmental microbiology 54(2):507-512.

Betzl, D., W. Ludwig and K.-H. Schleifer (1990). "Identification of Lactococci and Enterococci by colony hybridization with 23S rRNA-targeted oligonucleotide probes." Applied and Environmental Microbiology 56(9): 2927-2929.

Bidle, K. D. and M. Fletcher (1995). "Comparison of free-living and particle-associated bacterial communities in the Chesapeake Bay by stable low molecular-weight RNA analysis." Applied and Environmental Microbiology 61(3): 944-952.

Bisson, J. W. and V. J. Cabelli (1979). "Membrane filter enumeration method for *Clostridium perfringens*." Applied and Environmental Microbiology 37(1): 55-66.

Böcklemann, U., W. Manz, T. R. Neu and U. Szewzyk (2000). "Characterization of the microbial community of lotic organic aggregates ('river snow') in the Elbe river of Germany by cultivation and molecular methods." FEMS Microbiology Ecology 33: 157-170.

Böcklemann, U., U. Szewzyk and E. Grohmann (2003). "A new enzymatic method for the detachment of particle-associated soil bacteria." Journal of Microbiological Methods 55: 201-211.

Borst, M. and A. Selvakumar (2003). " Particle-associated microorganisms in stromwater runoff." Water Research 37: 215-223.

Bridge, P. D. and P. H. A. Sneath (1982). "*Streptococcus gallinarum* sp. nov. and *Streptococcus oralis* sp. nov." International Journal of Systematic Bacteriology 32(4): 410-415.

Burr, S. (2003). "Final Kaneohe Stream bioassessment." AECOS, INC. Prepared for the Hawaii State Department of Health.

Cabelli, V. J., et al. (1979). "Relationship of microbial indicators to health effects at marine bathing beaches." American Journal of Public Health 69(7): 690-696.

Cabelli, V. J., A. P. Dufour, L. J. McCabe and M. A. Levin (1982). "Swimming-associated gastroenteritis and water quality." American Journal of Epidemiology 115(4): 606-616.

Cabelli, V. J., A. P. Dufour, L. J. McCabe and M. A. Levin (1983). "A marine recreational water quality criterion consistent with indicator concepts and risk analysis." Journal WPCF 55(10): 1306-1314.

Cabelli, V. J. (1983). "Health effects criteria for marine recreational waters." EPA-600/1-90-031.

Cabelli, V. J. (1989). "Swimming-associated illness and recreational water quality criteria." Wat. Sci. Tech. 21(2): 13-21.

Camper, A. K., M. W. LeChevallier, S. C. Broadaway and G. A. McFeters (1985 a). "Evaluation of procedures to desorb bacteria from granular activated carbon." Journal of Microbiological Methods 3: 187-198.

Camper, A. K., M. W. LeChevallier, S. C. Broadaway and G. A. McFeters (1985 b). "Growth and persistence of pathogens on granular activated carbon filters." Applied and Environmental Microbiology 50(6): 1378-1382.

Camper, A. K., M. W. LeChevallier, S. C. Broadaway and G. A. McFeters (1986). "Bacteria associated with granular activated carbon particles in drinking water." Applied and Environmental Microbiology 52(3): 434-438.

Camper, A. K., Broadaway, S. C., M. W. LeChevallier and G. A. McFeters (1987). "Operational variables and the release of colonized granular activated carbon particles in drinking water." Journal AWWA May: 70-74.

Carlson, G. F. Jr., F. E. Woodward, D. F. Wentworth and O. J. Sproul (1968). "Virus inactivation on clay particles in natural waters." Journal WPCF 40(2): R89-R106.

Carrillo, M., E. Estrada and T. C. Hazen (1985). "Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed." Applied and Environmental Microbiology 50(2): 468-476.

Clarke, K. R. and I. R. Joint (1986). " Methodology for estimating numbers of free-living and attached bacteria in estuarine water." Applied and Environmental Microbiology 51(5): 1110-1120.

Collins, M. D., D. Jones, J. A. E. Farrow, R. Kilpper-Bälz and K. H. Schleifer (1984). "*Enterococcus avium* nom. rev., comb. nov.; *E. casseliflavus* nom. rev., comb. nov.; *E. durans* nom. rev., comb. nov.; *E. gallinarum* comb. nov. and *E. malodoratus* sp. nov." International Journal of systematic bacteriology 34(2):220-223.

Costerton, J. W., R. T. Irvin and K.-J. Cheng (1981). "The bacterial glycocalyx in nature and disease." Ann. Rev. Microbiol. 35: 299-324.

Costerton, J. W., et al. (1987). 'Bacterial biolfilms in nature and disease." Ann. Rev. Microbiol. 41: 435-464.

Cox, E. (2001). Coastal Intensive Site Network (CISNet): http://www.hawaii.edu/cisnet

CRM (Commissions romandes de mathématique, de physique et de chimie) (1992). "Formulaires et tables: mathématique, physique, chimie." Editions du Tricorne:Genève.

Crump, B. C. and J. A. Baross (1996). "Particle-attached bacteria and heterotrophic plankton associated with the Columbia river estuarine turbidity maxima." Marine Ecology Progress Series 138: 265-273.

Crump, B, C., E. V. Armbrust and J. A. Baross (1999). "Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia river, its estuary and the adjascent coastal ocean." Applied and Environmental Microbiology 65(7): 3192-3204.

De Carlo, E. H., V. L. Beltran and M. S. Tomlinson (2004). "Composition of water and suspended sediment in streams of urbanized subtropical watersheds in Hawaii." Applied Geochemistry 19:1011-1037.

DeLong, E. F., G. S. Wickham and N. R. Pace (1989). "Phylogenetic stains: ribosomal RNA-based probes for the identification of single cells." Science 28: 1360-1363.

Devriese, L. A., B. Pot and M. D. Collins (1993). "Phenotypic identification of the genus Enterococcus and differentiation of phylogenetically distinct enterococcal species and species groups." Journal of Applied Bacteriology 75: 399-408.

Devriese, L. A., B. Pot, K. Kersters, S. Lauwers and F. Haesebrouck (1996). "Acidification of Methyl-α-D-Glucopyranoside: a useful test to differentiate *Enterococcus casseliflavus* and *Enterococcus gallinarum* from *Enterococcus faecium* species group and from *Enterococcus faecalis*." Journal of Clinical Microbiology 34(10): 2607-2608.

Devriese, L. A., et al. (2002). "Differentiation and identification of *Enterococcus durans*, *E. hirae* and *E. villorum*." Journal of Applied Microbiology 92: 821-827.

El-Shaarawi, A. H., S. R. Esterby and B. J. Dutka (1981). "Bacterial density in water determined by poisson or negative binomial distributions." Applied and Environmental Microbiology 41(1): 107-116.

Emerick, R. W., F. J. Loge, D. Thompson and J. L. Darby (1999). "Factors influencing ultraviolet disinfection performance. Part II: Association of coliform bacteria with wastewater particles." Water Environment Research 71(6): 1178-1187.

EPA, Environmental Protection Agency (2001). Region 9: Water Programs. http://www.epa.gov/region09/water/tmdl/.

Favero, M. S. (1985). "Microbiologic indicators of health risks associated with swimming." American Journal of Public Health 75(9):1051-1054.

Ferguson, D. M., D. F. Moore, M. H. Zhowandai and M. A. Getrich (2003). "Occurrence of enterococci in ocean water, sediments, strom drains, soils and seagulls at Baby Beach, Dana Point Harbor, California." A poster presented at the ASLO/TOS conference, Honolulu.

Ferguson, D. M., et al. (2004). "Characterization of Enterococcus species in coastal waters exceeding water quality standards." A poster presented at the ASM conference, New Orleans.

Feynman, R. P., R. B. Leighton and M. Sands (1963). "The Feynman lectures on physics: Mainly mechanics, radiation and heat (volume I)." Addison-Wesley Publishing Company: Reading, Massachussets.

Fleisher, J. M. and R. T. McFadden (1980). "Obtaining precise estimates in coliform enumeration." Water research 14: 477-483.

Fleisher, J. M. (1990). "The effects of measurement error on previously reported mathematical relationships between indicator organism dentiy and swimming-associated illness: a quantitative estimate of the resulting bias." International Journal of Epidemiology 19(4): 1100-1106.

Fujioka, R. S. and L. K. Shizumura (1985). " *Clostridium perfringens*, a reliable indicator of stream water quality." Journal WPCF 57(10): 986-992.

Fujioka, R. S. and M. N. Byappanahalli (2001). "Microbial ecology controls the establishment of fecal bacteria in tropical soil environment." In: Advances in water and wastewater treatment technology. Eds: Matsumoto, T., K. Hanaki, S. Takizawa and H. Satoh. Elsevier: New-York 273-281.

Gale, P., R. Pitchers and P. Gray (2002). "The effect of drinking water treatment on the spatial heterogeneity of microorganisms: implications for assessment of treatment efficiency and health risk." Water Research 36: 1640-1648.

Geesey, G. G. and J. W. Costerton (1979). "Microbiology of a northern river: bacterial distribution and relationship to suspended sediment and organic carbon." Can. J. Microbiol. 25: 1058-1062.

Geldreich, E. E. (1978). "Bacterial populations and indicator concepts in feces, sewage, stormwater and solid wastes." In Berg, G. (Ed.) "Indicators of viruses in water and food" , pp51-97, Ann Arbor Science, Ann Arbor, MI.

Gerba, C. P. and G. E. Schaiberger (1975). "Effect of particulates on virus survival in seawater." Journal WPCF 47(1): 93-103.

Gerba, C. P. and J. S. McLeod (1976). "Effect of sediments on the survival of *Escherichia coli* in marine waters." Applied and Environmental Microbiology 32(1): 114-120.

Glöckner, F. O. et al. (1996). "An in situ hybridization protocol for detection and identification of planktonic bacteria. Syst. Appl. Microbiol. 19:403-406.

Gough, H. L. and D. A. Stahl (2003). "Optimization of direct cell counting in sediment." Journal of Microbiological Methods 52: 39-46.

Goyal, S. M., C. P. Gerba and J. L. Melnick (1977). "Occurrence and distribution of bacterial indicators and pathogens in canal communities along the Texas Coast." Applied and Environmental Microbiology 34(2): 139-149.

Grabow, W. O. K., J. C. de Villiers and N. Prinsloo (1991). "An assessment of methods for the microbiological analysis of shellfish." Wat. Sci. Tech. 24(2): 413-416.

Greenberg, M. D. (1998). "Advanced engineering mathematics."Second Edition. Prentice Hall: Upper Saddle River (NJ)

Griffith, P., F.-K. Shiah, K. Gloersen, H.W. Ducklow and M. Fletcher (1994). "Activity and distribution of attached bacteria in Cheasapeake Bay." Marine Ecology Progress Series 108:1-10.

Grossart, H.-P., T. Kiørboe, K. Tang and H. Ploug (2003). "Bacterial colonization of particles: growth and interactions." Applied and Environmental Microbiology 69(6): 3500-3509.

Haas, C. N. (1996). "How to average microbial densities to characterize risk." Wat. Res. 30(4): 1036-1038.

Haas, C. N., J. B. Rose and C. P. Gerba (1999). "Quatitative microbial risk assessment."Wiley & Sons: New York, 449 pp.

Hames, B. D. and S. J. Higgins, Editors (1995). "Gene probes 1-2 : a practical approach." Oxford; New York : Oxford University Press.

Hardina, C.M. and R. S. Fujioka (1991). "Soil: The environmental source of *Escherichia coli* and Enterococcus in Hawaii's streams." Environmental Toxicology and Water Quality: An International Journal 6: 185-195.

Hardie, J. M. and R. A. Whiley (1997). "Classification and overview of the genera Streptococcus and Enterococcus." Journal of Applied Microbiology Symposium Supplement 83:1S-11S.

Hawaii State Department of Health (1999). "Hawaii Implementation Plan for Polluted Runoff Control" http://www.state.hi.us/health/eh/cwb/prc/implan-index.html.

Hawaii State Department of Health (2002). "Rules amending Title 11 Hawaii administrative rules, Chapter 54: Water quality standards." Department of Heath.

Hawaii State Department of Health (2003). "Indicators of Environmental Quality." Environmental Planning Office, 919 Ala Moana Blvd. Room 312, Honolulu, HI, 96814.

Hawaii Coastal Zone Management Program (1996). "Hawaii's coastal nonpoint pollution control program: management plan." Volume 1. Office of State Planning, Honolulu, Hawaii.

Hoover, D.J. (2002). "Fluvial nutrient fluxes to Oahu coastal waters: land use, storm runoff and impacts on coastal ecosystems." Dissertation, Dep. Oceanography, University of Hawaii at Manoa, Honolulu.

Jagals, P., W. O. K. Grabow and J. C. de Villiers (1995). "Evaluation of indicators for assessment of human and animal faecal pollution of surface run-off." Wat. Sci. Tech. 31(5-6): 235-241.

Jagals, P. (1997). "Stormwater runoff from typical developed and developing South African urban developments: definitely not for swimming." Wat. Sci. Tech. 35(11-12): 133-140.

Jurtshuk, R. J., M. Blick, J. Bresser, G. E. Fox and P. Jurtshuk, Jr. (1992). "Rapid in situ hybridization technique using 16S rRNA segments for detecting and differentiating the closely related gram-positive organisms *Bacillus polymyxa* and *Bacillus macerans*." Applied and Environmental Microbiology 58(8): 2571-2578.

Kiørboe, T., K. Tang, H.-P. Grossart and H. Ploug (2003). "Dyanmics of microbial communities on marine snow aggregates; colonization, growth, detachment and grazing mortality of attached bacteria." Applied and Environmental Microbiology 69(6): 3036-3047.

Kirchman, D. and R. Mitchell (1982). "Contribution of particle-bound bacteria to total microheterotrophic activity in five ponds and two marshes." Applied and Environmental Microbiology 43(1): 200-209.

Kusuda, R., K. Kawai, F. Salati, C. R. Banner and J. L. Fryer (1991). "*Enterococcus seriolicida* sp. Nov., a fish pathogen." International Journal of systematic bacteriology 41(3):406-409.

Lam, P. and J. P. Cowen (2004). "Processing deep-sea particle-rich water samples for fluorescence in situ hybridization: consideration of storage effects, preservation, and sonication." Applied and Environmental Microbiology 70(1):25-33.

LeChevallier, M. W., T. M. Evans and R. J. Seidler (1981). "Effect of turbidity on chlorination efficiency and bacterial persistence in drinking water." Applied and Environmental Microbiology 42(1): 159-167.

LeChevallier, M. W., T. S. Hassenauer, A. K. Camper and G. A. McFeters (1984). "Disinfection of bacteria attached to granular activated carbon." Applied and Environmental Micrbiology 48(5): 918-923.

Leclerc, H., L. A. Devriese and D. A. A. Mossel (1996). "Taxonomical changes in intestinal (faecal) enterococci and streptococci: consequences on their use as indicators of faecal contamination in drinking water." Journal of Applied Bacteriology 81: 459-466.

Levin, M. A. J. R. Fisher and V. J. Cabelli (1975). "Membrane filter technique for enumeration of enterococci in marine waters." Applied Microbiology 30(1):66-71.

Loge, F. J., R. W. Emerick, T. R. Ginn and J. L. Darby (2002). "Association of coliform bacteria with wastewater particles: impact of operational parameters of the activated sludge process." Wat. Res. 36: 41-48.

Manz, W. et al. (1993). "In situ identification of bacteria in drinking water and adjoining biofilms by hybridization with 16S and 23S rRNA-directed fluorescent oligonuelotide probes." Applied and Environmental Microbiology 59(7): 2293-2298.

Mehrabi, Y. and J. N. S. Matthews (1995). "Likelihood-based methods for bias reduction in limiting dilution assays." Biometrics 51:1543-1549.

Meier, H. (1998). "Detection and identification of indicator bacteria using rRNA targeted probes." OECD Workshop Molecular Methods for Safe Drinking Water. Interlaken.

Meier, H. et al. (1997). "Detection of enterococci with rRNA targeted DNA probes and their use for hygienic drinking water control." Wat. Sci. Tech. 35(11-12): 437-444.

Muñiz, I., L. Jimenez, G. A. Toranzos and T. C. Hazen (1989). "Survival and activity of *Streptococcus faecalis* and *Escherichia coli* in tropical freshwater." Microb. Ecol. 18: 125-134.

Murrell, M. C., J. T. Hollibaugh, M. W. Silver and P. S. Wong (1999). "Bacterioplankton dynamics in northern San Francisco Bay: role of particle association and seasonal freshwater flow." Limnol. Oceanogr. 44(2): 295-308.

Norden, R. H. (1973). "A survey of maximum likelihood estimation." Int. Stat. Rev. 41(1):39-58.

Oki, D.S. (2003). "Surface water in Hawaii." U. S. Geological Survey Fact Sheet 045-03.

Palumbo, A. V., R. L. Ferguson and P. A. Rublee (1984). "Size of suspended bacterial cells and association of heterotrophic activity with size fraction of particles in estuarine and coastal waters." Applied and Environmental Microbiology 48(1): 157-164.

Parker, J. and J. L. Darby (1995). "Particle-associated coliform in secondary effluents: shielding from ultraviolet light disinfection." Water Envrion. Res. 67: 1065-1075.

Perez-Rosas, N. and T. C. Hazen (1988). "In situ survival of Vibrio cholerae and Escherichia coli in tropical coral reefs." Applied and Enviornmental Microbiology 54(1): 1-9.

Perdek, J. M. and M. Borst (2000). "Particle association effects on microbial indicator concentrations and CSO disinfection." ASCE's 2000 Joint Conference on Water Resources Engineering and Water Resources Planning and Management. July-August, Minneapolis, MN.

Pernitsky, D. J., G. R. Finch and P. M. Huck (1997). "Recovery of attached bacteria from GAC fines and implications for disinfection efficacy." Wat. Res. 31(3): 385-390.

Pielou, E. C. (1977). "Mathematical ecology." Wiley & Sons: New York, 385 pp.

Phillips, C. J., Z. Smith, T. M. Embley and J. I. Prosser (1999). "Phylogenetic Differences between Particle-Associated and Planktonic Ammonia-Oxidizing Bacteria of the β Subdivision of the Class *Proteobacteria* in the Northwestern Mediterranean Sea." Applied and Environmental Microbiology 65(2):779-786.

Plummer, D. H., N. J. P. Owens and R. A. Herbert (1987). "Bacteria-particle interactions in turbid estuarine environments." Continental Shelf Research 7(11-12): 1429-1433.

Qualls, R. G., M. P. Flynn and J. D. Johnson (1983). "The role of suspended particles in ultraviolet disinfection." Journal WPCF 55(10):1280-1285.

Ramsay, A. J. (1984). "Extraction of bacteria from soil: efficiency of shaking or ultrasonication as indicated by direct counts and autoradiography." Soil Biol. Biochem. 16(5):475-481.

Ridgway, H. F. and B. H. Olson (1981). "Scanning electron microscope evidence for bacterial colonization of a drinking-water distribution system." Applied and Environmental Microbiology 41(1): 274-287.

Ridgway, H. F. and B. H. Olson (1982). "Chlorine resistance patterns of bacteria from two drinking water distribution systems." Applied and Environmental Microbiology 44(4):972-987.

Roll, B. M. and R. S. Fujioka (1997). "Sources of faecal indicator bacteria in a brackish, tropical stream and their impact on recreational water quality." Wat. Sci. Tech. 35(11-12): 179-186.

Roller, C., M. Wagner, R. Amann, W. Ludwig and K.-H. Schleifer (1994). "In situ probing of gram-positive bacteria with high DNA G+C content using 23S rRNA-targeted oligonucleotides. Mircobiology 140: 2849-2858.

Rompré, A., P. Servais, J. Baudart, M.-R. de-Roubin and P. Laurent (2002). "Detection and enumeration of coliforms in drinking water: current methods and emerging approaches." Journal of Microbiological Methods 49: 31-54.

Rönner, S. G. E. and E. Stackebrandt (1994). "Identification of Clostridium perfringens by 16S and 23S rRNA oligonucleotide probes." System. Appl. Microbiol. 17: 425-432.

Roper, M. M. and K. C. Marshall (1974). "Modification of the interaction between *Escherichia coli* and bacteriophage in saline sediment." Microibal Ecology 1: 1-13.

Roper, M. M. and K. C. Marshall (1979). "Effects of salinity on sedimentation and of particulates on survival of bacteria in estuarine habitats." Geomicrobiol. J. 1:103-116.

Salama, I. A., G. C. Koch and H. D. Tolley (1978). "On the estimation of the most probable number in a serial dilution experiment." Communications in statistics—theory, methodology A7(13):1267-1281.

Sayler, G. S. and A. C. Layton (1990). "Environmental application of nucleic acid hybridization." Annu. Rev. Microbiol. 44: 625-648.

Schallenberg, M., J. Kalff and J. B. Rasmussen (1989). "Solutions to problems in enumerating sediment bacteria by direct counts." Applied and Environmental Microbiology 55(5): 1214-1219.

Schillinger, J. E. and J. J. Gannon (1985). "Bacterial adsorption and suspended particles in urban stormwater." Journal WPCF 57(5): 384-389.

Schramm, A., D. de Beer, M. Wagner and R. Amman (1998). Idenfification and activities in situ of Nitrosospira and Ntrospira spp. as dominant populations in a nitrifying fluidized bed reactor. Applied and Environmental microbiology 64(9):3480-3485.

Sloat, S. and C. Ziel (1991). "The use of indicator organisms to assess public water safety." Technical information series—Booklet No. 13. HACH technical center for applied analytical chemistry: Loveland, Colorado, USA.

Sorensen, D. L., S. G. Eberl and R. A. Dicksa (1989). "*Clostridium perfringens* as a point source indicator in non-point polluted streams." Wat. Res. 23(2):191-197.

Stahl, D. A. and R. Amann (1991). "Development and application of nucleic acid probes, p205-248. In E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons Ltd., Chichester, England.

Stein, S. K. and A. Barcellos. "Calculus and analytical geometry." Fifth Edition. McGraw-Hill, Inc.: New York.

Stewart M. H., R. L. Wolfe and E. G. Means (1990). "Assessment of the bacteriological activity associated with granular activated carbon treatment of drinking water." Applied and Environmental Microbiology 56(12):3822-3829.

Stoodley, P., et al. (2001). "Growth and detachment of cell clusters from mature mixed-species biofilms." Applied and Environmental Microbiology 67(12): 5608-5613.

Stringfellow, W. T., K. Mallon and F. A. DiGiano (1993). "Enumerating and disinfecting bacteria associated with particles released from GAC filter-adsorbers." Journal AWWA 70-80.

Teixeira, L. M. et al. (1996). "Phenotypic and genotypic characterization of atypical *Lactococcus garvieae* strains isolated from water buffalos with subclinical mastitis and

confirmation of *L. garvieae* as a senior subjective synonym of *Enterococcus seiolicida*." International Journal of Systematic Bacteriology 46(3):664-668.

Teunis, P. F. M., G. J. Medema, L. Druidenier and A. H. Havelaar (1997). "Assessment of the risk of infection by Cryptosporidium or Giardia in drinking water from a surface water source." Wat. Res. 31(6): 1333-1346.

Tillet, H. E. and C. P. Farrington (1991). "Inaccuracy of counts of organisms in water or other samples: effects of pre-dilution." Letter in Applied Microbiology 13: 168-170.

Vanek, N. N. et al. (1999). "Enterococcus faecalis aggregation substance promotes opsonin-independent binding to human neutrophils via a complement receptor type 3-mediated mechanism." FEMS Immunology and Medical Microbiology 26:49-60.

Velji, M. I. and L. J. Albright (1986). "Microscopic enumeration of attached marine bacteria of seawater, marine sediment, fecal matter, and kelp blade samples following pyrophosphate and ultrasound treatments." Can. J. Microbiol. 32: 121-126.

Wagner M., et al. (1998). "Combining fluorescent in situ hybridization (FISH) with cultivation and mathematical modeling to study population structure and function of ammonia-oxidizing bacteria in activated sludge." Wat. Sci. Tech. 37(4-5): 441-449.

Wahl, G. M., S. L. Berger and A. R. Kimmel (1987). "Molecular hybridization of immobilized nucleic acids: theoretical concepts and practical considerations." Methods in Enzymology 152: 399-407.

Wallner, G., R. Erhart and R. Amann (1995). "Flow cytometric analysis of activated sludge with rRNA-targeted probes." Applied and Environmental Microbiology 61(5): 1859-1866.

Weiss, P., B. Schweitzer, R. Amann and M. Simon (1996). "Identification in sity and dynamics of bacteria on limnetic organic aggregates (lake snow)." Applied and Environmental Microbiology 62(6): 1998-2005.

Wilber, K. (1998). "The marriage of sense and soul." Broadway Books: New York

Yoon, W. B. and Rosson, R. A. (1990). "Improved method of enumeration of attached bacteria for study of fluctuation in the abundance of attached and free-living bacteria in response to diel variation in seawater turbidity." Applied and Environmental Microbiology 56(3): 595-600.