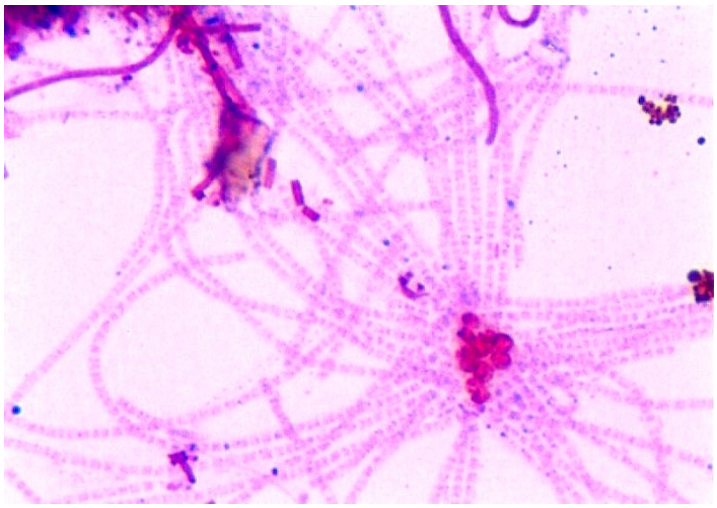
NATIONAL SURVEY OF FILAMENTOUS BACTERIAL POPULATIONS IN ACTIVATED SLUDGE

Report to the
Water Research Commission

by

PJ Welz, S Kumari-Santosh, C Uys, N van Blerk, A Smith, F Bux, T Conco, P Thobejane, N Sonjica
Cape Peninsula University of Technology



WRC Report No. 2471/1/22
ISBN 978-0-6392-0459-8

August 2022

WATER
RESEARCH
COMMISSION

Obtainable from
Water Research Commission
Private Bag X03
Gezina, 0031

orders@wrc.org.za or download from www.wrc.org.za

EXECUTIVE SUMMARY

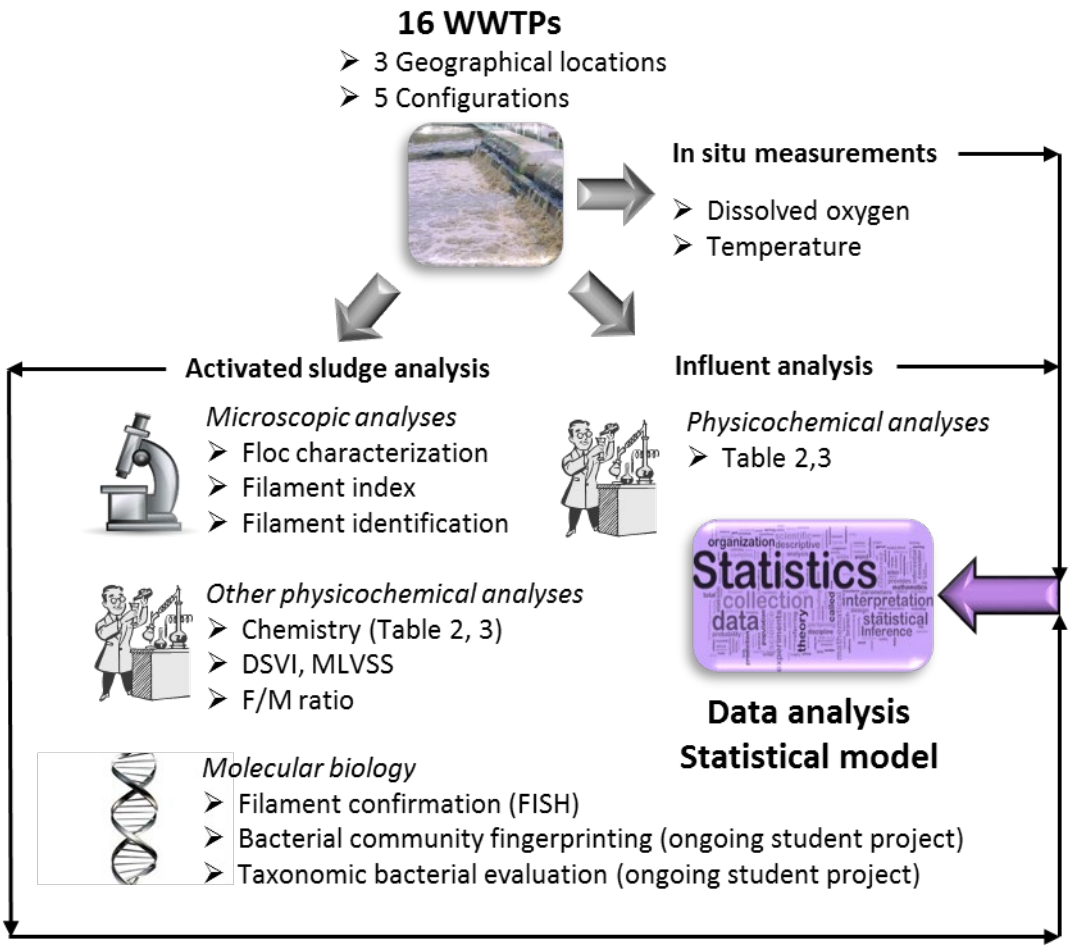
There is currently no universal strategy to overcome excessive growth of filamentous bacteria in activated sludge wastewater treatment plants. Chlorination, ozonation, and the addition of hydrogen peroxide are commonly applied physicochemical strategies, but they are unstable and unselective in action. Information about the specific physicochemical milieu that enhances the overgrowth of certain filaments over others may be useful when seeking new control measures. In some instances, this type of knowledge could be applied to adapt operations to control excessive filament growth, while maintaining optimal nutrient removal efficiencies in bioreactors affected by bulking and/or foaming. However, in-depth, local knowledge is required because multiple factors play a role in filament selection.

The primary aim of the project was to bridge the current gap in knowledge pertaining to the filamentous bacterial populations in South African activated sludge wastewater treatment plants with the following specific objectives:

- To identify the dominant filamentous bacteria in activated sludge samples using both molecular and non-molecular methodologies
- To co-ordinate and standardise the collection of physicochemical and microbiological data from 16 carefully selected wastewater treatment plants from three metropolitan areas in South Africa
- To interpret the data and correlate the filament species with measured parameters, wastewater treatment plant configuration, and geographical location
- To provide a validated statistical model for incorporation into the activated sludge BIOS tool currently under development at the Durban University of Technology

Sixteen wastewater treatment plants were identified for the study (six from the City of Cape Town, 6 from Gauteng, and 4 from Durban). Activated sludge samples were taken monthly from each site for a period of 20 months for microscopic analyses: determination of filament index, floc characterisation, and filament identification. *In-situ* measurements of temperature and dissolved oxygen were taken concurrently. For 14 of the 20 months, influent and activated sludge samples were also taken for an extensive range of physicochemical analyses.

Correlation studies and statistical models were screened as tools to assess the selective role of physicochemical parameters on filament morphotypes. The generalised linear model was chosen as the most suitable model and was applied separately to each of the seven most prevalent filaments.



Flow diagram outlining the methodologies used to generate the study outputs

The key findings of the study were:

- Type 0092 (68.9%), *M. parvicella* (39.8%), Type 0041 (33.9%), Type 021N (29.4%), Type 1851 (28.7%), Gram positive branching bacilli (21.5%), Type 0803 (5.9%), and *Thiothrix* spp. (4.5%) were the 8 most prevalent filaments found. Future studies pertaining to filamentous bulking in South Africa should therefore focus on these morphotypes.

- A seasonal distribution of *M. parvicella* was demonstrated in the Cape Town and Gauteng wastewater treatment plants. Contrary to European studies, no seasonal changes in the prevalence of any other filament morphotypes were noted, and there was no significant correlation between temperature and filament selection ($p < 0.05$). It was assumed that this was because the minimum temperature measured in the activated sludge was 14°C, and nutrient removal processes are typically affected at temperatures ≤ 10°C. A current study using amplicon sequencing will clarify whether this lack of seasonal distribution extends to other members of the bacterial consortia in the activated sludge samples.

- Filament selection was significantly ($p < 0.05$) correlated with wastewater treatment plant configuration. While this finding was not unexpected, it has not previously been described in literature. The dominant filament patterns were similar in the Gauteng and Cape Town wastewater treatment plants, while a different distribution was found in Durban. This could have been ascribed to differences in climate, and/or influent characteristics and/or wastewater treatment plant configuration and operation. Researchers and wastewater treatment plant operators should therefore guard against extrapolating results between differently-configured wastewater treatment plants and geographical locations.

- There was a weak but statistically significant ($p < 0.05$) positive correlation between the diluted sludge volume index and the filament index measured in the activated sludge. Amongst other factors, the accuracy of the former test when applied *ex-situ* was brought into question. A follow-up *in-situ* study focussing on the causes of bulking in South African wastewater treatment plants, and the use of laboratory and *in-situ* derived indices as bulking predictors is highly recommended.

- The readily to moderately biodegradable influent organic fractions were efficiently degraded in the Cape Town bioreactors, with only slowly and recalcitrant forms being found in the activated sludge (27% and 83%, respectively). In contrast, the activated sludge from the Durban bioreactors still contained 26% and 39% readily and moderately biodegradable organic fractions, respectively. Further research is needed to determine the reasons for the differences in qualitative organic removal efficiencies.

- The generalised linear model was valid for the filaments that were found at high prevalence in the Cape Town and Gauteng wastewater treatment plants (Type 0092, *M. parvicella* and Type 021N/*Thiothrix* spp.). However, the model was not a good fit for the less prevalent filaments, probably due to lack of data.
    - The only selective criterion for Eikelboom type 0092 was pH, with relative dominance increasing at higher pH values. This finding is supported by previous studies that have shown that Type 0092 does not have exacting nutritional growth requirements. Results also suggested that Type 0092 was not a significant causative agent of filamentous bulking.
    - The food to microorganism ratio was the most important selective parameter for Type 021N/*Thiothrix* spp., with the likelihood of dominance increasing more than 7-fold with each unit increase. In addition, the results strongly suggested that both the quality and quantity of the organic fraction is important for filament selection.

- Lower concurrent concentrations of influent sulphate and activated sludge ortho-phosphate were the most important selective criteria for *M. parvicella,* which was the filament most commonly associated with bulking.

This project successfully bridged the gap in knowledge pertaining to filament prevalence in South African wastewater treatment plants. Results showed conclusively that filament distribution differs according to location and wastewater treatment plant configuration, and that studies which include only limited numbers of wastewater treatment plants are of little value. Ongoing work using samples collected during the course of this project will allow inter-plant comparison of entire bacterial communities at a structural and function level using high throughput molecular methodologies. This will add significantly to the body of knowledge already generated during the course of WRC Project K5/2471/3.

ACKNOWLEDGEMENTS

The authors would like to thank the Water Research Commission (WRC) Research Manager and Administrator, as well as the reference group (RG) members whom attended meetings and provided guidance throughout this project:

Name	Role	Affiliation
Dr J Zvimba	Research Manager	WRC
Mr B Mokgonyane	Project Administrator	WRC
Dr W Rössle	RG member	City of Cape Town (CoCT)
Dr R Magoba	RG member	CoCT
Dr S Surujlal-Naicker	RG member	CoCT
Mr M Vulindlu	RG member	CoCT
Mr J Topkin	RG member	East Rand Water Care company
Dr P Biyela	RG member	University of the Witwatersrand
Mr K Esterhuyse	RG member	City of Tshwane Metropolitan Municipality
Mr G Brown	RG member	Dikubu Water and Environmental Services

Table of Contents

List of Tables

List of Figures

List of abbreviations and chemical formulae:

ANN:	artificial neural networks
AOB:	ammonium oxidising bacteria
AS:	activated sludge
BNR:	biological nutrient removal
BOD:	biological oxygen demand
BOD_t:	total biological oxygen demand
BOD_5:	five day biological oxygen demand
BOD_s:	soluble biological oxygen demand
C:	carbon
CAS:	continuous activated sludge
COD:	chemical oxygen demand
COD_p:	particulate chemical oxygen demand
COD_s:	soluble chemical oxygen demand
COD_t:	total chemical oxygen demand
COHNS:	organic matter
CPT:	Cape Town
DBN:	Durban
DGGE:	denaturing gradient gel electrophoresis
DO:	dissolved oxygen
DSVI:	diluted sludge volume index
EPBR:	enhanced biological phosphate removal
ERWAT:	East Rand Water Care Company
F:	food
FISH:	fluorescent in situ hybridisation
F/M ratio:	food to microorganism ratio
GAU:	Gauteng
M:	Microorganism
ML:	mixed liquor
MLE:	modified Ludzack-Ettinger
MLR:	mixed liquor recycle
MLVSS:	mixed liquor volatile suspended solids
N:	nitrogen
$N_2\uparrow$:	dinitrogen gas
NH_3:	ammonia
NH_4^+:	ammonium ion
NO_3^-:	nitrate
NO_2^-:	nitrite
NO_3^-/NO_2^-:	nitrates and nitrites
NOB:	nitrogen oxidising bacteria
P:	phosphate/phosphorus
PAO:	polyphosphate accumulating organisms
PHA:	poly-B-hydroxyalkanoate

PST:	primary settling tank
RAS:	return activated sludge
SO_4^{2-}:	sulfate
TKN:	Total Kjeldahl nitrogen
TP:	total phosphate
UCT:	University of Cape Town
VFA:	volatile fatty acid
WAS:	waste activated sludge
WWTP:	wastewater treatment plant
3SB:	3-stage Bardenpho
5SB:	5-stage Bardenpho
σ-P:	ortho-phosphate

CHAPTER 1: INTRODUCTION

1. Background

In South Africa, secondary biological treatment of domestic and certain industrial wastewaters typically takes place in centralised municipal facilities. In larger centres, activated sludge (AS) wastewater treatment plants (WWTPs) dominate. All AS systems are designed to reduce the concentrations of organic carbon (C) and ammonia (NH_3). Many are also intended to remove other forms of nitrogen (N), and sometimes phosphate/phosphorus (P). In AS bioreactors, nutrients are either volatilised or incorporated into microbial biomass. In functional WWTPs, the biomass coalesces with inorganic elements and forms flocs which are then removed from the supernatant by gravitational settling, resulting in 'cleaner' water. The biologically driven removal processes are reliant on the continued presence of robust, functionally active microbial consortia, including bacteria which grow in filamentous forms (filaments). Unless present in excess numbers, some filaments (e.g. Eikelboom type 0092) form an integral part of the floc structure, adding stability and assisting floc settling (Eikelboom, 2000; Jenkins et al., 2004). Other filaments, even in small numbers, form bridges between flocs or have hydrophobic cells walls that hamper floc settling (Eikelboom, 2000; Jenkins et al, 2004; Lakay et al., 1999). Over-abundance or selective dominance of filaments, particularly of 'undesirable' species, results in poor settling, causing bulking and/or foaming which negatively affects effluent quality. Despite the fact that enhanced biological nutrient WWTPs have been used in South Africa for decades, bulking and foaming continues to plague many of these facilities.

There is no universal strategy to overcome the excess growth of filamentous bacteria in full-scale WWTPs (Da Motta et al., 2003; Mielczaerk et al., 2012). Chlorination, ozonation and the addition of hydrogen peroxide are some of the common physicochemical strategies used to control filamentous bulking. However, these have been reported to be unstable and unselective in action (Caravelli et al., 2004; Leeuwen 1988; Saayman et al., 1999). Information about specific nutritional and operational requirements that enhance the growth of certain filaments may be useful when seeking means to control overgrowth of these bacteria (Gerardi, 2006). It is possible that in WWTPs affected by bulking and/or foaming, that operational changes can be made to control excessive filament growth, while maintaining optimal nutrient removal efficiencies. However, in-depth, local knowledge is required because multiple factors play a role in filament selection (e.g. Eikelboom, 2000; Casey et al., 1999; Jenkins et al., 2004; Martins et al., 2004; Noutsopoulis et al., 2006).

This project aimed to bridge the current gap in knowledge pertaining to the filamentous bacterial populations in South African AS WWTPs. To achieve this, the WWTP configuration and a number of physicochemical parameters obtained from the analysis of influent and AS

were correlated with the dominance of different filament morphotypes/species. Sixteen WWTPs from three primary geographical locations were included in the study cohort.

1.1 Microbial selection and floc formation in activated sludge wastewater treatment plants

In a functional AS system under aerobic conditions, the synthesis of biomass ($C_5H_7NO_2$) from organic matter (COHNS) takes place according to Equation 1 (Tchobanglous and Burton, 1991).

$$COHNS + O_2 + nutrients \rightarrow CO_2 + NH_3 + C_5H_7NO_2 + other\ end\ products \quad \textbf{Eq.1}$$

Decomposition of the biomass takes place by a process of endogenous respiration. This is enhanced when nutrients have been expended, and is represented by Equation 2 (Tchobanglous and Burton, 1991).

$$C_5H_7NO_2 + 5O_2 \rightarrow 5CO_2 + 2H_2O + NH_3 + energy \quad \textbf{Eq.2}$$

At the beginning of aeration, the organic nutrients, or food (F) is at its highest concentration, and the biomass concentration (M) increases rapidly. During aeration the food/microorganism (F/M) ratio decreases logarithmically until (if and when) the supply of nutrients is expended. At low F/M ratios, the cells of many strains of bacteria stick together forming flocs, which can be separated from the water by settling in a clarifier (Eikelboom, 2000).

For optimal settling results in the clarifier, microbial growth in the mixed liquor (ML) should be in the endogenous phase and excess sludge should be wasted to maintain the correct F/M ratio, as well as the desired concentration of suspended solids and sludge age (Veissman and Hammer, 1998).

For process stability, the aim is to create conditions that allow proliferation of 'desirable' bacterial strains. These strains should consume the largest part of the influent nutrients. Due to ongoing competition for available nutrients, the population changes constantly. The configuration and operation of the WWTP has an effect on the nutrient concentration and loading rate of the AS (Eikelboom, 2000).

According to Eikelboom, the operational and physico-chemical parameters that have the largest effect are on filament selection are: sludge load, influent composition, dissolved oxygen (DO) and temperature (Eikelboom, 2000).

1.2 Effect of aerobic, anoxic and anaerobic zones on microbial processes

Organic C, and sometimes P and/or N (mainly N) are removed from wastewater in WWTPs. The extent to which removal takes place is dictated by the presence or absence of three different zones:

Aerobic zone: In this zone, mechanical aeration is employed to increase the DO concentration. The rationale is that oxygen enhances bacterial uptake and utilisation of organic C for energy by fast-growing heterotrophic bacteria, and facilitates the conversion (oxidation) of NH_3 into nitrite (NO_2^-) and nitrate (NO_3^-) by slow-growing nitrifying bacteria. The nitrifying bacteria are autotrophic, meaning that they obtain their energy from the oxidation and/or reduction of inorganic molecules. Autotrophic metabolism is not as efficient as heterotrophic metabolism. Nitrifiers therefore grow more slowly, and are less robust than their heterotrophic counterparts.

Nitrification itself takes place by the sequential oxidation of NH_3 to NO_2^-, and NO_2^- to NO_3^- by different nitrifying bacterial species: ammonium oxidising bacteria (AOB) and nitrite oxidising bacteria (NOB), respectively. To ensure stable nitrification, process parameters need to be monitored and adjusted where possible to keep the ratio of chemical oxygen demand (COD) to total Kjeldahl N (TKN), as well as the alkalinity, DO, pH, retention time, and temperature within a specified range conducive to the establishment of a functional nitrifying population. Nitrifiers are also particularly susceptible to many toxins and metals, and population losses may take more than 14 days (1-2 sludge ages) to be restored.

Anoxic zone: This zone is not aerated and is defined by the presence of nitrates/nitrites (NO_3^-/NO_2^-) and low DO. Depending on the configuration of the WWTP (Section 1.1.3), an upstream aeration zone may or may not be present. If present, NH_3 levels in the anoxic zone should be lower, and NO_3^-/NO_2^- higher, than in configurations where there is no upstream aerobic zone.

Denitrification is the mineralisation of NO_3^- and NO_2^- to dinitrogen gas ($N_2\uparrow$), catalysed by denitrifying heterotrophic bacteria. In the absence of molecular oxygen, these bacteria can utilise NO_3^- and NO_2^- as terminal electron acceptors during respiration. Organic molecules from the wastewater and/or decaying biomass provide the energy for oxidation, so that N reduction is coupled with the oxidation of organic C. When the decaying biomass is the source of energy, the process is known as endogenous respiration. The COD:N ratio is a critical functional parameter because sufficient organic C is required for effective denitrification.

Anaerobic zone: This zone is defined by the lack of both oxygen and NO_3^-/NO_2^-, and is therefore always located at the influent side of the reactor. Under anaerobic conditions, some heterotrophic bacteria, known as polyphosphate-accumulating organisms (PAOs) can take up C sources such as volatile fatty acids (VFAs) and store them as poly-B-hydroxyalkanoates (PHAs). Energy for this is partly supplied by polyphosphate hydrolysis, and ortho-phosphate (σ-P) is released into the bulk liquid. When the wastewater enters the anoxic and aerobic zone/s, the PAOs have a competitive advantage because they can immediately metabolise the stored energy for nutrient uptake and growth. The PAOs therefore take up the σ-P from the bulk liquid at a high rate. In theory, the subsequent removal of PAOs with the rest of the biomass in the clarifier sludge results in the effective removal of phosphate from the wastewater. Systems that operate with anaerobic zones are known as enhanced biological phosphate removal (EBPR) systems.

1.3 Filamentous populations in activated sludge

To assess settling and microbial abundance in AS, laboratory analyses in SA are typically limited to quantitative *ex-situ* macroscopic tests. Settling indices [diluted sludge volume index (DSVI) and/or sludge volume index (SVI)] are commonly determined as indicators of settle-ability (discussed in detail in Section 4.1), and the mixed liquor volatile suspended solids (MLVSS) concentration is sometimes determined as a semi-quantitative predictor of microbial abundance. However, the underlying aetiology of settling problems can only be properly diagnosed using a combination of visual observations at the WWTPs (e.g. to check for rising sludge), and microscopic evaluation of the character of the AS to determine the floc and filament characteristics. If it is established that filamentous bulking is the 'culprit', it may be possible to limit the growth of the filament specie/s involved by modifying the operational conditions to adjust the physicochemical environment. However, the relationships between environmental parameters and filament species need to be properly established.

1.3.1 Common methods used for filament identification: pros and cons

Academia and advanced wastewater laboratories have systematically moved away from classical microscopy to molecular methods for the identification of microbial populations in AS (Morgan-Sagastume et al., 2007). Skilled staff are required for both, but the former may still be more suited to a routine setting as it is more cost effective, with a faster turn-around time, and can be used to identify protozoan, metazoan, and filamentous bacterial communities simultaneously (Eikelboom, 2000; Jenkins et al., 2004; Salvado et al., 1995). The drawbacks are that it cannot distinguish between different taxa that are morphologically the same, filaments from the same taxa may classed separately if they exhibit variable morphology, and filaments located inside the flocs are only visible with staining (although, for practical purposes, it is the filaments outside of the flocs that are responsible for bulking and foaming) (Liao et al., 2003; Hug et al., 2005; Gulez et al., 2008).

The most commonly employed molecular methods are fluorescent in-situ hybridization (FISH) and denaturing gradient gel electrophoresis (DGGE). 16S rRNA DGGE can be highly sensitive and has been shown to detect organisms representing as little as 1% of the population (Muyzer et al., 1992). However, several phylogenetic species can form part of a single band, extensive clone libraries need to be established for identification, and the clone frequencies in the libraries do not accurately reflect *in-situ* quantities (Eschenhagen et al., 2003; Sekiguchi et al. 2001).

FISH allows for the definitive identification of a range of microorganisms and is particularly useful in detecting non-filamentous organisms (Eschenhagen et al., 2003). Unlike DGGE, FISH can be used successfully to identify and quantitate species or groups in AS (Hug et al., 2005, Morgan-Sagastume et al., 2007). However, the technique is technically challenging, and separate probes and hybridisation conditions are needed for each filament type, (Liu et al., 2001).

1.4 Wastewater treatment plant configurations

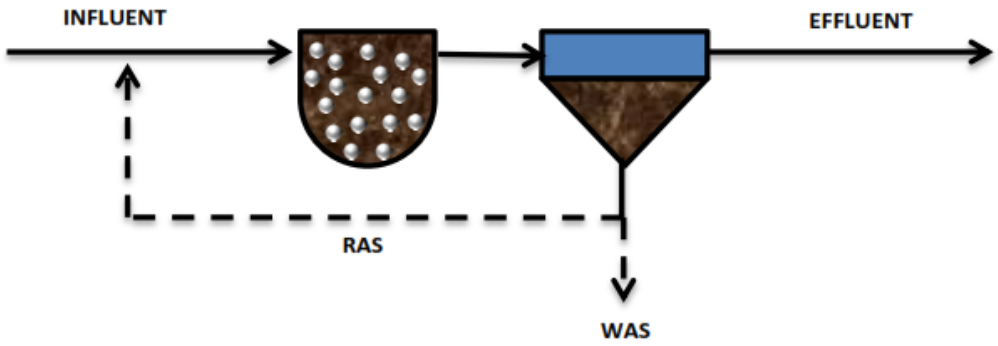
In different bioreactor configurations, the zones (number, size, location, type/s), and the physical structures for wastewater and sludge wasting and/or recycling are engineered differently. The major factors to be considered when selecting a configuration are the capital and operational costs, the character of the influent, and the final effluent quality requirements. The basic principles of the configurations of the WWTPs included in this project have been briefly described and shown schematically in Figure 1.

The continuous activated sludge (CAS) process configuration (Figure 1A) is limited to one aerobic zone and is designed to achieve removal of organic C, with some concurrent nitrification, but no denitrification.

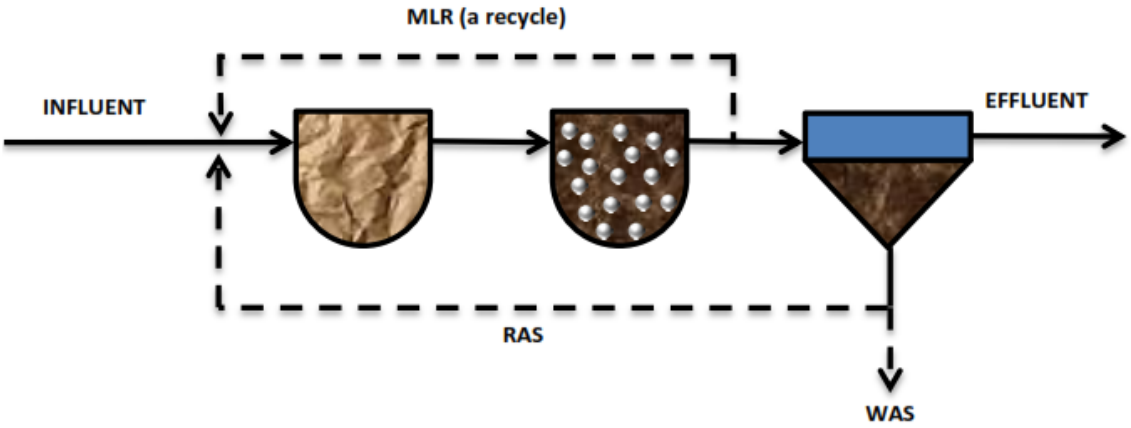
The Modified Ludzack-Ettinger (MLE) process configuration (Figure 1B) is comprised of an anoxic zone followed by an aerobic zone. In the aerobic zone, enhanced C utilisation takes place, as well as nitrification of NH_3 to NO_3^-/NO_2^-. The ML from the downstream aerobic zone is recycled back to the anoxic zone, where the NO_3^-/NO_2^- are denitrified and organic C is 'converted' into biomass. Only a fraction of ML is recycled, so not all of the NO_3^-/NO_2^- can be removed.

The 3-stage Bardenpho (3SB) and University of Cape Town (UCT) process configurations (Figure 1C, 1D) are biological nutrient removal (BNR) systems that are designed to remove C, N and P. They are comprised of concurrent anaerobic, anoxic and aerobic zones. The difference between the two lies in the paths for the (i) return activated sludge (RAS) from the clarifier and, (ii) ML recycle/s (MLR) from within the bioreactors. In both processes, a high fraction of the ML that has been nitrified in the aerobic zone is recycled back to the anoxic zone to be denitrified (a-recycle), and to provide electrons for denitrification via endogenous decay. In the case of the 3SB process, P removal is retarded to varying degrees by the presence of NO_3^-/NO_2^- from the RAS (Figure 1C). The UCT system strives to overcome this by: (i) returning the RAS to the anoxic zone, instead of the anaerobic zone, and (ii) the inclusion of an anoxic-anaerobic MLR path instead (r-recycle). However, due to the presence of NO_3^-/NO_2^- in the anoxic zone from the UCT aerobic-anoxic MLR, there is still a risk of adding these to the anaerobic zone. In the modified UCT system, the anoxic zone is therefore compartmentalised.

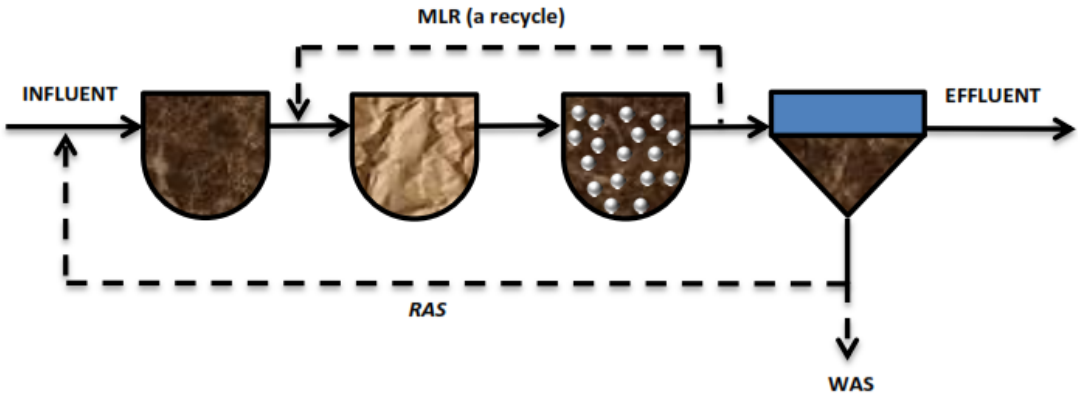
The 5-stage Bardenpho (5SB) process configuration (Figure 1E) is applied where very low concentrations of P and N are required in the final effluent, for example, in cases where the treated effluent will be discharged into environmentally sensitive areas. In comparison to the 3SB process configuration, the 5SB incorporates an additional anoxic and aerobic zone for denitrification and nitrification, and to ensure that no residual NO_3^-/NO_2^- are present in the clarifier which can enter the RAS and retard the growth of PAOs.



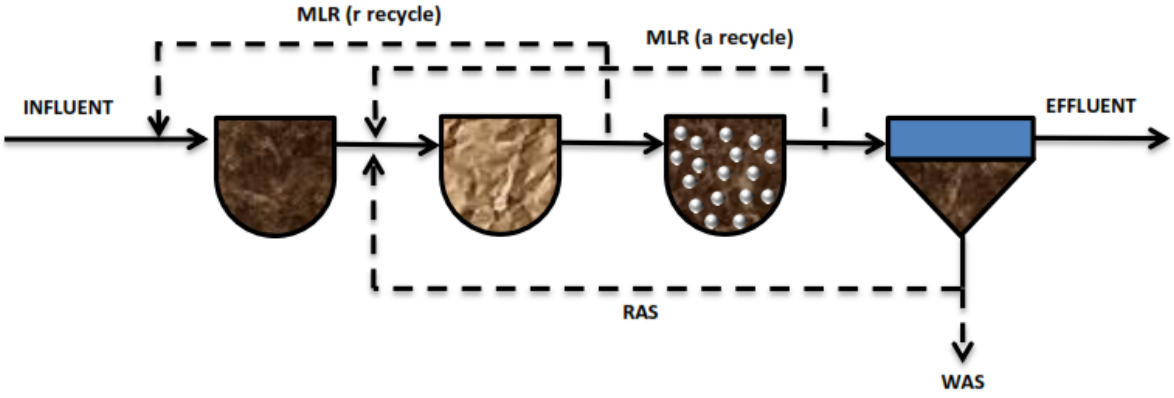
1A: Continuous activated sludge process configuration



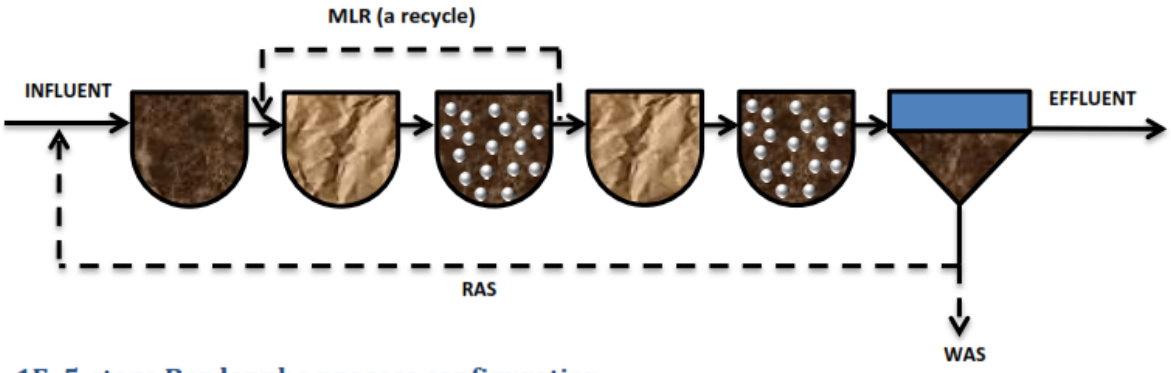
1B: Modified Ludzack-Ettinger process configuration



1C: 3-Stage Bardenpho process configuration



1D: University of Cape Town process configuration



1E: 5-stage Bardenpho process configuration

Figure 1: Schematic diagrams of the five configurations (A-E) of the wastewater treatment plants included in the study cohort. The waste activated sludge may be taken from the clarifiers (as per 1A-E), or directly from the aerobic reactors.

CHAPTER 2: MATERIALS AND METHODS

2.1 Selection of wastewater treatment plants

Comprehensive operational details of WWTPs in Gauteng (GAU) were obtained from the East Rand Water Care Company (ERWAT) (example given in Appendix 1). The City of Cape Town (CPT), and GAU WWTPs were compared and the closest matches were selected for inclusion in the study. The primary criteria used for selection and geographical matching were configuration and influent wastewater character. Other factors that were considered were the presence/absence of primary settling tanks and the type of aeration employed. It was recognised that any significant industrial influent would complicate a comparative study. It was therefore decided that, wherever possible, domestic effluent should be used. Where this was not possible, the WWTPs with the lowest fraction of industrial influent were included. In cases where there was more than one reactor with the same configuration, the same reactor was used for the duration of the study. To maintain anonymity, alphanumeric designations were given to the WWTPs (Table 1).

With the exception of C1 and C4, surface aeration was applied to all. At C1, fine bubble aeration was applied, while at C4, both fine bubble and surface aeration were applied, the latter being in the form of a float. Primary settling tanks (PSTs) were used in:

(i) Both of the 5SB-configured WWTPs in Cape Town (CPT), but not in the 5SB-configured WWTP in GAU.
(ii) Two of the four 3SB-configured WWTPs in Gauteng.
(iii) One of the three CAS-configured WWTPs in Durban (DBN; Table 2). At the rest of the cohort WWTPs, screened raw formed the reactor influent.

Table 1: Use of primary settling in the WWTP study cohort

	Gauteng		Cape Town		Durban	
Configuration	**WWTP**	**PST**	**WWTP**	**PST**	**WWTP**	**PST**
5SB	G1	no	C1	yes	none	
			C2	yes		
MLE	G2	no	C3	no	none	
			C4	no		
3SB	G3	yes	C5	no	none	
	G4	no				
	G5	yes				
	G6	no				
UCT	none	no	C6	no	D2	no
CAS	none		none		D1	yes
					D3	no
					D4	no

PST – primary settling tank

2.2 Physicochemical analyses

2.2.1 Sampling and in-situ analyses

Monthly samples were taken for microscopic evaluation (310 samples: Aug 2015-Mar 2016) and physicochemical analyses (224 samples: Feb 2016-Mar 2017). As far as possible, samples were taken during the same week of the month from each site. Influent samples and AS samples from each WWTP reactor were taken on the same day. In WWTPs with PSTs, grab samples were taken from the settled influent to the reactors. In WWTPs with no PSTs, 24 hr composite influent samples were taken. The AS (grab samples) were taken from the exits from the final aerobic zones of the reactors. DO and temperature readings were taken manually *in-situ* by submerging DO temperature probes in the AS at a minimum of three sites, and taking an average of the readings.

2.2.2 Laboratory tests and derived parameters

All influent chemical parameters were determined on whole, freshly mixed samples. The MLVSS concentration and DSVI were determined on whole, freshly mixed samples according to standard methods. Both total (whole) and soluble (filtered) forms of COD and BOD were determined [(COD_t and BOD_t) and (COD_s and BOD_s)]. Only soluble forms of sulfate (SO_4^{-2}-S), NH_4^+-N, and NO_3^-/NO_2^--N were determined (Table 2). Total alkalinity (T.alk) and VFA concentrations were not determined in AS samples.

Table 2: Chemical analyses performed on influent and activated sludge samples

Sample	pH	BOD	COD	T.alk	VFA	TP	σ-P	SO_4^{-2}-S	NH_4^+-N	NO_3^-/NO_2^--N
Influent	√	√	√	√	√	√	√	√	√	√
AS	√	√*	√*			√	√**	√**	√**	√**

*whole and filtered samples **filtered samples only

Table 3: Methods applied for the determination of chemical data

		CPT	GAU	
COD	**Method**	Colorimetric: Dichromate reflux and titration	Colorimetric: Dichromate reflux and spectrophotometry	Colorimetric :Dichromate reflux and spectroscopy
	Equipment	- Gerhardt (Köningswinter, Germany) COD digester - Metrohm Titrando auto titrator (Herisau, Switzerland)	- Hach (Loveland, USA) 45600-00 COD digester - Hach DR 5000 UV/VIS Spectrophotometer	- Merck Thermoreaktor TR300 digester (Massachusetts,USA) - Gallery Plus Automated Photometric Analyzer
	Reference	ISO 15705 & APHA 5220-B	APHA 5220-B	APHA 5520-B
BOD_5	**Method**	Respirometric	Respirometric	Respiromentric
	Equipment	Oxitop (WTW, Weilheim, Germany) benchtop measuring device	Loviond Oxidirect® BSB/BOD instrument (Amesbury, UK)	Oxitop (WTW, Welheim, Germany) benchtop measuring device
	Reference	According to manufacturers' instructions	According to manufacturers' instructions	According to manufacturer's instructions
T. alk	**Method**	Automated potentiometric titration	Potentiometric titration	Automated potentiometric titration
	Equipment	Aquakem Discreet Analyser [Thermo Fisher (Waltham, USA)]	Digital burette	Mettler Toledo(Columbus, USA) instrument
	Reference	ISO 9963 & APHA 2320	APHA 2320	According to manufacturers' instruction
TP	**Method**	Colorimetric: Persulfate digestion with ammonium molybdate/ascorbic acid	Colorimetric: Persulfate digestion with ammonium molybdate/ascorbic acid	Colorimetric: Persulphate digestion with ammonium molybdate/ascorbic Acid
	Equipment	Lachat (Milwaukee, USA) QuikChem 8000 series FIA	Lachat QuikChem 8000 series FIA	Gallery Plus Automated Photometric Analyzer (Thermo Fischer, Waltham, USA)
	Reference	According to manufacturers' instructions based on ISO 15681	According to manufacturers' instructions based on ISO 15681	EPA Method 365.1
σ-P	**Method**	Colorimetric: Ammonium molybdate/ascorbic acid	Colorimetric: Ammonium molybdate/ascorbic acid	N/A
	Equipment	Lachat QuikChem 8000 series FIA	Lachat QuikChem 8000 series FIA	N/A
	Reference	ISO 15681	ISO 15681	N/A
SO_4^{-2}	**Method**	Turbidimetric: Barium chloride/hydrochloric acid	Turbidimetric: Barium chloride/hydrochloric acid	Turbidimetric: Barium chloride/hrdrochloric acid
	Equipment	Lachat QuikChem 8000 series FIA	Lachat QuikChem 8000 series FIA	Gallery Plus Automated Photometric Analyzer
	Reference	According to manufacturers' instructions based on ISO 15923-1 & APHA 4500 SO_4-E	According to manufacturers' instructions based on ISO 15923-1 & APHA	EPA Method 375.4
NH_4^+	**Method**	Colorimetric: Bertholot method	Colorimetric: Bertholot method	Colorimetric: Hypochlorite
	Equipment	Lachat QuikChem 8000 series FIA	Lachat QuikChem 8000 series FIA	Gallery Plus Automated Photometric Analyzer
	Reference	According to manufacturers' instructions based on ISO 11732 & APHA 4500 NH3-F	According to manufacturers' instructions based on ISO 11732 & APHA 4500 NH3-F	Ammonia in Waters (1981) Methods for the examination of water and associated materials ISBN0117516139
NO_3^- /NO_2^-	**Method**	Colorimetric: hydrazine/sulfanilimide	Colorimetric: hydrazine/sulfanilimide	Colorimetric: Hydrazine/sulfanilimide
	Equipment	Lachat QuikChem 8000 series FIA	Lachat QuikChem 8000 series FIA	Gallery Plus Automated Photometric Analyzer
	Reference	According to manufacturers' instructions based on ISO 13395 & SM 4500 NO3-H	According to manufacturers' instructions based on ISO 13395 & SM 4500 NO3-H	EPA Method 353.1

ISO = International Standards Organisation FIA = flow injection analyser

APHA: American Public Health Association, American Water Works Association and Water Environment Federation (2005). Standard Methods for the examination of water and wastewater, 21st Edition. Washington DC.

2.3 Microscopic evaluation and filament identification

The procedures used to analyse the AS are described in Section 2.3.1 to 2.3.3

2.3.1 Wet mount

Preparation

The samples were thoroughly mixed by gently inverting 8 times, and two wet mounts of the correct thickness were prepared. If the samples were too dense, dilutions were made to ensure good visualisation of the floc and filament characteristics.

Determination of floc characteristics

The slides were examined under 100x and 400x magnification (10x and 40x phase contrast objective lenses), and the characteristics were noted according to the method described by Eikelboom (2000):

- Size: Small (<25 um), medium (25-250 um), large (>250 um). In most cases, the size of the flocs was variable, and each was reported if they constituted >25% of all flocs.
- Shape: Round or irregular – whichever was dominant.
- Structure: Compact or open – whichever was dominant.
- Strength: Firm or weak – whichever was dominant.

Determination of dominant filaments

The filaments were identified by classical microscopy using the methods, charts and electronic data published by Eikelboom (2000). Both slides were scanned rapidly from left to right and top to bottom. If filaments are absent, or only present in occasional flocs (index 0 to 1), the examination was not continued. If filaments were commonly observed, the characteristics of the dominant filaments were noted. If the characteristics were unclear at low magnification, the slides were examined using the 100x bright field objective lens (1000x magnification), with a drop of oil on the coverslip. The following characteristics were noted for each dominant filament morpho-type:

- Branching (present/absent).
- Filament shape (straight, curved, twisted).
- Attached growth (abundant/scanty/absent).
- Septa (present/absent).
- Cell shape (round, square, rectangular, etc.).
- Sheath (present/absent).
- Sulphur granules (present/absent).

2.3.2 Stains and confirmatory tests

Preparation

The samples were thoroughly mixed by gently inverting 8 times, and light smears were prepared and allowed to air dry. The slides were stained by both the Gram and the Neisser method.

Determination of filament indices

The Gram-stained slides were examined under 400x magnification (phase-contrast) and, if required, 1000x magnification (bright field/oil immersion). The filament index (FI) was recorded using the criteria shown in Table 4.

Table 4: Criteria used to determine the filament index of activated sludge samples

FI	Abundance	Explanation
0	None	No filaments present
1	Few	Filaments present, but only occasionally
2	Some	Filaments commonly observed, but not present in all flocs
3	Common	Filaments observed in all flocs, but at low density (average 1 to 5 per floc)
4	Very common	Filaments observed in all flocs at medium density (average 5-20 per floc)
5	Abundant	Filaments observed in all flocs at high density (average >20 per floc)
6	Excessive	Filaments appearing as more filaments than flocs and/or filaments growing in high abundance in the bulk solution

Determination of filament characteristics

Slides were examined under 400x magnification (phase-contrast) to determine which filaments were dominant, and then under 1000x magnification (bright field/oil immersion) to visualise the features. The filament characteristics were matched with those seen in the wet mount, and the staining characteristics were noted: Gram staining (Gram positive, Gram negative or Gram variable) and Neisser staining [Neisser positive (grey-violet), Neisser negative (colourless or light pink), presence of polyphosphate granules (blue-black).

A sulphur storage test was performed, and wet mounts were examined for the presence of yellowish, highly refractive intracellular granules. If it was not possible to discern whether sheaths were present or absent, sheath stains were performed.

Fluorescent in-situ hybridisation

The identity of the seven most prevalent filament types was confirmed using FISH according to the method described by Amann et al. (1990a, 1990b) and Nielsen (2009a) and the probes shown in Table 5, with the hybridisation conditions as described by the respective authors (Table 5).

Table 5: Oligonucleotide probes used to confirm the identity of the most prevalent filament types

	Probe	Sequence (5'-3')	Reference
Eubacteria	EUB338	GCT GCC TCC CGT AGG AGT	Amann et al., 1990
γ-Proteobacteria (all)	GAM42a	GCC TTC CCA CAT CGT TT	Manz et al., 1992
Type 021N, *Thiothrix* spp.	G123T G123T	CCT TCC GAT CTC TAT GCA (P) CCT TCC GAT CTC TAC GCA	Kanagawa et al., 2000
Chloroflexi	CFX_{mix}	CCA TTG TAG CGT GTG TGT MG & AAA CCA CAC GCT CCG CT	Bjornsson et al., 2002
Type 0092	CFX223 CFX197	GGT GCT GGC TCC TCC CAG TCC CGA AGC GCC TGA ACT	Spiers et al., 2009
Type 1851	CHL1851	AAT TCC ACG AAC CTC TGC CA	Beer et al., 2002
Type 0041/0675	TM7905	CCG TCA ATT CCT TTA TGT TTT A	Hugenholtz et al., 2001

CHAPTER 3: FILAMENTOUS POPULATIONS

3.1 Relative abundance and prevalence of filamentous bacteria

3.1.1 Relative abundance of filaments from all three locations combined

Monthly samples were taken from August 2015 to March 2017 (20 months). From a potential 320 samples, dominant filaments were identified in 289 (90%). Of the remaining 10%, the filament index (FI) was ≤1 in 6%, no dominant filaments were seen in 1%, and in 3% of cases samples were not taken because of logistical reasons. In each sample examined with FI>1, between 1 and 4 filaments dominated. Filaments were ranked in order of abundance. Secondary species were ignored.

A total of 696 filaments were identified. Of these, Type 0092 was the most abundant dominant filament, followed by *M. parvicella*, Type 0041, Type 021N, Type 1851, and Gram positive branching bacilli (GPBB) (Figure 2). The remainder constituted <2%.

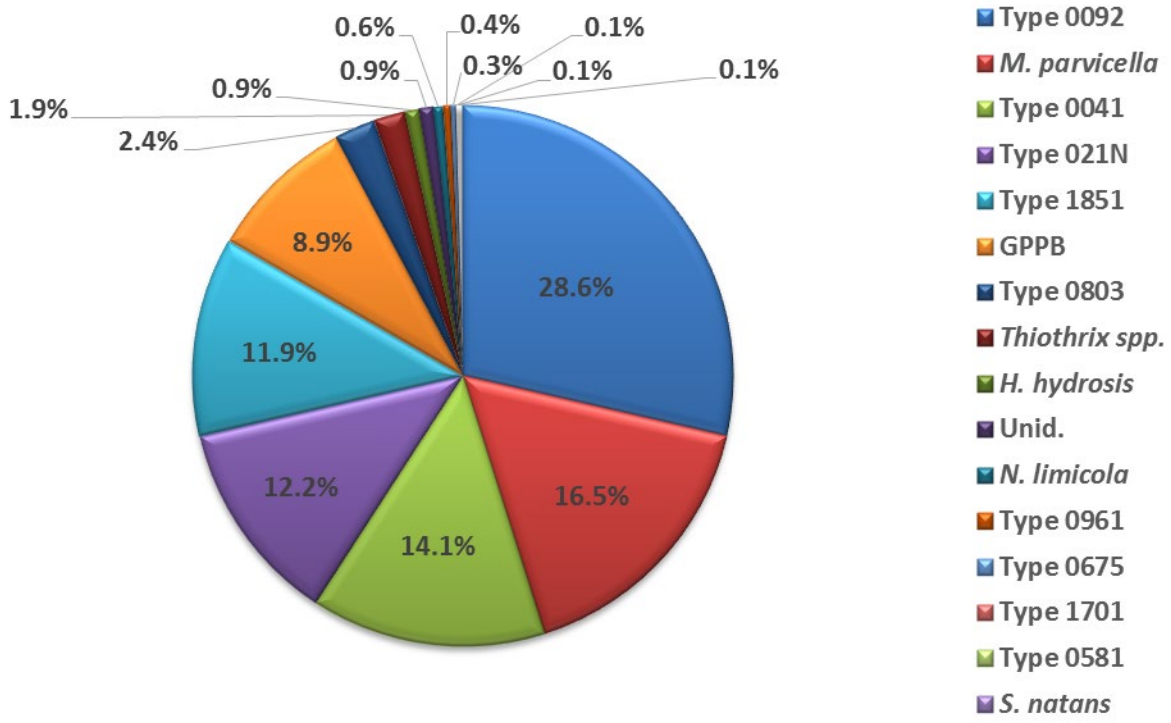


Figure 1: Relative abundance (%) of filaments identified in monthly samples from all three centres from August 2015 to March 2017 (*n* = 696 filaments identified)

3.1.2 Overall filament prevalence and comparison of prevalence from different locations

Type 0092 was found as a dominant filament in 68.9% of the 289 samples that were examined with FI>1. This was followed by *M. parvicella* (39.8%), Type 0041 (33.9%), Type 021N (29.4%), Type 1851 (28.7%) and GPBB (21.5%). The prevalence of other filaments was <6% (Figure 3).

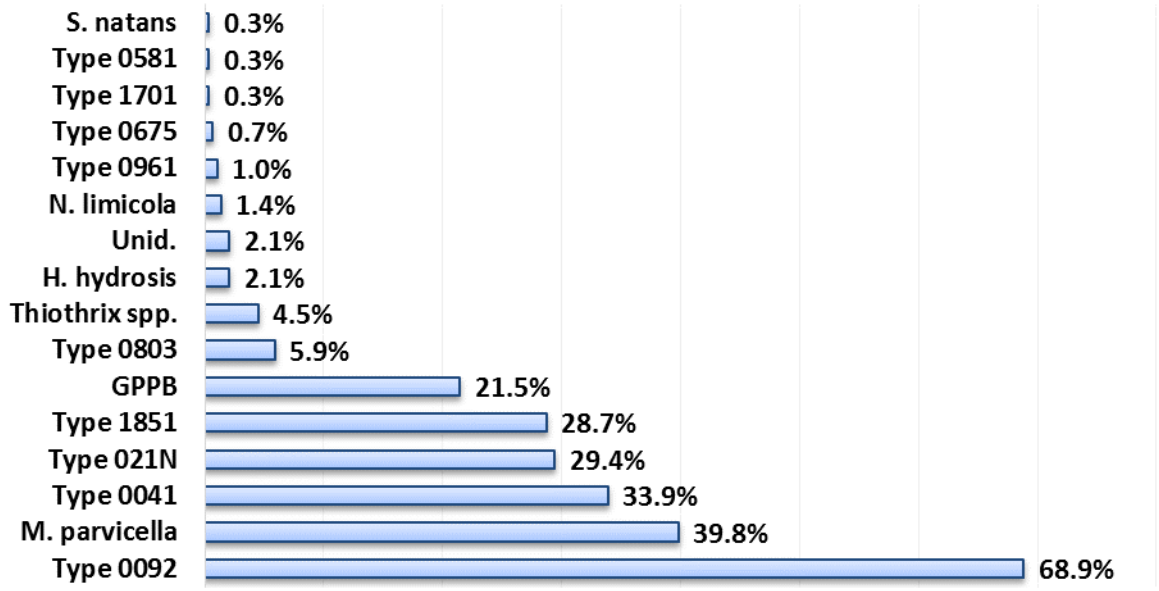


Figure 2: Prevalence of filaments identified in monthly samples with FI>1 from all three centres combined from August 2015 to March 2017 (*n* = 289 samples)

Major similarities and differences were noted in the prevalence of dominant filaments from the three different locations (Figure 4). In order of abundance of filaments constituting >1% abundance at any of the three locations, the most notable results were:

(i) The prevalence of both Type 0092 and *M. parvicella* was not significantly different in GAU and CPT, despite the sites having WWTPs with different reactor configurations and industrial/domestic wastewater inputs.

(ii) Although Type 0092 was the most prevalent filament in all three locations, the prevalence was significantly lower in DBN (57.5%) than in CPT (75.3%), or GAU (72.3%).

(iii) The prevalence of *Microthrix parvicella* was higher in DBN (56.3%) than in CPT (36.1%) or GAU (36.6%).

(iv) The prevalence of Type 1851 was significantly higher in DBN (52.5%) than in CPT (15.5%) or GAU (24.1%).

(v) There was a moderate prevalence of Type 021N at all 3 locations (46.4% in CPT, 36.6% in GAU, 28.8% in DBN).

(vi) Although prevalent in all 3 locations in >10% of samples, there were significant differences in the prevalence of GPBB and Type 0041, with the prevalence in DBN and CPT being highest

and lowest, respectively (38.8% and 51.3% in DBN, 16.1% and 36.6% in GAU, 10.3% and 16.5% in CPT for GPBB and Type 0041, respectively).

(vii) *Thiothrix* spp. were not present as dominant species in GAU. However, the relatively high prevalence in CPT could be ascribed to a single WWTP that was experiencing severe bulking from late in 2016 until the end of the study in March 2017 (C5).

(viii) *Nostocoida limicola* (3.8%) and Type 0675 (1.3%) were only present as dominant morphotypes in DBN. The former was found in 2 of the 4 DBN WWTPs.

(ix) Type 0961 (3.1%) was only present as a dominant filament in CPT, being found in 2 of the 6 WWTPs at this location.

(x) *Haliscomenobacter hydrosis* (5.4%) and Type 0803 were only found as dominant filaments in GAU. In the case of the former, it was found in only one WWTP (G4) from June to October 2016.

The inter-site differences in filament prevalence could be ascribed to a number of factors, including:
(i) Climatic differences – DBN has less severe winters than the other two centres, while CPT is the only winter rainfall area. To determine the role of climate, correlations of filament types and temperature were determined as part of the statistical analysis. Where applicable, dilution of influent by rainwater were reflected in the physicochemical data, obviating the need to measure rainfall and include in the statistical analysis.

(ii) Influent characteristics – for example, two of the four WWTPs in the DBN cohort accept >50% industrial effluent, while the GAU WWTPs ostensibly accept 100% domestic wastewater, mostly raw (unsettled). It must be noted that dumping of chemicals into toilets and sewers may occur, which may change the composition of domestic effluent.

(iii) WWTP configuration and operation (Section 3.4)

Despite the differences alluded to in (i-iii), the pattern of filament prevalence in GAU and CPT were remarkably similar.

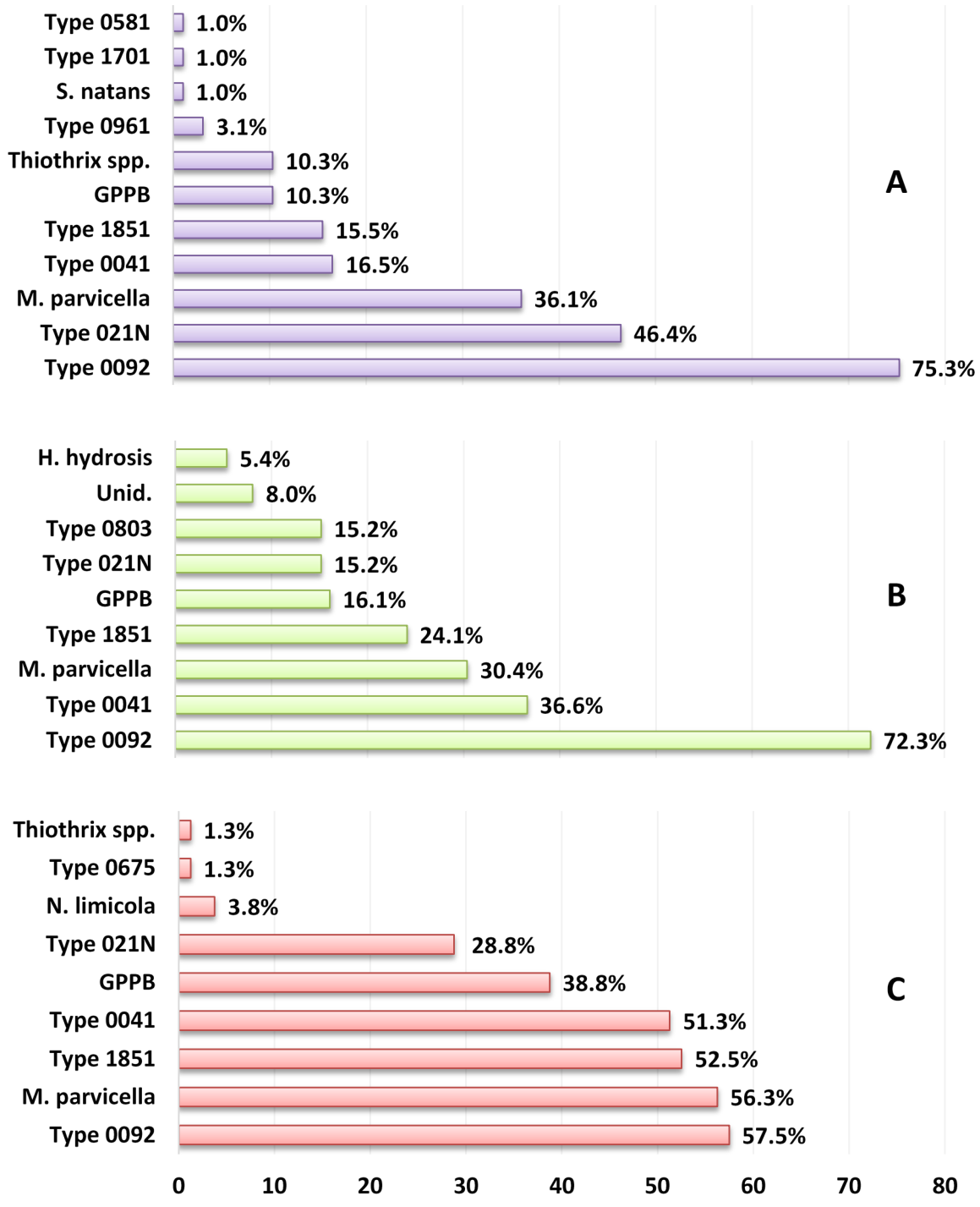


Figure 3: Prevalence of different filaments >1% from each site as dominant morphotypes in samples taken from the City of Cape Town (A), Gauteng (B), and Durban (C)

3.2 The effect of seasonality on the selection of dominant filament types

3.2.1 Overall seasonal dominance

In any outdoor environment, temporal shifts occur in the microbial populations. It has been shown that species succession occurs in the filamentous populations in the WWTPs in South Africa. However, no long term ecological studies have been conducted to ascertain to what degree these shifts are seasonal. Seasonal shifts can be climatic and/or dependent on other factors such as changes in human behaviour, and seasonal population changes (e.g. tourist season in Cape Town). During 2017-2018, there was a severe drought in Cape Town. Eventually, due to water-saving efforts, the hydraulic load to the WWTPs decreased. This was accompanied by an increase in the concentration of the influent (but not the organic loading rate). These changes were only demonstrated later in the drought period as water restrictions became more stringent. During the actual period of this study, the influent concentrations were compared on a temporal basis during the no drought/drought periods, and it was shown that they remained within similar ranges (Section 4).

One of the aims of this study was to establish the presence or absence of distinct seasonal shifts, and to determine to what extent the different climates in South Africa impact on these. To establish seasonality regarding the filamentous population, filament identification was performed over the period of 20 months to allow a year-on-year overlap from the end of winter to the beginning of autumn.

It must be stressed that for this study, only the dominant filaments were identified. In some instances only one or two filaments were dominant, although there were invariably a number of secondary filaments. This is not at all unexpected. Changes in process parameters or climatic conditions can lead to seasonal shifts and allow secondary filaments to become dominant.

During the study period, there were monthly changes in the dominant filamentous populations from an overall perspective (Figure 5), as well as from the three different geographical locations (Figure 6).

Some observations were made from an overall perspective of the six most prevalent dominant filaments, in order of prevalence:

(i) No distinct seasonal trend was observed in the prevalence of Type 0092.

(ii) *M. parvicella* appeared to be more prevalent in winter/spring and less prevalent in summer/autumn.

(iii) Although there were significant monthly fluctuations in the prevalence of Type 0041, Type 021N and Type 1851, this did not appear to be seasonal (year-on-year).

(iv) GPBB prevalence tended to decrease during the warmer months.

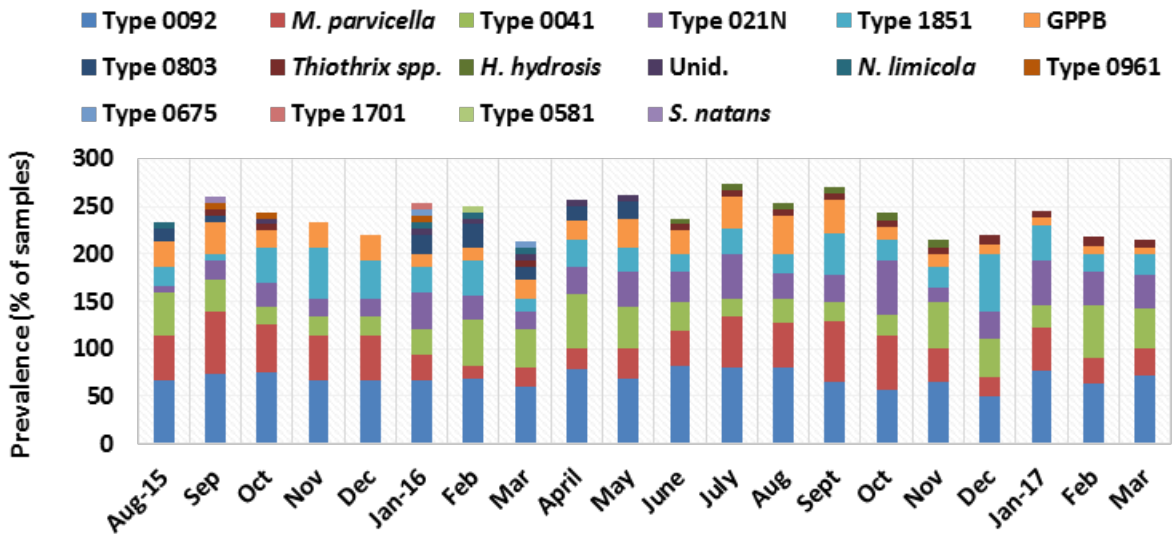


Figure 4: Monthly variations in the prevalence of dominant filaments from all wastewater treatment plants combined from August 2015 to March 2017

3.2.2 Seasonal dominance in different study sites

To ascertain whether there were any seasonal trends in the prevalence of dominant filaments in the individual locations that were masked when the results were analysed as combined data, the results were analysed separately from each site. Figure 5 shows the results of the filaments prevalent in >10% of samples from the individual WWTPs. *Thiothrix* spp. has been excluded from Figure 5A because it was only found at 1 WWTP, and prevalence was not seasonal. When comparing the data for each study site separately, the tendency of GPBB to decrease in the warmer months (Section 3.3.1), was not observed. The only filament that showed distinct seasonal prevalence was *M. parvicella*, but only in the WWTPs from GAU and CPT.

Some researchers have suggested that there is an inverse correlation between the presence of *M. parvicella* and Type 0092 (i.e. they are niche competitors). The results of this study do not support this hypothesis.

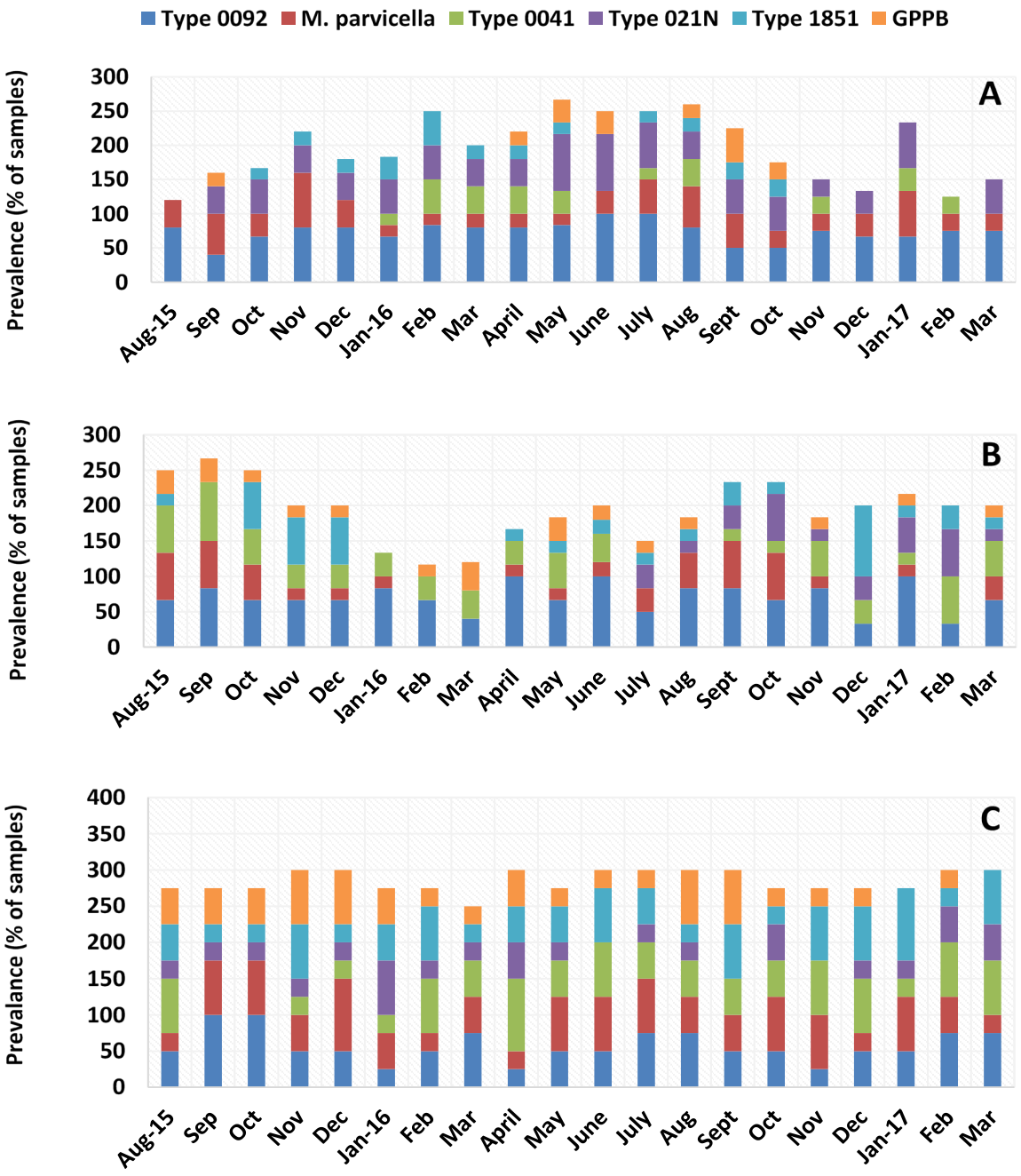


Figure 5: Monthly variations in the six most prevalent filaments from Cape Town (A), Gauteng (B) and Durban (C)

There were distinct seasonal trends in the prevalence of *M. parvicella* in CPT and GAU (Figure 7), with the highest prevalence being found in the colder months. This seems counter-intuitive because the prevalence of *M. parvicella* was >20% higher in DBN than in either of the other locations. In addition, some outliers were seen in the CPT cohort (Nov 2105, Jan 2016, encircled in red in Figure 7). These results suggested that temperature may play a significant role in selection of *M. parvicella*, but that other factors are also important selection criteria.

Due to non-normality of data, the Mann-Whitney non-parametric test was used to determine the statistical significance of the difference in temperature between the WWTPs where *M. parvicella* was either dominant or not-dominant (Appendix 3). Results showed that temperature was not a significant parameter ($p>0.05$) for selection of any filaments, including *M. parvicella*). Any further assessment about the aetiology of the seasonal pattern for *M. parvicella* prevalence would have been purely speculative.

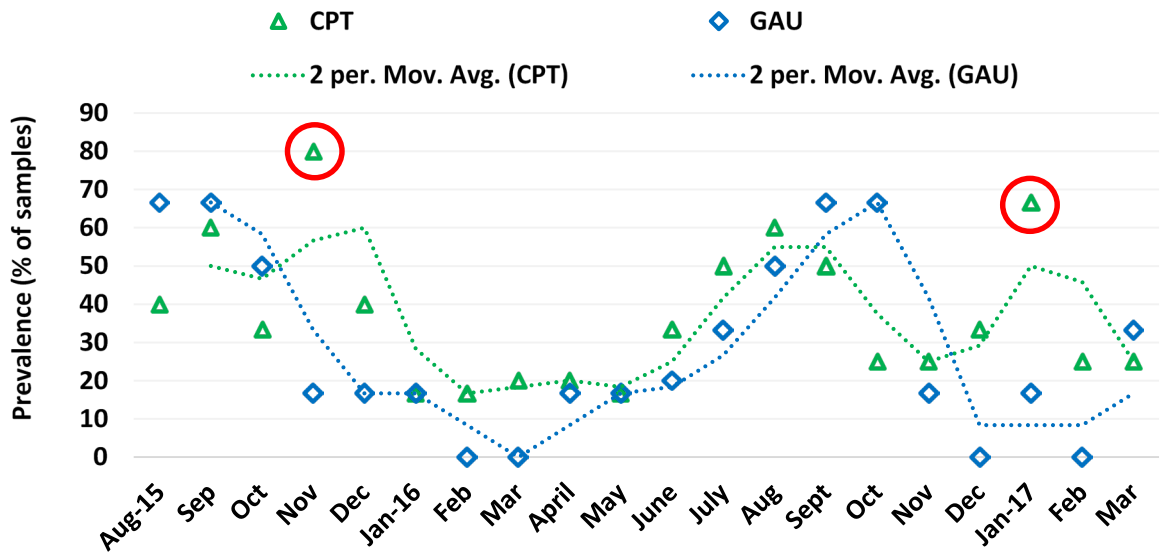


Figure 6: Seasonal variation in the prevalence of *Microthrix parvicella* as a dominant filament in the study cohorts from Cape Town and Gauteng

3.3 The effect of wastewater treatment plant configuration on the selection of dominant filament types

Visual analysis of graphs plotting the prevalence of different filaments in all the WWTPs with different configurations (Figure 8A) and the CPT and GAU WWTPs (Figure 8B) strongly suggested that configuration plays a significant role in filament selection. The DBN WWTPs were removed from the graph to attempt to discount the influence of industrial input. (Figure 8B). The following was noted:

- Eikelboom Type 0092 was the most prevalent dominant filament in WWTPs of all configurations, but configuration did not appear to be a selective criterion, probably because of the ubiquitous nature of the filament, discussed further in Chapter 5.
- *M. parvicella* dominated in 20% to 60% of WWTPs of different configurations, and was notably lower in the MLE (20%) and 3SB (24%) configured WWTPs.

- The MLE and 3SB configurations followed similar trends for the selection of Type 0092, *M. parvicella*, Type 0041 and Type 021N.
- The non-BNR WWTPs appeared to be more selective for GPBB than the BNR WWTPs strongly suggesting that nutrient removal discourages the overgrowth of these species of filaments.

The results obtained from DBN were inconsistent, possibly because of the high industrial input, and were removed from the statistical model (Chapter 5). To determine whether configuration played a significant role on filament selection, a chi-squared test was applied to a cross-tabulation of configuration types versus most dominant filaments in the CPT and GAU WWTPs. The results of the analyses are included in Appendix 2, and confirm that filament dominance is significantly influenced by WWTP process configuration ($p<0.01$). This is not surprising, but has not previously been described in literature.

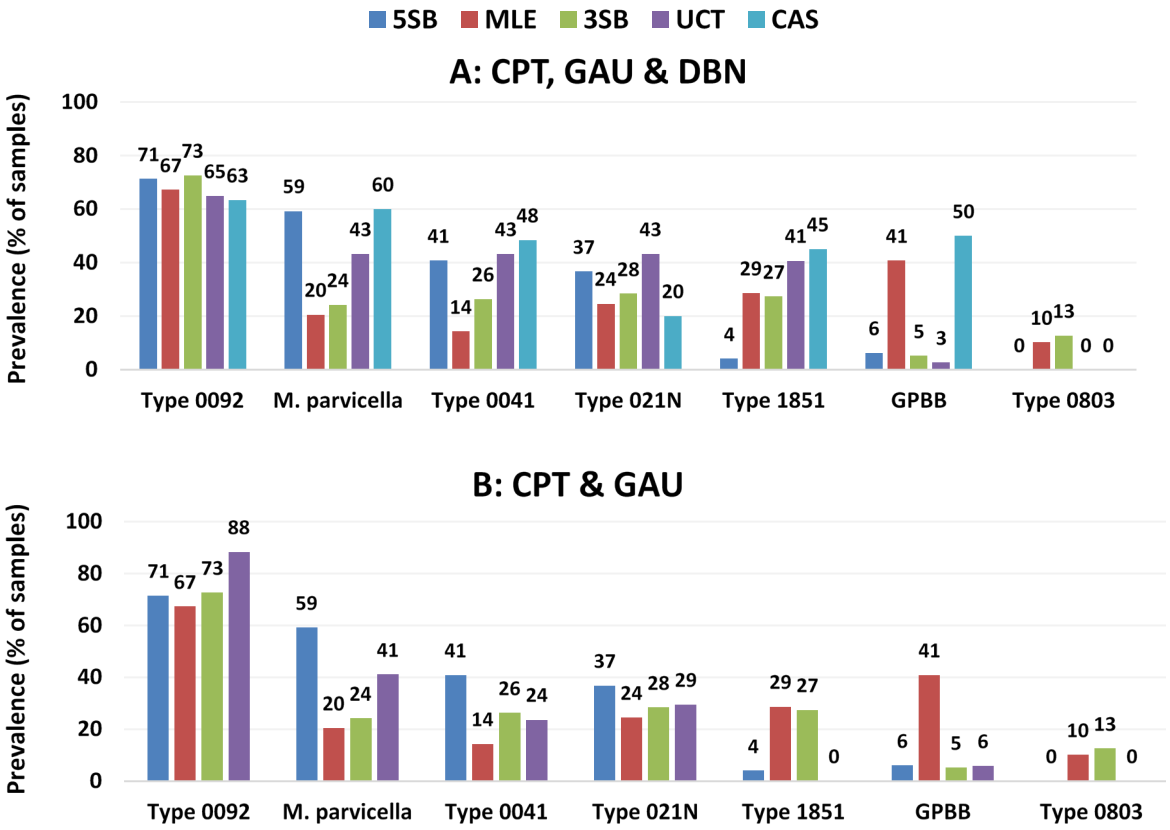


Figure 7: Prevalence of seven most prevalent dominant filaments seen in activated sludge samples from WWTPs with different process configurations. All results (A), and results with DBN data removed (B).

CAS = continuous activated sludge; MLE = Modified Ludzack-Ettinger; 3SB = 3-stage Bardenpho; UCT = University of Cape Town; 5SB = 5-stage Bardenpho

3.4 The effect of industrial influent on the selection of dominant filament types

The prevalence (%) of dominant filaments in WWTPs where the influent was ≥90% domestic in origin was compared with those WWTPs with >10% industrial component. The latter were spread between WWTPs from Durban (*n* = 2 of 4) and Cape Town (*n* = 3 of 6). Both of these WWTPs from Durban are CAS-configured, one with 50% and one with 65% industrial component. Two of the WWTPs from Cape Town are 5SB-configured (one with 15-20% and one with 30% industrial component) and one WWTP is MLE-configured with 80% industrial component. It appears that the proliferation of *M. parvicella* and GPBB may be enhanced by the addition of industrial influent, which could be the main reason for the higher prevalence of both of these filaments in the DBN WWTPs (Figure 9).

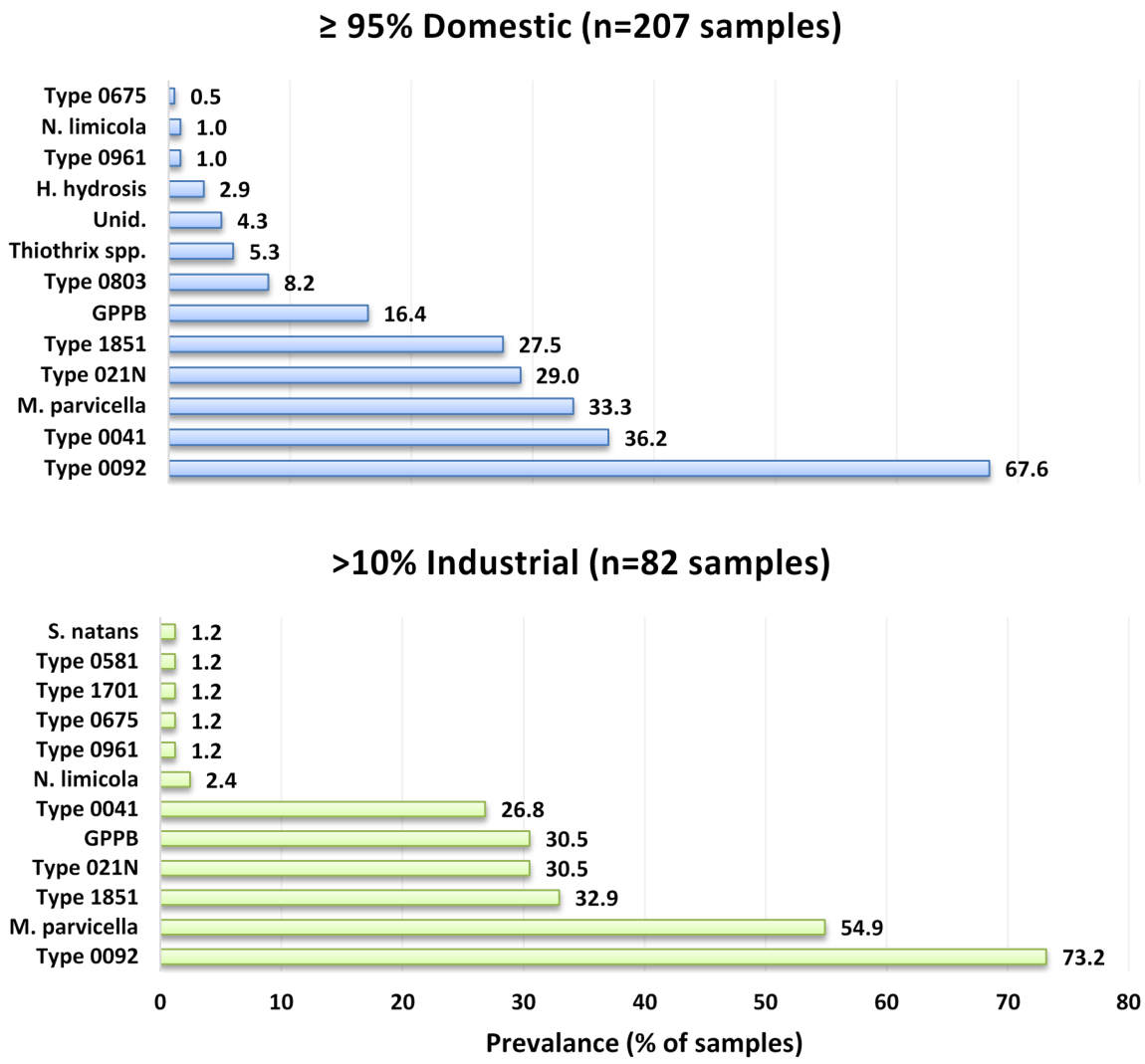


Figure 8: Comparison of the prevalence of dominant filamentous populations seen in activated sludge samples from wastewater treatment plants with industrial and domestic inflows

3.5 Most prevalent filaments: literature review

Although some filaments, such as *M. parvicella,* are typically associated with bulking, ***almost any filament type can cause bulking if conditions enhance particular qualitative (morphological) and/or quantitative growth patterns***. Seven filament morphotypes were prevalent in >2% of the samples examined (Sections 3.6.1-3.6.7).

3.5.1 Eikelboom type 0092

Type 0092 is a ubiquitous filament that dominates consistently in many WWTPs across the globe including Australia, Europe, North America and Asia (Martins et al., 2004; Mielczarek et al., 2012; Spiers et al., 2009). It is Gram negative and Neisser positive, with a typical ribbon-like structure when stained (Figure 10). Type 0092 includes members of the class *Caldilineae* and has been designated to the phylum *Chloroflexi* (Spiers et al., 2009; Yoon et al., 2010). These filaments typically form the backbone of the flocs and contribute to floc stability (Spiers et al., 2009; Yoon et al., 2010). Although Type 0092 is often present as a co-dominant filament, it is seldom implicated in bulking (Spiers et al., 2009; Yoon et al., 2010; Welz et al., 2014a).

Type 0092, Type 0041/0675 and *M. parvicella* have been grouped together as niche competitors (Eikelboom, 2000; Martins et al., 2004; Rossetti et al., 2005). It has been postulated by Martins et al. (2004) that these filaments can metabolise particulate substrates and therefore dominate in WWTPs with anaerobic- anoxic- aerobic- zones and high solids retention times. It has also been reported that a seasonal, temperature dependant, successional dominance takes place between *M. parvicella* and Type 0092, and it has been suggested that this is because both of these organisms have similar substrate utilisation profiles (Eikelboom, 2000; Martins et al., 2004; Noutsopolis et al., 2006). This seasonal link between *M. parvicella* and Eikelboom Type 0092 was not demonstrated in this study.

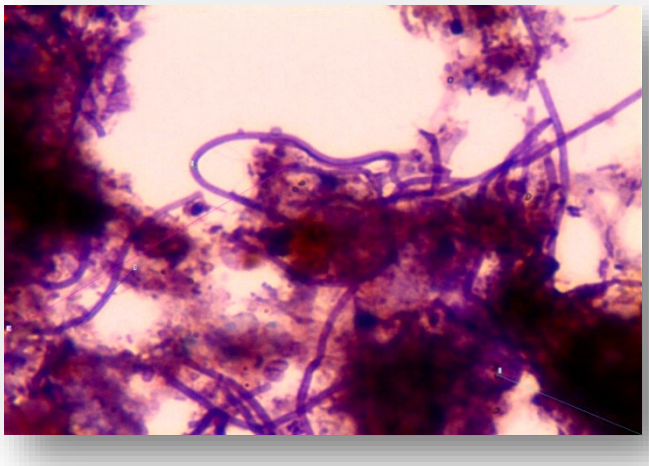


Figure 9: Neisser stain of Eikelboom type 0092

3.5.2 *Microthrix parvicella*

M. parvicella is a Gram positive coiled filament (Figure 11). Typical blue-black storage granules are present on staining with the Neisser method. It is a 'nuisance' organism, being the most common cause of bulking sludge and scum formation (Ekama et al., 1984; Eikelboom et al., 2000; Jenkins et al., 2004). It is thought that surface lipases allow *M. parvicella* to take up and store long chain fatty acids under anoxic as well as anaerobic conditions, providing the filament with a competitive advantage (Nielsen et al., 2002). Promising strategies to control overgrowth with *M. parvicella* include: reducing the solids retention time, reducing the free ammonia concentration, increasing the DO, reducing the lipid content with pre-flotation, and adding flocculants (Jenkins et al., 2004; Nielson et al., 2002; Roels et al., 2002; Tsai et al., 2003).

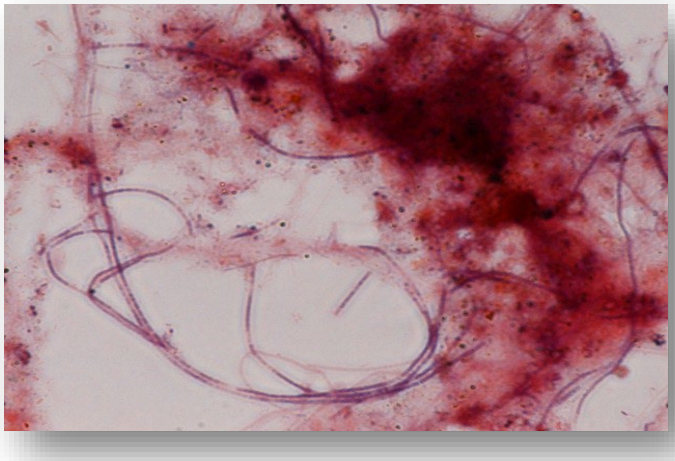


Figure10: Gram stain of *Microthrix parvicella*

3.5.3 Eikelboom type 1851

Type 1851 is a Gram positive filament that stains Gram variable. It forms bundles and may or may not have copious amounts of attached bacterial growth (Figure 12). An isolate of Type 1851 has been identified as belonging to the phylum *Chloroflexi,* with the most closely related species being *Roseiflexus castenholzii* (Beer et al., 2002). It is possible that the growth of this filament can be inhibited by intermittent feeding, the presence of anoxic or anaerobic conditions, or the use of a selector (Eikelboom, 2000; Jenkins et al., 2004). The fact that Type 1851 was less dominant in the 5SB configured WWTPs seems to indicate that this may hold true in the SA context.

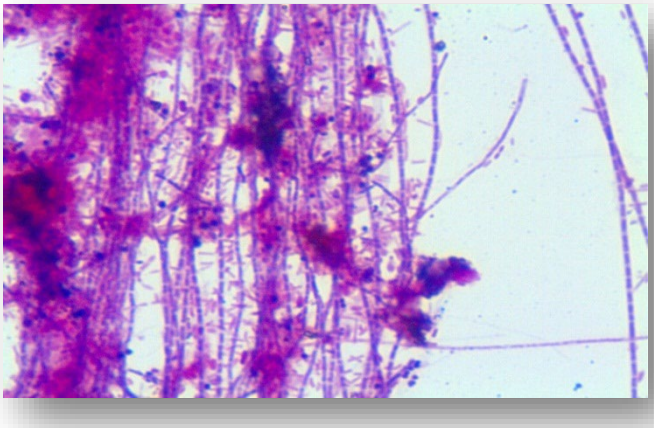


Figure 11: Gram stain of Eikelboom type 1851

3.5.4 Eikelboom type 021N/*Thiothrix* spp

Most filaments were originally described based on microscopic morphology alone. With many, including Eikelboom type 021N and *Thiothrix* spp., there is now a dispute with taxonomy and nomenclature. Historically, genotypic and phenotypic differences have led to the identification of two phylogenetic clusters within the *Thiothrix* genus, the *T. nivea* group (also known as *Thiothrix* I) and the Eikelboom Type 021N group (also known as *Thiothrix* II or *T. Eikelboomii*) (Aruga et al., 2002; Jenkins et al., 2004; Kanagawa et al., 2000; Wagner et al., 1994). It has been argued that each should be considered a separate genus. However, Chernousova et al. (2012) have recently found that each group forms an evolutionary branch within the same genus. In this study, the two groups have been named using the classical nomenclature of *Thiothrix* and Eikelboom Type 021N. According to Eikelboom (2000), Type 021N is a strict aerobe, metabolises easily biodegradable substrates, and can utilise hydrogen sulphide as a source of energy, so that it proliferates in 'septic' wastewater.

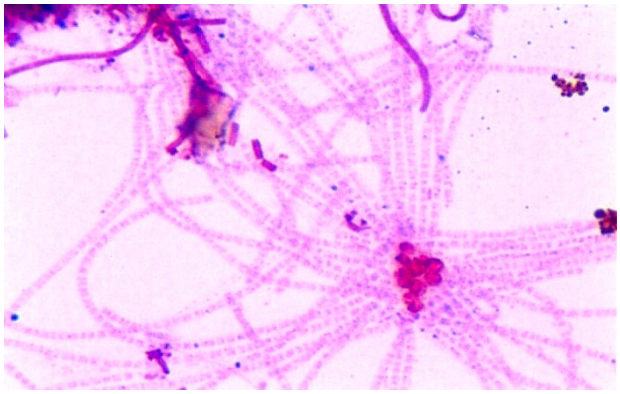


Figure 12: Gram stain of Eikelboom type 021N

3.5.5 Gram positive branching bacilli

This Gram positive branching filamentous group of organisms, currently and previously also known as *Gordonia amarae* like organisms (GALO), *Nocardia amarae* like organisms (NALO), *Nocardia* spp., and Actinomycetes, were the most common cause of bulking and foaming in the United States in the 1980's (Richard et al., 1982; Strom and Jenkins, 1984). Strictly speaking, none of the names applied previously are correct, so for the purposes of this study, in consultation with a taxonomist, it was decided to name these filaments according to their morphology and Gram positive cell wall. Their distinctive branching (Figure 14) means that they cannot be confused with any other morphotypes. Like *M. parvicella*, these filaments can utilize the fat fraction of wastewater for growth and their hydrophobic/lipophilic cell wall causes them to float and form scum (Eikelboom, 2000, Jenkins et al., 2004).

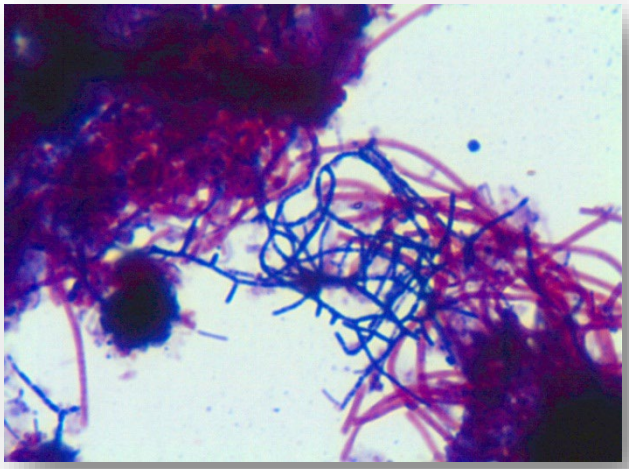


Figure 13: Gram positive branching bacilli (Gram stain)

3.5.6 Eikelboom type 0041

Eikelboom Type 0041 morphotype is characterised as being a Gram positive/variable, Neisser negative filament that proliferates in similar environments to Type 0092 and *M. parvicella* (Hugenholz, 2001; Nielsen, 2009b, Rosetti, 1994). The filaments have sheaths, and like Type 1851, this is usually accompanied by attached microbial growth. Although it has been shown that some members of this morphotype exhibit similar physiological characteristics, phylogenetic analysis has shown that is comprised of more than one phylogenetic group (Hugenholz, 2001). It is therefore possible that morphological characterisation by classical microscopy may be inadequate for linking Type 0041 to process conditions. However, to date, no comprehensive studies have been performed to elucidate the range of phylogenetic groups belonging to this morphotype (Hugenholz, 2001; Nielsen, 2009; Rosetti, 1994, Seviour, 1994).

3.5.7 Eikelboom type 0803

Although this filament is commonly observed in WWTPs, where it can over-proliferate and presumably causes bulking, there is very little information available about this morphotype. According to Eikelboom (2000), Type 0803 only causes bulking in industrial WWTPs, and pre-denitrification results in the disappearance of filament. In this study, Type 0803 was found as a dominant filament in three domestic WWTPs in GAU with 3SB configuration, which has pre-denitrification. The results of this study thus refute the findings of Eikelboom (2000).

More recently, Spiers et al. (2015) found that Type 0803 encompassed more than one member of the *Chloroflexi* phylum, and was distinct from the *Chloroflexi* of the Type 0092 and 1851 morphotypes. The authors found this filament in a range of EPBR and non-EPBR WWTPs.

CHAPTER 4: RESULTS AND DISCUSSION: PHYSICOCHEMICAL AND MICROSCOPIC DATA

In order for the final statistical model to be precise and accurate, spurious data needed to be excluded. It was important that outliers were scrutinised. In some instances, outliers may have arisen from sample inconsistencies or technical errors. In these cases, they were excluded from the statistical database. However, in other cases, outliers were true reflections of the physicochemical nature of samples from which they were measured. It was important that these outliers were maintained within the database. Results were therefore interrogated to ascertain whether there was an explanation for unusually high or low parameters, inconsistent monthly trends, or inter-and/or intra-site variation – for example, high levels of ammonia when there was a mechanical failure with aeration.

The physicochemical data from the three centres from samples taken between February 2016 and March 2017 were thoroughly analysed and are displayed either graphically and/or in tabular form in this section. It was clear that the results for DBN exhibited distinct differences. Amongst other factors, the influent to most of the DBN WWTPs contained high percentages of industrial effluent. The complexity of the influent increased the number of variables. It was decided, in consultation with the reference group, ***to exclude the DBN WWTPs from the statistical model*** (Chapter 5). None of the DBN results displayed in Chapter 4 were therefore included in the statistical database. In addition, spurious results from the CPT and GAU WWTPs were removed from the database – these are denoted by red arrows in the graphs.

4.1 Settling indices and filament index

The DSVI is the most widely used *ex-situ* laboratory method for measuring the settling propensity of solids in AS. The test is simple to perform, requires minimal expertise and is arguably robust. It has been determined that a DSVI >150 mLg^{-1} is indicative of bulking sludge. Previous research has shown that the DSVI is well correlated with the FI, and Jenkins et al. (2004) concluded that the former should be used because “filament counting of activated sludge is not an efficient process control method for sludge settling”. The authors also stated that “filament identification is an extremely useful tool for diagnosing and rectifying activated sludge solids settling problems”.

The direct correlation between FI and DSVI found by some researchers may be overly-simplistic: while it is true that filamentous overgrowth will result in a raised DSVI, there are additional causes of poor settling that translate into elevated DSVI values. These include the presence of numerous pin-point flocs, dispersed microbial growth, zoogleal overgrowth (viscous bulking), and rising sludge caused by $N_2\uparrow$ in cases where there is incomplete denitrification in the bioreactor/s (Jenkins, 2004). In addition, sub-optimal *in-situ* operational and process conditions can have a significantly deleterious effect on clarifier efficiency, although this will not be reflected by elevated DSVI values.

Since FI is a count variable, and DSVI is a non-normal variable, a Spearman correlation analysis was used to calculate the correlation coefficient between these variable for each site (Table 6). There was a weak but significant ($p < 0.01$) positive correlation for the samples from CPT ($r_s = 0.381$) and GAU ($r_s = 0.384$). When analysed separately, significant correlations were demonstrated for only 4 of the 28 months analysed (data not shown). For correlation studies, a minimum of two variables with at least 8 to 10 observations for each variable is recommended, suggesting that the monthly results (typically with 6 observations) should be interpreted with caution.

Apart from activated sludge samples with a high FI (≥4) and DSVI (≥150), Mielczarek et al. (2012) did not find a strong positive correlation between DSVI and FI in 28 WWTPs over a four year period (Figure 15A). This followed a similar trend to the results of this study (Figure 15B,C). It has already been shown that diurnal variations in the temperature and the MLSS concentrations occur *in-situ*, and these may exert a significant effect on the settling propensity of AS, as measured using the SVI. In addition, the temperature effects bring in to question the accuracy of *ex-situ* testing of settling in laboratories, where temperatures are unlikely to reflect those in the bioreactors at the time of sampling (RÖssle and Pretorius, 2008).

The weak correlations could also be explained by elevated DSVIs caused by other factors than filamentous overgrowth. In addition, some filamentous bacteria have a greater propensity to float, while other settle more easily so that filament assessment should be both qualitative and quantitative (Eikelboom, 2000).

Zooglea were not noted in abundance in any of the samples, so this form of bulking was discounted during the study period. The significance of each of the floc characteristics on the DSVI was determined (Appendix 4). It was found that as individual parameters, there was no significant link between the size or strength of the flocs and the DSVI ($p>0.05$). Open flocs were associated with significantly ($p<0.001$) higher DSVIs (151±89 ml/g) than compact flocs (115±49 ml/g), which is logical. However, round flocs were associated with significantly higher DSVIs (164±93 ml/g) than irregular flocs (125±70 ml/g), which was unexpected.

Table 6: Spearman's correlation analysis of diluted sludge volume and filament indices

Location			FI
CPT	DSVI	Correlation Coefficient	0.381**
		p-value (2-tailed)	< 0.001
GAU	DSVI	Correlation Coefficient	0.384**
		p-value (2-tailed)	< 0.001

**. Correlation is significant at the 0.01 level (2-tailed).

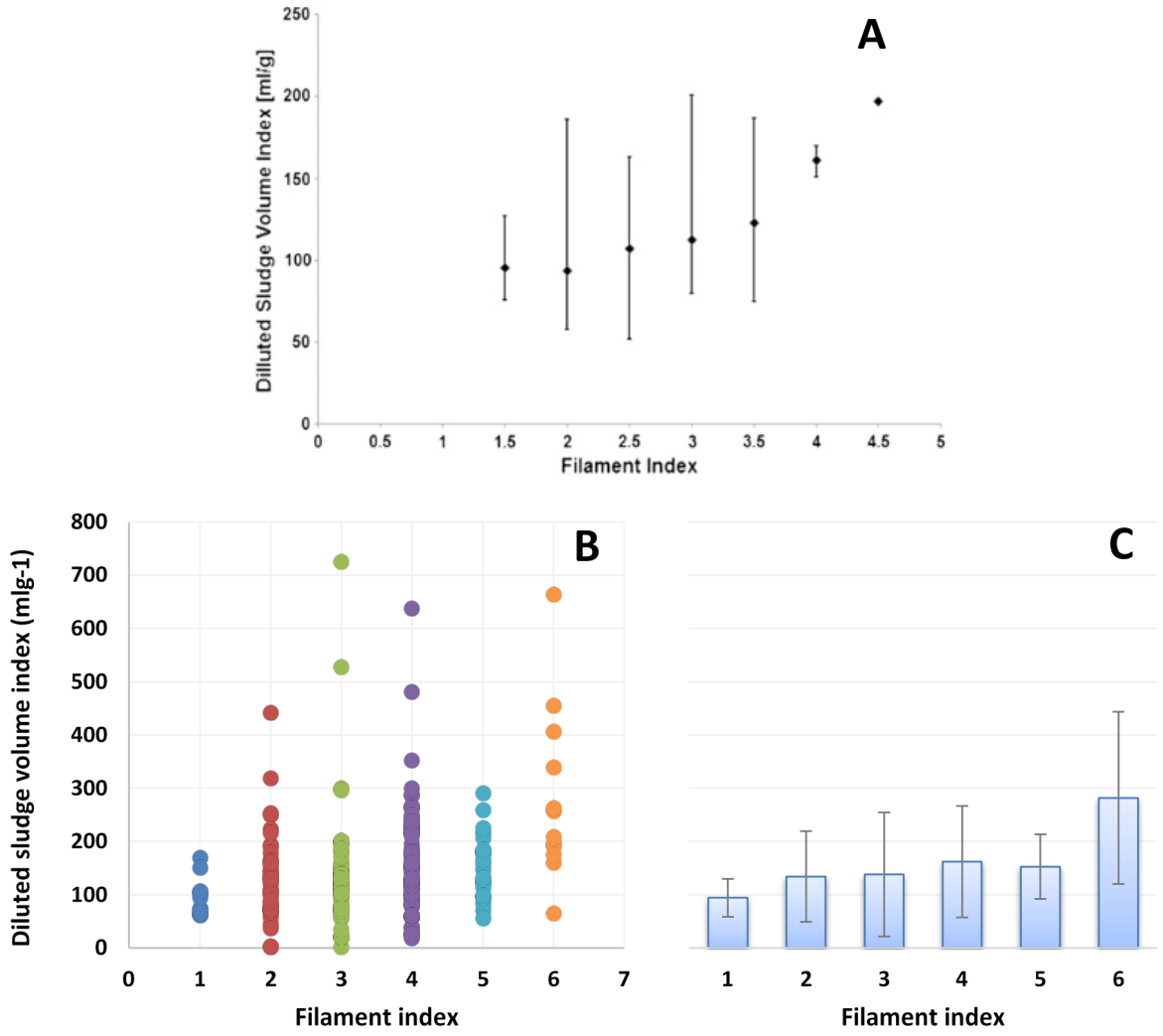


Figure 14: Overview of settling properties and abundance of filaments as determined by diluted sludge volume index and filament index. Average and range over four years (2008-2011) determined by Mielczarek et al., 2012 (A) and the results of this study (B, C)

Some of the WWTPs included in the study exhibited high DSVIs for the entire study period, or almost the entire study period, e.g. C5, G2, and the DBN WWTPs, especially D2 (Figure 16). In the case of C5, the WWTP experienced severe bulking, with an overgrowth of *Thiothrix* spp. and Eikelboom Type 021N, which are closely related genetically. In the case of G2, GPBB were found as dominant species throughout the study period, but were not found in any of the other GAU WWTPs. In contrast to the CPT and GAU WWTPs, there was a constant succession of different dominant filament species over much of the study period at the DBN WWTPs, indicating greater process instability, possibly related to fluctuating industrial inputs.

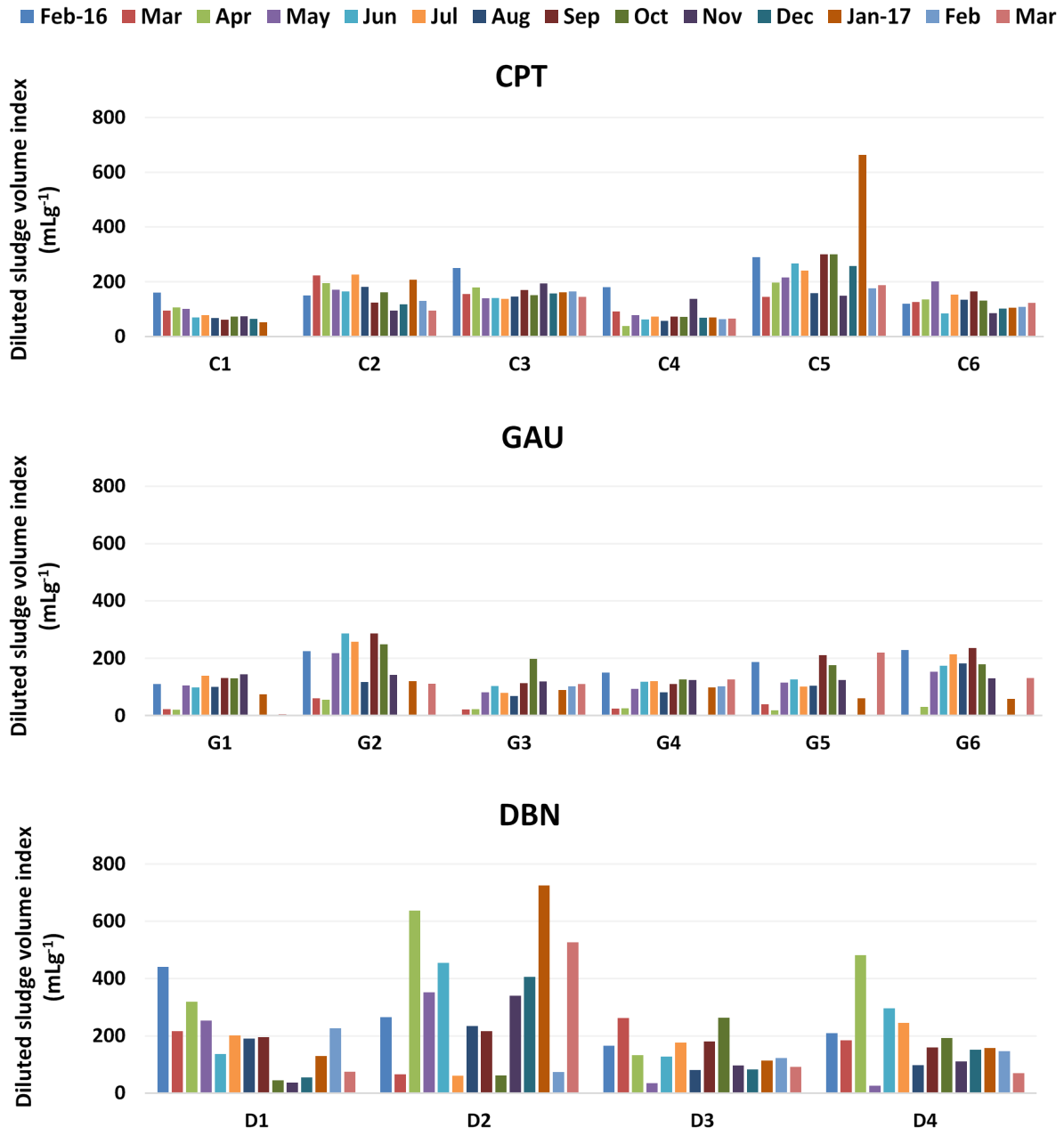


Figure 15: Diluted sludge volume indices obtained for the individual wastewater treatment plants during the study period

4.2 Dissolved oxygen, temperature, pH, total alkalinity and volatile fatty acid concentrations

Dissolved oxygen concentrations in the aerobic reactors in CPT were obtained from historical results from *in-situ* probe readings from Feb to June 2016 (Figure 17). *In-situ* AS temperature and DO measurements were taken by the project team in GAU (Figure 3B) and DBN (Figure 3C), as well as from Oct 2016 in CPT using the same methodology.

Operators typically strive to maintain DO levels in the aerobic zone(s) of AS WWTPs between 1-2 mgL^{-1} to ensure both adequate organic degradation and nitrification, and to maintain good floc structure. However, a number of factors, including the configuration and the MLVSS, COD and NH_3/NH_4^+ concentrations, determine the optimal DO requirements for each WWTP. In this study, the DO readings from some WWTPs were consistently low, e.g. in C5 from CPT (<1 mgL^{-1}), and G2 from GAU (~1 mgL^{-1} from April to November 2016), which both experienced high DSVI values. In the case of C5, there was severe filamentous bulking due to overgrowth of *Thiothrix* spp. and Eikelboom Type 021N. The overgrowth of these filaments is consistent with the fact that they have been associated with 'septic' sludge which can occur when there is insufficient aeration. In the case of G2, the ubiquitous presence of GPBB in this WWTP, but not in any of the other WWTPs from GAU suggests that the low DO may have led to the overgrowth of this filament, which was reflected by a high DSVI.

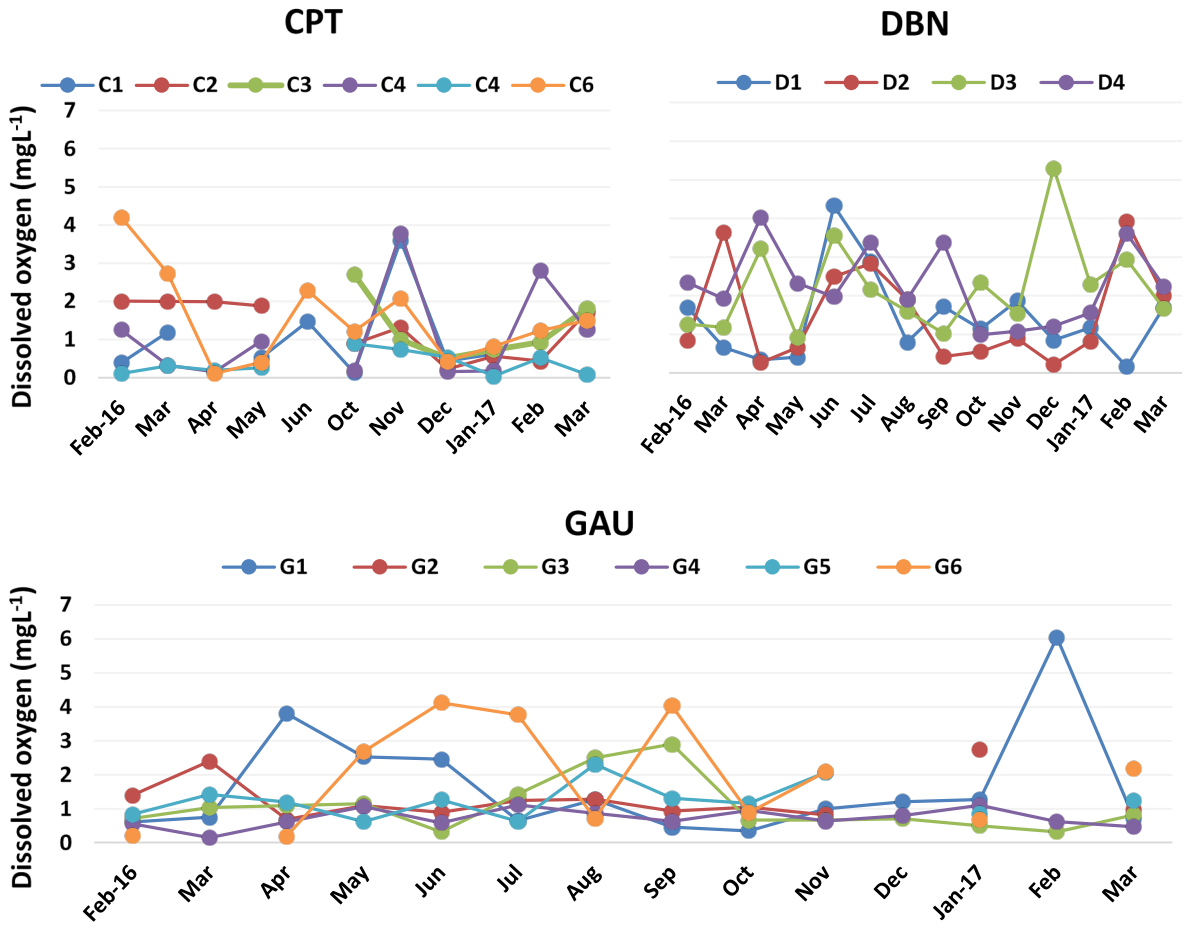


Figure 16: Dissolved oxygen concentrations measured *in-situ* in the activated sludge

The temperature of the AS was consistently and significantly ($p<0.01$) highest in the DBN WWTPs (Figure 18). Temperature is an important contributing factor to microbial selection, so this finding, although not unexpected, is noteworthy. There was little variability between the temperatures of the AS in the GAU WWTPs as evidenced by the magnitude of the error

bars signifying the standard deviation from the mean in Figure 18A. The largest variations in temperatures were seen in the WWTPs from CPT (Figure 18B) and to a lesser extent, DBN. The differences in intra-site variation within the three geographical locations are most likely due to climatic factors. GAU is inland and experiences little wind. In contrast, CPT has high winds at certain times of the year, the strength of which differs from area to area. Apart from wind effects, the proximity of the WWTPs to the ocean also affects the temperature. In summer, the temperatures at the CPT WWTPs close to the sea tend to be lower than those further away, with the converse being true in winter.

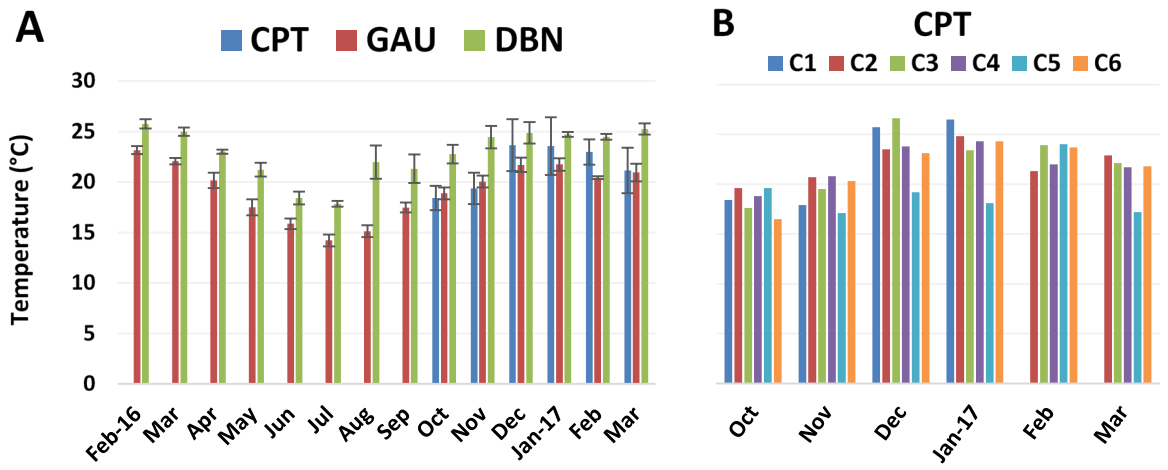


Figure 17: *In-situ* temperatures measured in activated sludge during the study period: the mean and standard deviation for all wastewater treatment plants from each geographical location (A) and individual measurements from the Cape Town wastewater treatment plants

Like temperature, pH plays a major role in microbial selection. Nitrification can be inhibited at influent pH values <6.5 (Li and Irvin, 2007; Rasool et al., 2014). In this study, all pH measurements fell in the range 6.1-8.9 (Figure 19). While the pH was stable in some of the WWTPs (e.g. C1), it was more erratic in others.

It is expected that the pH should decrease during bioremediation of domestic wastewater, mainly due to the consumption of alkalinity during nitrification (Li and Irvin, 2007; Rasool et al., 2014). Almost without exception, the pH measured in the AS of the WWTPs in CPT and GAU was less than the influent pH taken on the same day. In the CPT WWTPs, the influent and AS pH ranged from 6.1 (C4, Mar 2016) to 8.3 (C3, Aug-Sep 2016), and 6.3 (C3, Feb 2017) to 7.3 (C2, Feb 2017, C5 Nov-Dec 2016), respectively. In the GAU WWTPs, the influent and AS pH ranged from 6.6 (G2, Nov 2016) to 8.6 (G4, Jan 2017) and 6.3 (G4, Mar 2017) to 8.9 (G6, Feb 2016), respectively.

In contrast to the GAU and CPT WWTPs, the pH in the AS was not significantly lower than the influent pH in the DBN WWTPs ($p>0.5$). The pH in the influent and AS ranged from 6.9 (D3,

Nov 2016) to 8.3 (D2 and D3, Jun 2016) and 6.9 (D3 Jun 2016) to 7.9 (D3, Apr and Dec 2016), respectively.

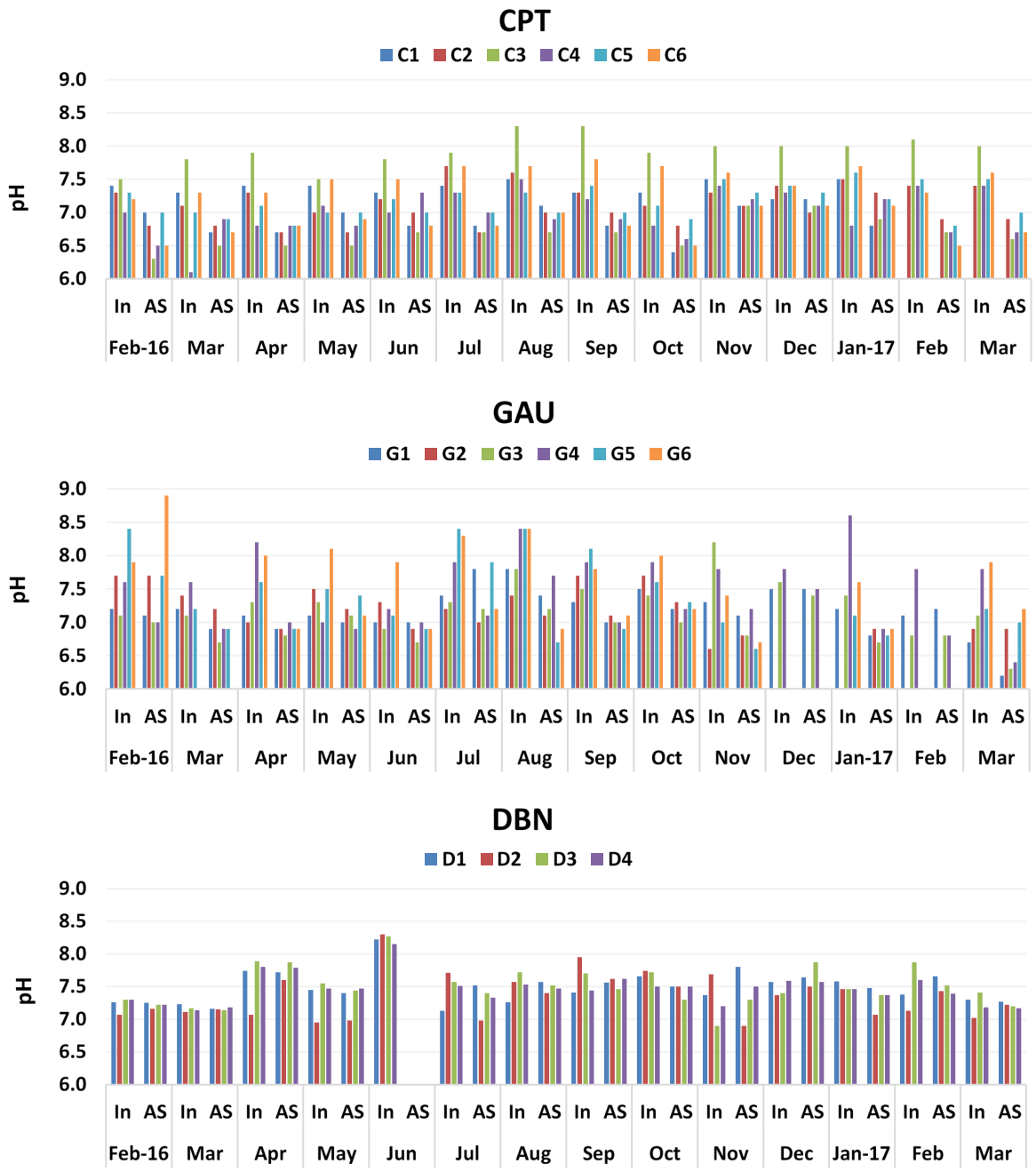


Figure 18: The pH measured in the influent (In) and activated sludge (AS) of the individual wastewater treatment plants from each geographical location

The total alkalinity (Figure 20) in the influent was not measured in the samples from the DBN WWTPs until August 2016. The concentrations were highest in the CPT WWTPs, and were typically >300 $CaCO_3$ mgL^{-1}.

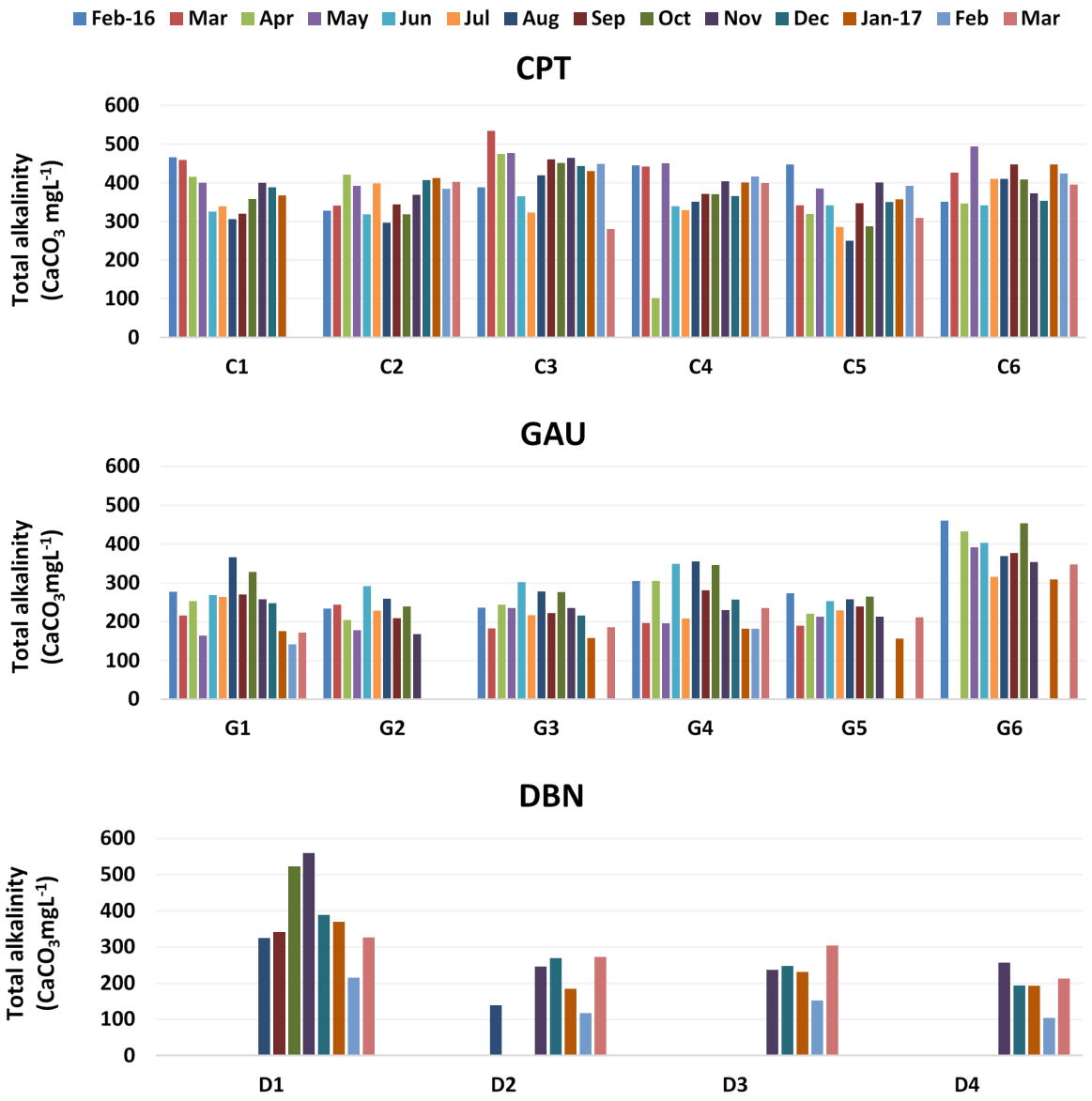


Figure 19: Total alkalinity concentrations measured in the influent of the wastewater treatment plants from each geographical location

Short-chain VFAs play a selective role in determining the microbial community structure in WWTPs, and are important substrates for PAOs in EBPR WWTPs (Begum and Batista, 2014). In addition, high concentrations of VFAs can decrease the pH, and high concentrations of acetate in the influent have previously been associated with filamentous bulking (Guo et al., 2014). In this study, the VFA concentrations in the influent to the GAU WWTPs were lower than at the WWTPs from the other locations, but were the COD and BOD concentrations (Section 4.3). There were a few instances when the concentrations of VFAs in the influent exceeded 200 mgL^{-1}.

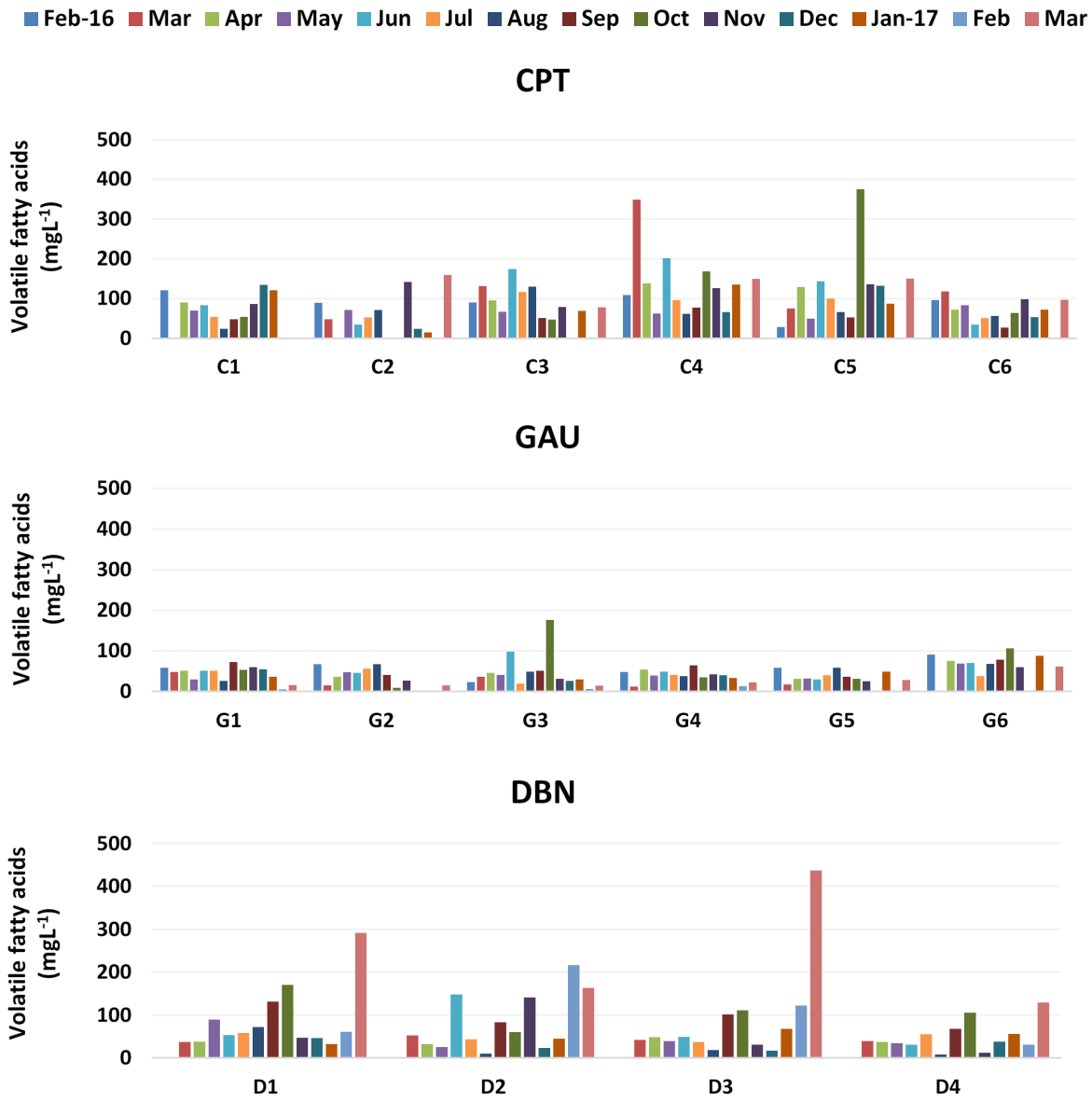


Figure 20: The VFA concentrations measured in the influent of the wastewater treatment plants from each geographical location

4.3 Analyses of organics and biomass in influent and activated sludge

4.3.1 Introduction

The organic composition in wastewater is generally measured using COD and/or biochemical oxygen demand (BOD) as a proxy. While the former is a robust chemical test and therefore widely used, it only provides a quantitative measurement of how much oxygen is required to oxidise the organic matter (and some inorganics) to carbon dioxide, ammonia and water. Conversely, the biologically-based BOD test is less reproducible, but indicates how much oxygen is actually used during the biological oxidation of the biodegradable fraction of

wastewater. The BOD therefore also plays a semi-qualitative role (Dhall et al., 2012). Bearing these points in mind, it was decided to include both of these parameters in the study because both the quantity and the quality of the nutrients are important for microbial selection (Eikelboom, 2000).

4.3.2 Chemical oxygen demand and mixed liquor volatile suspended solids

The influent COD was significantly higher in the CPT WWTPs than in the WWTPs from GAU ($p < 0.001$; Table 7). The influent COD was highest at C4 and was relatively consistent at each CPT WWTP on a month-to-month basis (Figure 22).

Table 7: Influent chemical oxygen demand from February 2016 to March 2017

	Average COD (mgO_2L^{-1})	Range (mgO_2L^{-1})	*n*
CPT			
C1	971 ± 178	700-1273	12
C2	924 ± 106	632-1044	13
C3	1157 ± 282	682-1646	14
C4	1870 ± 336	1348-2537	14
C5	1034 ± 119	918-1288	14
C6	907 ± 184	680-1396	14
GAU			
G1	597 ± 242	19-876	13
G2	641 ± 233	313-1154	11
G3	309 ± 178	105-743	14
G4	377 ± 184	95-607	14
G5	344 ± 83	116-476	12
G6	801 ± 106	592-928	11
DBN			
D1	1065 ± 734	472-3109	12
D2	873 ± 530	425-2201	14
D3	1006 ± 1000	293-3920	13
D4	830 ± 755	200-3109	14

The average total COD (COD_t) of the AS at the CPT WWTPs was 3-12 times higher than the influent COD, presumably because of the contribution of recycled biomass (endogenous COD). For the GAU cohort, apart from an outlier from G3 in Nov 2016 [AS COD_t = 71 mgL^{-1} (omitted from the statistical database)], and G1 and G3 results in Feb and Mar 2017, where the COD_t of the AS and influent AS were similar, the COD_t of the AS was also 3-12 times higher than the influent concentration. The pattern in the DBN WWTPs differed in that there were many instances where the COD_t of the AS was similar, or even lower than the influent COD.

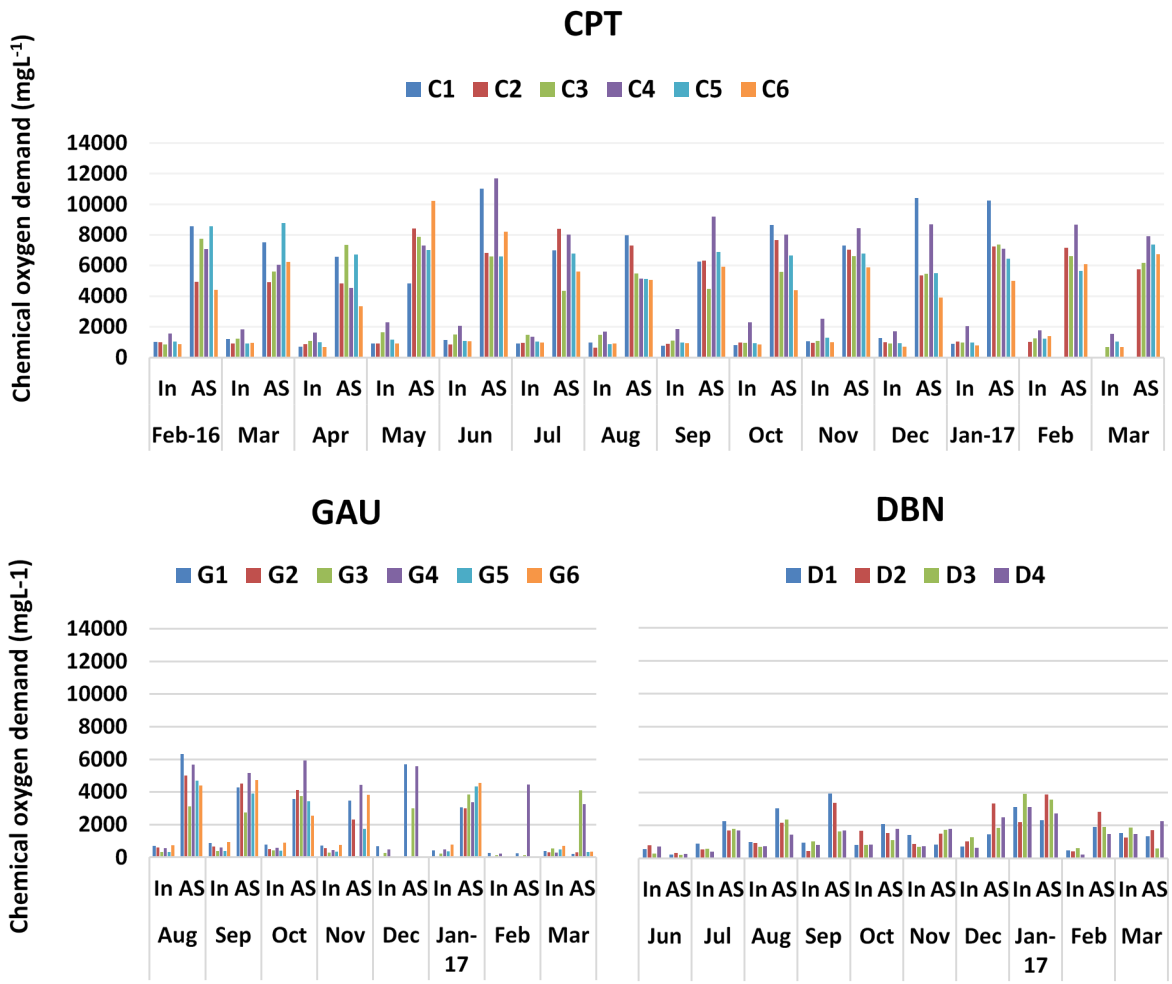


Figure 21: Chemical oxygen demand concentrations in the influent and activated sludge

Access to organic nutrients plays a major role in microbial selection. Both the quality and quantity of the organic molecules in the wastewater are important selection criteria (Ramond et al., 2013, Welz et al., 2012, 2014b, 2014c). However, these become depleted to varying degrees as the influent flows through the bioreactors, so measuring the influent concentrations may be simplistic if one needs to understand the factors affecting filament selection. This area has been poorly researched to date.

The rate of organic depletion within a bioreactor is dependent on many variables (e.g. hydraulic retention time, sludge recycle ratio, influent characteristics, and microbial population dynamics). The COD_t (unsettled, unfiltered) in AS is comprised of (i) residual (soluble and particulate) organics and biomass (minimal) from the influent, (ii) the retained residual organics (if any) and biomass in the reactor, and (ii) residual organics (if any) and biomass in the reactor from the return flows. Endogenous decay of biomass in the reactor also takes place, particularly when there is insufficient 'food' for the microbial population (e.g. in some BNR WWTPs).

In most instances therefore, the COD is significantly higher in the AS than in the influent. In a small scale study, Contreras et al. (2002) showed that there was a good correlation between the MLVSS and the particulate COD fraction (COD_p), suggesting that the quantity of microbial biomass can be indirectly deduced by calculating the difference between COD_t and soluble COD (COD_s).

In this study, there were a limited number of samples where both COD_t and COD_s had been determined in the AS samples. In the case of the CPT WWTPs, the COD_s constituted <2% of the COD_t in all the samples (*n*=47, Table 8). With the exception of C4, there was a significant correlation between the COD_t and the MLVSS in the all of the WWTPs from the CPT cohort. In the WWTPs from GAU, the COD_s constituted <3% of the COD_t in 80% of the samples (*n*=35), and there was a significant correlation between COD_t and MLVSS in the AS from G2 and G5. In stark contrast to the CPT WWTPs, the average % contribution and (range) of COD_s to COD_t from the DBN WWTPs D1, D2, D3 and D4, respectively, were 13 ± 8% (2-25%), 9 ± 7% (2-12%), 14 ± 13 (1-43%), and 11 ± 4% (6-16%) (*n*=32). There was no correlation between COD_t and MLVSS in the AS of any of the WWTPs. Neither was there any correlation between the COD_p and the MLVSS, as described by Contreras et al. (2012).

Table 8: Pearson's correlation coefficients between the total and particulate chemical oxygen demand and the mixed liquor volatile suspended solids concentrations in the activated sludge samples

	COD_t and MLVSS	COD_p and MLVSS
CPT (*n*=47)		
C1	0.822**	ND
C2	0.604*	ND
C3	0.812**	ND
C4	0.229	ND
C5	0.549*	ND
C6	0.971**	ND
GAU (*n*=35)		
G1	0.840*	ND
G2	0.559	ND
G3	0.266	ND
G4	0.885*	ND
G5	0.524	ND
G6	-0.083	ND
DBN (*n*=32)		
D1	0.084	0.083
D2	-0.088	-0.323
D3	0.350	0.437
D4	0.266	-0.014

COD_t = total chemical oxygen demand **MLVSS = mixed liquor volatile suspended solids**

COD_p = particulate chemical oxygen demand

***significant correlation at 0.05 level, **significant correlation at 0.01 level**

As with many of the other parameters, the MLVSS concentrations in the CPT WWTPs exhibited a relative temporal stability, while the concentrations in the GAU and DBN cohorts fluctuated (Figure 23). The MLVSS concentrations in the GAU WWTPs were typically lower

than those in the WWTPs from the other locations, but so was the influent COD. The F/M ratio is discussed in Section 3.3.6. There was insufficient data available to speculate on the selective roles that the COD and MLVSS concentrations and ratios may have played on the microbial community structure in activated sludge, including the filamentous species. However, the relationship between the MLVSS and AS COD, and the reason/s for good correlation between these parameters in some WWTPs and poor correlation in others merits interrogation and study, and is a recommended avenue for future research.

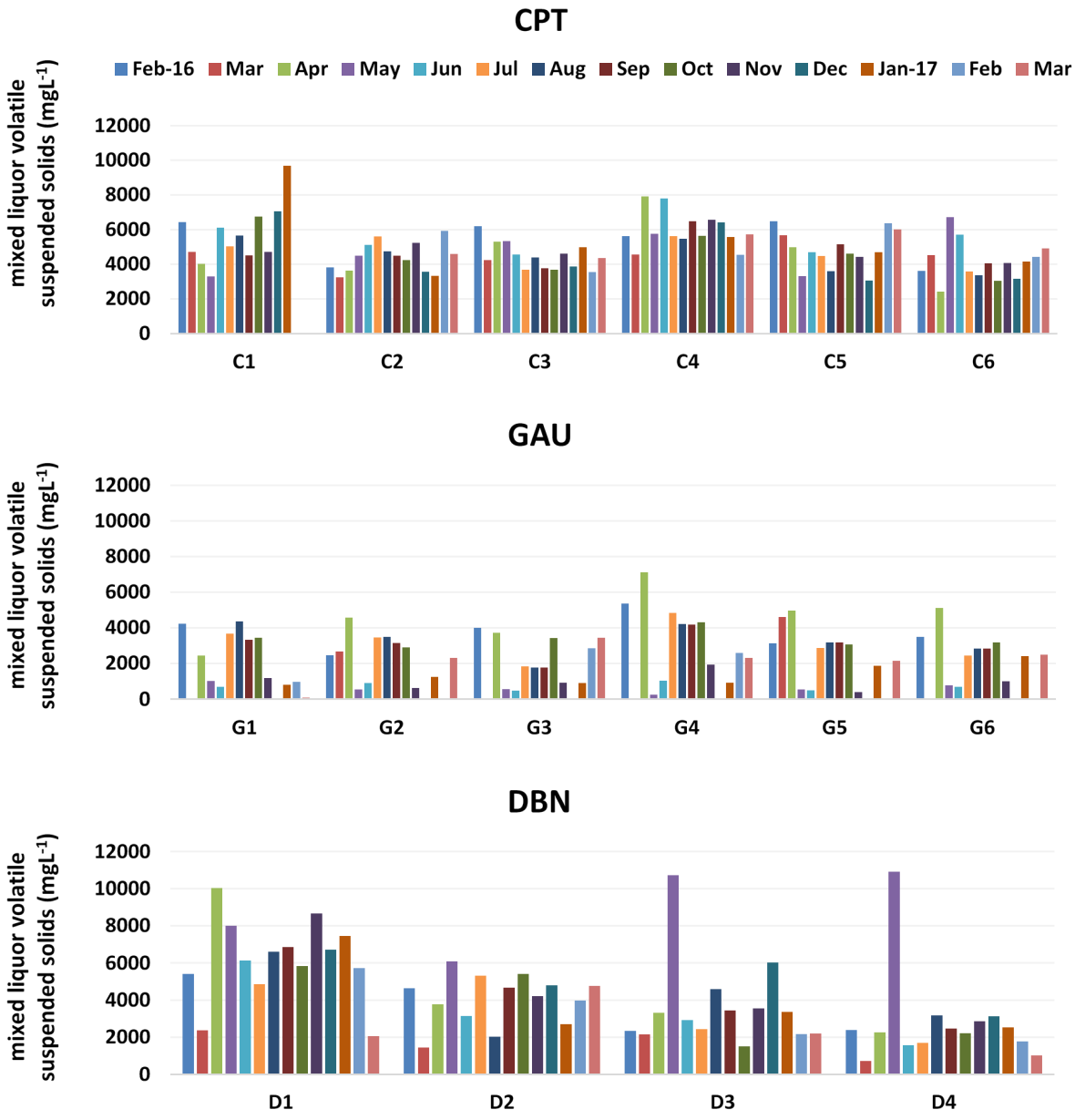


Figure 22: Mixed liquor volatile suspended solids concentrations in the activated sludge

4.3.3 Biological oxygen demand

From the influent BOD results, there were 9 low outliers that were omitted from the database for the statistical analyses (6 from GAU, 2 from CPT and 1 from DBN). The influent BOD was significantly lower in the GAU than the CPT and DBN WWTPs ($p < 0.01$, Table 9).

Table 9: Influent biological oxygen demand from February 2016 to March 2017

	Average BOD (mgL^{-1})	Range (mgL^{-1})	*n*
CPT			
C1	339 ± 126	75-475	11
C2	391 ± 200	45-775	12
C3	523 ± 210	250-1025	13
C4	683 ± 187	325-900	13
C5	490 ± 138	200-725	12
C6	427 ± 183	200-725	13
GAU			
G1	247 ± 104	1-400	14
G2	237 ± 109	30-420	11
G3	128 ± 61	46-210	13
G4	158 ± 128	2-470	14
G5	150 ± 51	1-200	11
G6	324 ± 117	30-440	11
DBN			
D1	469 ± 196	152-715	10
D2	488 ± 278	34-1126	13
D3	438 ± 129	268-660	10
D4	376 ± 155	51-586	12

The results from CPT and GAU that were removed from the statistical database are denoted by red arrows in Figure 24. The total BOD (BOD_t) in the AS was typically higher than the influent BOD (Figure 24), but the AS BOD_t to influent BOD ratios were lower than the COD counterparts. The most reasonable explanation is that there was a decrease in the fraction of readily biodegradable substrates during retention of wastewater in the bioreactor. Further information is provided in Section 4.3.4.

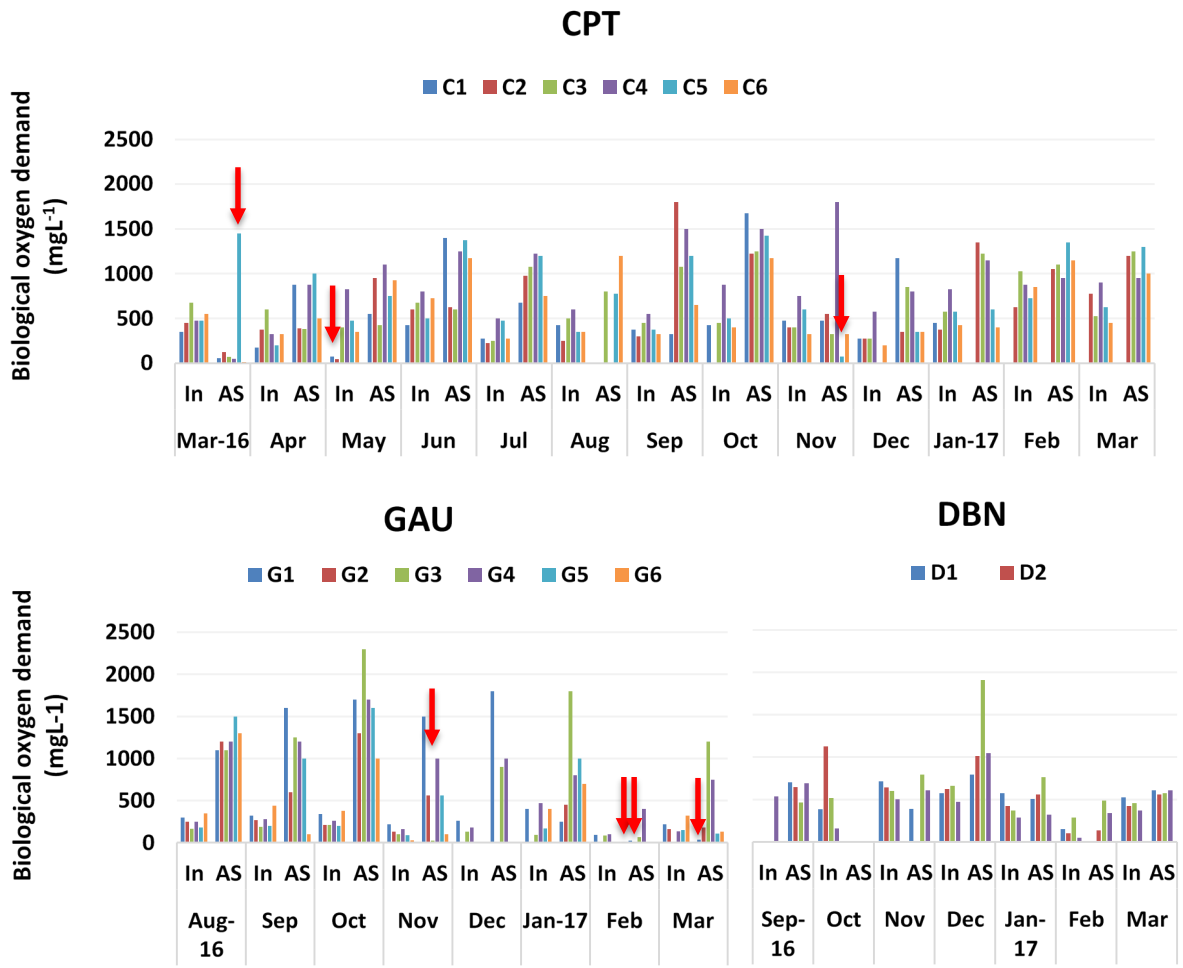


Figure 23: Biological oxygen demand concentrations in the influent and activated sludge

4.3.4 Ratio of biological to chemical oxygen demand

The ratio of BOD:COD serves as a proxy to determine the biodegradability of the organic fraction of wastewater. The following BOD:COD ratios are used as general measures:

> 0.5	Readily biodegradable
< 0.5 to > 0.4	Moderately biodegradable
< 0.4 to > 0.2	Slowly biodegradable
< 0.2	Recalcitrant

In this study, seven results from the CPT and GAU were deemed spurious, either due to the COD, BOD or COD:BOD ratio. These have been denoted by red arrows on Figure 25, and removed from the statistical database.

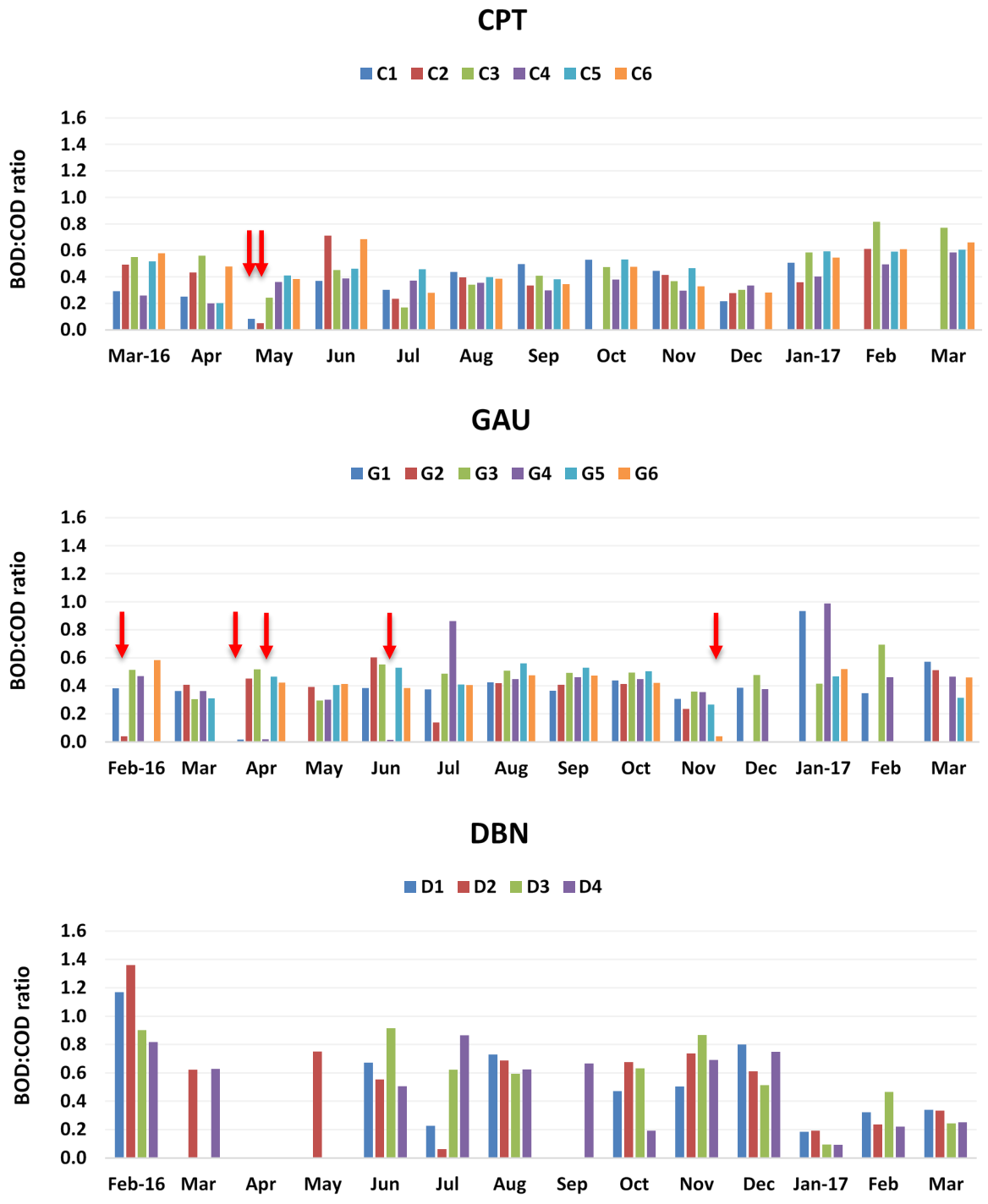


Figure 24: Influent biological oxygen demand to chemical oxygen demand ratios

With only one exception (C3, July 2016) the BOD:COD ratios in the influent samples from the CPT WWTP samples were notably higher than in the AS samples (Figure 26). This trend was expected, because the most biodegradable substrates from the influent should be utilised first. However, the trend seen in the CPT WWTPs were not always noted in the WWTPs from the other locations. It is possible that this may have been due to technical challenges. For example, the BOD test is notoriously demanding, as described by Jouanneau et al. (2014).

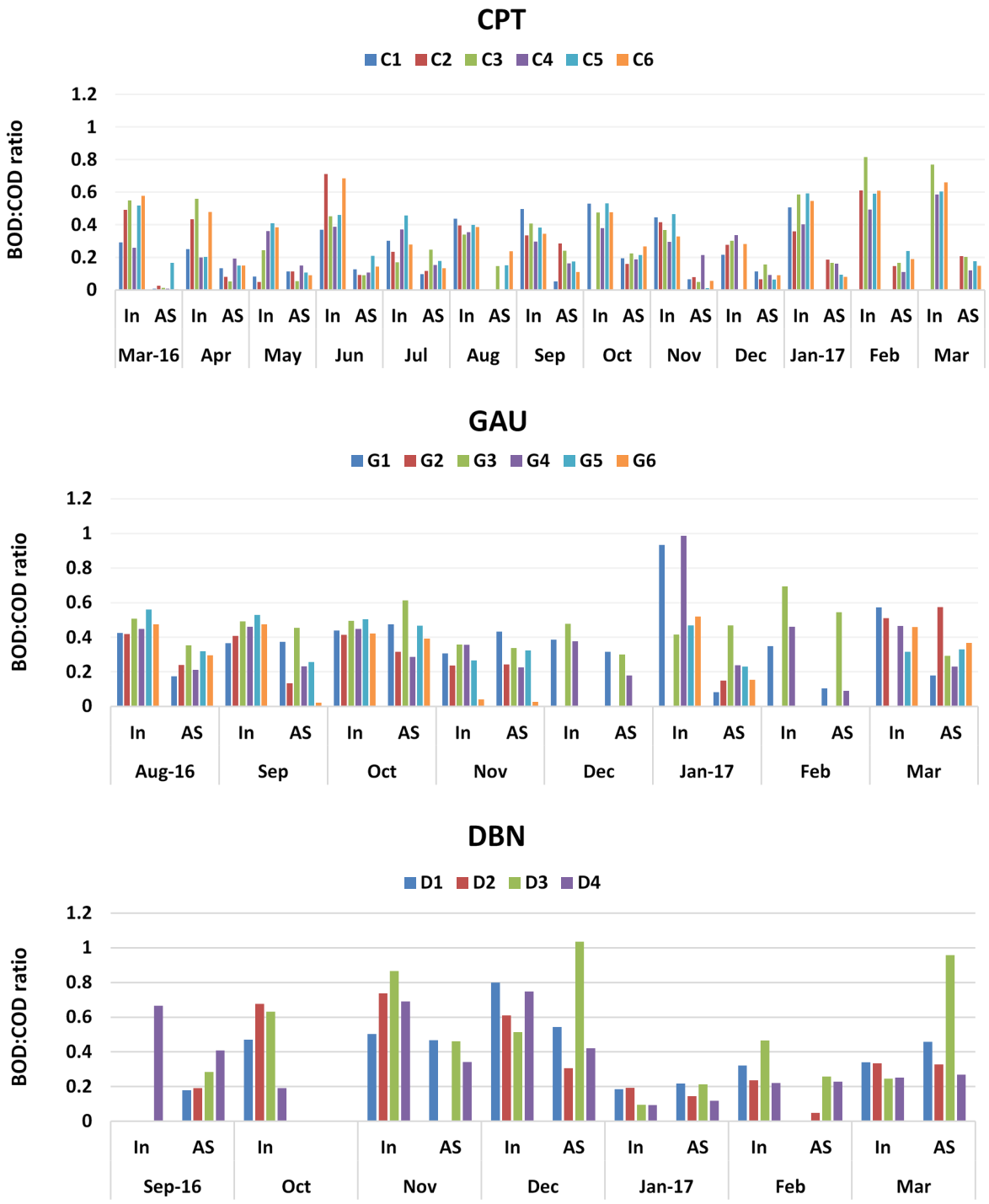


Figure 25: Comparison between the biological oxygen demand to chemical oxygen demand ratios in the influent and activated sludge

Despite the fact that many of the BOD:COD ratios in the AS of the GAU and DBN WWTPs were higher in the influent, there was still a general reduction in biodegradability from influent to AS (Figure 27). This was most noticeable in the CPT WWTPs, where the organic fraction could be classified as recalcitrant in 83% of AS samples as opposed to 1.4% in the influent samples. In addition, there were no AS samples where the organic fraction could be classified as either

moderately or readily biodegradable. On the other end of the spectrum, samples with the highest fraction of readily biodegradable organics was found in the DBN WWTPs (73.5%), of which 21% were still classified at readily biodegradable in the AS. These results are noteworthy from a WWTP performance perspective.

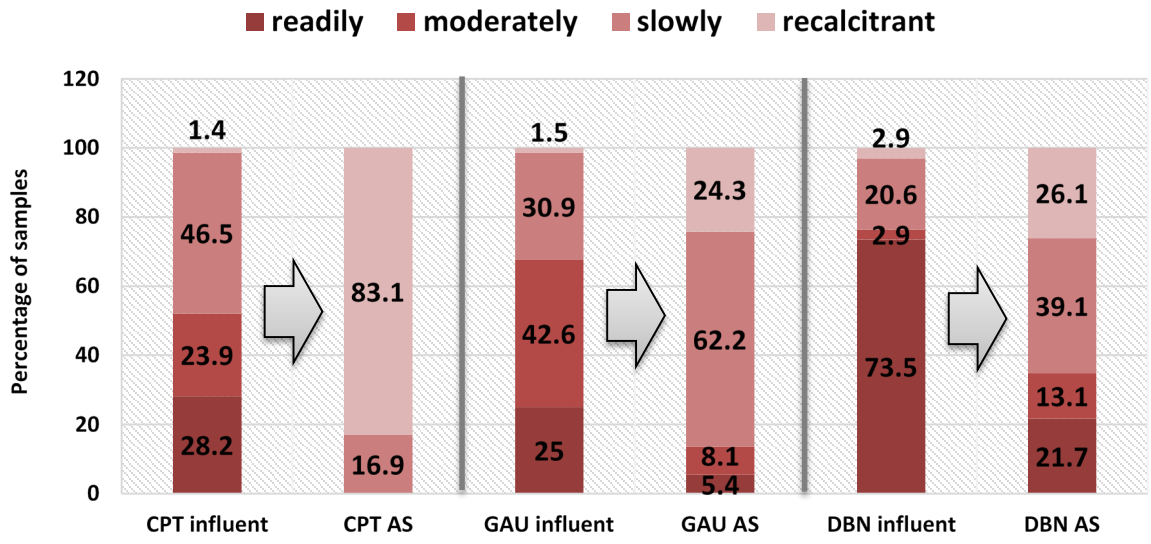


Figure 26: Decrease in biodegradability from influent to activated sludge using the biological to chemical oxygen demand ratio as a proxy: >0.5 = readily biodegradable; 0.4-0.5 = moderately biodegradable; 0.2-0.4 = slowly biodegradable; <0.2 = recalcitrant

4.3.5 Particulate and soluble biological oxygen demand and chemical oxygen demand in influent and AS

The BOD_5 is a technically challenging method because it is reliant on the growth of a microbial inoculum. Due to the variability of the microbial growth, there is a 20% inter-laboratory variance in results (Jouanneau et al., 2014). This is exacerbated when there are low concentrations of recalcitrant organics present (Jouanneau et al., 2014). It is therefore expected that if a COD method specific to low ranges is performed, results will be comparatively more accurate than BOD values. In this study, the average BOD_s concentrations were <50 mgL^{-1} in 13 of the WWTPs, and 50-100 mgL^{-1} in the other 3 WWTPs (Figure 28). The average COD_s concentrations were <100 mgL^{-1} in all the CPT and GAU WWTPs, but higher (150-200 mgL^{-1}) in all the DBN WWTPs (Figure 28). In terms of biodegradability, in contrast to the total organic fraction of the AS (shown previously in Figure 27), the soluble fraction in the GAU and CPT WWTPs were all moderately to highly biodegradable when using the BOD_s:COD_s ratio as a proxy (0.42 to 1.54), and highly recalcitrant in all the DBN WWTPs (0.01 to 0.04). The BOD_s:COD_s ratios in two of the CPT WWTPs (C1 and C2) were ≥1, which is theoretically impossible. When high BOD_s outliers (one for each WWTP) were removed, these values decreased to 0.84 and 0.42, respectively.

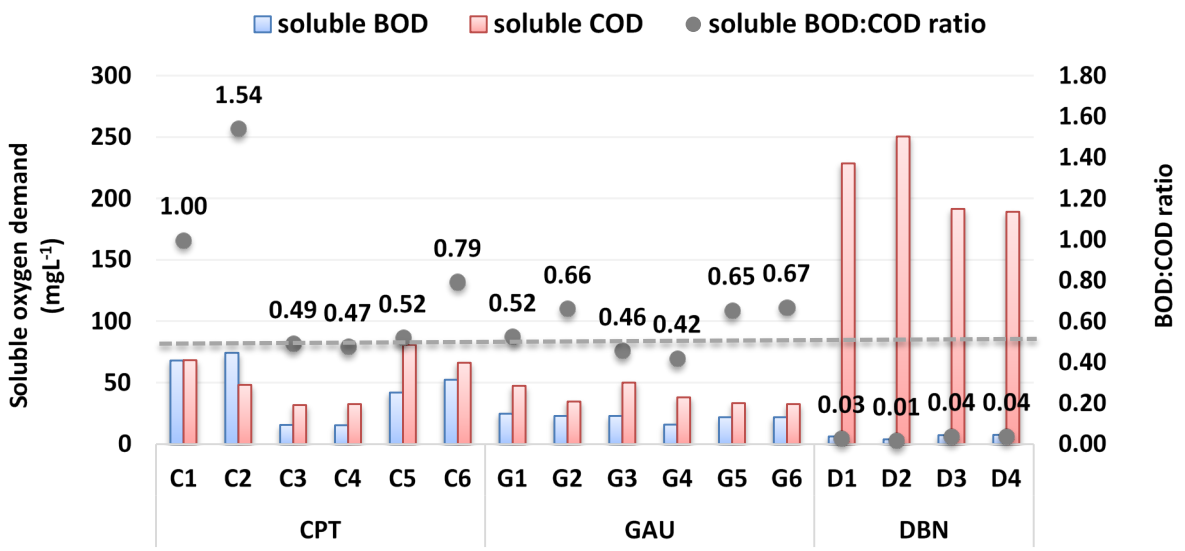


Figure 27: The average soluble biological and chemical oxygen demand concentrations measured in the activated sludge samples of all the study wastewater treatment plants. The dotted line denotes the ratio at which the soluble organic fraction is considered to be readily biodegradable

4.3.6 Food to microorganism ratio and biomass

It has been shown that the food to microorganism (F/M) is an extremely important factor for filament selection (Eikelboom, 2000; Jenkins et al., 2004). This parameter is often expressed using BOD or COD as a proxy for 'food'. The 'microorganism' concentration is often expresses using the mixed liquor suspended solids (MLSS) concentration as a proxy, i.e. the F/M ratio is expressed as mgBOD/mgMLSS.day^{-1} or mgCOD/mgMLSS.day^{-1} (e.g. Eikelboom, 2000; Jenkins et al., 2004)]. It is less complex to determine the MLSS concentration than the MLVSS concentration, because the latter requires additional incineration and weighing steps. However, MLVSS provides a better approximation of the microbial content because it represents only the organic fraction of the suspended solids. The use of MLVSS is therefore more technically correct for calculating F/M ratios and was therefore used in the F/M calculations in this study. The drawback is that it does not allow comparison of results with most literature values.

Even with the use of MLVSS, it is still unclear what portion (range) of the organic fraction is comprised of viable microorganisms, and will differ according to operating parameters in each WWTP. Results from a study conducted in Italy using flow cytometry showed that the active viable biomass constituted around 11% of the COD in AS, and that around 20% of the cells were non-viable (Foladori et al., 2010). Although the methodology used for this study was accurate and comprehensive, the study was limited to one WWTP. Further work using large numbers of WWTPs is therefore required to better understand the quantification of active biomass in different WWTPs.

In this study, the F/M ratio was generally lowest in the CPT WWTPs and highest in the GAU WWTPs (Table 10, Figure 29). The F/M ratios in some of the the GAU and DBN WWTPs were extremely erratic on a temporal basis, while they were more stable in the CPT cohort. This scenario was seen with many of the other parameters. The F/M ratios at G5 were particularly erratic, and have been excluded from Figure 29.

Table 10: Food to microorganism ratio calculated for each wastewater treatment plant used in the study

	F/M calculated with BOD (mgBOD/mg MLVSS.day^{-1})			F/M calculated with COD (mgCOD/mg MLVSS.day^{-1})		
	Average	Range	*n*	Average	Range	*n*
CPT						
C1	0.09 ± 0.03	0.04-0.13	9	0.22 ± 0.07	0.09-0.32	12
C2	0.11 ± 0.04	0.05-0.18	11	0.24 ± 0.06	0.16-0.33	11
C3	0.50 ± 0.02	0.03-0.11	13	0.11 ± 0.03	0.03-0.17	14
C4	0.17 ± 0.07	0.07-0.26	13	0.45 ± 0.12	0.25-0.64	14
C5	0.09 ± 0.07	0.05-0.29	12	0.19 ± 0.10	0.10-0.49	14
C6	0.06 ± 0.02	0.03-0.08	12	0.13 ± 0.04	0.07-0.20	14
GAU						
G1	0.38 ± 0.32	0.11-0.89	8	1.09 ± 1.00	0.25-0.64	8
G2	0.24 ± 0.29	0.04-0.76	10	0.61 ± 0.65	0.14-1.95	11
G3	0.13 ± 0.09	0.01-0.25	11	0.30 ± 0.22	0.01-0.70	12
G4	0.16 ± 0.27	0.02-0.91	10	0.24 ± 0.24	0.09-0.38	11
G5	0.74 ± 0.88	0.03-2.33	10	1.69 ± 2.05	0.11-5.75	11
G6	0.21 ± 0.16	0.07-0.49	9	0.52 ± 0.41	0.17-1.78	10
DBN						
D2	0.13 ± 0.10	0.001-0.18	10	0.27 ± 0.13	0.20-0.27	11
D2	0.13 ± 0.11	0.002-0.37	11	0.27 ± 0.25	0.02-0.60	13
D3	0.18 ± 0.09	0.05-0.32	10	0.25 ± 0.24	0.06-0.87	12
D4	0.21 ± 0.17	0.04-0.58	12	0.48 ± 0.60	0.02-2.22	13

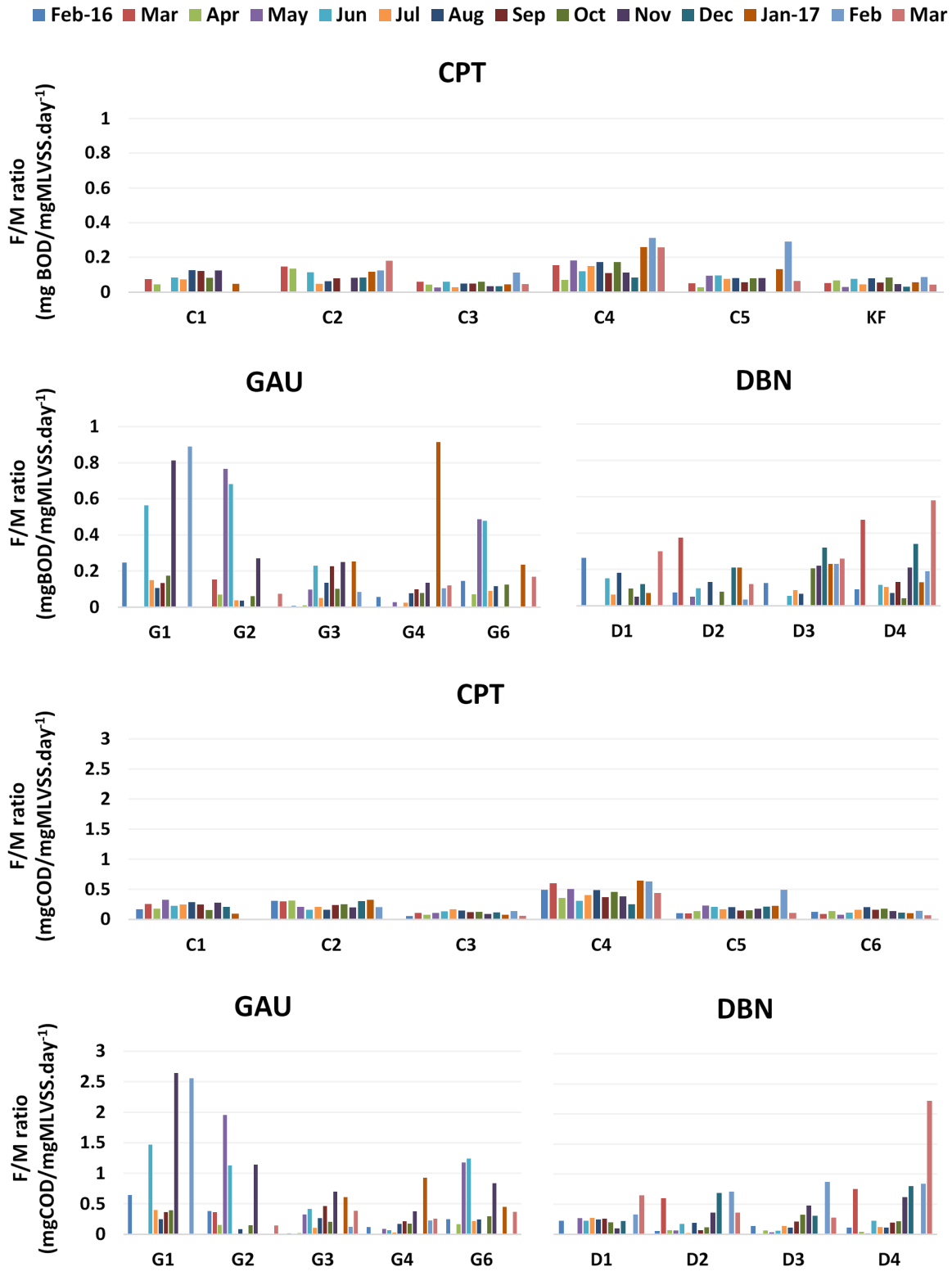


Figure 28: The food to microorganism ratios for biological and chemical oxygen demand calculated for the wastewater treatment plants during the study period

4.4 Inorganics: nitrogen, phosphorus and sulphate

4.4.1 Nitrogen

4.4.1.1 Ammonia

In the CPT WWTPs, the influent NH_3/NH_4^+ (as N) concentrations were relatively consistent at each WWTP during the study period. The highest values were found at SD (50-104 mgL^{-1}) and KF (48-92 mgL^{-1}). Using the influent to AS NH_3/NH_4^+-N values as a proxy for nitrification, it appeared that no nitrification took place at C5, but good nitrification took place in the other CPT WWTPs over most of the study period. Likewise, good nitrification took place at the GAU WWTPs, even with the consistently high influent NH_3/NH_4^+ experienced at G6 (Figure 30).

In contrast to CPT and GAU, the NH_3/NH_4^+-N concentrations in the samples from the DBN WWTPs were inconsistent. For example, in July 2016, the influent NH_3/NH_4^+ concentrations were >80 mgL^{-1}, but in most other months <20 mgL^{-1}. In many months, the influent NH_3/NH_4^+-N concentrations were low when compared to those at CPT and GAU. In D2 (2 instances and D3 (3 instances), the concentrations in the AS were higher than in the influent (Figure 30).

4.4.1.2 Nitrates/nitrites

The concentration of NO_3^-/NO_2^- (as N) in the reactors and effluent of WWTPs is complex as it is dependent on both nitrification ($NH_3/NH_4^+ \rightarrow NO_3^-/NO_2^-$) and/or denitrification ($NO_3^-/NO_2^- \rightarrow N_2\uparrow$). Furthermore, less NO_3^-/NO_2^- is expected in the effluent of BNR systems that are designed for nitrification/denitrification than CAS or MLE systems that are designed with nitrification as the only priority in terms of N. Influent N is removed by incorporation into biomass or via denitrification, with the remainder being present in the final effluent. A notable portion of denitrification can take place in the clarifiers (Koch et al., 1999; Monti et al., 2006).

In this study, as expected, the influent NO_3^-/NO_2^- (as N) concentrations were typically low from the WWTPs in GAU and CPT, with the exception of one sample (C4 in Apr 2016; 12.3 mgL^{-1}). In 91% of the samples the concentrations were ≤2 mgL^{-1}, and ≤3.5 mgL^{-1} in the remaining samples. Influent to C4 is approximately 80% industrial in origin, so intermittent chemical spikes are possible. No particular trends relating to any period and/or WWTP in GAU or CPT were noted. While the influent NO_3^-/NO_2-N concentrations were still generally low in the samples from the DBN WWTPs, only 61% were ≤2 mgL^{-1}. The highest values (4.0 mgL^{-1}) were obtained in samples from D1 (July 2016) and D3 (Oct 2016).

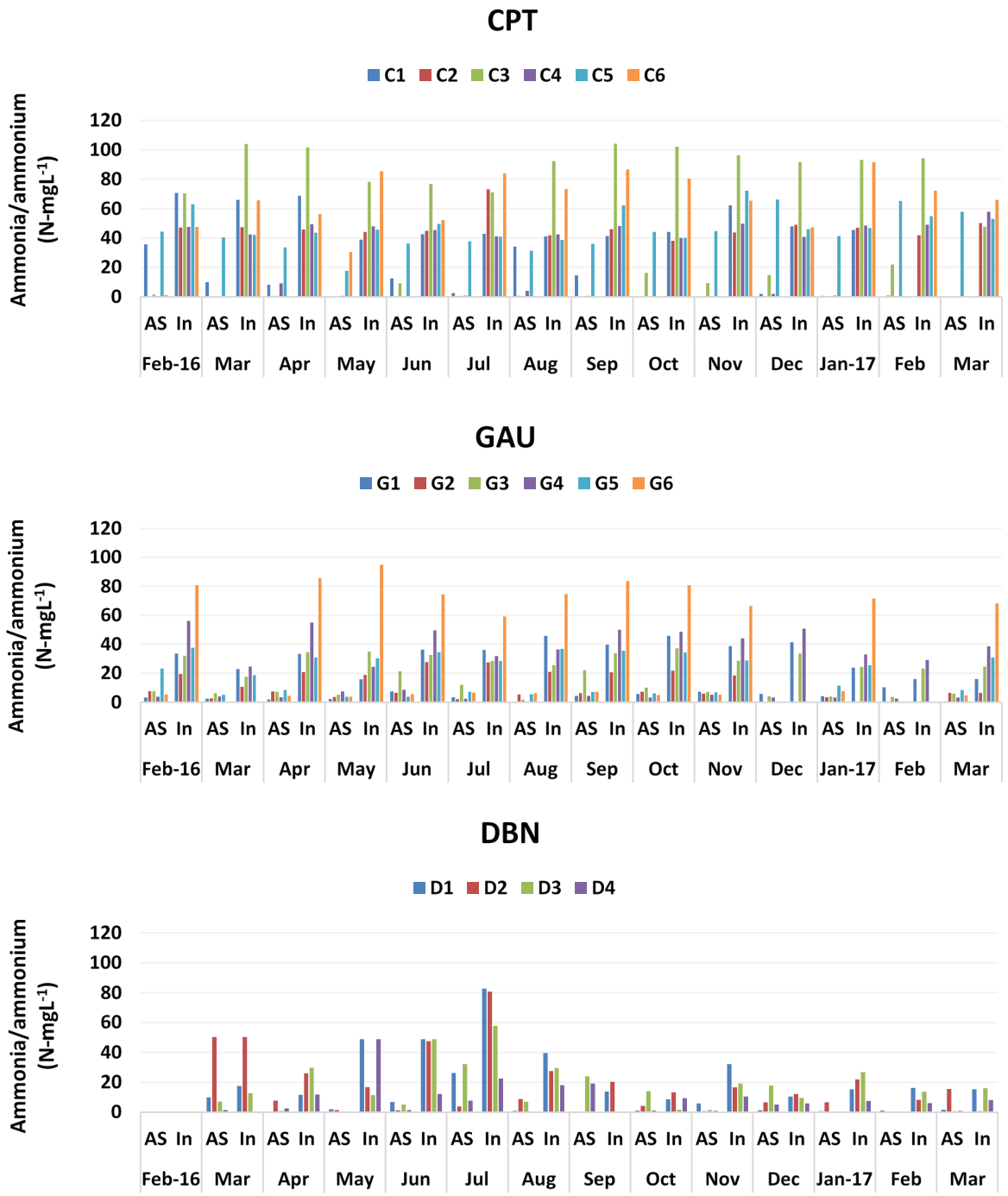


Figure 29: Monthly ammonia concentrations measured in samples during the study period

The NO_3^-/NO_2^--N concentrations measured in the AS samples were typically low during the study period (Figure 31), for the most part reflecting good denitrification. The lowest values were obtained at WV and in this instance reflect poor nitrification (Section 3.4.1.1), not good denitrification. In the CPT WWTPs, the most consistently high results were obtained at C4, followed by C3. These two WWTPs were the only non-BNR WWTPs in the CPT cohort, both being designed and operated with the MLE configuration. In contrast, comparatively low NO_3^-/NO_2^--N concentrations were found in the samples from the only non-BNR WWTP in the GAU

cohort. This may be explained by the low influent NH_3/NH_4^+-N concentrations (Section 4.4.1.1). KB was the only BNR WWTP in DBN, and apart from an anomalous result in June 2016 (13.6 mgL-1), the concentrations of NO_3^-/NO_2^--N in the monthly AS samples was ≤1.1 mgL-1. The high concentration in June can be explained by the concomitant high anomalous influent NH_3/NH_4^+-N concentration (Section 4.4.1.1).

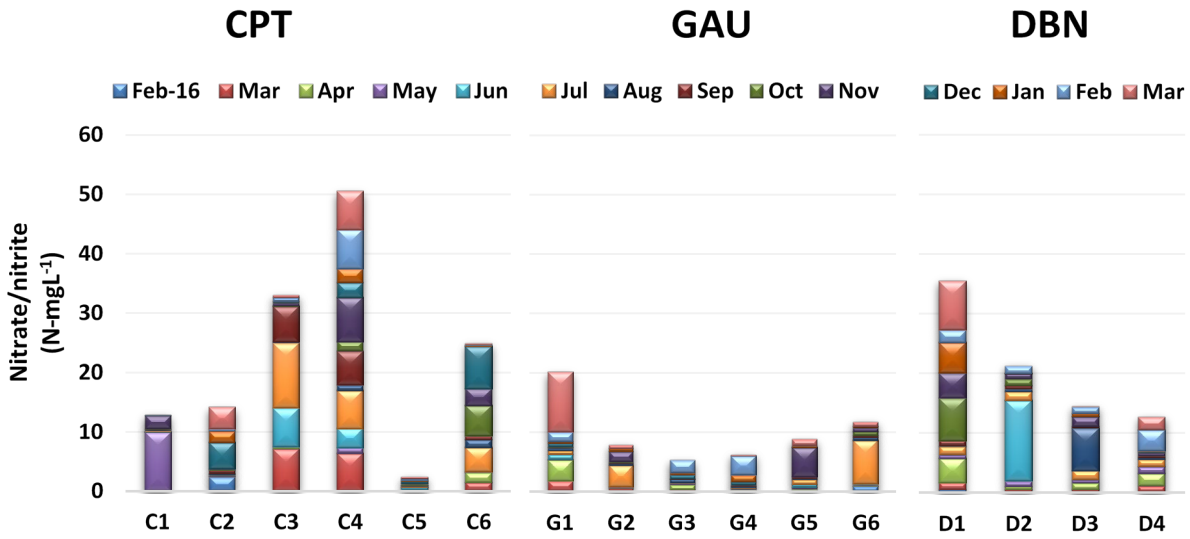


Figure 30: Monthly nitrate/nitrite concentrations measured in the activated sludge during the study period

4.4.2 Phosphate

The total phosphate (TP, measured as P) concentrations were not determined in the samples from the Durban WWTPs. Although there were some spikes in the influent TP concentrations in the samples from C3 from May to July 2016, the month-to-month variability in the influent TP and ratio of ortho-phosphate (σ-P, measured at P) to TP was typically low in the samples from the CPT WWTPs (Figure 32A,B). The influent σ-P:TP ratios of most samples from the CPT WWTPs were between 0.4 and 0.8. As with the other nutrients, the TP concentrations in the influent samples from the GAU WWTPs were lower than in the CPT samples, and there were some notable spikes in the samples from G1, G3 and G4. The lone spike of 193 mgL^{-1} in the sample from G1 in July was designated as spurious and removed from the statistical database. There was significantly more variability in the influent σ-P:TP, and in 2 instances, ratios >1 were obtained, which is theoretically impossible. These results were also removed from the statistical database.

The TP concentrations in the AS samples were significantly higher than those in the influent samples from the CPT WWTPs (Figure 32C). As with most influent parameters, the concentrations of TP in the AS from the GAU were notably lower, and in contrast to the CPT results, there were many instances when TP concentrations in the AS samples were lower

than those in the influent. At such low concentrations, this could be explained by fluctuating influent concentrations and/or a 'snapshot' sampling bias and/or low phosphate uptake by PAOs and/or intermittent low RAS recycle ratios or hydraulic retention times. The intermittent occurrence of low MLVSS concentrations at the GAU WWTPs suggests the latter may have played a role.

It is expected that σ-P in the bulk water will be taken up (at least to some extent) by PAOs in the aerobic zone (Ge et al., 2010). The relatively stable and high MLVSS concentrations measured in the AS samples from the CPT WWTPs indicated high biomass retention (Kumar et al., 2014). The results obtained (Table 11), support the theory that P was taken up by the active biomass in the AS and concentrated by recycling of RAS from the clarifiers in the CPT WWTPs.

Table 11: Influent:activated sludge ratio of the concentrations of total phosphate (as P) measured in the samples from the City of Cape Town study cohort

WWTP	Total phosphate concentration (influent:activated sludge ratio)		
	Average	Range	*n*
C1	0.11 ± 0.08	0.05-0.31	10
C2	0.07 ± 0.07	0.04-0.31	11
C3	0.17 ± 0.08	0.07-0.22	11
C4	0.20 ± 0.07	0.13-0.34	11
C5	0.09 ± 0.03	0.05-0.16	10
C6	0.08 ± 0.02	0.05-0.11	11

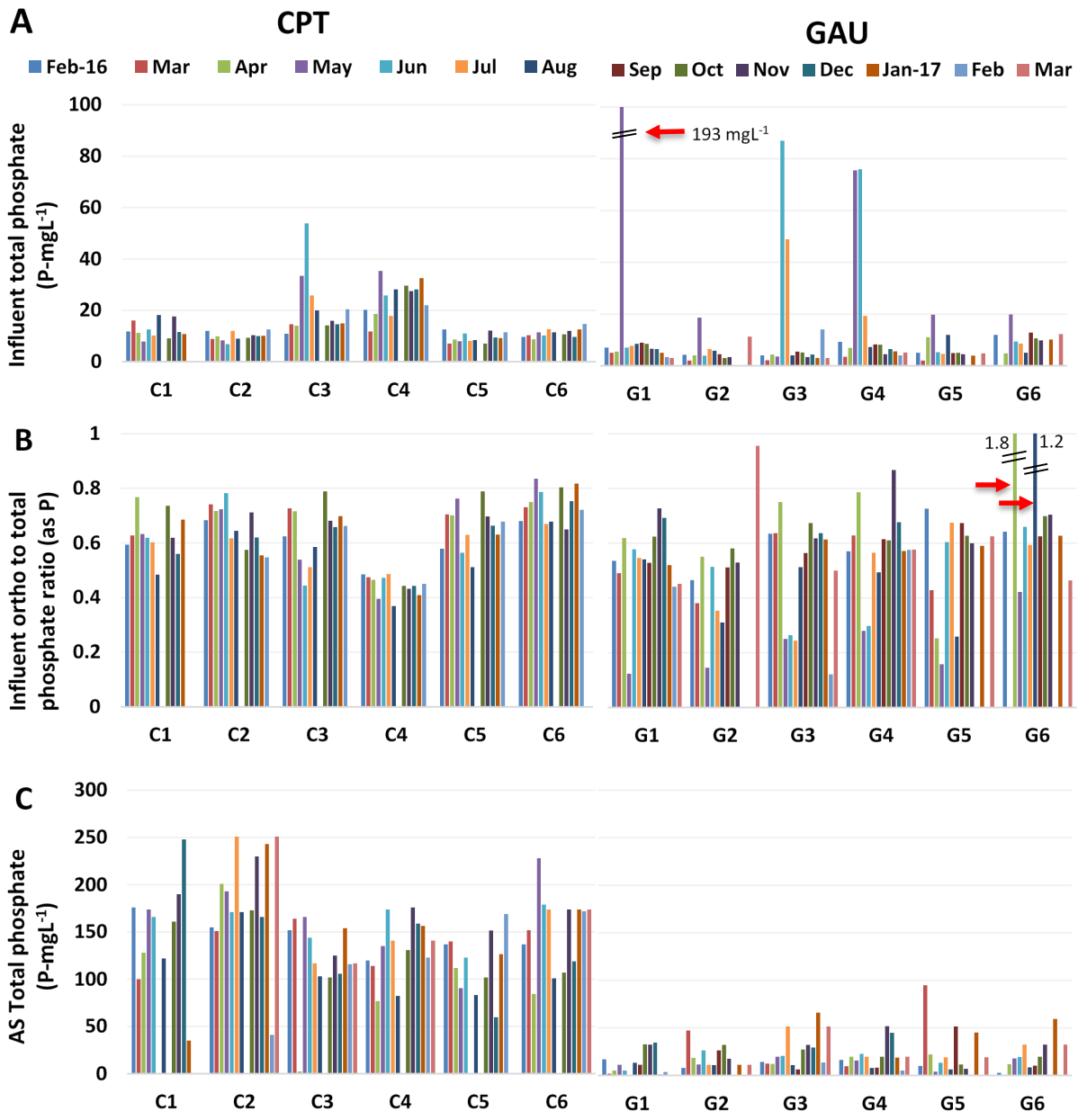


Figure 31: Influent total phosphate concentrations (A), ratio of ortho-phosphate to total phosphate (B), and AS total phosphate concentrations (C) in samples from the Cape Town and Gauteng study cohort

The σ-P:TP ratio in the influent samples from the CPT and GAU WWTPs were relatively consistent (Table 12). The σ-P:TP ratios in the AS samples from the CPT WWTPs were significantly lower than the influent σ-P:TP ratios ($p<0.001$). As highlighted in red in Table 12, the average σ-P:TP ratios in the AS samples from SD (0.109±0.274) were higher than those from the other CPT WWTPs (0.07±0.009-0.064±0.081), but fell into a similar range when a high outlier was removed (influent TP: 3 mgL^{-1}). The results from CPT further support the hypothesis that the influent σ-P was taken up by the active biomass in the AS, thereby moving from the bulk liquid onto the MLVSS.

In theory, high concentrations of P associated with solids, and concurrent low concentrations in the bulk water should ensure good P removal in clarifiers. However, long retention times in the clarifiers could lead to release of P by the PAOs back into the bulk water and have the opposite effect. Without effluent results to assess system performance, it is not possible to speculate further on this theory.

In contrast to the CPT results, the σ-P:TP ratios in the AS samples from the GAU WWTPs were not significantly different to the influent ratios ($p>0.05$). In 4 of the 6 WWTPs, the ratios calculated from the AS samples were higher than those from the influent samples. This could be ascribed to one or more of the reasons alluded to previously, i.e. fluctuating influent concentrations and/or a 'snapshot' sampling bias and/or low phosphate uptake by PAOs and/or intermittent low RAS recycle ratios or hydraulic retention times.

Table 12: Ortho-phosphate:total phosphate ratios measured in the influent and activated sludge samples from the Cape Town and Gauteng wastewater treatment plants

WWTP	Influent samples			Activated sludge samples		
	Average	Range	*n*	Average	Range	*n*
CPT						
C1	0.63 ± 0.08	0.48-0.77	11	0.061 ± 0.073	<0.01-0.230	10
C2	0.67 ± 0.08	0.55-0.78	12	0.007 ± 0.009	<0.01-0.026	11
C3	0.63 ± 0.10	0.44-0.79	12	0.109 ± 0.274	<0.01-0.961	12
				0.031 ± 0.059	<0.01-0.207	11
C4	0.44 ± 0.04	0.40-0.49	12	0.064 ± 0.081	<0.01-0.255	12
C5	0.66 ± 0.08	0.51-0.79	12	0.049 ± 0.066	<0.01-0.236	11
C6	0.74 ± 0.06	0.65-0.82	12	0.020 ± 0.040	<0.02-0.145	12
GAU						
G1	0.54 ± 0.14	0.44-0.72	14	0.385 ± 0.264	<0.01-0.799	12
G2	0.43 ± 0.20	0.31-0.99	11	0.806 ± 0.508	0.05-2.192	12
G3	0.53 ± 0.20	0.25-0.67	14	0.621 ± 0.367	<0.01-1.147	14
G4	0.58 ± 0.16	0.28-0.89	14	0.448 ± 0.274	0.01-0.923	14
G5	0.51 ± 0.19	0.16-0.73	11	0.682 ± 0.501	0.11-1.870	12
G6	0.80 ± 0.39	0.46-1.77	12	0.861 ± 1.156	0.22-4.286	12

4.4.3 Sulfates

Sulfides (S^{2-}) are typically found in concentrations orders of magnitude lower than sulfates (SO_4^{2-}) in domestic wastewater and AS. In addition, the former species is less stable and concentrations should ideally be measured immediately after sampling, which is not practical. Due to these factors, only SO_4^{2-} concentrations were ascertained in the influent and AS.

The presence of SO_4^{2-}, especially in instances where there are low concentrations of DO and NO_3^-, can lead to the over-proliferation of SO_4^{2-} reducing bacteria (SRB) such as Eikelboom Type 021N (Yamamoto-Ikemoto et al., 1994). This filament has been found ubiquitously in the South African WWTPs. The presence of S^{2-} on the other hand, may select for S^{2-} oxidising bacteria, such as the filamentous *Thiothrix* spp. These organisms are also capable of storing

and utilising thiosulfate (Nielsen et al., 2000). *Thiothrix* spp. have rarely been found as dominant species in South African WWTPs, but were very prevalent along with Type 021N in C5 during the study period. A number of spurious results were identified and omitted from the statistical database and this data analysis (Figure 33).

When outliers, and non-paired values were removed, there was no significant difference ($p > 0.1$) in the overall SO_4^{2-} concentrations measured in the influent and AS of the samples from the CPT WWTPs (Influent = 63 ± 26 mgL^{-1}; AS = 67 ± 22 mgL^{-1}) or the DBN WWTPs (Influent = 50 ± 38 mgL^{-1}; AS = 46 ± 27 mgL^{-1}). However, the concentrations of SO_4^{2-}measured in the samples from the GAU WWTPs were significantly higher ($p<0.001$) in the influent (80 ± 38 mgL^{-1}) than in the AS (57 ± 15 mgL^{-1}).

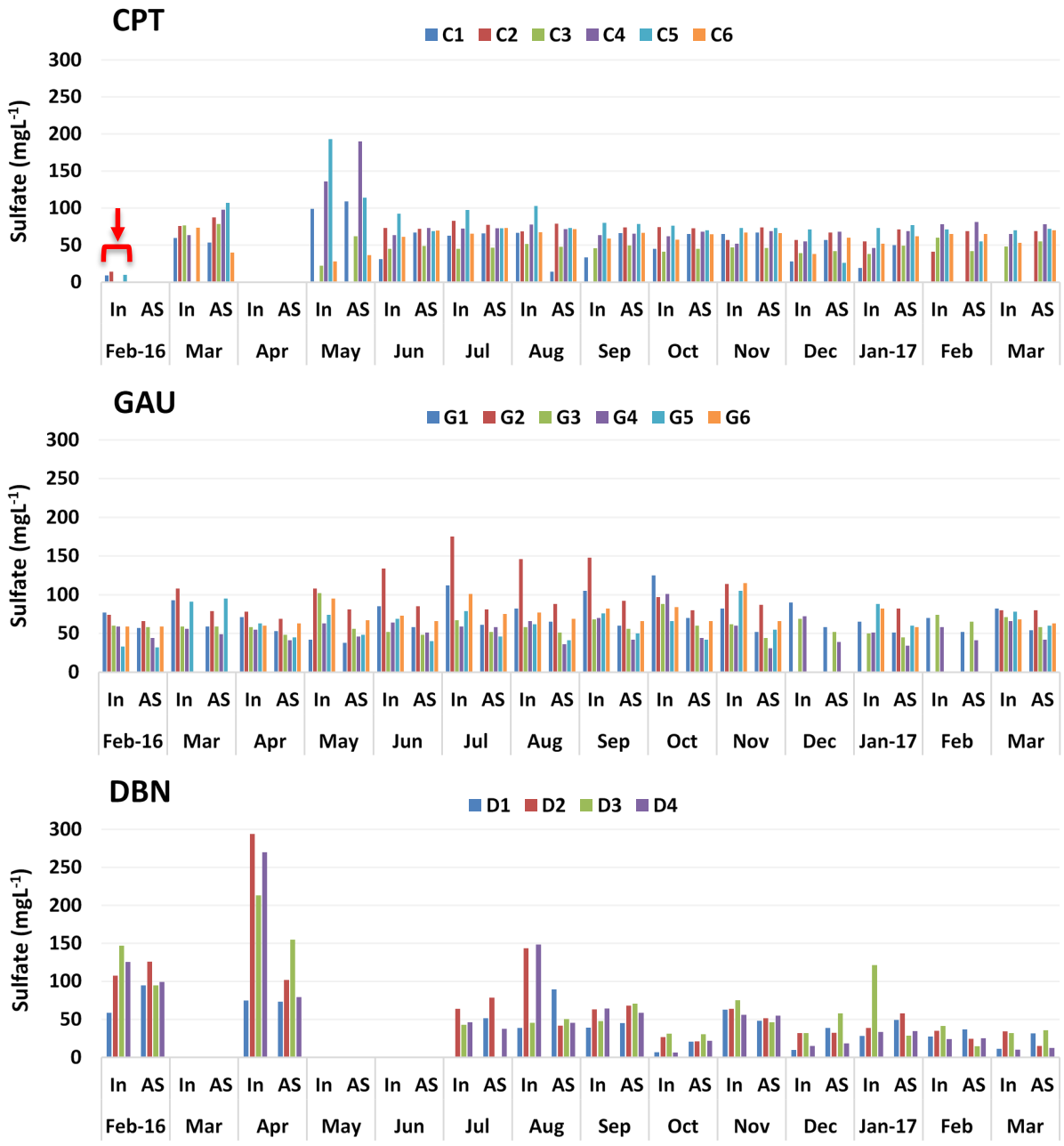


Figure 32: Sulfate concentrations measured in influent and activated sludge samples

CHAPTER 5: STATISTICAL MODELS

5.1 Approach

The results from all of the locations (CPT, GAU, DBN) were included in the analyses discussed in Chapter 4. After collating the data, it became clear that the results for DBN were erratic and exhibited distinct differences from those from CPT and GAU. It was hypothesised that this was because the influent to most of the DBN WWTPs contained high percentages of industrial effluent. There was a strong likelihood that the complexity of the influent would increase the number of variables that were not measured (e.g. toxic chemicals). It was therefore decided, in consultation with the reference group, to exclude the DBN WWTPs from the statistical model.

Due to the lack of a strong correlation between DSVI and FI, prediction of bulking was not included in the models because it was impossible to substantiate the presence of filamentous bulking. Only 8 samples simultaneously exhibited a DSV≥150 and FI≥5, strongly suggestive of filamentous bulking (Table 13). Although Type 0092 was present in 62.5% of these samples, this was less than the overall prevalence of this filament, probably indicating incidental presence in the 'bulking' samples as previously described by Welz et al. (2104) in a survey of the filamentous populations in 11 WWTPs in Cape Town. It is recommended that further research is conducted using the dichotomous key presented in Section 5.4 to establish the aetiology of bulking in different WWTPs in SA.

Table 13: Filaments found in samples with elevated diluted sludge volume index and filament index

DSVI	FI	Most dominant	2nd most dominant	3rd most dominant
290	5	Type 0092 & Type 021N	2 TYPE 0041	
216	5	Type 021N & Type 0092	GPBB and Type 0041	
226	5	*M. parvicella*	Type 0092 and Type 021N	
182	5	Type 0092	*M. parvicella*	
257	6	*Thiothrix* spp. & Type 021N	*M. parvicella*	
207	5	*M. parvicella*	Type 021N	Type 0092
663	6	*Thiothrix* spp. & Type 021N	*M. parvicella*	
176	6	*Thiothrix* spp.		

5.2 Model application and validation

5.2.1 Comparison of existing correlation analyses and models

Complex models have been used to correlate bulking indicators with WWTP operational parameters. Lou and Zhao (2012) used principal component analysis (PCA) and artificial neural networks (ANN) to correlate SVI with operational parameters, and Singh et al. (2008) used ANN and support vector regression to predict clarifier efficiency.

Limited research has been conducted to establish the link between physicochemical parameters and filament selection. Milobędska et al. (2016) used simple correlations (Pearson's, Spearman's rank) supported by analysis of variance (ANOVA), PCA, and cluster analysis to determine statistical linkages between operational variables and filament selection in domestic and mixed industrial/domestic WWTPs. Samples were taken bi-annually for 2 years. The number of WWTPs (*n*=5), samples (*n*=45), and physicochemical parameters [TP, TN (influent); BOD COD (AS), sludge age, SVI] were low. Some weak correlations were noted, but the authors conceded that larger studies were required.

Dos Santos et al. (2015) took samples from 16 WWTPs in Portugal every 3 months for a 2 year period. The total number of samples (*n*=128) was notably lower than those taken during the course of Project K5/2471/3 (*n*=224). Data was missing from the dataset of each study, which decreased the number of samples included in the statistical analysis. In the case of Project K5/2471/3, the DBN results were also excluded from the database, reducing the number (*n*=142). This was still notably higher than the comparative study (*n*=78). In addition, for both studies, the data applicable to each filament type decreased with the prevalence of that filament, and for some parameters, less data was available. For example, there were relatively large data gaps for F/M ratio (Project K5/2471/3). Table 14 shows the final number of samples included in the comparative statistical analyses (Table 15) per dominant filament type. Only the seven most prevalent filaments found in the South African cohort (with Type 021N and *Thiothrix* spp. being classified together) are shown in Table 14. Some filaments dominated in the Portuguese, but not in the South African WWTPs, and were excluded [*Sphaerotilus natans* (1% prevalence), Type 0581 (3% prevalence), Type 0914 (2% prevalence)]. Despite the data gaps, the number of samples was significantly higher for Project K5/2471/3.

Table 14: Number of samples containing specific dominant filaments that were correlated with physicochemical parameters

Dominant filaments	Number of samples	
	Project K5/2471/3*	Dos Santos et al. (2015)
Type 0092	80 (F/M ratio), 97-110 (other parameters)	7
M. parvicella	36 (F/M ratio), 44-50 (other parameters)	8
Type 0041	29 (F/M ratio), 35-40 (other parameters)	7
Type 021N/*Thiothrix* spp.	50 (F/M ratio), 62-70 (other parameters)	4
Type 1851	43 (F/M ratio), 52-59 (other parameters)	10
GPBB	14 (F/M ratio), 17-20 (other parameters)	2
Type 0803	8 (F/M ratio), 10-11 (other parameters)	N/A

*Different numbers due to data gaps. Only parameters presented in Table 15 were assessed.

There were some differences between the two studies, including the scale, but there were also many similarities. Dos et al. (2015) measured the following parameters: pH, COD, BOD, TSS, TP, TN, NH_4-N (influent); pH, SVI, SRT, F/M ratio, DO, TSS (aeration tank); pH, COD, BOD, TSS, TP, TN, NH_4-N, NO_3-N (effluent). The authors did not specify at which point in the 'aeration tank' the samples were collected. The F/M ratios for both studies was based on the MLVSS, not the MLSS. This is preferable for reasons alluded to in Section 4.3.6. In contrast to

Project K5/2471/3, temperature, BOD:COD ratio, total alkalinity, σ-P, SO_4^{2-}, VFAs were not determined, and SVI was used *in lieu* of DSVI. The inclusion of effluent parameters by dos Santos et al. (2015) is questionable because this has no bearing on filament selection in the bioreactors. Only dominant filaments were included in the database for Project K5/2471, whereas dos Santos et al. (2015) included all filaments, irrespective of dominance.

To compare results, the methodology, used by dos Santos (2015) was applied to the data generated during project K5/2471/3. This was based on determining simple Kendall correlation coefficients (Appendix 5). The only result that was mutually significant was the correlation between influent pH and Type 0092 (given in red).

It was concluded that the use of simple correlations may not adequately describe the influence of physicochemical parameters on filament selection. In addition, the small database for the comparative study, and for the less prevalent dominant filaments in Project K5/2471/3 was unlikely to yield statistically valid correlations.

Table 15: Comparison of the significance of the Kendall correlation coefficients between selected filaments and operational parameters from the comparative study (dos Santos et al., 2015, shaded) and this study (non-shaded)

	Type 0092		*M. parvicella*		Type 0041		Type 1851		GPBB	
Influent										
pH	*	*	NS	NS	NS	NS	NS	NS	NS	*
COD	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
BOD	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TP	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS
TN	NS	NS	NS	NS	NS	NS	NS	*	NS	NS
NH_4-N	NS	NS	NS	NS	NS	NS	NS	*	NS	NS
Activated sludge/aeration tank										
pH	NS	NS	NS	*	NS	NS	**	NS	NS	NS
DSVI/SVI	NS	**	NS	NS	NS	NS	NS	*	NS	NS
F/M ratio	NS	**(-)	NS	NS	NS	NS	NS	NS	NS	NS
DO	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS

Levels of significance: **<0.01≥*0.05 NS = not significant

Mielczarek et al. (2012) conducted a larger study, which compared favourably to Project K5/2471/3 in terms of database size. The researchers took seasonal samples (*n*=147) from 28 operationally stable domestic and mixed industrial/domestic nutrient removal WWTPs in Denmark. Similar to dos Santos et al. (2015), they attempted to determine statistical linkages by using simple correlations (Pearson's) supported by PCA, multivariate ANOVA, and cluster analyses. They found no significant intra-WWTP differences in the filamentous consortia, but highly significant ($p<0.001$) inter-WWTP differences. From these results they concluded that although some intra-WWTP seasonal succession took place, the filamentous consortia in each WWTP were unique. They only found weak correlations between the filaments analysed and operational variables, apart from a strong correlation between "alternating operation" and *Thiothrix*/Type 021N.

The project team (Project K5/2471/3) hypothesised that although only weak correlations were demonstrated between filaments and physicochemical parameters in previous studies, this does not necessarily mean that physicochemical parameters are not important criteria for filament selection. The correlation studies suggested that many physicochemical factors may act in concert to select the filamentous consortia. The team hypothesised that further insight may be provided by the use of different statistical models. This was based on a logistic model that was recently proposed as a tool to understand the factors that influence the selection of different filaments in WWTPs (Deepnarain et al., 2015). One of the five aims of Project K5/2471/3, as outlined in the proposal, was to validate this particular model using a larger data set.

The model was based on very limited data [results from one WWTP (Kingsburgh, DBN)]. Results obtained from samples taken over the period of a year from the anaerobic, anoxic, and aerobic zones, as well as the clarifier and RAS stream were used to generate data for the model. After screening for the most significant parameters, those included in the model were limited to the influent COD, influent NH_4-N, PO_4-P (aerobic zone), F/M ratio, DO, pH and temperature. A cut-off value was applied to this data, which was then categorised as high (1) or low (0).

The same methodology was applied to the data obtained during Project K5/2471/3. An ordinal logistic model with a cumulative logit link function was used to assess whether there were any significant relationships between particular filaments and selected operational parameters as described by Deepnarain et al. (2015). A comparison of the results of the original outputs (shaded), together with the results generated with the data from this study (not shaded) are summarised in Table 16. The original statistical analyses using this model for Project K4/2471/3 are included in Appendix 5.

With the exception of *Thiothrix* spp./Type 021N and pH, the model did not predict any significant relationships between filaments and the selected parameters when applied to the larger data set. The use of cut-off values and the small data set was questioned. It was clear that the original model could not be validated with a more inclusive data set.

A number of different models were then applied to the dataset generated during Project K5/2471/3. Temperature was excluded because there were some data gaps, and it was shown that it did not play a significant role on the selection of any of the major filaments (Section 3.2.2). All of the influent parameters, as well as the DO, DSVI, MLVSS and F/M ratios (COD and BOD) were included. Due to spurious results, probably due to experimental inaccuracy at low concentrations, the BOD_t, COD_s and BOD:COD ratios in the AS were excluded, but all other AS parameters were included.

It was found that the generalised linear model (GLM) using a binary distribution for the response variable (each of the indicator variables), and a logit link function, was the most promising model.

Table 16: Cumulative logistic outputs comparing the significance of relationships between selected filaments and operational parameters from the original study (Deepnarain et al., 2015, shaded) and this study (non-shaded)

	Type 0092		Type 1851		Type 0041		*Thiothrix* spp./Type 021N	Type 021N	*Thiothrix* spp.	*M. parvicella*	GPBB	*S. natans*
$COD_{inf.}$	NS	*	NS	*	NS	NS	NS	*	*	NS	NS	*
NH_4-$N_{inf.}$	NS	*	NS	*	NS	*	NS	**	NS	NS	NS	**
DO	NS	*	NS	*	NS	NS	NS	NS	NS	NS	NS	**
F/M	NS	*	NS	*	NS	NS	NS	*	*	NS	NS	*
Temp.	ND	*	ND	NS	ND	NS	ND	NS	NS	ND	ND	NS
pH	NS	NS	NS	NS	NS	*	*	NS	NS	NS	NS	NS
PO_4-$P_{aer.}$	ND	NS	ND	*	ND	NS	ND	NS	NS	ND	ND	NS

Levels of significance (Pr > ChiSq): **<0.01≥*0.05 NS = not significant ND = not determined

COD_{inf} = influent chemical oxygen demand; NH_4-$N_{inf.}$ Influent NH_4 as N; PO_4-P_{aer} = PO_4 as P in aerobic zone

5.3 Application and explanation of the generalised linear model

The GLM was applied separately to each of the seven most prevalent filaments. Filaments were ranked according to abundance in the samples. A new binary variable was created for each filament. These variables indicated which filament was most dominant. For example, if Type 0092 was most dominant for the specific case, then the indicator binary variable was assigned a "1". Conversely, if the Type 0092 was not the most dominant for the specific case, then the indicator binary variable for Type 0092 was assigned a "0".

The GLM is of the family of linear models that includes analysis of variance and regression models. It is a generalised form of the classic linear model.

The classic linear model has the form:

$$\boldsymbol{E(Y) = a + bx} \quad \text{or} \quad \boldsymbol{Y = \mu + \epsilon}$$

Where a = the intercept

b = the slope

x= independent variable

Y = dependent variable

Classic linear models assume that all observations are independent of each other and are normally distributed. Variables that indicate prevalence of each of the filaments at a time are dichotomous (i.e. of binomial distribution), and are, as such, not normally distributed.

The GLM consists of three components, a random component, a systematic component and a link function. The assumptions of the classic linear model, as outlined by McCullagh and Nelder (1989), are:

1) each component of the dependent variable, Y, is independent and normally distributed, having a common variance (random component),

2) the covariates are combined to give the linear predictor (systematic component), $\boldsymbol{\eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}}$

3) a link function, $\boldsymbol{g(\cdot)}$, which specifies the relationship between the random component and the systematic component
$\boldsymbol{g(p_i) = \eta_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_k X_{ik}}$

In the case of the GLM, the first assumption is relaxed such that the dependent values may be from one of the exponential family of distributions, the variance does not have to be common, and the link function mentioned in the third assumption is monotonic and differentiable. The exponential family is the class of distributions that includes the normal, Poisson, gamma, inverse Gaussian, binomial, exponential and other distributions. Link functions are chosen according to the data type and the context of the data. In this study, the dependent variable is a binary variable, thus a logit link function $\boldsymbol{g(p) = \ln \frac{p}{1-p}}$ was selected, where p, for example, is the probability (p) of a specific profile making a specific selection (Simonoff, 2003).

In this study, many of the independent variables showed evidence of multicollinearity with influent BOD:COD_t, pH, NH_3/NH_4^+-N, NO_3^-/NO_2-N, TP, SO_4^{-2}-S and VFA concentrations, and AS pH, NH_3/NH_4^+-N, NO_3^-/NO_2-N, and σ-P. Multicollinearity occurs when at least one explanatory variable is closely related to one or more other explanatory variables. In the case of this study, the presence of multicollinearity was not surprising, as the concentrations of many parameters are positively related to the 'strength' of domestic wastewater. However, the results obtained for the GLM were similar when the model was limited to these variables only, or was applied with all the chosen variables included.

Although the model was a good fit for the more prevalent filaments where the data set was larger (5.3.1-5.3.3), additional data is needed to further validate this model, especially for the less prevalent filaments (Section 5.3.4).

5.4 Results obtained with the generalised liner model

5.4.1 Eikelboom type 0092

Eikelboom Type 0092 was the most prevalent filament, being found as a dominant filament in 68.9% of AS samples. The results of the statistical analyses (Appendix 6) indicated that the model could be used (deviance = 1.003).

Although the model was a good fit, the predictability of the model was not good because multicollinearity caused measurements to be eliminated. Only one variable, namely pH, was retained by the model and the model intercept was significant ($p < 0.05$). The odds ratio showed that each increase in pH by one unit doubled the chances of Type 0092 being present.

Although less AS samples exhibited pH values in the low (6.1-6.9) or high (8.0-8.6) ranges, there was an increase in the % of samples in which Type 0092 was the most dominant filament in each range from low to high (Figure 34), substantiating the results of the statistical model. In addition, dos Santos et al. (2015), also demonstrated a significant link between pH and Type 0092 (Kendall correlation co-efficient).

The fact that only one variable was 'singled out' is not completely unexpected: Type 0092 is ubiquitous in SA WWTPs (Welz et al., 2014) which indicates that it can flourish in most WWTPs, i.e. it does not have exacting growth requirements. There is only limited information available from pure culture studies, which suggest that this filament can utilise a variety of short chain fatty acids and sugars as sources of carbon (Nielsen et al. 2009).

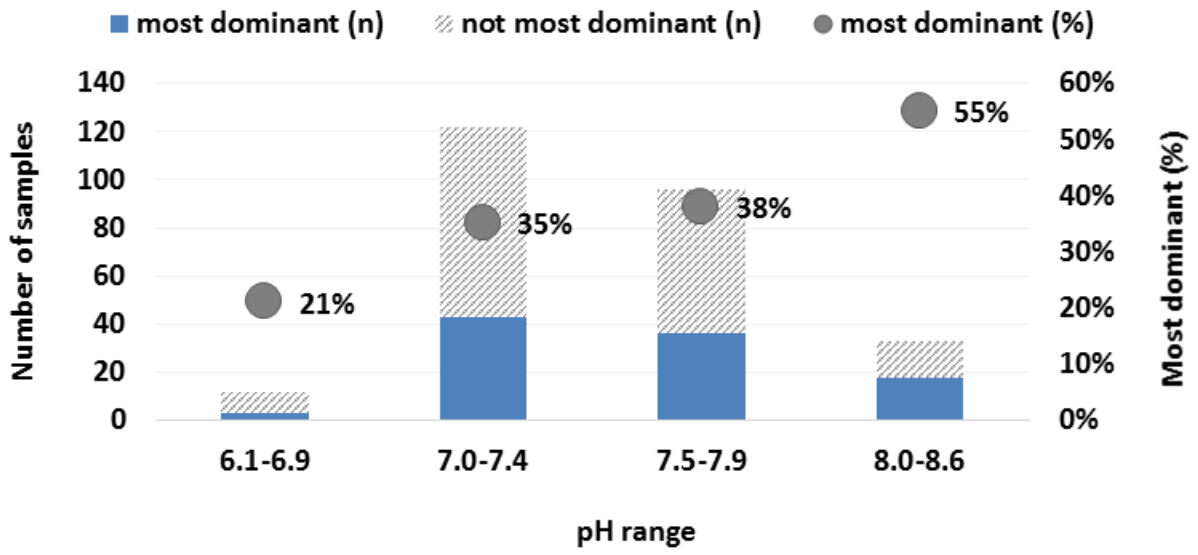


Figure 33: Relationship between dominance of Eikelboom type 0092 and pH

5.4.2 *Microthrix parvicella*

M. parvicella was the second most prevalent dominant filament found in this study. It is a well-known 'nuisance' filament, being the most common cause of filamentous bulking in AS WWTPs.

The statistical analyses (Appendix 7) showed that the GLM was significant and a good fit for *M. parvicella*, with a p-value of <0.05 and deviance value of 0.773 (a deviance of ~1 is seen as ideal). According to the model (Appendix 7), only two covariates were important for filament selection, namely the influent SO_4^{-2} concentration, and the AS σ-P concentration. When comparing results to previously described models, the SO_4^{2-} concentration was not measured, and was therefore not included in the model described by Deepnarain et al. (2015) and the correlations described by other researchers (Section 5.1). In addition, the σ-P in the aeration

tank was not measured by dos Santos et al. (2015), and *M. parvicella* was not prevalent in the WWTP included in the study conducted by Deepnarain et al. (2015).

In this study, the parameter estimates [exp(B)s in Appendix 7], were close to, but smaller than 1, indicating that an increase in either of the covariates would result in a small decrease in the other, determined by Equation 3 (Garson and Davids, 2013):

- Exp(β_0)= e^{β_0} = the odds that the characteristic is present in an observation i when $X_i = 0$

- Exp(β_1)= e^{β_1}= for every unit increase in Xi, the odds that the characteristic is present is multiplied by exp(β_1). This is an estimated odds ratio.

$$e^{\frac{\beta_0+\beta_1(X_i+1)}{\beta_0+\beta_1 X_i}} = e^{\beta_1} \qquad \textbf{Eq. 3}$$

The logistic model stipulates that the effect of a covariate on the chance of 'success' is linear on the log-odds scale, or multiplicative on the odds scale.

- If $\beta_j > 0$, then Exp(β_j) > 1, and the odds increase.

- If $\beta_j < 0$, then Exp(β_j) < 1, and the odds decrease.

In light of the GLM results, and because *M. parvicella* did not dominate at high influent SO_4^{-2} concentrations (Figure 35), it was concluded that a combination of low influent SO_4^{-2} and AS σ-P were important for robust growth of this filament in the CPT and GAU WWTPs. This is a novel finding.

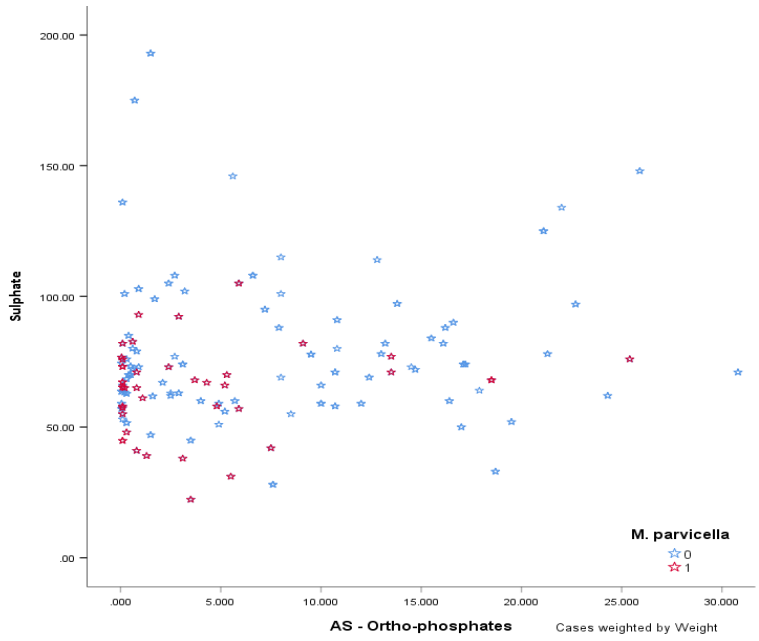


Figure 34: Scatter plot of influent sulphate and activated sludge ortho-phosphate: selection of *Microthrix parvicella*

Filaments are difficult to culture under laboratory conditions, and currently only three pure culture studies on *M. parvicella* have been described (Rosetti et al., 2005; Nielsen et al., 2009). These studies suggested that *M. parvicella* is a facultative heterotroph, capable of utilising either oxygen or nitrate as a terminal metabolic electron acceptor. It has therefore been hypothesised that it may survive in a wide range of oxygen partial pressures in WWTPs. Indeed, previous research has indicated that low DO environments are selective for *M. parvicella* (Rosetti et al., 2005). In this study (Project K5/2471/3), no statistical association between DO and *M. parvicella* selection was demonstrated. However, measurements were only taken once per day, and it is possible that diurnal variation in DO occurred.

Low AS temperature (<10°C) has also been reported as an important selection criterion for *M. parvicella*. However, the result of this study showed that this was not relevant in GAU, DBN or CPT, because even in winter, the lowest temperatures measured were ~15°C (Figure 18). In fact, the prevalence of *M. parvicella* as a dominant filament increased in winter (Figure 7). In addition, no statistical link between temperature and *M. parvicella* was demonstrated (Appendix 3).

In pure culture studies, *M. parvicella* shows a preference for long-chain fatty acids as carbon sources, being incapable of utilising sugars. The filament stores lipids in intracellular granules, analogous to PAOs with phosphates. They have therefore been called "lipid accumulating organisms" (LAOs) (Rosetti et al., 2005). Unfortunately, due to the fact that the method for determining the concentration of fats, oils and grease (FOG) is expensive method, and not widely performed, it was not included the suite of parameters for this study or in other studies. It is recommended that if bulking by *M. parvicella* is definitively established, that the

FOG, and/or other methods to determine the concentration of long-chain fatty acids is conducted in future studies (Chapter 6). Pre-flotation to remove fats could be a useful non-chemical method to control bulking by *M. parvicella*.

5.4.3 Eikelboom type 021N and *Thiothrix* spp.

The overall prevalence of *Thiothrix* spp. (which includes Eikelboom type 021N) was 34%, but was 75% in samples with elevated DSVI and FI (Table 13). In one case, the *Thiothrix* spp. was the only dominant filament. Type 021N was also previously found to cause long-term bulking in a CPT WWTP (Welz et al., 2014).

As with Type 0092 and *M. parvicella*, the GLM was a good fit for Type 021N/*Thiothrix* spp. (deviance = 0.720) and significant ($p < 0.05$) (Appendix 8). Seven significant co-variables were identified (influent BOD:COD ratio, NH_3/NH_4^+-N, and VFAs, AS TP, and NO_3^-/NO_2-N, and F/M (BOD and COD). The test of model effects showed that the model was valid ($p > 0.000$). The odds ratios (ExpB) was particularly high (7.839) for F/M (COD), indicating a large positive effect on the outcome (Appendix 8 – parameter estimates), i.e. for every unit increase in the F/M (COD) ratio, the odds of Type 021N/*Thiothrix* spp. being dominant is multiplied by 7.839. The odds ratios for the other parameters ranged from 0.003 [F/M (BOD)] to 1.007 (VFAs)].

Type 021N and *Thiothrix* spp. were not highly prevalent in the study conducted by dos Santos et al. (2015). When the methodology described by these authors was applied to the data obtained during Project K5/2471/3, it was found that the influent BOD and AS COD and BOD:COD showed significant correlation (Kendall) with these filaments (Table 16).

Deepnarain et al. (2015) found that at Kingsburgh WWTP, influent COD and F/M ratio were significant selective criteria for *Thiothrix* spp. and 021N, and that NH_4^+ was also selective for 021N. These findings substantiated the interpretation of the results obtained by the GLM that the organic fraction is a highly important parameter for selection of Type 021N and *Thiothrix* spp.

However, contradictory results were obtained when using the data generated during Project K5/2471/3 using the 'Deepnarain' model (Table 15), where only pH was seen as significant parameter for selection of Type 021N and *Thiothrix* spp., bringing the fitness of the model (Deepnarain et al. 2015) based on cut-off values into question.

As with most filaments, very few pure culture studies have been conducted (Nielsen et al., 2000). Results of these indicated that most Type 021N/*Thiothrix* spp. can utilise short chain fatty acids, and some, but not all, can utilise sugars as a source of carbon. The positive ExpB value for VFAs obtained in this study supports this. These filaments are metabolically versatile, being able to utilise heterotrophic, mixotrophic or chemolithotrophic metabolic pathways, although when utilising the latter, growth rates are slow. Under anaerobic conditions, acetate is taken up and intracellular sulphur globules are formed from

thiosulphate. According to Nielsen et al. (2009), the potential for mixotrophic growth, and strong stimulation by thiosulphate for acetate and bicarbonate uptake, are the "key properties" that can offer *Thiothrix* spp. a competitive advantage.

It was concluded that the association of *Thiothrix* spp. with F/M (COD) ratio, F/M (BOD), BOD:COD ratio and influent VFA concentration are noteworthy and novel, and show that both the quantity and quality of the carbon source are important for filament selection. Detailed chemical analysis of the organic content of the influent is recommended for future studies. For example, using high performance liquid chromatography to characterise the carbon substrates in WWTPs where bulking by Type 021N/*Thiothrix* spp. occurs.

5.4.4 Eikelboom type 0041 and Gram positive branching bacilli

Although the model gave some significant results for Eikelboom types 0041 and 1851, and GPBB, it was judged to be invalid due to lack of data: for Type 0041, 71% of the samples were excluded from the model, and of the 88 remaining samples, only 9 contained Type 0041 (Appendix 9). Similarly, for GPBB (Appendix 10) and Type 1851 (Appendix 11), 67% and 59% of samples were excluded from the models, and of the remaining samples, only 7 and 13 were GPBB, and Type 1851, respectively. The decision to judge the model as invalid for these filaments was supported by the deviance values for goodness of fit (0 for Type 0041 and GPBB, and 0.197 for Type 1851).

CHAPER 6: CONCLUSIONS, RECOMMENDATIONS AND FUTURE WORK

In terms of prevalence, Type 0092 was found to dominate in almost 70% of the study WWTPs, followed by *M. parvicella* (40%) and Type 021N/*Thiothrix* spp (34%). Three additional filaments dominated in > 6% of WWTPs, but the GLM was not deemed valid due to insufficient data. The most important findings and recommendations for future work (bold) were identified:

- According to the GLM, the only selective criterion for Eikelboom type 0092 was pH, with dominance increasing at higher pH values. This substantiates the hypothesis that Type 0092 does not have exacting nutritional growth requirements in WWTPs. Although this filament was the most prevalent of the dominant species, results strongly suggest that it does not commonly cause bulking in SA. This is supported by the fact that it grows preferentially in the flocs (forming the 'backbone' of the flocs) as opposed to the bulk liquid.

  In future studies, unless: (i) Type 0092 is the only dominant filament, (ii) the presence of Type 0092 is accompanied by a high DI and FI, (iii) most Type 0092 filaments are present outside of the flocs, and (iv) there is evidence of bulking *in-situ*, it should be assumed that its presence its incidental in SA WWTPs.

- *M. parvicella* appeared to be filament most commonly associated with bulking, followed by *Type 021N/Thiothrix spp*. The model identified the influent SO_4^{-2}-S concentration, and the AS σ-P as the most important selective criteria for *M. parvicella*, with a combination of low concentrations of both appearing to select for robust filament growth. Unlike reports from elsewhere in the world, temperature was not correlated with selection of this filament. This is almost certainly because temperatures <10°C were not experienced in any of the WWTPs. The F/M ratio (COD) was the most important selective parameter for Type 021N/*Thiothrix* spp., with the likelihood of dominance increasing > 7-fold with each unit increase. In addition, the co-variable results strongly suggested both the quality and quantity of the organic fraction is important.

  If bulking by *M. parvicella* and/or Type 021N/*Thiothrix* spp. is identified at a particular WWTP, amongst other parameters, full characterisation of the organic fraction of the influent to WWTPs should be performed. If bulking is chronic, it should be ascertained whether abundance increases or decreases with concentrations of FOG (or other organics). If chronic bulking by *M. parvicella* is identified, the addition of an upstream floatation system to reduce the fat content may be considered.

- It was difficult to ascertain from laboratory-based studies whether filamentous bulking was present or not at WWTPs because: (i) the FI and DSVI only showed a weak correlation, (ii) ex-situ testing for DSVI was brought into question, (iii) as expected, the floc character also influenced settling, and (iv) other causes of bulking, such as rising sludge due to nitrification were not accounted for.

  It is recommended that plant operators are vigilant and are trained to detect bulking, after which they follow a protocol such as outlined in Figure 36, to properly diagnose bulking, and if possible, apply corrective measures.

- This work has shown that even with a large survey such as this, it is very difficult to collect sufficient data to validate a model to predict the selection criteria for all filamentous bacteria. It is also extremely costly as many parameters need to be included. Even so, important parameters may be overlooked (for example, FOG for the selection of *M. parvicella)*.

  It is recommended that a follow-up study is performed, building on the results of this work, and ongoing work being conducted by students involved in this project. Now that data is available for all the WWTPs (bulking/no bulking), it is recommended that selected WWTPs are observed and monitored. An (unrefined) protocol is provided in Figure 36 by way of example. By focusing only on the WWTPs where bulking is taking place, costs would be minimal. Such a project would potentially be able to definitively identify the major causes of bulking in SA WWTPs and hopefully, increase our knowledge of remedial action.

  It is further recommended that the initial focus should be on the 'top' three prevalent filaments, for which the GLM appeared to be valid, namely Type 0092 (to prove unequivocally whether it causes bulking or not), and *M. parvicella* and Type 021N/*Thiothrix* spp.

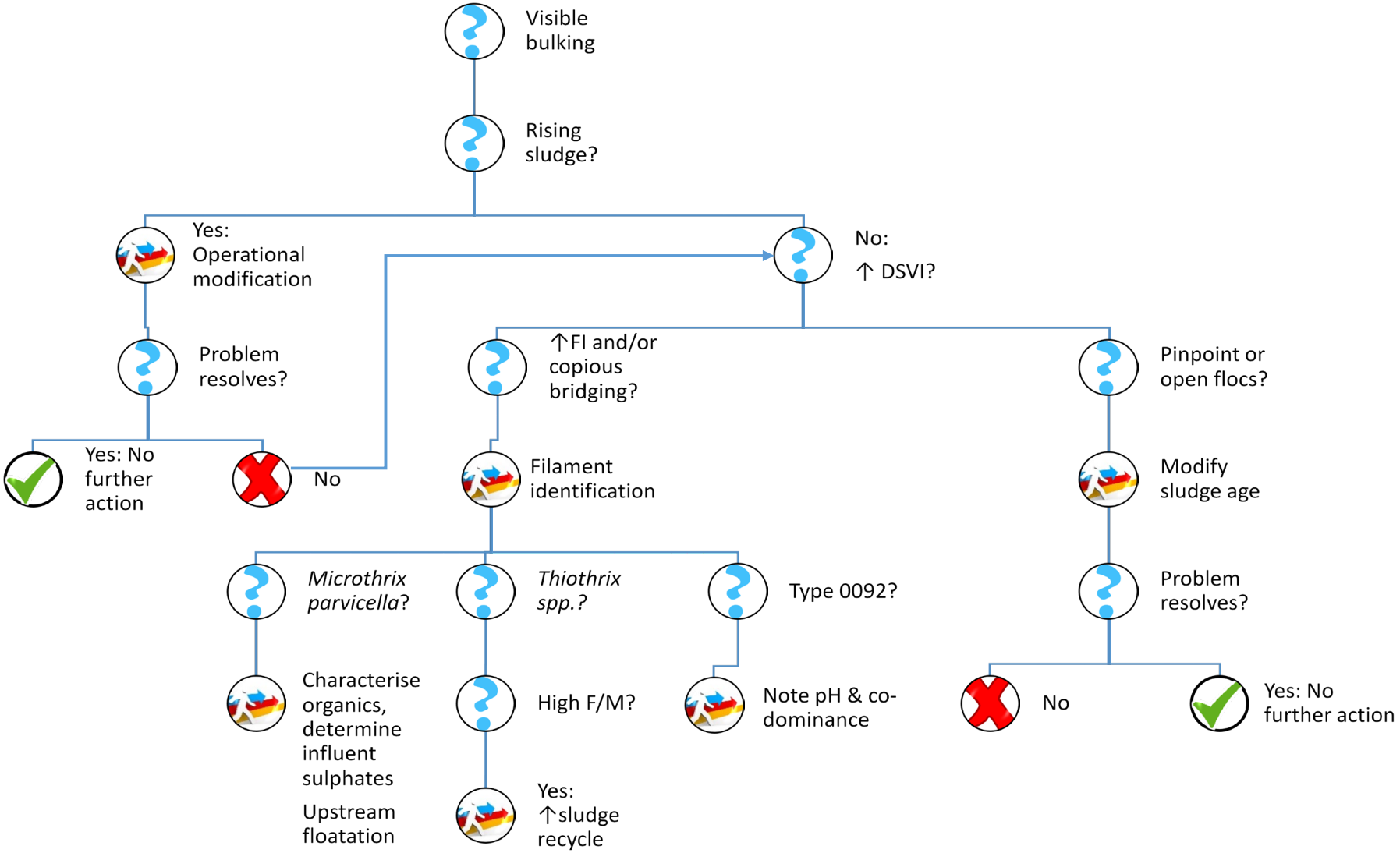


Figure 35: Example (unrefined) of action to diagnose & remedy bulking in South African wastewater treatment plants

References

Amann, R.I., Binder, B.J., Olson, R.J., Chisholm, S.W., Devereux, R., Stahl, D.A. (1990a) Combination for 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. Applied and Environmental Microbiology, 56: 1919-1925.

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

American Public Health Association (APHA), American Water Works Association (AWWA) & Water Environment Federation (WEF). 2005. Standard Methods for the Examination of Water and Wastewater, 21st Edition, Washington D.C.

Aruga, S., Kamagata, Y., Kohno, T., Hanada, S., Nakamura, K., Kanagawas, T. (2002) Characterization of filamentous Eikelboom type 021N bacteria and description of *Thiothrix disciformis* sp. nov. and *Thiothrix flexilis* sp. International Journal of Systematic Evolutionary Microbiology, 52: 1309-1316.

Beer, M., Seviour, E.M., Kong, Y., Cunningham, M., Blackall, L.L., Seviour, R.J. (2002) Phylogeny of the filamentous bacterium Eikelboom Type 1851, and design and application of a 16S targeted oligonucleotide probe for its fluorescence in situ identification in activated sludge. FEMS Microbiology Letters, 207:179-183.

Bjornsson, L., Hugenholtx, P., Tyson, G.W., Blackall, L.L. (2002) Filamentous *Chloroflexi* (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. Microbiology, 148: 2309-2318.

Blackbeard, J.R., Gabb, D.M.D., Ekama, G.A., Marais, G.R.V. (1986) Identification of filamentous organisms in nutrient removal activated sludge plants in South Africa. Water SA, 14: 29-33.

Breiman, L. (2001). "Decision-tree forests." Machine Learning, 45: 5-32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984) Classification and regression trees. Pacific Grove, CA: Wadsworth Inc.

Caravelli, A., Giannuzzi, L. and Zaritzky, N. (2004) Effect of chlorine on filamentous microorganisms present in activated sludge as evaluated by respirometry and INT-Dehydrogenase activity. Water Research, 38: 2394-404.

Casey, T.G., Wentzel, M.C. and Ekama, G.A. (1999) Filamentous Organism Bulking in Nutrient Removal Activated Sludge Systems: Paper 9, 10, 11. Water SA, 25: 409-451.

Casey, T.G., Wentzel, M.C., Loewenthal, R.E., Ekama, G.A., Marais, G.v.R. (1992) A hypothesis for the cause of low F/M filament bulking in nutrient removal activated sludge systems. Water Research, 26:867-769.

Chernousova, E.Y., Belousova, E.V., Gavrish, E.Y., Dubinina, G.A., Tourova, T.P, Grabovich, M.Y. (2012) Molecular phylogeny and taxonomy of colorless, filamentous sulfur bacteria of the genus *Thiothrix*. Microbiology, 81: 332-341.

Chua, H., Yu, P.F.H., Sin, S.N., Tan, K.N. (2000) Effect of Food:Microorganism ratio in activated sludge foam control. Applied Biochemistry and Biotechnology, 84: 1127-1135.

Contreras, E.M., Bertola, N.C., Giannuzzi, L., Zaritzky, N.F. (2002) A modified method to determine biomass concentration as COD concentrations in pure cultures and an activated sludge system. Water SA, 28: 463-467.

Da Motta, M., Pons, M. N. and Roche, N. (2003) Monitoring filamentous bulking in activated sludge systems fed by synthetic or municipal wastewater. Bioprocess Biosystems Engineering, 25: 387-93.

Deepnarain, N., Kumari, S., Ramjith J., Swalaha, F.M., Tandoi, V., Pillay, K., Bux, F. (2015) A logistic model for the remediation of filamentous bulking in a biological nutrient removal wastewater treatment plant. Water Science and Technology, 181: 391-405.

Deurinck, J., Smets I.Y., Van Impe, J.F. (2006). Modelling of filamentous bulking in activated sludge: a literature review.

Dhall, P., Siddiq, T.O., Ahmad, A., Kumar, R., Kumar, A. (2012). Restructuring BOD:COD ratio of dairy milk industrial wastewaters in BOD analysis by formulating a specific microbial seed. The Scientific World Journal Volume 2012, Article ID 105712, 7 pages doi:10.1100/2012/105712.

Dos Santos, L.A., Ferreira, V., Neto, M.M., Pereira, M.A., Mota, M., Nicolau, A. (2015). Study of 16 Portuguese activated sludge systems based on filamentous bacteria populations and their relationships with environmental parameters. Environmental Biotechnology, 99: 5307-5316.

Eikelboom, D.H. (2000). Process Control of Activated Sludge Plants by Microscopic Investigation. IWA Publishing, London.

Ekama, G.A., Marais, G.V.R., Siebritz, I.P., Pitman, A.R., Keay, G.R.P., Gerber, A., Smollen, M. (1984) Theory, Design and Operation of Nutrient Removal Activated Sludge Processes. WRC Commission report no TT 16/84.

Eschenhagen, M., Schuppler, M and Roske, I. (2003) Molecular characterization of the microbial community structure in two activated sludge systems for the advanced treatment of domestic effluents. Water Research, 37: 3224-3232.

Foladori, P., Bruni, L., Tamburini, S., Ziglio, G. (2010) Direct quantification of bacterial biomass in influent, effluent and activated sludge of wastewater treatment plants by flow cytometry. Water Research 44: 3807-3818.

Garson, G., and Davids. Generalized Linear Models & Generalized Estimating Equations 2013 (Statistical Associates Blue Book Series 26) (Kindle Locations 730-731). Statistical Associates Publishers. Kindle Edition.

Ge, S., Peng, Y., Wang, S., Guo, J., Ma, B., Zhang, L., Cao, X. (2010) Enhanced nutrient removal in a modified step feed process treating municipal wastewater with different inflow distribution ratios and nutrient ratios. Bioresource Technology, 101: 9012-9019.

Gerardi, M. H. (2006). Bacterial groups. In: Gerardi, M. H. (ed.) Wastewater microbiology. Canada: John Wiley and Sons.

Gulez, G. and De lost Reyes III F.L. (2009) Multiple approaches to assess filamentous bacterial growth in activated sludge under different carbon sources. Journal of Applied Microbiology, 106: 682-691

Henze, M., Grady Jr C.P.L., Gujer, W., Marais, G.v.R., Matsuo, T. (1987) Activated Sludge Model No. 1. Technical report, IAWPRC Task Group on Mathematical Modelling for Design and Operation of Biological Wastewater Treatment.

Hug, T., Gujer, W. and Siegrist, H. (2005) Rapid quantification of bacteria in activated sludge using fluorescence in situ hybridization and epifluorescence microscopy. Water Research. 39: 3837-3848.

Hugenholtz, P., Tyson, G.W., Webb, R.I., Wagner, A.M., Blackall, L.L. (2001) Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. Applied and Environmental Microbiology, 67: 411-419.

Jenkins, D. Richard, M.G. Diagger, G.T. (2004). Manual on the Causes and Control of Activated Sludge Bulking, Foaming and Other Solids Separation Problems. CRC Press: Florida and IWA Publishing: London.

Jouanneau, S., Recoules, L., Durand, M.J., Boukabache, A., Picot, V., Primault, Y., Lakel, A., Sengelin, M., Barillon, B., Thouand, G. (2014) Methods for assessing biological oxygen demand (BOD): a review. Water Research, 49: 67-82.

Kanagawa, T., Kamagate, Y., Aruga, S., Kohno, T., Horn, M., Wagner, M. (2000) Phylogenetic analysis of and oligaonucleotide probe development for Eikelboom Type 021N filamentous bacteris isolated form bulking activated sludge. Applied and Environmental Microbiology, 66: 5043-5053.

Koch, G., Piantra, R., Krebs, P., Siegrist, H. (1999) Potential denitrification and solids removal in the rectangular clarifier. Water Research, 33: 309-318.

Kumar, K., Singh, G.K., Dasatidar, M.G., Sreekrishnan, T.R. (2014). Effect of mixed liquor volatile suspended solids (MLVSS) and hydraulic retention time (HRT) on the performance of activated sludge process during the biotreatment of real textile wastewater. Water Resources and Industry, 5: 1-8.

Lakay, M.T. Ketley, D., Warburton, C., de Villiers, M., Casey, T.G., Wentzel, M.C., and Ekama, G.A. (1999). Filamentous organism bulking in nutrient removal activated sludge systems: Paper 7 – Exploratory experimental investigations. Water SA, 25: 283-396.7

Lacko, N., Bux, F., Kasan, H.C. (1999) Survey of filamentous bacteria in activated sludge plants in KwaZulu-Natal. Water SA, 25: 63-68.

Leeuwen, J. (1988) Bulking control with ozonation in a nutrient removal activated sludge system. Water SA, 14: 119-124.

Li, B., Irvin, S. (2007) The comparison of alkalinity and ORP as indicators for nitrification and denitrification in a sequencing batch reactor. Biochemical Engineering Journal, 34: 248-255.

Liao, J., Lou, I and de los Reyes III F.L. (2004) Relationship of species-specific filament levels to filamentous bulking in activated sludge. Applied and Environmental Microbiology, 70: 2420-2428.

Liu, R.L. and Seviour, R.J. (2001) Design and application of oligonucleotide probes for fluorescent in situ identification of the filamentous bacterial monotype *Nostocoida limicola* in activated sludge. Environmental Microbiology, 3: 551-560.

Lou, I., Zhao, Y. (2012) Sludge bulking prediction using principle component regression and artificial neural network. Mathematical Problems in Engineering, 1: 1-17.

Manz, W., Amann, R., Ludwig, W., Wagner, M., Schleifer, K.H. (1992) Phylogenetic oligodeoxynecleotides probes for the major subclasses of proteobacteria: problems and solutions. Systematic and Applied Microbiology, 15: 593-600.

Martins, A.M.P., Pagillal, K. Heijnen, J.J., van Loosdrecht, M.C.M. (2004) Filamentous bulking sludge – a critical review. Water Research, 38: 793-817.

McCullagh, P. and Nelder, J. A. (1989) *Generalized linear models.* 2nd ed. London: Chapman & Hall/CRC.

Mielczarek, A.T., Kragelund, C., Eriksen, P.S., Nielsen, P.H. (2012) Population dynamics of filamentous bacteria in Dutch wastewater treatment plant with nutrient removal. Water Research, 46: 3781-3795.

Milobędska, A., Witeska, A., Muszyęski, A. (2016) factors affecting population of filamentous bacteria in wastewater treatment plants with nutrient removal. Water Science and Technology, 73: 790-797.

Monti, A., Hall, E.R., Dawson, R.N., Husain, H., Kelly, H.G. (2006) Comparative study of biological nutrient removal (BNR) processes with sedimentation and membrane-based separation. Biotechnology and Bioengineering, 94: 740-752.

Morgan-Sagastume, F., Larsen, P., Nielsen, J.L. and Nielsen, H.N. (2008) Characterization of the loosely attached fraction of activated sludge bacteria. Water Research, 42: 843-854.

Nielsen, P.H., Kragelund, C., Seviour, R.J., Nielsen, J.L. (2009) I) Identity and ecophysiology of filamentous bacteria in activated sludge. FEMS Microbiology Reviews, 33: 969-998.

Nielson, P. H., Lemmer, H. and Daims, H. (2009) FISH handbook for biological wastewater treatment. UK: IWA Publishing.

Nielsen, P.H., Roslev, P., Dueholm, T.E., Nielsen, J.L. (2002) *Microthrix parvicella*, a specialized lipid consumer in anaerobic aerobic activated sludge plants. Water Science and Technology, 46: 73-80.

Nielsen, P.H., de Muro, M.A., Nielsen (2000) Studies on the physiology of *Thiothrix* spp. present in activated sludge. Environmental Microbiology, 24, 389-398.

Noutsopolis, C., Mamais, D., Andreadakis, A. (2006) Effect of solids retention time on *Microthrix parvicella* growth. Water SA, 32: 315-321.

Ramond J-B., Welz P.J., Tuffin, M., Burton S.G., Cowan D.A. (2013) Selection of diazotrophic bacterial communities in biological sand filter mesocosms used for the treatment of phenolic-laden wastewater. Microbial Ecology, 66: 563-570.

Rasool, K., Ahn, D.H., Lee, D.S. (2014) Simultaneous organic carbon and nitrogen removal in an anoxic-oxic activated sludge system under various operating conditions. Bioresource Technology, 162: 373-378.

Report No. TT 16/84. Water Research Commission, Pretoria.

Richard, M.G., Jenkins, D., hao, O., Shimzu, G. (1982) The isolation and characterization of filamentous micro-organisms from activated sludge bulking. Report 81-82. Sanitary Engineering and Environmental Health Research Laboratory, University of California, Berkeley.

Roels, T., Dauwe, F., van Damme, S., de Wilde, K., Roelandt, E. (2002) The influence of PAX-14 on activated sludge systems and in particular Microthrix parvicella. Water Science and Technology, 46: 487-490.

Rosetti, S., Tomei, M.L., Nielsen, M., Tandoi, V. (2005) *Microthrix parvicella*, a filamentous bacterium causing bulking and foaming in activated sludge: a review of current knowledge. FEMS Microbiology Research, 29: 49-64.

Rosetti, S., Carucci, A., and Rolle, E. (1994) Survey on the occurrence of filamentous organisms in municipal wastewater treatment plants related to their operating conditions. Water Science and Technology, 29: 305-308.

RÖssle, W.H., Pretorius, W.A. (2008) Batch and automated SVI measurements based on short term temperature variations. Water SA, 34: 237-243.

Saayman, G. B., Schutte, C. F. and van Leeuwen, J. (1999). Chemical control of filamentous sludge bulking in a full-scale biological nutrient removal activated sludge plant. Ozone: Science and Engineering, 20: 1-15.

Salvado, H.S., Garcia, M.P. and Amigó J.M. (1995) capability of ciliated protozoa as indicators of effluent quality in activated sludge plants. Water Research, 29: 1041-1050.

Sekiguchi, H., Tomioka, N., Nakahara, T and Uchiyama, H. (2001) Biochemistry Letters, 23: 1205-1208.

Seviour, E.M., Williams, C., DeGrey, B., Soddell, J.A., Seviour, R.J., and Lindrea, K.C. (1994) Studies on filamentous bacteria from Australian sludge plants. Water Science and Technology, 28: 235-2342.

Simonoff, J. S. (2003) *Analyzing categorical data.* New York: Springer.

Singh, K.K., Pal, M., Ojha, C.S.P. Singh, V.P. (2008) Estimation of removal efficiencies for settling biomass using neural networks and support vector machines. Journal of Hydraulic Engineering, 13: 146.

Spiers, L.B.M., Tucci, J., Seviour, R.J. (2015) The activated sludge bulking filament Eikelboom morphotype 0803 embraces more than one member of the Chloroflexi. Microbial Ecology, 91: 1-9.

Spiers, L., Nittamie, T., McIlroy, S., Schroeder, S., Seviour, R.J. (2009) Filamentous bacterium Eikelboom Type 0092 in activated sludge plants in Australia is a member of the Chloroflexi. Applied and Environmental Microbiology 75: 2446-2452.

Strom, P.F., Jenkins, D. (1984) Identification and significance of filamentous microorganisms in activated sludge. Journal of Water Pollution Control, 56: 449-459.

Tchobanoglous, G., and Burton, F.L. (1991) Wastewater engineering: treatment, disposal and reuse. 6th ed. McGraw-Hill: New York.

Tsai, M-W., Wentzel, M.C., Ekama, G.A. (2003) The effect of residual ammonia concentration under aerobic conditions on the growth of *Microthrix parvicella* in biological nutrient removal plants. Water Research, 37: 3009-3015.

Veissman, W. jr. and Hammer, M.J. (1998) Water supply and pollution control. 6th ed. Addison Wesley Longman, inc., Menlo Park.

Wagner, M., Amann, R., Kaempeer, P., Assmus, B., Hartmann, A., Hutzler, P., Springer, N., Schleifer, K-H. (1994) Identification and in situ detection of Gram negative filamentous bacteria in activated sludge. Systematic and Applied Microbiology, 17: 405-417.

Wanner, J. (1994) Activated sludge bulking and foaming control. Technomic Publishing Company.

Welz, P.J., Ramond, J.B, Cowan, D.A., Burton, S.G. (2012) Phenolic removal processes in biological sand filters, sand columns and microcosms. Bioresource Technology, 119: 262-269.

Welz, P.J., Esterhuysen, A., Vulindlu, M., Bezuidenhout, C. (2014a) Microbial characterization and dominance of Eikelboom type 0092 in activated sludge from wastewater treatment facilities in Cape Town, South Africa. Water SA, 40: 649-656.

Welz, P.J., Palmer, Z., Isaacs, S., Kirby, B., Le Roes-Hill, M. (2014b) Analysis of substrate degradation, metabolite formation and microbial community responses in sand bioreactors treating winery wastewater: a comparative study. Journal of Environmental Management, 145: 147-156.

Welz, P.J., Le Roes-Hill, M. (2014c) Biodegradation of organics and accumulation of metabolites in experimental biological sand filters used for the treatment of synthetic winery wastewater: a mesocosm study. Journal of Water Process Engineering, 3C: 155-163.

Yamamoto-Ikemoto, R., Matsui, S., Komoir, T. (1994) Ecological interactions among denitrification, poly-P accumulation, sulphate reduction, and filamentous sulphur bacterial in activated sludge. Water Science and Technology, 30: 201-210.

Yoon, D-N., parks, S-J., Kim, S-J., Jeon, C.O., Chae, J-C., Rhee, S-K. (2010) Isolation, characterization and abundance of filamentous members of *Caldilineae* in activated sludge. Journal of Microbiology, 48: 275-283.

APPENDICES

Appendix 1: example of criteria used during WWTP selection

Carl Grundlingh WWTP

WWTW configuration:
1 x BNR module
BNR module: Pasveer ditch

Configuration modifications: None

Raw influent: 95% domestic sewage + 5% industrial wastewater
Type of aeration: surface
Number of reactors: 1
Reactor capacity: (4 800m^3)

Notes on sampling points:

- Currently no samples collected at PST
- Currently no samples collected at aerobic zone, only at weir overflow
- Currently no samples collected at SST (clarifier) before disinfection, only after disinfection

Notes on laboratory analyses:

- Only MLSS performed
- Only SVI performed
- Only acetic acid performed

Satellite image of Carl Grundlingh WCW (GPS co-ordinates: 26°23.053' S 28°28.179' E)

Plant : Carl Grundlingh

Physical Address: Vorsterkroon, Nigel

The Carl Grundlingh works is situated north of Nigel and falls within the DD5 drainage district. It was built in 1977 and upgraded in 1984. It is designed to treat 2,5 megalitres per day of industrial effluent and raw sewage from the Dunnottar, Sharon Park and Marievale areas. Activated sludge is employed as the main treatment process.

A central inlet works, incorporating the processes of screening and grit removal serves the works. The BNR activated sludge process includes a three stage reactor basin and final clarification. The final effluent is chlorinated before discharging it into the Nigel Dam. The outflow of the dam flows into the Blesbokspruit.

Capacity : 2.00 Mℓ/d

Appendix 2: Statistical analyses to determine the significance of configuration on filament selection

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
FirstFilamentsCode Most DOMINANT FILAMENTS WITH FI≥2 * ConfigurationCodeNBNBNB Configuration Code	264	100.0%	0	0.0%	262.500	100.0%

Most DOMINANT FILAMENTS WITH FI≥2 * Configuration Code Crosstabulation

			Configuration Code				
			1 3 stage Bard.	2 5 stage Bard.	3 MLE	4 UCT	Total
Most DOMINANT FILAMENTS WITH FI≥2	1 TYPE 0092	Count	44	22	19	14	99
		% within Configuration Code	38.9%	33.8%	30.2%	60.9%	37.5%
	2 Type 021N & Thiothrix spp.	Count	21	18	10	1	50
		% within Configuration Code	18.6%	27.7%	15.9%	4.3%	18.9%
	3 M. parvicella	Count	11	12	6	6	35
		% within Configuration Code	9.7%	18.5%	9.5%	26.1%	13.3%
	4 TYPE 0041	Count	14	10	7	2	33
		% within Configuration Code	12.4%	15.4%	11.1%	8.7%	12.5%
	5 Type 1851	Count	9	2	9	0	20
		% within Configuration Code	8.0%	3.1%	14.3%	0.0%	7.6%
	6 GBPP	Count	8	1	7	0	16
		% within Configuration Code	7.1%	1.5%	11.1%	0.0%	6.1%
	7 TYPE 0803	Count	6	0	5	0	11
		% within Configuration Code	5.3%	0.0%	7.9%	0.0%	4.2%
Total		Count	113	65	63	23	264
		% within Configuration Code	100.0%	100.0%	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	36.633[a]	18	.006
Likelihood Ratio	42.896	18	.001
Linear-by-Linear Association	.078	1	.779
N of Valid Cases	264		

a. 13 cells (46.4%) have expected count less than 5. The minimum expected count is .96.

Crosstabs

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Most DOMINANT FILAMENTS WITH FI≥2 * Configuration Code	264	100.0%	0	0.0%	262.500	100.0%

Most DOMINANT FILAMENTS WITH FI≥2 * Configuration Code Crosstabulation

			Configuration Code				
			1 3 stage Bard.	2 5 stage Bard.	3 MLE	4 UCT	Total
Most DOMINANT FILAMENTS WITH FI≥2	1 TYPE 0092	Count	44	22	19	14	99
		% within Most DOMINANT FILAMENTS WITH FI≥2	44.4%	22.2%	19.2%	14.1%	100.0%
	2 Type 021N & Thiothrix spp.	Count	21	18	10	1	50
		% within Most DOMINANT FILAMENTS WITH FI≥2	42.0%	36.0%	20.0%	2.0%	100.0%
	3 M. parvicella	Count	11	12	6	6	35
		% within Most DOMINANT FILAMENTS WITH FI≥2	31.4%	34.3%	17.1%	17.1%	100.0%
	4 TYPE 0041	Count	14	10	7	2	33
		% within Most DOMINANT FILAMENTS WITH FI≥2	42.4%	30.3%	21.2%	6.1%	100.0%
	5 Type 1851	Count	9	2	9	0	20
		% within Most DOMINANT FILAMENTS WITH FI≥2	45.0%	10.0%	45.0%	0.0%	100.0%
	6 GBPP	Count	8	1	7	0	16
		% within Most DOMINANT FILAMENTS WITH FI≥2	50.0%	6.3%	43.8%	0.0%	100.0%
	7 TYPE 0803	Count	6	0	5	0	11
		% within Most DOMINANT FILAMENTS WITH FI≥2	54.5%	0.0%	45.5%	0.0%	100.0%
Total		Count	113	65	63	23	264
		% within Most DOMINANT FILAMENTS WITH FI≥2	42.8%	24.6%	23.9%	8.7%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	36.633[a]	18	.006
Likelihood Ratio	42.896	18	.001
Linear-by-Linear Association	.078	1	.779
N of Valid Cases	264		

a. 13 cells (46.4%) have expected count less than 5. The minimum expected count is .96.

Appendix 3: Statistical analysis of effect of temperature on filament selection

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_1 1 TYPE 0092	304	.35	.477	0	1

Mann-Whitney Test

Ranks

	Fil_1 1 TYPE 0092	N	Mean Rank	Sum of Ranks
Temperature	0	134	99.37	13315.50
	1	66	102.80	6784.50
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	4270.500
Wilcoxon W	13315.500
Z	-.394
p-value (2-tailed)	.694

a. Grouping Variable: Fil_1 1 TYPE 0092

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_2 2 Type 021N & Thiothrix spp.	304	.20	.399	0	1

Mann-Whitney Test

Ranks

	Fil_2 2 Type 021N & Thiothrix spp.	N	Mean Rank	Sum of Ranks
Temperature	0	163	102.45	16699.00
	1	37	91.92	3401.00
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	2698.000
Wilcoxon W	3401.000
Z	-.999
p-value (2-tailed)	.318

a. Grouping Variable: Fil_2 2 Type 021N & Thiothrix spp.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_3 3 M. parvicella	304	.14	.352	0	1

Mann-Whitney Test

Ranks

	Fil_3 3 M. parvicella	N	Mean Rank	Sum of Ranks
Temperature	0	172	100.39	17267.00
	1	28	101.18	2833.00
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	2389.000
Wilcoxon W	17267.000
Z	-.067
p-value (2-tailed)	.947

a. Grouping Variable: Fil_3 3 M. parvicella

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_4 4 TYPE 0041	304	.13	.335	0	1

Mann-Whitney Test

Ranks

	Fil_4 4 TYPE 0041	N	Mean Rank	Sum of Ranks
Temperature	0	176	102.03	17957.50
	1	24	89.27	2142.50
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	1842.500
Wilcoxon W	2142.500
Z	-1.013
p-value (2-tailed)	.311

a. Grouping Variable: Fil_4 4 TYPE 0041

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_5 5 Type 1851	304	.08	.270	0	1

Mann-Whitney Test

Ranks

	Fil_5 5 Type 1851	N	Mean Rank	Sum of Ranks
Temperature	0	183	99.75	18254.00
	1	17	108.59	1846.00
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	1418.000
Wilcoxon W	18254.000
Z	-.602
p-value (2-tailed)	.547

a. Grouping Variable: Fil_5 5 Type 1851

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_6 6 GBPP	304	.07	.248	0	1

Mann-Whitney Test

Ranks

	Fil_6 6 GBPP	N	Mean Rank	Sum of Ranks
Temperature	0	183	99.76	18256.50
	1	17	108.44	1843.50
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	1420.500
Wilcoxon W	18256.500
Z	-.592
p-value (2-tailed)	.554

a. Grouping Variable: Fil_6 6 GBPP

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Temperature	200	19.66200	2.855543	13.300	26.600
Fil_7 7 TYPE 0803	304	.04	.187	0	1

Mann-Whitney Test

Ranks

	Fil_7 7 TYPE 0803	N	Mean Rank	Sum of Ranks
Temperature	0	189	99.74	18850.50
	1	11	113.59	1249.50
	Total	200		

Test Statistics[a]

	Temperature
Mann-Whitney U	895.500
Wilcoxon W	18850.500
Z	-.772
p-value (2-tailed)	.440

a. Grouping Variable: Fil_7 7 TYPE 0803

Appendix 4: Statistical analyses to determine the relationship of floc characteristics on the diluted sludge volume index

FLOC SHAPE:
T-Test

Group Statistics

	FlocShape Floc Shape	N	Mean	Std. Deviation	Std. Error Mean
DSVI_NBNBNB AS - Dissolved sludge volume index	1 Round	150	164.34667	93.640179	7.645689
	2 Irregular	144	125.40278	69.824287	5.818691

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DSVI_NBNBNB AS - Dissolved sludge volume index	Equal variances assumed	.789	.375	4.030	292	.000	38.943889	9.664307	19.923360	57.964417
	Equal variances not assumed			4.053	275.340	.000	38.943889	9.608003	20.029411	57.858367

FLOC DENSITY:
T-Test

Group Statistics

	FlocDensity Floc Density	N	Mean	Std. Deviation	Std. Error Mean
DSVI_NBNBNB AS - Dissolved sludge volume index	1 Compact	48	114.91667	49.065667	7.082019
	2 Open	246	151.19512	89.193819	5.686790

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DSVI_NBNBNB AS - Dissolved sludge volume index	Equal variances assumed	6.898	.009	-2.736	292	.007	-36.278455	13.260668	-62.377059	-10.179852
	Equal variances not assumed			-3.994	117.759	.000	-36.278455	9.082653	-54.264961	-18.291949

FLOC STRENGTH:
T-Test

Group Statistics

	FlocStrength Floc Strength	N	Mean	Std. Deviation	Std. Error Mean
DSVI_NBNBNB AS - Dissolved sludge volume index	1 Weak	29	136.75862	68.338786	12.690194
	2 Strong	265	146.20377	86.646683	5.322663

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DSVI_NBNBNB AS - Dissolved sludge volume index	Equal variances assumed	.511	.475	-.568	292	.571	-9.445153	16.637493	-42.189759	23.299453
	Equal variances not assumed			-.686	38.592	.497	-9.445153	13.761242	-37.289313	18.399008

FLOC SIZE:
Oneway

Descriptives

AS - Dissolved sludge volume index

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Small	34	157.32353	58.711271	10.068900	136.83820	177.80886	70.000	290.000
Medium	130	148.30000	82.344017	7.222048	134.01100	162.58900	4.000	500.000
Large	130	139.09231	92.994161	8.156127	122.95521	155.22941	2.000	500.000
Total	294	145.27211	84.963679	4.955182	135.51985	155.02437	2.000	500.000

Test of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
AS - Dissolved sludge volume index	Based on Mean	1.833	2	291	.162
	Based on Median	1.758	2	291	.174
	Based on Median and with adjusted df	1.758	2	274.635	.174
	Based on trimmed mean	1.583	2	291	.207

ANOVA

AS - Dissolved sludge volume index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11094.598	2	5547.299	.767	.465
Within Groups	2104021.633	291	7230.315		
Total	2115116.231	293			

Appendix 5: Statistical results for comparison with previous studies

Binary: Using cut-off values given by Deepnarain et al. (2015) to create binary predictor variables.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-.602	.6034	-1.785	.580	.997	1	.318	.548	.168	1.787
[pHBin=1]	.247	.3730	-.484	.978	.439	1	.507	1.280	.616	2.660
[FMCodBin=1]	-.747	.4616	-1.651	.158	2.617	1	.106	.474	.192	1.171
[AveDOBin=1]	.331	.3166	-.290	.951	1.090	1	.296	1.392	.748	2.588
[CODBin=1]	-.171	.3611	-.879	.537	.224	1	.636	.843	.415	1.711
[NH4Bin=1]	.437	.4228	-.392	1.265	1.068	1	.301	1.548	.676	3.545
(Scale)	1[b]									

Dependent Variable: **1 TYPE 0092**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-2.022	.8876	-3.762	-.282	5.189	1	.023	.132	.023	.754
[pHBin=1]	-.939	.3933	-1.710	-.168	5.701	1	.017	.391	.181	.845
[FMCodBin=1]	1.043	.7857	-.497	2.582	1.761	1	.184	2.837	.608	13.230
[AveDOBin=1]	-.656	.4155	-1.470	.159	2.491	1	.115	.519	.230	1.172
[CODBin=1]	-.004	.4478	-.882	.873	.000	1	.993	.996	.414	2.395
[NH4Bin=1]	.685	.5557	-.404	1.774	1.519	1	.218	1.984	.668	5.895
(Scale)	1[b]									

Dependent Variable: **2 Type 021N & Thiothrix spp.**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-5.297	1.4078	-8.056	-2.538	14.156	1	.000	.005	.000	.079
[pHBin=1]	2.025	1.0402	-.014	4.064	3.790	1	.052	7.577	.986	58.192
[FMCodBin=1]	.511	.7902	-1.038	2.059	.418	1	.518	1.667	.354	7.842
[AveDOBin=1]	.670	.4531	-.218	1.558	2.184	1	.139	1.954	.804	4.748
[CODBin=1]	-.132	.4957	-1.104	.839	.071	1	.790	.876	.332	2.315
[NH4Bin=1]	1.056	.7037	-.323	2.435	2.252	1	.133	2.875	.724	11.420
(Scale)	1[b]									

Dependent Variable: **3 *M. parvicella***

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-1.663	.8181	-3.267	-.060	4.134	1	.042	.189	.038	.942
[pHBin=1]	.348	.5326	-.696	1.392	.426	1	.514	1.416	.498	4.022
[FMCodBin=1]	-.438	.6060	-1.626	.750	.522	1	.470	.645	.197	2.117
[AveDOBin=1]	-.588	.4784	-1.526	.350	1.511	1	.219	.555	.217	1.418
[CODBin=1]	.141	.5099	-.858	1.141	.077	1	.782	1.152	.424	3.129
[NH4Bin=1]	-.005	.5916	-1.165	1.154	.000	1	.993	.995	.312	3.171
(Scale)	1[b]									

Dependent Variable: **4 TYPE 0041**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-2.601	1.1873	-4.928	-.274	4.801	1	.028	.074	.007	.760
[pHBin=1]	-.444	.5291	-1.481	.593	.704	1	.401	.641	.227	1.809
[FMCodBin=1]	.718	1.0614	-1.362	2.798	.458	1	.499	2.051	.256	16.418
[AveDOBin=1]	-.141	.5315	-1.183	.900	.071	1	.791	.868	.306	2.461
[CODBin=1]	.591	.6589	-.701	1.882	.804	1	.370	1.805	.496	6.568
[NH4Bin=1]	-.419	.7156	-1.822	.983	.343	1	.558	.658	.162	2.674
(Scale)	1[b]									

Dependent Variable: **5 Type 1851**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-2.236	1.2000	-4.588	.116	3.472	1	.062	.107	.010	1.123
[pHBin=1]	-.111	.6229	-1.332	1.110	.032	1	.858	.895	.264	3.033
[FMCodBin=1]	.474	1.0735	-1.630	2.578	.195	1	.659	1.606	.196	13.165
[AveDOBin=1]	.092	.5712	-1.027	1.212	.026	1	.872	1.097	.358	3.360
[CODBin=1]	-.060	.7201	-1.472	1.351	.007	1	.933	.941	.230	3.862
[NH4Bin=1]	-1.001	.7033	-2.380	.377	2.026	1	.155	.367	.093	1.458
(Scale)	1[b]									

Dependent Variable: **6 GBPP**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

Ordinal: Using FI as a dependent variable for each filament. 0 = filament not dominant.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Intercept	[1 TYPE 0092=0]	.624	.5612	-.515	1.763	1.153	1	.283	1.866	.597	5.831
	[1 TYPE 0092=2]	1.036	.5651	-.109	2.184	3.145	1	.076	2.822	.897	8.884
	[1 TYPE 0092=3]	1.450	.5905	.292	2.608	6.023	1	.014	4.263	1.339	13.570
	[1 TYPE 0092=4]	3.450	.6954	2.087	4.813	24.615	1	.000	31.509	8.063	123.139
	[1 TYPE 0092=5]	5.270	1.1449	3.026	7.514	21.191	1	.000	194.494	20.624	1834.176
[pHBin=1]		.216	.3662	-.502	.934	.346	1	.555	1.241	.606	2.544
[FMCodBin=1]		-.530	.4251	-1.363	.304	1.552	1	.213	.589	.256	1.355
[AveDOBin=1]		.166	.3007	-.422	.757	.310	1	.578	1.182	.656	2.132
[CODBin=1]		-.371	.3447	-1.047	.304	1.159	1	.282	.690	.351	1.356
[NH4Bin=1]		.462	.4129	-.347	1.271	1.251	1	.263	1.587	.707	3.565
(Scale)		.993[b]									

Dependent Variable: **1 TYPE 0092**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Intercept	[2 Type 021N & Thiothrix spp.=0]	2.205	.8349	.569	3.842	6.975	1	.008	9.071	1.766	46.597
	[2 Type 021N & Thiothrix spp.=2]	2.343	.8367	.703	3.983	7.840	1	.005	10.410	2.019	53.664
	[2 Type 021N & Thiothrix spp.=3]	3.181	.8551	1.505	4.857	13.839	1	.000	24.071	4.504	128.628
	[2 Type 021N & Thiothrix spp.=4]	4.208	.9020	2.441	5.976	21.769	1	.000	67.250	11.479	393.967
	[2 Type 021N & Thiothrix spp.=5]	5.081	.9929	3.135	7.027	26.188	1	.000	160.917	22.966	1126.505
[pHBin=1]		-.856	.3622	-1.567	-.146	5.607	1	.018	.424	.209	.863
[FMCodBin=1]		1.066	.7362	-.377	2.509	2.097	1	.148	2.904	.686	12.291
[AveDOBin=1]		-.653	.3676	-1.412	.107	3.636	1	.092	.521	.244	1.113
[CODBin=1]		.110	.4145	-.702	.923	.071	1	.790	1.117	.496	2.516
[NH4Bin=1]		.730	.5163	-.286	1.746	1.995	1	.159	2.075	.751	5.732
(Scale)		.691[b]									

Dependent Variable: **2 Type 021N & Thiothrix spp.**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Intercept	[3 M. parvicella=0]	5.414	1.0010	3.452	7.376	29.256	1	.000	224.556	31.572	1597.166
	[3 M. parvicella=2]	5.796	1.0065	3.824	7.771	33.161	1	.000	329.560	45.805	2371.002
	[3 M. parvicella=3]	6.692	1.0255	4.675	8.706	42.301	1	.000	805.553	107.236	6051.264
	[3 M. parvicella=4]	7.420	1.0617	5.339	9.501	48.841	1	.000	1666.653	208.263	13366.362
	[3 M. parvicella=5]	8.126	1.1202	5.931	10.322	52.622	1	.000	3382.095	376.396	30389.689
[pHBin=1]		2.062	.7383	.615	3.509	7.800	1	.005	7.861	1.849	33.417
[FMCodBin=1]		.629	.5582	-.465	1.723	1.271	1	.260	1.876	.628	5.603
[AveDOBin=1]		.624	.3182	.001	1.246	3.846	1	.050	1.867	1.001	3.483
[CODBin=1]		-.157	.3480	-.839	.526	.202	1	.653	.855	.432	1.691
[NH4Bin=1]		1.072	.4979	.097	2.046	4.640	1	.031	2.923	1.101	7.754
(Scale)		.504[b]									

Dependent Variable: **3 M. parvicella**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

Parameter Estimates

Parameter		B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
				Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
Intercept	[4 TYPE 0041=0]	1.734	.6776	.406	3.062	6.547	1	.011	5.661	1.500	21.363
	[4 TYPE 0041=2]	2.021	.6821	.684	3.358	8.779	1	.003	7.545	1.982	28.726
	[4 TYPE 0041=3]	2.937	.7126	1.540	4.334	16.975	1	.000	18.857	4.664	76.247
	[4 TYPE 0041=4]	5.181	1.0562	3.107	7.255	23.973	1	.000	177.643	22.353	1414.939
[pHBin=1]		.382	.4401	-.460	1.245	.755	1	.385	1.466	.619	3.473
[FMCodBin=1]		-.360	.4962	-1.357	.596	.562	1	.446	.684	.258	1.816
[AveDOBin=1]		-.580	.3950	-1.355	.194	2.159	1	.142	.560	.258	1.214
[CODBin=1]		.152	.4244	-.680	.984	.129	1	.720	1.164	.507	2.675
[NH4Bin=1]		-.026	.4954	-.997	.945	.003	1	.958	.974	.369	2.572
(Scale)		.656[b]									

Dependent Variable: **4 TYPE 0041**

Model: (Intercept), pHBin, FMCodBin, AveDOBin, CODBin, NH4Bin

a. Set to zero because this parameter is redundant.

b. Computed based on the Pearson chi-square.

Dos Santos et al. (2015) Kendall's tau_b		TYPE 0092	Type 021N & Thiothrix spp.	M. parvicella	TYPE 0041	Type 1851	GBPP	TYPE 0803
Average Dissolved Oxygen	Correlation Coefficient	-0.023	-0.026	0.113	-0.080	0.004	0.047	-0.001
	Sig. (2-tailed)	0.729	0.695	0.091	0.229	0.954	0.481	0.988
	N	149	149	149	149	149	149	149
Temperature	Correlation Coefficient	-0.018	-0.081	-0.056	-0.075	.143*	0.060	0.103
	Sig. (2-tailed)	0.789	0.221	0.399	0.262	0.033	0.368	0.127
	N	149	149	149	149	149	149	149
Biological Oxygen Demand	Correlation Coefficient	0.078	-0.054	0.077	-0.099	-0.046	0.052	-0.043
	Sig. (2-tailed)	0.268	0.447	0.284	0.166	0.522	0.469	0.550
	N	132	132	132	132	132	132	132
Chemical Oxygen Demand	Correlation Coefficient	0.012	-0.001	0.044	-0.044	-0.033	0.057	-0.036
	Sig. (2-tailed)	0.863	0.993	0.524	0.515	0.634	0.404	0.604
	N	142	142	142	142	142	142	142
Total BOD:COD ratio	Correlation Coefficient	0.073	-0.059	0.092	-.143*	0.031	-0.037	0.033
	Sig. (2-tailed)	0.297	0.404	0.199	0.044	0.666	0.602	0.647
	N	130	130	130	130	130	130	130
pH_	Correlation Coefficient	.152*	-0.119	0.045	-0.072	-0.025	0.061	-0.072
	Sig. (2-tailed)	0.025	0.083	0.516	0.297	0.720	0.378	0.298
	N	147	147	147	147	147	147	147
Total alkalinity	Correlation Coefficient	0.080	0.010	0.048	-0.072	-0.035	0.005	-0.075
	Sig. (2-tailed)	0.234	0.877	0.483	0.290	0.607	0.944	0.276
	N	144	144	144	144	144	144	144
Ammonia	Correlation Coefficient	0.119	0.037	0.065	-0.065	-0.060	0.005	-.169*
	Sig. (2-tailed)	0.071	0.577	0.332	0.332	0.375	0.939	0.012
	N	147	147	147	147	147	147	147
Nitrates / Nitrites	Correlation Coefficient	-0.069	-0.021	0.059	-0.054	0.068	-0.042	0.153
	Sig. (2-tailed)	0.368	0.782	0.449	0.488	0.383	0.590	0.050
	N	143	143	143	143	143	143	143
Total phosphates	Correlation Coefficient	0.067	-0.050	0.077	-0.059	-0.046	0.056	-0.055
	Sig. (2-tailed)	0.321	0.458	0.256	0.382	0.501	0.409	0.423
	N	144	144	144	144	144	144	144
Ortho-phosphates	Correlation Coefficient	0.103	-0.026	0.082	-0.009	-0.092	0.032	-0.129
	Sig. (2-tailed)	0.128	0.700	0.232	0.890	0.182	0.645	0.062
	N	142	142	142	142	142	142	142
Sulphate	Correlation Coefficient	-0.030	0.099	-0.088	-0.101	0.015	0.103	0.000
	Sig. (2-tailed)	0.648	0.140	0.192	0.133	0.822	0.129	0.997
	N	147	147	147	147	147	147	147
Volatile fatty acids	Correlation Coefficient	0.036	-0.006	0.031	0.012	-0.070	0.056	-0.067
	Sig. (2-tailed)	0.591	0.928	0.641	0.864	0.303	0.411	0.324
	N	147	147	147	147	147	147	147
AS - Biological Oxygen Demand	Correlation Coefficient	-0.135	.224*	-0.155	.208*	0.128	-0.174	0.011
	Sig. (2-tailed)	0.154	0.018	0.104	0.031	0.189	0.071	0.908
	N	75	75	75	75	75	75	75
AS - Chemical Oxygen Demand	Correlation Coefficient	0.082	-0.087	-0.009	-0.036	-0.023	0.098	-0.094
	Sig. (2-tailed)	0.398	0.368	0.923	0.718	0.817	0.320	0.347
	N	70	70	70	70	70	70	70

Dos Santos et al. (2015) Kendall's tau_b		TYPE 0092	Type 021N & Thiothrix spp.	M. parvicella	TYPE 0041	Type 1851	GBPP	TYPE 0803
AS - Soluble Chemical Oxygen Demand	Correlation Coefficient	-0.111	.177*	-0.024	0.041	0.027	-0.016	-0.064
	Sig. (2-tailed)	0.103	0.010	0.732	0.559	0.702	0.817	0.359
	N	139	139	139	139	139	139	139
AS - Total BOD:COD ratio	Correlation Coefficient	-0.169	.235*	-0.145	.206*	0.108	-0.196	0.072
	Sig. (2-tailed)	0.087	0.017	0.144	0.039	0.282	0.050	0.475
	N	68	68	68	68	68	68	68
AS - pH	Correlation Coefficient	-0.062	-0.035	-0.016	-0.081	.199**	-0.028	0.132
	Sig. (2-tailed)	0.365	0.612	0.817	0.245	0.005	0.686	0.060
	N	149	149	149	149	149	149	149
AS - Ammonia	Correlation Coefficient	-0.055	0.116	-.139*	-0.011	-0.021	0.089	0.037
	Sig. (2-tailed)	0.404	0.082	0.037	0.868	0.752	0.185	0.579
	N	149	149	149	149	149	149	149
AS - Nitrates / Nitrites	Correlation Coefficient	0.068	-0.017	0.046	-0.098	-0.011	0.049	-0.071
	Sig. (2-tailed)	0.323	0.805	0.512	0.162	0.874	0.485	0.314
	N	149	149	149	149	149	149	149
AS - Total phosphates	Correlation Coefficient	-0.037	0.086	-0.103	0.026	0.032	0.033	-0.031
	Sig. (2-tailed)	0.569	0.196	0.124	0.693	0.629	0.626	0.648
	N	149	149	149	149	149	149	149
AS - Ortho-phosphates	Correlation Coefficient	-0.047	0.059	-0.114	0.023	0.053	0.042	-0.011
	Sig. (2-tailed)	0.470	0.373	0.087	0.735	0.433	0.530	0.870
	N	149	149	149	149	149	149	149
AS - Sulphate	Correlation Coefficient	-0.088	0.019	-0.059	-0.096	0.033	.139*	0.125
	Sig. (2-tailed)	0.183	0.773	0.383	0.154	0.622	0.040	0.066
	N	149	149	149	149	149	149	149
AS - F/M ratio BOD	Correlation Coefficient	-0.029	0.016	-0.056	0.034	0.069	0.035	-0.092
	Sig. (2-tailed)	0.703	0.837	0.478	0.661	0.386	0.653	0.242
	N	108	108	108	108	108	108	108
AS - F/M ratio COD	Correlation Coefficient	-0.016	0.045	-0.081	0.050	0.041	0.040	-0.112
	Sig. (2-tailed)	0.832	0.555	0.294	0.516	0.600	0.599	0.148
	N	113	113	113	113	113	113	113
AS - Mixed liquor volatile suspended solids	Correlation Coefficient	0.021	-0.009	0.015	-0.042	-0.004	-0.008	0.059
	Sig. (2-tailed)	0.760	0.895	0.829	0.548	0.951	0.912	0.398
	N	136	136	136	136	136	136	136
AS - Dissolved sludge volume index	Correlation Coefficient	0.027	0.078	-0.035	-0.111	0.035	0.058	-0.083
	Sig. (2-tailed)	0.687	0.254	0.611	0.106	0.613	0.396	0.232
	N	141	141	141	141	141	141	141

Validation of dos Santos (2015) correlations: Rotated Component Matrix[a]

	Component									
	1	2	3	4	5	6	7	8	9	10
Ortho-phosphates	0.968									
Total phosphates	0.931									
Chemical Oxygen Demand	0.897									
Biological Oxygen Demand	0.856							0.307		
Total alkalinity	0.819									
AS - Mixed liquor volatile suspended solids	0.725		-0.532							
Ammonia	0.679				-0.352			-0.316		
AS - Chemical Oxygen Demand	0.667		-0.396				-0.457			
AS - Ortho-phosphates	-0.443			0.434		0.340				
AS - Ammonia		0.935								
AS - Dissolved sludge volume index		0.846				0.309				
Volatile fatty acids	0.365	0.637								
AS - F/M ratio BOD			0.917							
AS - F/M ratio COD			0.881							
AS - Biological Oxygen Demand				0.864						
AS - Total BOD:COD ratio	-0.351			0.735			0.367			
AS - Nitrates / Nitrites				-0.597						
2 Type 021N & Thiothrix spp.		0.376		0.443	0.344			-0.350		
pH_NBNBNB					-0.856					
Total BOD:COD ratio		0.355			-0.775					
AS - Sulphate					0.629	0.462				
AS - pH						0.849				
Filament Index		0.393				0.697			-0.310	
Temperature							-0.652			
Sulphate	-0.321				0.367	0.341	0.626			
AS - Total phosphates	0.557						-0.627			
AS - Soluble Chemical Oxygen Demand				0.326	0.334	0.316	0.401			
5 Type 1851	0.360							0.711		
Average Dissolved Oxygen			0.426					-0.559		
4 TYPE 0041									0.861	
Nitrates / Nitrites	0.363			-0.320					0.517	
3 M. parvicella										-0.685
1 TYPE 0092				-0.317	-0.434					0.626

Colour coded parameters have high correlations with one another

Appendix 6: Results of generalised linear model applied with Type 0092

Model Information

Dependent Variable	1 TYPE 0092[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models 1 as the response, treating 0 as the reference category.

Case Processing Summary

	N	Percent	Unweighted N
Included	302.000	99.3%	302
Excluded	2.000	0.7%	2
Total	304.000	100.0%	304

Categorical Variable Information

			N	Percent	Unweighted N
Dependent Variable	1 TYPE 0092	0	196.000	64.9%	196
		1	106.000	35.1%	106
		Total	302.000	100.0%	302

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation	Unweighted N
Covariate	pH_NBNBNB	302.000	6.100	8.600	7.47947	.396130	302

Goodness of Fit[a]

	Value	df	Value/df
Deviance	19.063	19	1.003
Scaled Deviance	20.464	19	
Pearson Chi-Square	17.699	19	.932
Scaled Pearson Chi-Square	19.000	19	
Log Likelihood[b,c]	-33.989		
Adjusted Log Likelihood[d]	-36.489		
Akaike's Information Criterion (AIC)	71.979		
Finite Sample Corrected AIC (AICC)	72.019		
Bayesian Information Criterion (BIC)	79.399		
Consistent AIC (CAIC)	81.399		

Dependent Variable: 1 TYPE 0092
Model: (Intercept), pH_NBNBNB
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.
c. The log likelihood is based on a scale parameter fixed at 1.
d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	p-value
5.446	1	.020

Dependent Variable: 1 TYPE 0092
Model: (Intercept), pH_NBNBNB

Tests of Model Effects

Source	Type III Wald Chi-Square	df	p-value
(Intercept)	6.631	1	.010
pH_NBNBNB	5.335	1	.021

Dependent Variable: 1 TYPE 0092
Model: (Intercept), pH_NBNBNB

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square	df	p-value	Exp(B)	95% Wald Confidence Interval for Exp(B) Lower	95% Wald Confidence Interval for Exp(B) Upper
(Intercept)	-5.801	2.2529	-10.217	-1.386	6.631	1	.010	.003	3.655E-5	.250
pH_NBNBNB	.692	.2996	.105	1.279	5.335	1	.021	1.998	1.110	3.594
(Scale)	.932[a]									

Dependent Variable: 1 TYPE 0092
Model: (Intercept), pH_NBNBNB
a. Computed based on the Pearson chi-square.

Appendix 7: Results of generalised linear model applied with M. parvicella

Model Information

Dependent Variable	3 M. parvicella[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models 1 as the response, treating 0 as the reference category.

Case Processing Summary

	N	Percent	Unweighted N
Included	269.000	88.5%	269
Excluded	35.000	11.5%	35
Total	304.000	100.0%	304

Categorical Variable Information

			N	Percent	Unweighted N
Dependent Variable	3 M. parvicella	0	227.000	84.4%	227
		1	42.000	15.6%	42
		Total	269.000	100.0%	269

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation	Unweighted N
Covariate	Sulphate	269.000	22.300	193.000	74.40186	26.548181	269
	AS - Ortho-phosphates	269.000	.050	30.800	6.43810	7.330802	269

Goodness of Fit[a]

	Value	df	Value/df
Deviance	91.935	119	.773
Scaled Deviance	122.686	119	
Pearson Chi-Square	89.173	119	.749
Scaled Pearson Chi-Square	119.000	119	
Log Likelihood[b,c]	-74.296		
Adjusted Log Likelihood[d]	-99.147		
Akaike's Information Criterion (AIC)	152.591		
Finite Sample Corrected AIC (AICC)	152.637		
Bayesian Information Criterion (BIC)	159.781		
Consistent AIC (CAIC)	161.781		

Dependent Variable: 3 M. parvicella
Model: Sulphate, AS - Ortho-phosphates
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.
c. The log likelihood is based on a scale parameter fixed at 1.
d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	p-value
205.108	2	.000

Dependent Variable: 3 M. parvicella
Model: Sulphate, AS - Ortho-phosphates
a. Compares the fitted model against the null model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	p-value
Sulphate	53.351	1	.000
AS - Ortho-phosphates	5.727	1	.017

Dependent Variable: 3 M. parvicella
Model: Sulphate, AS - Ortho-phosphates

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square	df	p-value	Exp(B)	95% Wald Confidence Interval for Exp(B) Lower	95% Wald Confidence Interval for Exp(B) Upper
Sulphate	-.020	.0027	-.025	-.014	53.351	1	.000	.981	.976	.986
AS - Ortho-phosphates	-.063	.0262	-.114	-.011	5.727	1	.017	.939	.892	.989
(Scale)	.749[a]									

Dependent Variable: 3 M. parvicella
Model: Sulphate, AS - Ortho-phosphates
a. Computed based on the Pearson chi-square.

Appendix 8: Results of generalised linear model applied with Type 021N and Thiothrix spp.

Model Information

Dependent Variable	2 Type 021N & Thiothrix spp.[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models 1 as the response, treating 0 as the reference category.

Case Processing Summary

	N	Percent	Unweighted N
Included	211.000	69.4%	211
Excluded	93.000	30.6%	93
Total	304.000	100.0%	304

Categorical Variable Information

			N	Percent	Unweighted N
Dependent Variable	2 Type 021N & Thiothrix spp.	0	170.000	80.6%	170
		1	41.000	19.4%	41
		Total	211.000	100.0%	211

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation	Unweighted N
Covariate	Total BOD:COD ratio	211.000	.169	.987	.43087	.126285	211
	Ammonia	211.000	6.500	104.000	48.85024	22.164006	211
	Volatile fatty acids	211.000	.500	375.000	71.61137	58.557893	211
	AS - Nitrates / Nitrites	211.000	.050	11.000	1.20403	2.124426	211
	AS - Total phosphates	211.000	1.000	243.000	81.22867	70.904935	211
	AS - F/M ratio BOD	211.000	.007	2.331	.20421	.355859	211
	AS - F/M ratio COD	211.000	.014	5.750	.51185	.912364	211

Goodness of Fit[a]

	Value	df	Value/df
Deviance	65.563	91	.720
Scaled Deviance	95.995	91	
Pearson Chi-Square	62.152	91	.683
Scaled Pearson Chi-Square	91.000	91	
Log Likelihood[b,c]	-60.417		
Adjusted Log Likelihood[d]	-88.460		
Akaike's Information Criterion (AIC)	134.833		
Finite Sample Corrected AIC (AICC)	135.385		
Bayesian Information Criterion (BIC)	158.296		
Consistent AIC (CAIC)	165.296		

Dependent Variable: 2 Type 021N & Thiothrix spp.
Model: Total BOD:COD ratio, Ammonia, Volatile fatty acids, AS - Nitrates / Nitrites, AS - Total phosphates, AS - F/M ratio BOD, AS - F/M ratio COD
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.
c. The log likelihood is based on a scale parameter fixed at 1.
d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	p-value
155.576	7	.000

Dependent Variable: 2 Type 021N & Thiothrix spp.
Model: Total BOD:COD ratio, Ammonia, Volatile fatty acids, AS - Nitrates / Nitrites, AS - Total phosphates, AS - F/M ratio BOD, AS - F/M ratio COD
a. Compares the fitted model against the null model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	p-value
Total BOD:COD ratio	6.710	1	.010
Ammonia	7.870	1	.005
Volatile fatty acids	8.381	1	.004
AS - Nitrates / Nitrites	4.657	1	.031
AS - Total phosphates	3.666	1	.056
AS - F/M ratio BOD	4.056	1	.044
AS - F/M ratio COD	4.828	1	.028

Dependent Variable: 2 Type 021N & Thiothrix spp.
Model: Total BOD:COD ratio, Ammonia, Volatile fatty acids, AS - Nitrates / Nitrites, AS - Total phosphates, AS - F/M ratio BOD, AS - F/M ratio COD

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square	df	p-value	Exp(B)	95% Wald Confidence Interval for Exp(B) Lower	95% Wald Confidence Interval for Exp(B) Upper
Total BOD:COD ratio	-2.324	.8970	-4.082	-.566	6.710	1	.010	.098	.017	.568
Ammonia	-.023	.0084	-.040	-.007	7.870	1	.005	.977	.961	.993
Volatile fatty acids	.007	.0024	.002	.012	8.381	1	.004	1.007	1.002	1.012
AS - Nitrates / Nitrites	-.224	.1037	-.427	-.021	4.657	1	.031	.799	.652	.980
AS - Total phosphates	.005	.0025	.000	.010	3.666	1	.056	1.005	1.000	1.010
AS - F/M ratio BOD	-5.765	2.8624	-11.375	-.155	4.056	1	.044	.003	1.148E-5	.857
AS - F/M ratio COD	2.059	.9371	.222	3.896	4.828	1	.028	7.839	1.249	49.194
(Scale)	.683[a]									

Dependent Variable: 2 Type 021N & Thiothrix spp.
Model: Total BOD:COD ratio, Ammonia, Volatile fatty acids, AS - Nitrates / Nitrites, AS - Total phosphates, AS - F/M ratio BOD, AS - F/M ratio COD
a. Computed based on the Pearson chi-square.

Appendix 9: Results of generalised linear model applied with Type 0041

Warnings

A quasi-complete separation may exist in the data. The maximum likelihood estimates do not exist.
The GENLIN procedure continues despite the above warning(s). Subsequent results shown are based on the last iteration. Validity of the model fit is uncertain.

Model Information

Dependent Variable	4 TYPE 0041[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models 1 as the response, treating 0 as the reference category.

Case Processing Summary

	N	Percent	Unweighted N
Included	88.000	28.9%	88
Excluded	216.000	71.1%	216
Total	304.000	100.0%	304

Categorical Variable Information

			N	Percent	Unweighted N
Dependent Variable	4 TYPE 0041	0	79.000	89.8%	79
		1	9.000	10.2%	9
		Total	88.000	100.0%	88

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation	Unweighted N
Covariate	Average Dissolved Oxygen	88.000	.040	2.900	.93580	.670873	88
	Chemical Oxygen Demand	88.000	217.000	2303.000	841.54545	459.440987	88
	Total BOD:COD ratio	88.000	.236	.987	.43543	.109393	88
	pH_NBNBNB	88.000	6.100	8.600	7.51250	.437814	88
	Total alkalinity	88.000	156	494	328.47	89.133	88
	Ammonia	88.000	18.500	96.300	47.45227	19.150313	88
	Nitrates / Nitrites	88.000	.010	1.600	.09659	.240306	88
	Total phosphates	88.000	3.100	35.400	9.54091	5.977364	88
	Ortho-phosphates	88.000	1.800	14.000	5.73636	2.942834	88
	Sulphate	88.000	28.000	193.000	77.18636	33.145961	88
	Volatile fatty acids	88.000	9.000	375.000	76.42045	74.290626	88
	AS - Chemical Oxygen Demand	88.000	1734.000	11025.000	5467.87500	2058.396691	88
	AS - pH	88.000	6.400	7.700	7.00227	.254150	88
	AS - Ammonia	88.000	.100	44.600	8.75114	12.181086	88
	AS - Nitrates / Nitrites	88.000	.050	6.300	.98068	1.590172	88
	AS - Total phosphates	88.000	5.200	243.000	86.16648	78.207853	88
	AS - Ortho-phosphates	88.000	.050	22.700	5.85170	6.508255	88
	AS - Sulphate	88.000	34.000	190.000	65.31250	25.796364	88
	AS - F/M ratio BOD	88.000	.030	1.179	.17702	.217151	88
	AS - F/M ratio COD	88.000	.077	4.428	.45884	.731425	88
	AS - Mixed liquor volatile suspended solids	88.000	404.000	6720.000	3607.50000	1476.514263	88
	AS - Dissolved sludge volume index	88.000	58.000	500.000	152.70455	99.030259	88

Goodness of Fit[a]

	Value	df	Value/df
Deviance	.000	22	.000
Scaled Deviance	44.000	22	
Pearson Chi-Square	.000	22	.000
Scaled Pearson Chi-Square	22.000	22	
Log Likelihood[b,c]	-6.762		
Adjusted Log Likelihood[d]	-2408800328.001		
Akaike's Information Criterion (AIC)	51.523		
Finite Sample Corrected AIC (AICC)	62.700		
Bayesian Information Criterion (BIC)	98.593		
Consistent AIC (CAIC)	117.593		

Dependent Variable: 4 TYPE 0041

Model: (Intercept), Average Dissolved Oxygen, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Total phosphates, Ortho-phosphates, Sulphate, Volatile fatty acids, AS - Chemical Oxygen Demand, AS - pH, AS - Ammonia, AS - Nitrates / Nitrites, AS - Total phosphates, AS - Ortho-phosphates, AS - F/M ratio BOD, AS - F/M ratio COD, AS - Mixed liquor volatile suspended solids

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	p-value
10071001918.577	18	.000

Dependent Variable: 4 TYPE 0041

Model: (Intercept), Average Dissolved Oxygen, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Total phosphates, Ortho-phosphates, Sulphate, Volatile fatty acids, AS - Chemical Oxygen Demand, AS - pH, AS - Ammonia, AS - Nitrates / Nitrites, AS - Total phosphates, AS - Ortho-phosphates, AS - F/M ratio BOD, AS - F/M ratio COD, AS - Mixed liquor volatile suspended solids

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	p-value
(Intercept)	257.980	1	.000
Average Dissolved Oxygen	1398.768	1	.000
Chemical Oxygen Demand	303.053	1	.000
Total BOD:COD ratio	876.686	1	.000
pH_NBNBNB	90.287	1	.000
Total alkalinity	5.663	1	.017
Total phosphates	331.124	1	.000
Ortho-phosphates	38.932	1	.000
Sulphate	12.001	1	.001
Volatile fatty acids	161.144	1	.000
AS - Chemical Oxygen Demand	42.362	1	.000
AS - pH	61.399	1	.000
AS - Ammonia	147.261	1	.000
AS - Nitrates / Nitrites	163.933	1	.000
AS - Total phosphates	5.396	1	.020
AS - Ortho-phosphates	106.597	1	.000
AS - F/M ratio BOD	1315.238	1	.000
AS - F/M ratio COD	1404.319	1	.000
AS - Mixed liquor volatile suspended solids	96.303	1	.000

Dependent Variable: 4 TYPE 0041

Model: (Intercept), Average Dissolved Oxygen, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Total phosphates, Ortho-phosphates, Sulphate, Volatile fatty acids, AS - Chemical Oxygen Demand, AS - pH, AS - Ammonia, AS - Nitrates / Nitrites, AS - Total phosphates, AS - Ortho-phosphates, AS - F/M ratio BOD, AS - F/M ratio COD, AS - Mixed liquor volatile suspended solids

Appendix 10: Results of generalised linear model applied with Gram positive branching bacilli

Warnings

A quasi-complete separation may exist in the data. The maximum likelihood estimates do not exist.
The GENLIN procedure continues despite the above warning(s). Subsequent results shown are based on the last iteration. Validity of the model fit is uncertain.

Model Information

Dependent Variable	6 GBPP[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models 1 as the response, treating 0 as the reference category.

Case Processing Summary

	N	Percent	Unweighted N
Included	100.000	32.9%	100
Excluded	204.000	67.1%	204
Total	304.000	100.0%	304

Categorical Variable Information

			N	Percent	Unweighted N
Dependent Variable	6 GBPP	0	93.000	93.0%	93
		1	7.000	7.0%	7
		Total	100.000	100.0%	100

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation	Unweighted N
Covariate	Average Dissolved Oxygen	100.000	.040	2.900	.90570	.642418	100
	Biological Oxygen Demand	100.000	90	1025	394.45	218.724	100
	Chemical Oxygen Demand	100.000	217.000	2303.000	879.30000	446.469914	100
	Total BOD:COD ratio	100.000	.236	.987	.45378	.125442	100
	pH_NBNBNB	100.000	6.100	8.600	7.52600	.425837	100
	Total alkalinity	100.000	156	494	338.65	88.519	100
	Ammonia	100.000	18.500	96.300	49.81600	20.374929	100
	Nitrates / Nitrites	100.000	.010	1.600	.09470	.228276	100
	Total phosphates	100.000	3.100	35.400	10.11300	5.907992	100
	Ortho-phosphates	100.000	1.800	14.000	6.16300	3.097397	100
	Sulphate	100.000	28.000	193.000	74.56400	32.214968	100
	AS - Chemical Oxygen Demand	100.000	1734.000	11025.000	5559.37000	1958.818582	100
	AS - pH	100.000	6.400	7.700	6.97900	.254771	100
	AS - Ammonia	100.000	.100	65.200	10.42800	15.220879	100
	AS - Total phosphates	100.000	5.200	243.000	90.01250	76.184603	100
	AS - Ortho-phosphates	100.000	.050	22.700	5.61650	6.297322	100
	AS - Sulphate	100.000	34.000	190.000	64.17500	24.679540	100
	AS - F/M ratio BOD	100.000	.030	1.179	.17297	.206286	100
	AS - F/M ratio COD	100.000	.077	4.428	.43265	.691306	100
	AS - Mixed liquor volatile suspended solids	100.000	404.000	6720.000	3779.80000	1514.028528	100
	AS - Dissolved sludge volume index	100.000	58.000	500.000	152.16000	93.240725	100

Goodness of Fit[a]

	Value	df	Value/df
Deviance	.000	29	.000
Scaled Deviance	58.000	29	
Pearson Chi-Square	.000	29	.000
Scaled Pearson Chi-Square	29.000	29	
Log Likelihood[b,c]	-3.753		
Adjusted Log Likelihood[d]	-1319673357.577		
Akaike's Information Criterion (AIC)	41.507		
Finite Sample Corrected AIC (AICC)	48.970		
Bayesian Information Criterion (BIC)	85.795		
Consistent AIC (CAIC)	102.795		

Dependent Variable: 6 GBPP
Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, Sulphate, AS - Chemical Oxygen Demand, AS - pH, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD
a. Information criteria are in smaller-is-better form.
b. The full log likelihood function is displayed and used in computing information criteria.
c. The log likelihood is based on a scale parameter fixed at 1.
d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	p-value
11986579277.882	16	.000

Dependent Variable: 6 GBPP
Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, Sulphate, AS - Chemical Oxygen Demand, AS - pH, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD
a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	p-value
(Intercept)	1163.784	1	.000
Average Dissolved Oxygen	254.947	1	.000
Biological Oxygen Demand	2694.077	1	.000
Chemical Oxygen Demand	1470.221	1	.000
Total BOD:COD ratio	924.869	1	.000
pH_NBNBNB	13.464	1	.000
Total alkalinity	39.623	1	.000
Ammonia	1290.959	1	.000
Nitrates / Nitrites	63.667	1	.000
Total phosphates	805.626	1	.000
Ortho-phosphates	3349.702	1	.000
Sulphate	5.430	1	.020
AS - Chemical Oxygen Demand	519.588	1	.000
AS - pH	481.098	1	.000
AS - Sulphate	55.977	1	.000
AS - F/M ratio BOD	430.024	1	.000
AS - F/M ratio COD	540.053	1	.000

Dependent Variable: 6 GBPP
Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, Sulphate, AS - Chemical Oxygen Demand, AS - pH, AS - Sulphate, AS - F/M ratio BOD, AS F/M ratio COD

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	p-value		Lower	Upper
(Intercept)	737.930	21.6311	695.534	780.327	1163.786	1	.000	.[a]	1.166E+302	.[a]
Average Dissolved Oxygen	11.178	.7001	9.806	12.551	254.948	1	.000	71568.989	18147.074	282255.986
Biological Oxygen Demand	.641	.0124	.617	.665	2694.101	1	.000	1.898	1.853	1.945
Chemical Oxygen Demand	-.239	.0062	-.251	-.227	1470.223	1	.000	.787	.778	.797
Total BOD:COD ratio	-788.743	25.9353	-839.575	-737.911	924.885	1	.000	.000	.000	3.384E-321
pH_NBNBNB	-14.738	4.0165	-22.610	-6.866	13.464	1	.000	3.976E-7	1.516E-10	.001
Total alkalinity	.116	.0185	.080	.152	39.623	1	.000	1.123	1.083	1.165
Ammonia	2.881	.0802	2.724	3.038	1290.975	1	.000	17.826	15.234	20.860
Nitrates / Nitrites	-89.520	11.2191	-111.509	-67.531	63.668	1	.000	1.324E-39	3.733E-49	4.695E-30
Total phosphates	9.694	.3415	9.024	10.363	805.627	1	.000	16213.830	8301.973	31665.760
Ortho-phosphates	-44.682	.7720	-46.195	-43.169	3349.732	1	.000	3.935E-20	8.667E-21	1.787E-19
Sulphate	.048	.0208	.008	.089	5.430	1	.020	1.050	1.008	1.093
AS - Chemical Oxygen Demand	.011	.0005	.010	.012	519.584	1	.000	1.011	1.010	1.012
AS - pH	-68.210	3.1097	-74.305	-62.115	481.105	1	.000	2.382E-30	5.370E-33	1.057E-27
AS - Sulphate	.515	.0689	.380	.651	55.977	1	.000	1.674	1.463	1.917
AS - F/M ratio BOD	387.148	18.6692	350.557	423.739	430.032	1	.000	1.368E+168	1.758E+152	1.065E+184
AS - F/M ratio COD	-120.598	5.1894	-130.769	-110.427	540.062	1	.000	4.218E-53	1.614E-57	1.102E-48
(Scale)	2.844E-9[b]									

Dependent Variable: 6 GBPP

Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, Sulphate, AS - Chemical Oxygen Demand, AS - pH, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD

a. Set to system missing due to overflow

b. Computed based on the Pearson chi-square.

Appendix 11: Results of generalised linear model applied with Type 1851

Model Information

Dependent Variable	5 Type 1851[a]
Probability Distribution	Binomial
Link Function	Logit

a. The procedure models 1 as the response, treating 0 as the reference category.

Case Processing Summary

	N	Percent	Unweighted N
Included	153.000	50.3%	153
Excluded	151.000	49.7%	151
Total	304.000	100.0%	304

Categorical Variable Information

			N	Percent	Unweighted N
Dependent Variable	5 Type 1851	0	140.000	91.5%	140
		1	13.000	8.5%	13
		Total	153.000	100.0%	153

Continuous Variable Information

		N	Minimum	Maximum	Mean	Std. Deviation	Unweighted N
Covariate	Average Dissolved Oxygen	153.000	.040	6.050	1.11464	.964227	153
	Biological Oxygen Demand	153.000	36	1025	331.49	209.044	153
	Chemical Oxygen Demand	153.000	95.000	2303.000	744.64706	435.430481	153
	Total BOD:COD ratio	153.000	.236	.987	.45101	.120274	153
	pH	153.000	6.100	8.600	7.53072	.428934	153
	Total alkalinity	153.000	142	494	312.51	93.006	153
	Ammonia	153.000	10.700	96.300	46.07908	21.821298	153
	Nitrates / Nitrites	153.000	.010	3.500	.13889	.435717	153
	Total phosphates	153.000	2.100	87.000	11.80065	12.567190	153
	Ortho-phosphates	153.000	.800	23.000	5.90261	3.902512	153
	AS - pH	153.000	6.400	8.900	7.03856	.331687	153
	AS - Ammonia	153.000	.100	65.200	9.03922	12.628004	153
	AS - Nitrates / Nitrites	153.000	.050	7.300	.81961	1.465045	153
	AS - Total phosphates	153.000	1.000	243.000	65.26307	71.170976	153
	AS - Sulphate	153.000	34.000	190.000	63.00980	21.501979	153
	AS - F/M ratio BOD	153.000	.007	2.331	.25706	.406281	153
	AS - F/M ratio COD	153.000	.014	5.750	.63187	1.045967	153

Goodness of Fit[a]

	Value	df	Value/df
Deviance	10.848	55	.197
Scaled Deviance	59.052	55	
Pearson Chi-Square	10.103	55	.184
Scaled Pearson Chi-Square	55.000	55	
Log Likelihood[b,c]	-14.265		
Adjusted Log Likelihood[d]	-77.654		
Akaike's Information Criterion (AIC)	64.530		
Finite Sample Corrected AIC (AICC)	69.634		
Bayesian Information Criterion (BIC)	119.078		
Consistent AIC (CAIC)	137.078		

Dependent Variable: 5 Type 1851

Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, AS - pH, AS - Ammonia, AS - Nitrates / Nitrites, AS - Total phosphates, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

c. The log likelihood is based on a scale parameter fixed at 1.

d. The adjusted log likelihood is based on an estimated scale parameter and is used in the model fitting omnibus test.

Omnibus Test[a]

Likelihood Ratio Chi-Square	df	Sig.
217.644	17	.000

Dependent Variable: 5 Type 1851

Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, AS - pH, AS - Ammonia, AS - Nitrates / Nitrites, AS - Total phosphates, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD

a. Compares the fitted model against the intercept-only model.

Tests of Model Effects

Source	Type III Wald Chi-Square	df	Sig.
(Intercept)	11.234	1	.001
Average Dissolved Oxygen	5.015	1	.025
Biological Oxygen Demand	12.943	1	.000
Chemical Oxygen Demand	13.961	1	.000
Total BOD:COD ratio	12.566	1	.000
pH_NBNBNB	15.200	1	.000
Total alkalinity	8.719	1	.003
Ammonia	10.886	1	.001
Nitrates / Nitrites	16.152	1	.000
Total phosphates	9.778	1	.002
Ortho-phosphates	12.054	1	.001
AS - pH	18.415	1	.000
AS - Ammonia	14.365	1	.000
AS - Nitrates / Nitrites	11.909	1	.001
AS - Total phosphates	9.548	1	.002
AS - Sulphate	15.604	1	.000
AS - F/M ratio BOD	3.242	1	.072
AS - F/M ratio COD	2.561	1	.110

Dependent Variable: 5 Type 1851

Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Dem ratio, pH, Total alkalinity, Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, AS - pH - Nitrates / Nitrites, AS - Total phosphates, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test			Exp(B)	95% Wald Confidence Interval for Exp(B)	
			Lower	Upper	Wald Chi-Square	df	Sig.		Lower	Upper
(Intercept)	-50.971	15.2072	-80.776	-21.165	11.234	1	.001	7.305E-23	8.304E-36	6.426E-10
Average Dissolved Oxygen	1.601	.7149	.200	3.002	5.015	1	.025	4.958	1.221	20.128
Biological Oxygen Demand	-.122	.0339	-.188	-.055	12.943	1	.000	.885	.828	.946
Chemical Oxygen Demand	.063	.0168	.030	.096	13.961	1	.000	1.065	1.030	1.100
Total BOD:COD ratio	98.145	27.6863	43.881	152.410	12.566	1	.000	4.207E+42	114120030753 37052000.000	1.551E+66
pH_NBNBNB	-14.279	3.6624	-21.457	-7.100	15.200	1	.000	6.293E-7	4.802E-10	.001
Total alkalinity	.078	.0263	.026	.129	8.719	1	.003	1.081	1.026	1.138
Ammonia	-.264	.0800	-.421	-.107	10.886	1	.001	.768	.657	.898
Nitrates / Nitrites	6.267	1.5594	3.211	9.323	16.152	1	.000	526.917	24.797	11196.447
Total phosphates	.626	.2002	.234	1.018	9.778	1	.002	1.870	1.263	2.768
Ortho-phosphates	-3.413	.9831	-5.340	-1.486	12.054	1	.001	.033	.005	.226
AS - pH	15.076	3.5131	8.190	21.962	18.415	1	.000	3526966.648	3605.856	3449803672.647
AS - Ammonia	.256	.0675	.123	.388	14.365	1	.000	1.291	1.131	1.474
AS - Nitrates / Nitrites	-6.963	2.0178	-10.918	-3.009	11.909	1	.001	.001	1.812E-5	.049
AS - Total phosphates	.037	.0121	.014	.061	9.548	1	.002	1.038	1.014	1.063
AS - Sulphate	-.210	.0532	-.315	-.106	15.604	1	.000	.810	.730	.899
AS - F/M ratio BOD	8.167	4.5359	-.723	17.057	3.242	1	.072	3523.407	.485	25579913.456
AS - F/M ratio COD	-3.098	1.9357	-6.892	.696	2.561	1	.110	.045	.001	2.006
(Scale)	.184[a]									

Dependent Variable: 5 Type 1851

Model: (Intercept), Average Dissolved Oxygen, Biological Oxygen Demand, Chemical Oxygen Demand, Total BOD:COD ratio, pH_NBNBNB, Total alkalinity , Ammonia, Nitrates / Nitrites, Total phosphates, Ortho-phosphates, AS - pH, AS - Ammonia, AS - Nitrates / Nitrites, AS - Total phosphates, AS - Sulphate, AS - F/M ratio BOD, AS - F/M ratio COD

a. Computed based on the Pearson chi-square.