POLITECNICO DI TORINO

Master Degree course in Environmental And Land Engineering

Master Degree Thesis

Influent Characterization and Primary Clarifier Modelling for Digital Twin Enhancement at Viikinmäki WWTP

Supervisors
Prof. Anna Mikola
D.Sc.(Tech) Henri Haimi
Prof. Barbara Ruffino

Candidate
Sergio Cozzoli

Academic Year 2024-2025

Acknowledgements

I would like to thank professors Anna Mikola and Henri Haimi for supervising, advising, guiding, and inspiring me throughout this thesis. I am also grateful to Professor Barbara Ruffino for kindly supporting me throughout this process.

My sincere thanks go to all the staff at the Water and Environmental Engineering research group for making me feel welcome from the very first day of my exchange and for creating a healthy, productive, and safe working environment. In particular, I would like to thank Antti, Essi, Milla and Marina for all their help, patience and for sharing with me their indispensable knowledge to carry out this work.

My deepest gratitude goes to all those dear to me, who encourage me and teach me every day how to become a better person.

Abstract

This thesis presents the results of a sampling and analysis campaign conducted at the Viikinmäki wastewater treatment plant (Helsinki, Finland). The collected data were made available to the DIGICARBA project team, which used them to calibrate the plant digital twin currently under development. For this reason, the sampling strategy, laboratory analyses, and data processing were designed and carried out in accordance with guidelines, standards, and methods specifically developed for wastewater treatment plant modelling. During six sampling days, a 24-hour sampling programme was implemented, consisting of eight 3-hour composite samples per day. These samples were analysed to assess the diurnal variation of organic matter and suspended solids loads at three strategic points within the plant. These measurements enabled a detailed influent characterisation consistent with the requirements of digital twin modelling tools. In parallel, the availability of dimensional parameters and operational information of the primary clarifier allowed the development of prototype models of this unit. The modelling work was performed using SUMO simulation software, testing three different approaches: the Volumeless Point Separator (VPS), the Layered Flux Model (LFM), and the Three Compartment Model (TCM). Their performances were evaluated using the Mean Absolute Percentage Error (MAPE). The models were calibrated using the data from the present campaign and validated against those from a previous one. This comparison allowed the identification of modelling strategies that could serve as a foundation for extending the digital twin model to include the primary clarifier.

Contents

Symbols and Abbreviations

ASM	Activated Sludge Model
BCOD	Biodegradable Chemical Oxygen Demand
BOD	Biochemical Oxygen Demand
BOD_5	5-day Biochemical Oxygen Demand
BOD_7	7-day Biochemical Oxygen Demand
BOD_{tot}	Total Biochemical Oxygen Demand
COD	Chemical Oxygen Demand
COD_{tot}	Total COD
CSTR	Continuous-flow Stirred-Tank Reactor
DT	Digital Twin
fCOD	Filtered COD
ffCOD	Filtered Flocculated COD
GHG	Greenhouse gas
HSY	Helsinki Region Environmental Services
LFM	Layered Flux Model
PC	Primary Clarifier
PI	Primary Influent
PE	Primary Effluent
SE	Secondary Effluent
SI	Unbiodegradable Soluble Organics
SS	Soluble Substrate
TSS	Total Suspended Solids
TCM	Three Compartment Model
VPS	Volumeless Point Separator
VSS	Volatile Suspended Solids
WWTP	Wastewater Treatment Plant
XI	Unbiodegradable Particulate Organics
XS	Particulate Substrate

Chapter 1

Introduction

This thesis was written as part of the DIGICARBA project. Thanks to the efforts of researchers from Aalto University Water and Wastewater Engineering Group, this project aims to create a digital twin (DT) of the wastewater treatment process at the Viikinmäki wastewater treatment plant (WWTP) in Helsinki, Finland. The digital twin framework is implemented through SUMO, an open-source simulator developed by Dynamita (2022a), which enables to implement the process models forming the basis of the digital twin.

1.1 Background

The preservation and management of water resources are of fundamental importance in ensuring both human well-being and ecosystem stability. Within this framework, wastewater management represents a crucial step. Wastewater treatment plants (WWTPs) are currently facing numerous challenges, ranging from the emergence of new pollutants to the continuous growth of urban populations; from the impacts of climate change to nutrient overloading in surface waters. Policymakers increasingly emphasize the need to embed these processes within a more sustainable circular economy, extracting value from waste streams while minimizing harmful emissions (Wang et al., 2024).
The digitalisation of wastewater treatment plants is emerging as a promising strategy to open up the possibility of relying on digital tools capable of running simulations under varying operational and environmental conditions, enabling proactive plant management. Through the evaluation of alternative scenarios, operators could identify areas of improvement and vulnerabilities anticipating the real system. This could involve running simulations to identify optimal configurations that minimise emissions, energy consumption, or carbon footprint (Valverde-Pérez et al., 2021).
The ideal tool to accomplish these objectives is represented by the Digital Twin (DT). A DT can be defined as a model that can replicate the processes occurring in a physical plant almost in real time, receiving continuous data and automatically calibrating as conditions change (Torfs et al., 2022). For this reason, routine laboratory analyses and on-line sensors are often either too slow or too inaccurate for the development and operation of a DT (PUB, Singapore's National Water Agency, 2020; Yang, 2024). To address this issue, two complementary approaches must be taken: the development of soft sensors

and the implementation of an accurate characterization (Johnson et al., 2023).
Soft sensors, or virtual sensors, utilise data from reliable and easily accessible online sensors and, through data-driven or mechanistic algorithms, are capable of estimating process variables that would otherwise be too slow, costly, or impractical to measure directly (Haimi, 2016). In parallel, wastewater characterization provides, although for a limited period and at higher analytical cost, a detailed and dynamic picture of the influent composition and its transformation along the treatment line. These data are essential for calibrating the core model of the DT, ensuring that its predictions accurately reflect the real plant behaviour (Roeleveld and van Loosdrecht, 2002). One of the main focusses of this thesis is to contribute to this aspect by providing a comprehensive influent characterization to support model calibration and validation.

1.2 Aim and objectives of the study

The characterization campaign for the development of a DT mainly aims to obtain information on the fractions that make up organic matter and suspended solids that enter the plant. This information quantitatively describes the composition of the water, enabling calibration, an essential step in the construction of any model. Moreover, an additional sampling and analysis campaign was implemented to investigate the variation of the main nitrogen fractions across the activated sludge process. This profiling also contributed to the DT model calibration.
Furthermore, data obtained from the sampling campaign, along with other parameters provided by the WWTP operator Helsinki Region Environmental Services (HSY), were used to develop a prototype model of the primary clarifier (PC). Given that the current DT model starts only downstream of the PC, this study took the opportunity to lay the foundations for the potential future integration of this additional unit into the present DT model. This also offered the opportunity to test the data collected during this study in a modelling environment.

Based on these considerations, the present work addresses the following research questions:

Research questions

- What is the most effective way to characterize the influent for the digital twin modelling purpose?
- How do the data compare with the previous sampling campaign and with the routine analysis by HSY?
- Is it possible to create a representative model of the primary clarifier based on data collected during an influent characterization campaign?
- Which of the primary clarifier modelling approaches would be the most suitable in the given circumstances?

- How can influent characterization and primary clarifier modelling be integrated into the digital twin framework to improve process understanding and prediction accuracy?

1.3 Contribution and significance of the study

This thesis focuses primarily on the sampling campaign and characterization, providing useful data for both DT development and soft sensor modelling. Similarly to the work by Petäjä (2025), which provided the project with data on the composition of the primary influent and primary effluent during periods of warm and dry conditions, this work originally aimed to characterize the influent during the colder and snowy months. Therefore, in this manner, it would have been possible to obtain a representation of the boundary climatic conditions typical of the Nordic countries.

Although this thesis does not focus on biological processes and gaseous emissions, it is worth noting that one of the main objectives of the DIGICARBA project is to develop a tool that can accurately simulate GHG production at various treatment stages and in different scenarios and, consequently minimise them. The characterization of the influent carried out in this work directly supports this goal, since these parameters are fundamental for modelling biological nitrogen removal processes and predicting N_2O formation with greater precision.

Such direction also reflects the regulatory framework of the Urban Wastewater Treatment Directive (EU) 2024/3019 (European Union, 2024), which entered into force on 1 January 2025 and sets new and more stringent requirements for monitoring and reducing pollutants from urban WWTP. In particular, Article 21 mandates the quantification of GHG emissions, including at least CO_2, N_2O, and CH_4 from plants above 10,000 p.e. (population equivalent), using analyses, calculations or models, where appropriate. Therefore, works such as the one carried out by the DIGICARBA project become necessary, not only for scientific research purposes, but also to anticipate compliance with the evolving European legislation.

1.4 Structure of the thesis

This thesis is organized to guide the reader through the background, methodology, results, and discussion of the work. The Literature review chapter 2 provides the essential and relevant references needed to present the context and theoretical basis of the study. The Research material and methods 3.2 provide a detailed description of the sampling campaign, laboratory analyses, and target parameters, as well as the modelling approach adopted for the PC and the description of the wastewater treatment plant itself. The Results chapter 4 presents and discusses the findings, first from the laboratory analyses and then from the PC modelling. In the Discussion chapter 5, the results are interpreted and compared with both the previous sampling campaign and the routine analyses performed by HSY. Finally, the modelling of the PC is discussed in light of these findings.

Chapter 2

Literature review

2.1 Characterization of influent wastewater

For the development of this DT, a hybrid approach combining data-driven and mechanistic models was adopted. The data-driven component is used for predictive modelling, soft-sensor creation, and data pre-processing (Kiran, 2025). Once the data has been cleaned and validated, it is transferred to the mechanistic model, which is based on differential-algebraic equations. The mathematical models were implemented using SUMO, an open-source process simulation software developed by Dynamita (2022a). The platform is grounded on the Activated Sludge Models (ASM) proposed by the IWA Task Group on Mathematical Modelling for the Design and Operation of Biological Wastewater Treatment Plants (Henze et al., 2000). Optimal model performance requires accurate input data; however, the Task Group provides no specific guidelines to characterize the influent, and various research groups have proposed alternative approaches to address this limitation. In this study, the influent characterization was carried out following the COD fractionation method developed by the Dutch Foundation for Applied Water Research (STOWA). This method provides simple and reproducible guidelines for wastewater characterization, designed to support the development of reliable full-scale plant models (Roeleveld and van Loosdrecht, 2002).

2.1.1 Chemical oxygen demand fractionation

Chemical oxygen demand (COD) is an indicator that allows the organic matter content in wastewater to be quantified in terms of oxygen units (Chen et al., 2020), enabling mass balances to be calculated. COD fractions can be linked to sludge production, incoming loads and activated sludge operation in accordance with mass balances (Rieger et al., 2012). As shown in Fig. 2.1, total COD is typically subdivided, based on its physical and biochemical characteristics, into soluble (S) and particulate (X) fractions, as well as biodegradable ($_s$) and non-biodegradable ($_i$) components. The soluble biodegradable fraction (S_s) is considered readily biodegradable, while the particulate biodegradable fraction (X_s) is considered slowly biodegradable (Chen et al., 2020).

This fractionation is essential in process modelling, as it determines the rate and extent of substrate degradation and biomass growth within biological treatment units. In

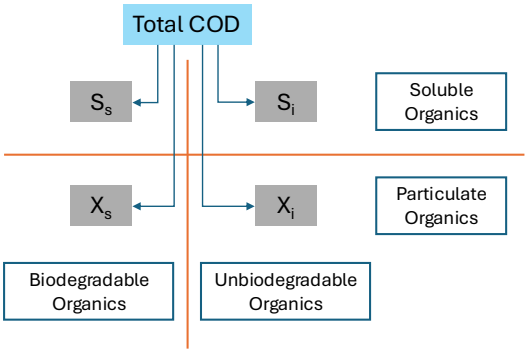


Figure 2.1: COD fractionation scheme. Figure adapted from Chen et al. (2020).

particular, the distinction between readily and slowly biodegradable COD enables a more accurate representation of substrate availability and microbial activity over time, which is crucial for reliable simulation and optimization of activated sludge systems (Chen et al., 2020).

Ideally, the biodegradable COD fractions consist of a soluble component (S_s) and a particulate component (X_s). During a wastewater treatment process, the soluble biodegradable fraction (S_s) is directly consumed during biomass growth, while a portion of the particulate biodegradable fraction (X_s) undergoes hydrolysis, forming S_s again, contributing indirectly to the biological processes. The remaining part of X_s, together with the inert particulate fraction (X_i), is expected to settle and end up in the waste sludge. Only the inert soluble fraction (S_i) passes through the activated sludge process and the secondary clarifier, ultimately reaching the secondary effluent. Based on these assumptions, Roeleveld and van Loosdrecht (2002) proposed the following expression:

$$S_i = 0.9 \cdot COD_{\text{eff,sol}}, \tag{2.1}$$

where the soluble inert fraction (S_i) is assumed to represent 90% of the soluble COD. This soluble fraction can be experimentally determined by flocculating and precipitating the colloidal material, and subsequently removing it together with the remaining particulate matter through filtration using a 0.45 µm filter (Mamais et al., 1993), and in this manner, even residual particulate or colloidal matter in the effluent sample, is removed. The detailed procedure for this measurement is described in the following section.

For high-loaded WWTPs, such as the one under investigation, Roeleveld and van Loosdrecht (2002) suggested refining the estimation by taking into account the biodegradable soluble fraction that still contributes to the effluent BOD_5. In this case, the effluent BOD_5 is subtracted, multiplied by a correction coefficient of 1.5, leading to the following formulation:

$$S_i = 0.9 \cdot COD_{\text{eff,sol}} - 1.5 \cdot BOD_{5,\text{eff}} \ . \tag{2.2}$$

Once the unbiodegradable soluble COD (S_i) is known, following the procedure proposed by Roeleveld and van Loosdrecht (2002), the readily biodegradable soluble COD (S_s) can be estimated as follows:

$$S_s = COD_{\text{inf,sol}} - S_i, \tag{2.3}$$

where (S_s) is calculated as the difference between soluble COD, measured in the influent, and unbiodegradable soluble COD fraction (S_i).

At this stage, the slowly biodegradable particulate fraction (X_s) can be estimated as

$$X_s = BCOD - S_s. \tag{2.4}$$

In this context, $BCOD$ denotes the total biodegradable organic fraction, which encompasses both the readily (S_s) and slowly biodegradable (X_s) components. The $BCOD$ value refers to the influent and is obtained as described below.

The last of the four fundamental COD fractions can be obtained as:

$$X_i = COD_{\text{inf,tot}} - (S_s + S_i + X_s). \tag{2.5}$$

This value is derived by difference, subtracting all previously determined fractions from the total influent COD ($COD_{\text{inf,tot}}$) . However, this method inevitably propagates the uncertainties and inaccuracies of the preceding estimations, particularly affecting this last fraction (Roeleveld and van Loosdrecht, 2002).

The biodegradable COD fraction (BCOD) represents another key indicator for influent characterization (Roeleveld and van Loosdrecht, 2002). It is composed of readily biodegradable soluble COD (S_s) and slowly biodegradable particulate COD (X_S). BCOD is estimated through biochemical oxygen demand analysis. Typically, BOD_5 is used, but it only reflects part of the biodegradable COD, since 50–95% is oxidised within five days. Although 95–99% may be reached after 20 days, BOD_{20} is considered unreliable and is therefore not recommended(Roeleveld and van Loosdrecht, 2002).

A more accurate approach is to follow the BOD progression over time and estimate the total BOD (BOD_{tot}) by fitting the measured data to the corresponding Eq. (2.6)Roeleveld and van Loosdrecht (2002) as shown in Fig. 2.2. The first-order rate constant k_{BOD}, typically ranging from 0.15 to 0.8 d^{-1} for municipal wastewater, can be obtained through linear regression using the least squares method.

$$BOD_{tot} = \frac{1}{1 - e^{-k_{BOD}\, t}}\, BOD_t. \tag{2.6}$$

As a result of biomass growth and decay during BOD measurements, the initial BCOD concentration is higher than the measured BOD_{tot}, and a correction factor f_{BOD} must be applied. A typical value of f_{BOD} is 0.15 (range $0.1 - 0.2$)(Roeleveld and van Loosdrecht, 2002). The corrected BCOD concentration is therefore calculated as:

$$BCOD = \frac{1}{1 - f_{BOD}}\, BOD_{tot}. \tag{2.7}$$

For illustrative purposes, those presented in Fig. 2.2, are the results of the BOD data from the water sample collected between 5:00 am and 7:00 am on the first sampling day (with duplicate). The values of f_{BOD}, BOD_{tot} and BCOD were determined by implementation of linear regression. The same procedure was applied to all BOD measurements taken during the campaign.

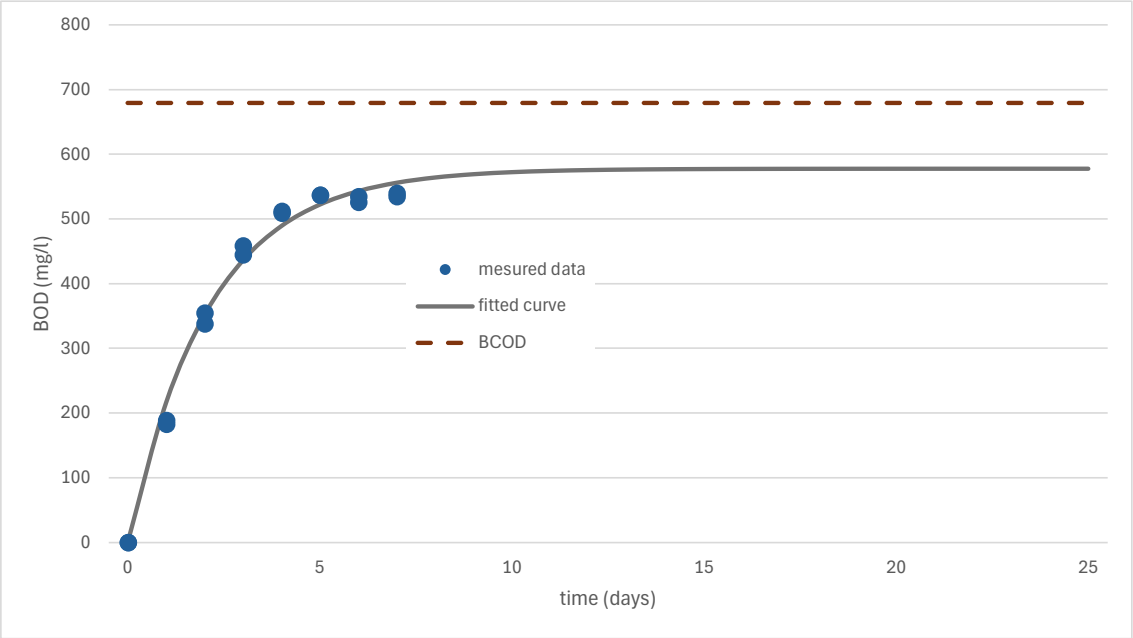


Figure 2.2: Fitted BOD curve for BCOD estimation (adapted from Roeleveld and van Loosdrecht (2002)).

2.1.2 Colloids

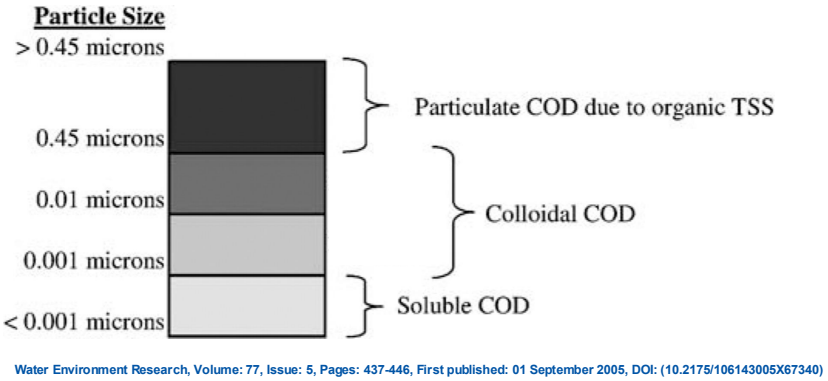
Colloids are defined as substances exhibiting physico-chemical characteristics intermediate between soluble and particulate matter (Chen et al., 2020). They are generally considered non-settleable, or at least require a very long time to settle. In practical wastewater treatment and modelling contexts, this subtle distinction is often overlooked: all material retained by a 0.45 µm filter is typically classified as particulate, while the portion passing through is collectively defined as soluble or colloidal (Chen et al., 2020).

However, when a more detailed distinction is required, as in the modelling framework proposed by *SUMO* and applied in this study, the colloidal material can be separated through a flocculation step followed by filtration ($ffCOD$)(Dynamita, 2022a). During this procedure, the addition of a coagulant causes the colloidal particles to agglomerate and settle at the bottom of the sample. The supernatant is then collected and filtered to remove any remaining suspended solids and residual colloids. Measuring the COD of this treated sample provides an estimate of the soluble organic matter ($ffCOD$). In contrast, the COD measured in samples subjected to simple filtration (without flocculation) represents the oxygen demand associated with both soluble and colloidal organic matter ($fCOD$). Based on these assumptions, the COD associated with the colloidal fraction can be estimated as:

$$colloidalCOD = fCOD_{\text{inf}} - ffCOD_{\text{inf}}. \tag{2.8}$$

In the influent characterization procedure proposed by *SUMO*, a 1.5 µm filtration is

adopted to obtain the input parameters of the model. In contrast, Jimenez et al. (2005),as shown in 2.3, identify 0.45 µm as the upper dimensional limit for colloids, matching the pore size of the filters used in the present work.



Water Environment Research, Volume: 77, Issue: 5, Pages: 437-446, First published: 01 September 2005, DOI: (10.2175/106143005X67340)

Figure 2.3: COD fractions based on particulate, colloidal, soluble distinction. (Adapted from Jimenez et al. (2005)).

2.2 Primary clarifier modelling

The primary clarifier is a fundamental unit of a WWTP that uses gravitational settling to separate suspended solids from the wastewater stream. Settleable solids with high organic content is removed in the form of sludge, ensuring optimal operation of downstream units and preventing clogging. However, Bachis et al. (2015) asserts that the role of the primary clarifier has been neglected in the modelling of WWTPs. Simple models, considered sufficiently robust, failed to capture the heterogeneous nature of the particles involved in this settling process, their behaviour, and the significant modification of COD fractionation that occurs across this unit (Bachis et al., 2015). These phenomena are of critical importance for whole plant modelling and resource recovery. Alternatively, models are developed from the primary effluent, but this means foregoing a broader understanding of the plant and the inputs of the model closely linked to the dynamics of the PC (Bachis et al., 2015).

The PC is not only crucial for enhancing the efficiency of subsequent treatment processes and improving the quality of the effluent before its discharge into the environment, but it is also an important unit for valorizing wastewater treatment by-products (Chen et al., 2020). Indeed, the anaerobic digestion of the primary sludge enables the production of biogas, which can be converted to thermal and electrical energy to support the operation of the plant (Chen et al., 2020). For these reasons, developing a model capable of accurately representing this unit and predicting its outputs can serve as a valuable tool to optimize both treatment performance and resource recovery strategies (Polorigni et al., 2021).

Polorigni et al. (2021), in comparing previous primary sedimentation models, points out that their main limitation lies in the fact that they are calibrated on the overall variations of TSS. This simplification neglects the underlying fractions of TSS identified as: unbiodegradable particulate organic (UPO), biodegradable particulate organic (BPO), and inorganic settleable solids (ISS), each of which displays distinct settling dynamics. Therefore, adequate fractionation of TSS and VSS, together with appropriate particle settling velocity distribution, are the foundation for realistic modelling of a PC.
Consistently with these considerations, this study employed organic matter, TSS and VSS fractions as input data for the primary clarifier model. Three different modelling options offered by the SUMO software were evaluated, each representing the sedimentation process with a different level of complexity. Their theoretical background is briefly outlined below.

2.2.1 Volumeless point separator

The Volumeless Point Separator (VPS) is an ideal non-reactive separator; it is based on an algebraic model and allows various combinations to be set as input parameters. In this unit, the user can define either the sludge concentration or a fixed underflow rate, while effluent characteristics can be specified in terms of solids and/or colloids removal efficiency either as percentages or fixed values (Dynamita, 2022b).

2.2.2 Layered flux model

The Layered Flux Model (LFM) divides the tank into horizontal layers that exchange solids with each other. Depending on the solids concentration in each layer, a different settling velocity regime is applied, which is governed by exponential equations. These equations are calibrated using parameters obtained from Vesilind's zone settling velocity lab tests. In this way, the model can predict the solids concentration in both the effluent and the sludge. Indeed, in addition to specifying the tank dimensions, the user has the option of entering the settling parameters. These are the maximum settling velocity (v_0) [$\mathrm{m\,d^{-1}}$] and the hindered settling parameter (r_{hin}) [$\mathrm{L\,g^{-1}}$], which govern the Vesilind equation

$$v_s(X) = v_0 \cdot e^{-r_{\mathrm{hin}} \cdot X}, \tag{2.9}$$

where (v_s) is the sedimentation rate at concentration X of suspended solids (Dynamita, 2022b).

2.2.3 Three compartment model

The Three-Compartment Model (TCM) conceptualises the primary clarifier as three interconnected zones (feed well, clean water, and sludge blanket) connected by a VPS as shown Fig. 2.4. The feed well represents the inlet zone where the influent flow is evenly distributed and the initial mixing and energy dissipation take place, promoting a more uniform settling process. The clean water compartment represents the clarified liquid layer above the sludge blanket, accounting for the elutriation flow, i.e., the natural or forced circulation of liquid that allows the diffusion of soluble components from the sludge blanket into the overlying water. Each compartment is modelled as a CSTR and the clarifier performance can be tuned by adjusting their relative volumes. It is also possible to define underflow specifications, such as sludge flow rate or concentration, solids removal efficiency, and the increase in VSS/TSS ratio. Any dosing of polymers would be done in the feed well unit. Feed well and sludge blanket are necessarily reactive zones in this model. (SUMO, model-specific documentation (Dynamita, 2022b)).

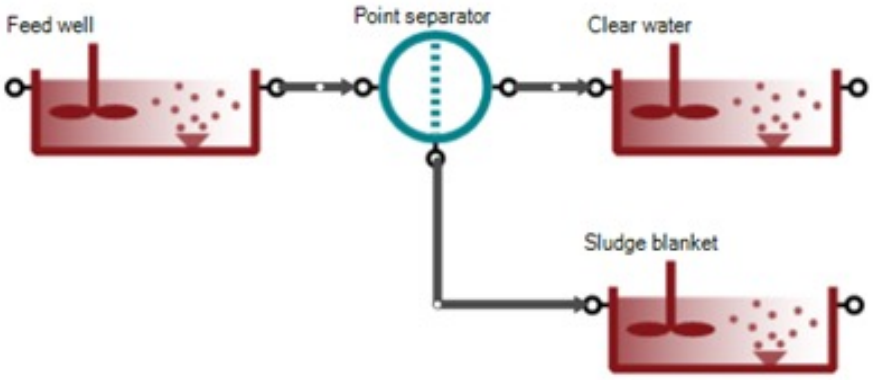


Figure 2.4: Three compartment model scheme. Figure adapted from Dynamita (2022b).

Chapter 3

Materials and Methods

3.1 Description of the Viikinmäki wastewater treatment plant

The Viikinmäki wastewater treatment plant serves Helsinki and seven of its neighbouring municipalities. The wastewater is sourced 15% from industrial sites and 85% from domestic use.
Central Helsinki, which includes the historic and densely built-up areas, operates with a combined sewer system in which wastewater and stormwater are conveyed through the same network to the Viikinmäki treatment plant. This configuration contributes to significant variations in influent flow and composition, particularly during rainfall events and snowmelt events.
According to the HSY annual report on wastewater treatment in the Helsinki Metropolitan Area 2024 (Urho et al., 2025), the Viikinmäki wastewater treatment plant treats wastewater from approximately 930,000 inhabitants of Helsinki and surrounding municipalities. The equivalent population covered by the plant is estimated at 1.18 million considering a load contribution of 70 g of $BOD_{7(ATU)}$/person. (Urho et al., 2025) The same document indicates an average daily flow of approximately $290\,000\,m^3\,d^{-1}$, with peak values reaching up to $700\,000\,m^3\,d^{-1}$ (Urho et al., 2025).
The treatment process, as outlined in Fig. 3.1, integrates both chemical and biological stages. Phosphorus removal is achieved through parallel chemical precipitation with ferrous sulphate, dosed at multiple stages of the process. Nitrogen removal follows a two-step configuration: Denitrification-Nitrification (DN) in the activated sludge process, complemented by post-treatment in biological filters. Lime is added to maintain sufficient alkalinity during nitrification, while methanol is supplied to the post-denitrifying filters as an external carbon source to enhance post-denitrification process. Organic pollutants (BOD) are partly eliminated during primary sedimentation by solid separation and further degraded biologically in the activated sludge stage.
Viikinmäki WWTP is largely located underground in excavated rock caverns. The underground construction enables to avoid freezing winter temperatures maintaining more stable process conditions. The treated effluent is discharged into the Gulf of Finland through a 16 km tunnel, the outlet being situated approximately 8 km offshore from

southern Helsinki at a depth exceeding 20 m. (Urho et al., 2025)

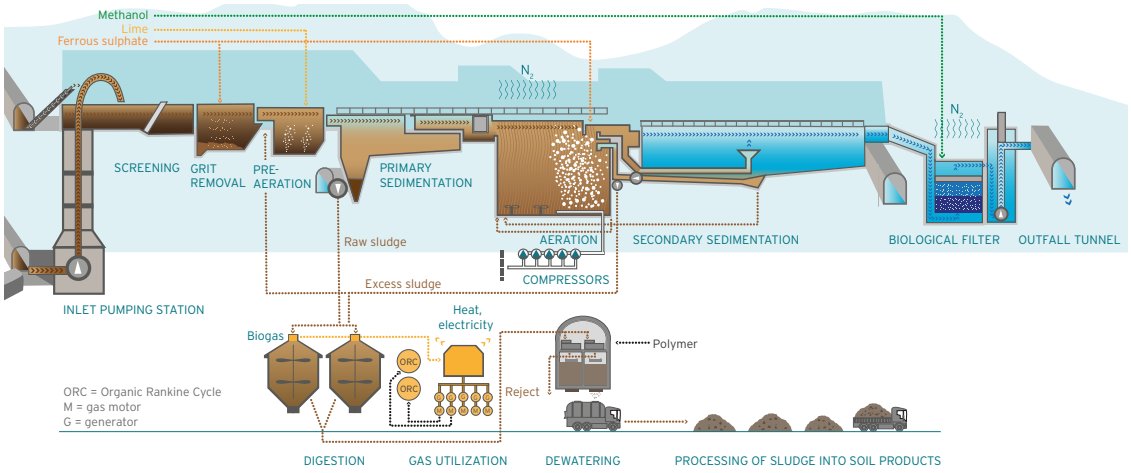


Figure 3.1: Scheme of the Viikinmäki WWTP. Adapted from HSY (2017).

3.2 Sampling campaign

The sampling campaign was conducted over six weeks between March and April 2025; see Fig. 3.2. In this period, seven sampling days were selected according to specific criteria, i.e., avoiding sampling on the same day of the week and including weekends to ensure that the collected data represented a range of operational and environmental conditions. The selection of days was also influenced by a number of practical limitations, such as the availability of personnel and equipment. In order to ensure the comparability of those data with those collected by Petäjä (2025), an attempt was made to maintain methodological consistency between the two works whenever possible. For the seven designated sampling days, three automated samplers were used to simultaneously collect samples at various stages of the plant. The samplers, which were equipped with 24 plastic bottles in the container base, were programmed to collect four samples of fixed volume per hour per bottle over a 24-hour period. It was necessary to find a compromise between the sampling rate and the volume sampled, with the objective of maintaining adequate tracking of events in the flow and avoiding the use of too small and error-prone volumes. In addition, the final volume had to be large enough to carry out all planned analyses.
On sampling days, the samplers were activated between 07:30 and 08:00 a.m., and 24 hours later, the samples were collected. At the time of sample collection, the contents of each third consecutive bottle were mixed, thus reducing the number of bottles from 24, each representing one hour, to 8 composite samples, each representing three hours of flow.

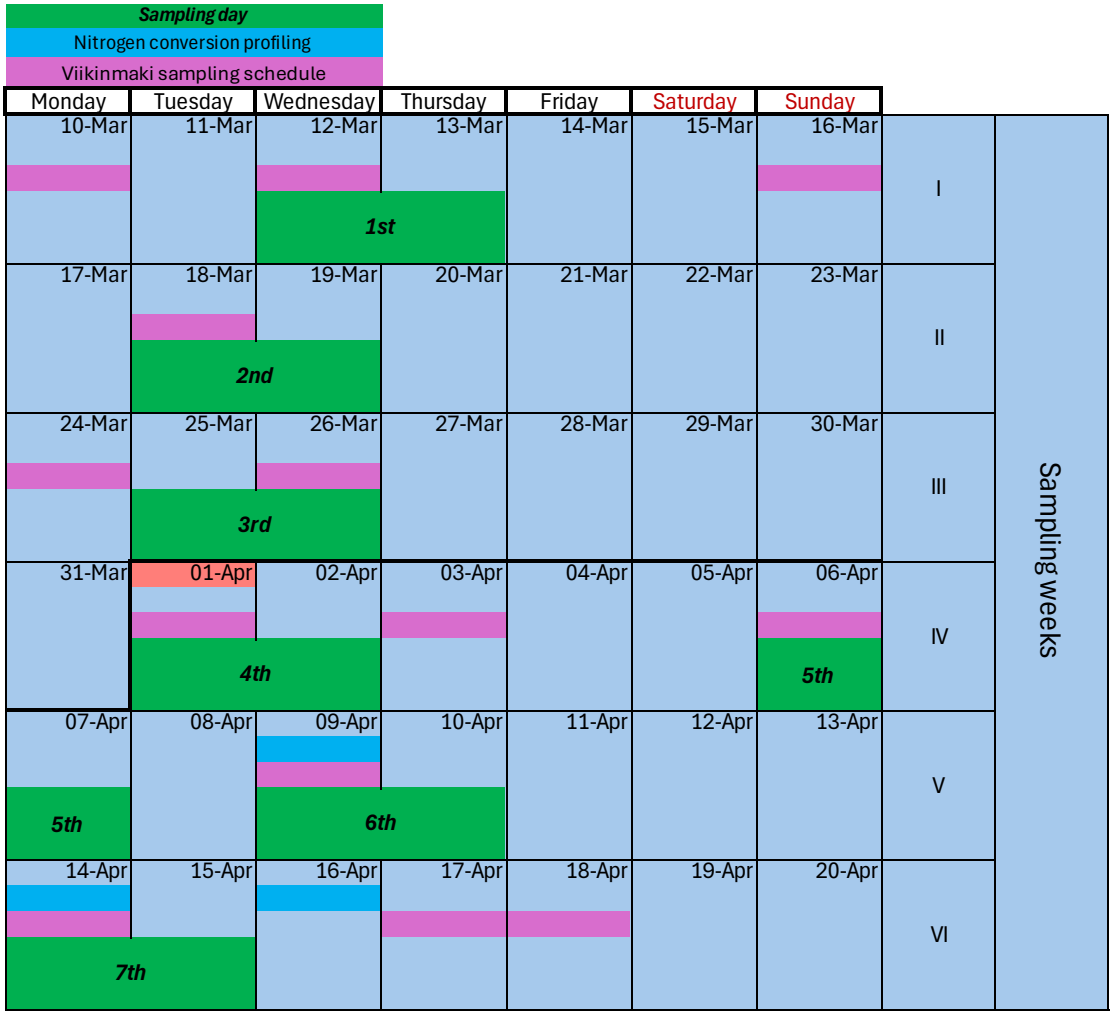


Figure 3.2: Samplings schedule. The 7 sampling days are highlighted in green, each slot represents the two-day period during which the 24-hour sampling was carried out. Blue marks the three days of sampling for the nitrogen conversion profiling. Magenta corresponds to the days when the Viikinmäki environmental monitoring campaign was carried out by HSY during the period covered by this study. All dates refer to the year 2025.

In the samplers used (Figures 3.3), a peristaltic pump is operated by a computer according to the user's settings. A suction tube connected to the peristaltic pump was submerged in the wastewater, while a mechanical arm directed the flow into one of the 24 bottles in the collection base, depending on the setup. All samplers were configured to fill a bottle every hour, with a fixed volume of water being pumped every 15 minutes. To obtain a precise characterization of the influent and understand the impact of different processes on wastewater quality, samples must be taken at various stages of the treatment

process.

The **Isco 3700 Portable Sampler** (Figure 3.3 (a)) was used to take a sample of the rawest wastewater, it was placed immediately downstream of the pumping station and upstream of the bar screens section, which is henceforth called the primary influent (PI). The wastewater leaving the PC designated as primary effluent (PE) was sampled using the **Isco 6712 full-size portable sampler** (Figure 3.3 (b)). The third sampler, dedicated to the sampling of wastewater leaving the secondary clarifier, Secondary Effluent (SE), was the **Sigma 900 Composite Sampler from Hach** (Figure 3.3 (c)). Both Isco samplers were equipped with 24 one-liter bottles each. The capacity of the Sigma bottles was 0.5 liters. The three samplers were all equipped with a perforated weight at the end of the pipe, ensuring that it remained below the water level even during periods of high flow and also preventing clogging caused by coarse material. To minimise contamination between consecutive samples, the samplers were programmed to automatically rinse the suction tube before each sampling event. The samplers were placed so that the end of the pipe, from which the wastewater was collected, was approximately 50 centimetres below the surface.
It should be noted that due to a battery failure of the ISCO 3700 sampler on the third day of the campaign, the data collected that day are incomplete. Therefore, where required, six days of samplings were considered, instead of seven.

(a) Isco 3700 Portable Sampler used for sampling PI.

(b) Isco 6712 full-size portable sampler used for sampling PE.

(c) Sigma 900 Composite Sampler from Hach used for sampling SE.

Figure 3.3: Samplers used during the sampling campaign at the Viikinmäki WWTP.

3.3 Laboratory analysis

This section describes the analytical procedures adopted to determine the concentrations of the main target parameters selected for influent characterization and model calibration.

3.3.1 Chemical oxygen demand (COD)

The analysis of COD provides a reliable and relatively rapid indicator of the amount of organic matter in the sample. The standard method Suomen Standardoimisliitto SFS (1986) involves the use of potassium dichromate $K_2Cr_2O_7$ as an oxidizing agent and the titration process was carried out using a solution of ammonium iron (II) sulfate $(NH_4)_2Fe(SO_4)_2\,6\,(H_2O)$. The titration method is used to determine the excess oxidizing agent that has not contributed to the degradation of organic matter. From this value, the COD concentration expressed in milligrams per liter (mg/L) is derived.
In addition to total COD, filtered COD (fCOD) and filtered flocculated COD (ffCOD) were measured from each sample. To obtain filtered COD, it is necessary to filter the samples through an acetate cellulose syringe filter with a pore size of 0.45 µm. These membranes retain biodegradable and non-biodegradable particles while allowing soluble matter to pass through.
ffCOD is obtained by subjecting the sample to a flocculation process: first, a zinc sulfate solution ($ZnSO_4\,7\,H_2O$) is dosed as an oxidizing agent and then the pH is raised to 10.5 for optimal flocculation. The supernatant of the settled sample is then collected using a syringe and filtered as described for fCOD.
Before use, syringe filters were rinsed by filtering approximately 5 mL of reverse osmosis water.
The analytical accuracy of COD measurements varied depending on the concentration range, as specified by the standard method:

- for results $\leq$ 50 mg/L COD_{Cr}, the accuracy was $\pm$1 mg/L;
- for 50 mg/L $<$ result $\leq$ 100 mg/L COD_{Cr}, the accuracy was $\pm$5 mg/L;
- for 100 mg/L $<$ result $\leq$ 1000 mg/L COD_{Cr}, the accuracy was $\pm$10 mg/L;
- for results $>$ 1000 mg/L COD_{Cr}, the accuracy was $\pm$100 mg/L.

3.3.2 Biological oxygen demand (BOD)

The analyses of biochemical oxygen demand (BOD) were performed following the standard ISO (2019).
BOD is a measure of the amount of oxygen required by microorganisms to degrade organic matter in a water sample. The duration of this process can vary; in this study was measured the amount of O_2, expressed in milligrams per liter (mg/L), consumed during a 7-day period at a constant temperature of 20 degrees in the dark.

However, due to historical and geographical differences in standard practices, the 5-day test (BOD_5) has traditionally been more widely adopted, while Nordic countries have commonly used the 7-day version (BOD_7). As the incubation time increases, a greater

portion of the biodegradable organic matter is degraded, leading to a more accurate estimate of the total oxygen demand.(Chen et al., 2020). Allylthiourea (ATU) was added to inhibit nitrification and its subsequent oxygen demand, so that only carbonaceous BOD was recorded.
Due to the limited availability of analytical devices, it was not possible to measure BOD for all samples. At least four of the 8 PI samples were always measured, in addition to one sample made by a combination of all daily SE samples, with a duplicate. When feasible, duplicates of the PI samples were made and the BOD of the PE was measured. However, given the tendency of this measurement method to result in unreliable readings due to potential air infiltration or alterations in the metabolic activity of microorganisms, all available measuring devices were consistently used.
For PI samples, the BOD bottles were filled with a volume of 164 ml, which is optimal for a measurement range between 0 and 400 mg/L. In contrast, SE samples require a maximum permissible volume of 432 ml to detect low BOD values, ranging from 0 to 40 mg/L. Two models of the BOD measuring system with similar characteristics were used: Ox-iTop®-i and OxiTop®-C.

3.3.3 Total suspended solids and volatile suspended solids

The standard Suomen Standardoimisliitto SFS (2005) was used to determine the concentration of total suspended solids (TSS) and volatile suspended solids (VSS) of the samples, expressed in milligrammes per liter (mg/L). In accordance with this method, prewashed glass-fibre filters (Whatman GF/A) are used to perform a vacuum filtration of a variable sample volume. This volume variability is due to the need to find a compromise between a large volume that would well represent the sample, and the risk of clogging the filter invalidating the analysis. The volume of the filtered PI sample varies between 12 mL and 50 mL due to the substantial variability in the amount of solids in the sample depending on the time of day under analysis. In contrast, the filtered sample volume for PE and SE rarely deviates from the mean value of 50 and 250 millilitres, respectively.
Considering the elevated quantity of solids present in PI samples, the recommended filtration duration of 1-2 minutes was not feasible. After drying the filters at 105 °C, the TSS concentration is determined by subtracting the initial weight of the filter and dividing by the filtered volume.
Since VSS are thermally degradable, they are estimated as the mass defect after incineration of the filters in the furnace at 550 ° C. Fig. 3.4 shows how filters appear after this process. However, as reported by Petäjä (2025), a proportion of the glass fibre filter itself also undergoes a weight reduction during the incineration and drying process. Consequently, a positive weight correction was applied to compensate for this loss. The values estimated by the laboratory staff are 713 µg and 212 µg for incineration and drying, respectively.

Figure 3.4: Some of the filters in the crucibles after incineration in the furnace at 550 °C.

3.3.4 Flow rate weighted concentration calculation

Given that the samplers used in this study took a constant volume of water for each designated time interval, subsequent analyses were able to provide only absolute concentrations of target substances. Although these results are suitable for time patterns and qualitative concentration analyses, they do not represent actual loads, especially under variable flow conditions such as those that are being analysed.
Based on the hourly flow rates, the measured concentrations of COD, BOD, TSS and VSS were converted into flow-weighted values. Each concentration was weighted by the corresponding flow rate and subsequently normalised to the 24-hour average flow rate of the respective sampling day, as shown in the following equations:

$$C_i^* = C_i \cdot \frac{Q_i}{\overline{Q}}. \tag{3.1}$$

Daily averages of the flow-weighted concentrations were then computed as:

$$\overline{C}^* = \frac{1}{n} \sum_{i=1}^{n} C_i^*. \tag{3.2}$$

where

- C_i = concentration measured in the i-th time slot
- Q_i = wastewater flow rate in the i-th time slot
- $\overline{Q}$ = daily average flow rate
- C_i^* = flow-weighted concentration for the i-th time slot
- $\overline{C}^*$ = daily average flow-weighted concentration
- $n = 8$, corresponding to eight 3-hour time slots per day

3.3.5 Nitrogen conversion profiling

Nitrogen conversion profiling was performed to provide the research team with accurate data on the actual behaviour of nitrogen components in the activated sludge process. Nitrogen conversion profiling is defined as the analysis of changes in the concentration of nitrogen chemical species as the activated sludge process progresses.
An analysis of the concentrations of ammonium (NH_4^+), nitrate (NO_3^-) and nitrite (NO_2^-) was carried out in each of the six zones of the Denitrification-Nitrification (DN) process, which are indicated by white numbers in Fig. 3.5, during each sampling round. A total of five rounds were conducted; on the first day samples were collected only in the morning, while on the following two days, samples were taken both during morning and afternoon. These three days are marked in blue in the table 3.2. The initial and sole sampling of 9 April 2025 was conducted on line 2, while the remaining four were carried out on line 9. In order to prevent any alteration of the samples due to the continuation of the nitrification-denitrification processes, the analyses were carried out in situ immediately after sampling.

Nitrite (NO_2^-)

Following the standard SFS 3029 (1976), the concentrations of NO_2^- were determined. The method is outlined in the following sequence of steps:

- Filtering the sample through a 0.45 µm filter
- Dosing of 125 µL of sulfanilamide (SA) solution in 5 ml of filtered sample or its dilution.
- Dosing of 125 µL of N-(1-Napthyl)ethylenediamine (NED)

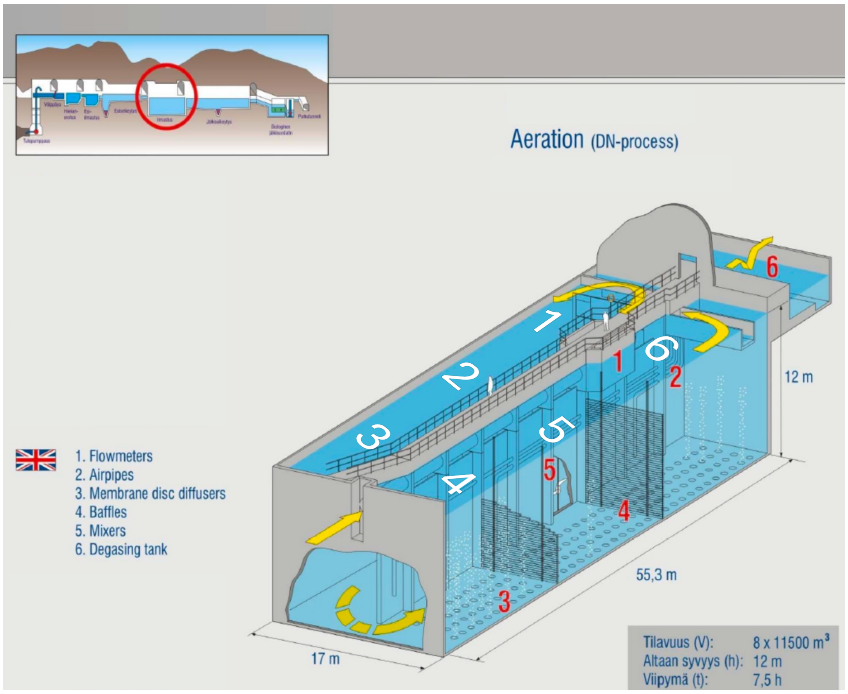


Figure 3.5: Schematic diagram of the activated sludge process. The white numbers represent the six zones through which the sludge flows and the Denitrification-Nitrification (DN) process takes place. Figure adapted from HSY (2017).

- Measuring the absorbance at 545 nm by UV spectrophotometer (Shimadzu UV-VIS Spectrophotometer UV-1201) after leaving the samples in the dark for a period of 20 minutes to two hours.

The NO_2^- content in the sample was determined according to Eq. 3.3. If the sample was diluted prior to analysis, the dilution factor was applied to derive the concentration of NO_2^- in the raw wastewater sample.

$$C_{NO_2^-} = 294.1 \times \text{abs}^{-0.794} \tag{3.3}$$

where:

- $C_{NO_2^-}$ is the concentration of nitrite in the sample, expressed in [µg L^{-1}].
- abs is the absorbance measured at 545 nm (range 0–1.8) [–].

Nitrate (NO_3^-)

Following the standard procedure APHA 4500-NO_3^- (2005), a nitrate analysis was performed. To briefly describe the procedure, each sample was filtered through a 0.45 µm membrane filter, and 10 mL of the filtrate were transferred into a volumetric flask. Then, 0.2 mL of 1 M hydrochloric acid were added. The absorbance was measured using a

UV–VIS spectrophotometer (Shimadzu UV-1201) at a wavelength of 220 nm to determine the NO_3^- concentration, and the result was corrected for organic matter interference by subtracting the absorbance at 275 nm, as shown in Eq. 3.4.

$$\text{abs}_{NO_3^-} = \text{abs}_{220\,\text{nm}} - 2 \cdot \text{abs}_{275\,\text{nm}} \quad \text{with} \quad (2 \cdot \text{abs}_{275\,\text{nm}} < 0.1\,\text{abs}_{220\,\text{nm}}) \tag{3.4}$$

Ammonium (NH_4^+)

Ammonium concentrations in activated sludge samples were measured in the six aeration zones using an ion-selective portable meter (Orion Model 250A) electrode. The procedure was performed as follows:

- Standard solutions of 4 and 40 mg/L N were used to plot a calibration curve on a semi-logarithmic scale, with the mV reading on the x-axis and the logarithm of ammonium concentration on the y-axis.
- For each of the six samples, 0.5 mL of ionic strength adjustor (ISA) solution was added and the electrode was placed in a magnetically stirred beaker to ensure homogeneity.
- Stable mV readings were then recorded and the ammonium concentration was read by interpolating the values on the calibration curve.

3.4 Primary clarifier modelling

This section describes the development of the prototype model of the primary clarifier, detailing its configuration in SUMO, the data sources used for calibration and validation, and the approach adopted for assessing model performance.
The modelling work was carried out using the wastewater treatment process simulation software SUMO (Dynamita, 2022a) (version 22.1.0; model SUMO1). The model was designed to investigate the influence of the primary clarifier on the concentrations of total COD, its fractions, TSS, and VSS. To calibrate the model, the dataset obtained from the present sampling campaign was combined with dynamic flow rate measurements and with the physical and operational characteristics of the system provided by HSY. Model validation was then performed using an independent dataset collected over three weeks in June 2024 during Petäjä (2025)'s sampling campaign.
The performance of the model was evaluated using the Mean Absolute Percentage Error (MAPE) between observed and simulated data, calculated as follows:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3.5}$$

where y_i are the observed values, $\hat{y}_i$ the predictions of the model and n is the number of experimental data points. It expresses the average percentage deviation between the measured and simulated data. The metric was computed both during calibration, to

assess the model fit to the estimation dataset, and during validation, to quantify its predictive ability on independent data.
In order to represent the PC in a way that is consistent with the actual plant, the model was developed as shown in Fig. 3.6. The process units included are described below:

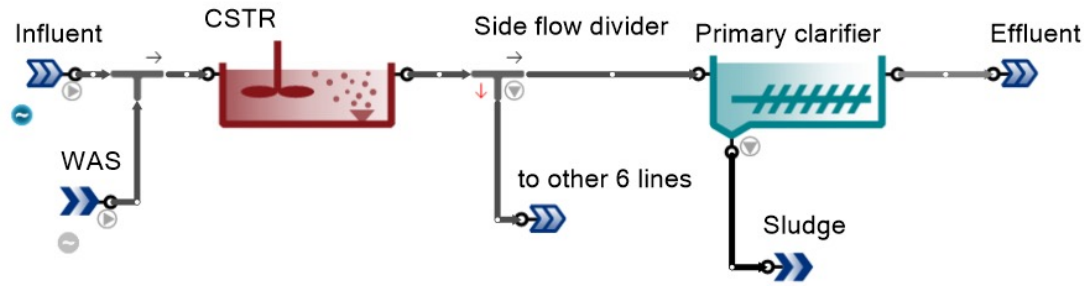


Figure 3.6: model of the plant segment that comprises the primary settler.

Influent: This unit corresponds to the point where the PI sampler is positioned, thus immediately downstream of the pumping station.
During the calibration phase, all dynamic parameters collected during the sampling process were entered through the influent configuration.
For both calibration and validation, all dynamic parameters obtained from non-consecutive sampling days, as illustrated in Figure 3.2, were combined to form a single continuous time series.
SUMO requires the data to be entered in a specific way in order to uniquely characterise the influent without having redundancies. Specifically, the data provided were:

- Total COD ($g\,m^{-3}$) from PI
- Fraction of filtered COD in Total COD from PI
- Fraction of filtered flocculated COD in Total COD from PI
- Fraction of soluble unbiodegradable organics in filtred COD[1]
- Fraction of particulate unbiodegradable organics in total COD
- VSS and TSS ratio

The dynamic flow rate is also provided to the software in the influent unit, allowing it to calculate the actual load of the incoming substance. The average value during this sampling period was about $260\,000\,m^3\,d^{-1}$.

[1]SUMO requires filtered COD to be obtained by filtering the samples through 1.5 µm filters; however, 0.45 µm filters were used in the laboratory. To address this inconsistency, a correction factor of 1.16 was estimated by Sieranen (2024) and applied.

Waste activated sludge (WAS): Part of the activated sludge, separated from the secondary sedimenter, is recirculated in the aeration tanks (RAS), the excess part (WAS), on the other hand, is mixed in the pre-aeration tank with the raw sludge and removed from the PC.
While in the actual plant, the WAS flow rate is manipulated to control the sludge retention time and ensure optimal biomass sludge concentration in aeration basins, in the model a constant value of $4500\,m^3\,d^{-1}$ is established. SUMO allows to specify the type of input sludge, to automatically define default parameters. In this case the 'Thickened waste activated sludge' option was chosen and the available informations were provided: VSS/TSS fraction 69% and Total suspended solids $7000\,g\,m^3$.

Pre-aeration tank (CSTR): The pre-aeration tank is 70 m long, 18 m wide and 7.7 m deep. It has a total volume of $9700\,m^3$ and a hydraulic retention time of about 50 minutes. In the plant, the pre-aeration basin serves to equalise water quality by mixing raw water with WAS, reject water from sludge dewatering unit, ferrous sulphate and lime. Submerged aerators at the bottom are also activated for 10 minutes every two hours to prevent premature settling. In the model, this unit is represented by an anaerobic CSTR of equal size. In this way, the longitudinal mixing effect can be reproduced, dampening the peaks of substance concentration entering the PC.

Side flow divider: The Side Flow Divider ensures that the correct water flow rate is directed to the single primary clarifier line considered in the model, as only one of the seven identical units is simulated.

Primary clarifier: The PC is composed of twin tanks arranged in parallel, each with a surface area of $525\,m^2$, 4.7 m deep, 69 m long and 8.5 m wide. The surface load is $1.4\,m\,h^{-1}$ for average flow rates up to $3.4\,m\,h^{-1}$ at high flow rates. Hydraulic retention time is 3 - 3.5 hours. The thickening pocket, where the sludge is dragged by the scrapers, has a depth of 15.7 meters. This raw sludge is removed and sent to anaerobic digesters. Similarly, skimmers remove the floating sludge.
SUMO offers three modelling configurations for PC hydraulics with increasing levels of modelling complexity, all of them were tested in this work:

- Volumeless point separator
- Layered flux model
- Three compartment model

Implementation details are provided below; theoretical description of the models are discussed in chapter 2.2.

Effluent: This conceptual point corresponds to the physical location of the second sampler, at the PC effluent. The activated sludge process begins straight downstream.

3.4.1 Volumeless point separator

The VPS unit is a non-reactive separator based on an algebraic model, allowing the specification of underflow sludge concentration or flow rate and the definition of effluent solids removal. The model was calibrated by slightly adjusting the suspended solids removal efficiency. Although the measured average removal efficiency was about 70%, a value of 75% provided a better fit between simulated and observed effluent concentrations, compensating for the simplifications of the VPS model. A sludge flow rate of $800\,m^3\,d^{-1}$, corresponding to the plant's average operating conditions, was used in the simulations. Furthermore, as no significant increase in the VSS/TSS ratio was observed in the effluent, this parameter was set to 0%.
During calibration, the pre-aeration tank volume was increased by a factor of ten to improve agreement between simulated and observed effluent characteristics. This adjustment resulted in a volume approximately three times larger than that of the primary clarifier, suggesting that it compensates for additional behaviours not captured by the simplified VPS model. In particular, this empirical correction likely accounts for effects such as hydraulic equalisation, internal dead zones or unmodelled recirculation dynamics that occur in the real system. Consequently, the adjusted volume should be interpreted as a modelling proxy for these combined effects rather than as a direct representation of the physical tank volume.

3.4.2 Layered flux model

In this study, the LFM was implemented as a non-reactive unit since the focus of this work is not on nutrient removal but on the difference in suspended solids and COD fractions before and after the clarifier. Since Vesilind settling velocity tests were not available, these parameters were maintained as default data. The previously mentioned dimensional data specific to the PC were used.

3.4.3 Three compartment model

Following Amin (2024), the implemented three compartment model included a feed well and sludge blanket with respective volumes of $250\,m^3$ and $1250\,m^3$, accounting for 10% and 50% of the total volume. These volumes were slightly adjusted to achieve an optimal calibration fit. The increase in VSS/TSS was set at 0%.

3.4.4 Simplifications and assumptions

In order to keep the focus on the physical wastewater treatment that occurs in the PC and to avoid increasing the complexity of the model, the following simplifications were made.
To simplify the model, the chemical and biological reactions affecting nitrogen and phosphorus dynamics were not explicitly considered, although the analysis of COD and solid fractions indirectly reflects their behaviour during sedimentation.
The dosing of ferrous sulphate (Fe_2SO_4), which in the real plant takes place in the pre-aeration basin, was not included in the model. Indeed, in SUMO, this chemical agent does

not influence the sedimentation of solids in general, but only the chemical precipitation of phosphorous, which was not taken into account in this study. Similarly, in the model option section, it was selected to ignore the pH calculation, since no detailed modelling of chemical reactions or monitoring of pH and alkalinity was required.
Although SUMO provides the option to simulate polymer dosing in the primary clarifier, in Viikinmäki WWTP polymers are only dosed when, under high-flow conditions, biological treatments must be partially bypassed (Haimi, 2016). Therefore, no polymer dosing is applied in the model.
It should also be mentioned that in the actual plant segment considered in the model, the screening, grit removal, and recirculation of water from the sludge dewatering processes, also take place. The role of these units was deemed negligible for the purposes of the investigation.
All simulations were initialised at steady state in order to allow direct comparison between the different scenarios, avoiding the need to discard the initial transient phase before reaching equilibrium.

3.5 Data received from HSY

To construct the digital twin, the DIGICARBA team was provided with access to historical Viikinmäki data by HSY. The dataset under consideration comprises measurements from the various sensors along the treatment line (e.g. flow rate, pH, temperature, dissolved oxygen) as well as laboratory data from the environmental permit monitoring campaign (e.g. BOD, COD, SS) at various stages of the process. The data from the online sensors are available at one-hour intervals, whereas the laboratory data are daily representative and are measured by flow-paced samplers only on predefined days (see Fig. 3.2).
In this work, data were requested from HSY with the aim of obtaining a reference for the laboratory analysis data, supporting the construction of the PC model and to calculate the input loads of the different components with hourly flow rates.

Chapter 4

Results

This chapter presents the main results of the influent characterization and primary clarifier modelling carried out during the study. The first part focusses on the analysis of measured concentrations of target parameters, as well as the outcomes of COD fractionation analysis, while the second part shows the outcomes of the model via the three configurations.

4.1 Laboratory analysis results

4.1.1 Chemical oxygen demand results

The following sections present the results obtained measuring total COD, filtered COD and filtered flocculated COD. Particular attention is paid to the daily dynamics of total COD and its fractions.
Figure 4.1 shows the hourly variation of the concentration of COD, fCOD and ffCOD through PI, PE, and SE. Each parameter is represented through box plots made up of data collected during the six sampling days, to highlight intra-hour variability, central tendency, and possible outliers. These values are weighted according to the flow rate; this allows the evaluation of the actual COD load at a given stage of the plant in each time interval, as expressed in the equation 3.1.
In the top-left panel, the morning peak in COD_{tot} is particularly evident. This is due to the way the plant operators carry out the pumping. In fact, during the early hours of the day, when the load should be minimal, pumping is gradually increased to prepare the tunnel for the typical diurnal peaks of domestic patterns. This early morning pumping, under low flow conditions in the influent tunnel, sends wastewater with a high organic and inorganic load, previously deposited, to the plant. The difference between COD and fCOD confirms the abundant presence of particulate matter in wastewater, especially in the morning hours, due to the process described above. The distinction between fCOD and ffCOD, which is less clear but nevertheless discernible, is indicative of the presence of colloidal material.
To understand the second panel describing what comes out of the primary sedimentation tank, it is important to note that there is a hydraulic retention time of approximately

4.5 hours between the first and second sampling points. It can be seen how the PI fCOD is similar to the PE CODtot but shifted one time step to the right. This shows how the clarifier has the effect of settling the particulate matter that was removed by filtration in the PI analysis. However, there is still a difference between total COD and fCOD, but it is considerably reduced in comparison to the first panel. Analysis of the ffCOD results, reveals a lack of significant variations, confirming that the colloidal fraction does not undergo substantial removal during this phase.
Finally, in the secondary effluent (third panel), a further general reduction can be observed in all fractions. Both the median values and the variability decrease. The COD, fCOD, ffCOD gradient is more compressed and scarcely visible, suggesting a progressive removal of both particulate and soluble organic substances thanks to the combined action of biological processes and secondary sedimentation.
The uncertainty of the COD measurement method is discussed in Section 3.3.1.

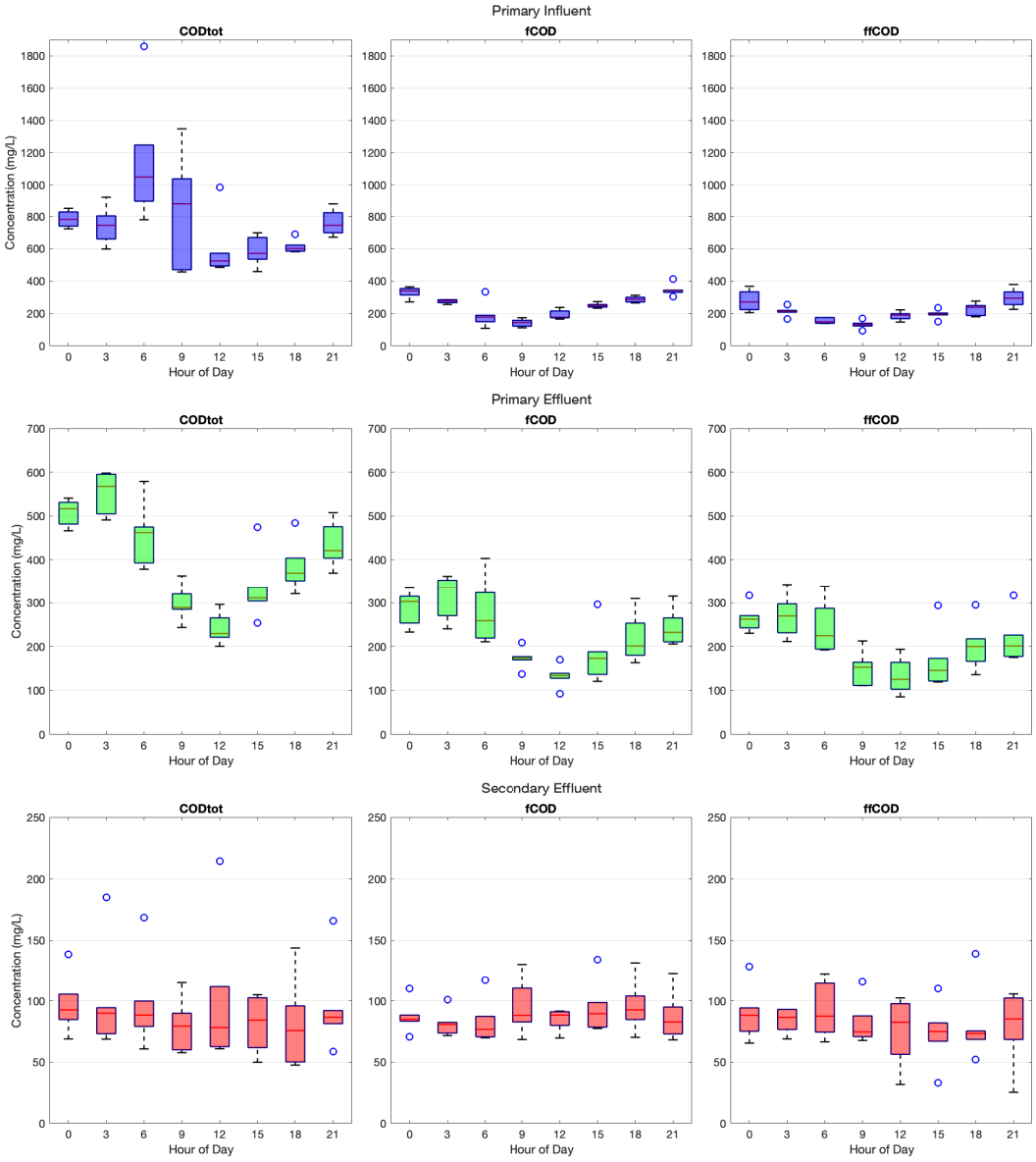


Figure 4.1: Flow-weighted COD measurements (total, filtered, and filtered–flocculated), at different stages of the treatment process (primary influent, primary effluent, secondary effluent). Box plots represent data collected over six sampling days.

4.1.2 Biological oxygen demand results

The box plot in Fig. 4.2, show the diurnal variability of biodegradable COD (BCOD) of the PI. Each box represents the flow-weighted BCOD distribution relative to a specific time interval across the six sampling days. The graph highlights a temporal pattern, with higher BCOD loads typically occurring between 21:00 and 3:00, coinciding with peak flow rate.
It should be noted that the accuracy of this analysis is affected by the time elapsed between the sampling and filling of the BOD bottles. Since the implemented setup involved the use of non-refrigerated samplers over a 24-hour period, it is to be expected that the samples collected first may present low accuracy and underestimation of the measurement issues, as organic matter is biodegraded continuously before the collection. Fig. 4.2 reveals that the BCOD for 9 o'clock, which corresponds to the first samplings of the day, has a maximum inter-quartile range. In contrast, the BCOD of 6 o'clock, corresponding to the last samplings before collection, shows a narrower inter-quartile box.
Regarding SE, BOD_7 measurements were performed to estimate the residual biodegradable organic matter downstream of the secondary settler, to support the calculation of the soluble inert fraction (S_i), as suggested by the STOWA guidelines (Roeleveld and van Loosdrecht, 2002). The average BOD_7 measured across all sampling days was 7.4 mg/L, but results showed a very high variability, with a standard deviation of 6.3 mg/L. This dispersion likely reflects the difficulty in accurately measuring such low concentrations of BOD.

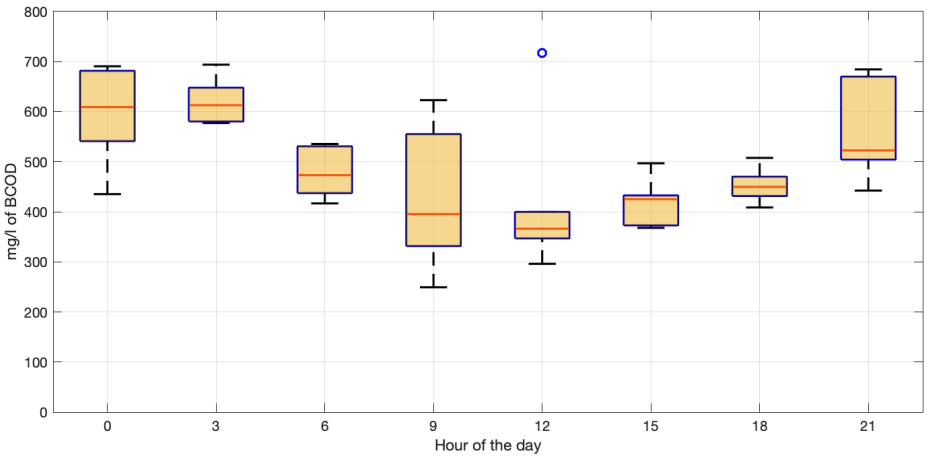


Figure 4.2: Flow-weighted BCOD concentrations in the primary influent at different hours of the day, based on measurements from six sampling days.

4.1.3 Total suspended solids and volatile suspended solids results

The box plots in Fig. 4.3 show the 24-hour profiles of TSS and VSS at the three monitoring points: PI, PE, and SE. They are based on the six sampling days and emphasise the variability observed from day to day in suspended solids concentrations. The values are weighted with the corresponding flow rates, thus representing the actual loads of suspended solids. The plots allow an easy comparison between TSS and VSS, providing an indication of how the proportion between organic (volatile) and inorganic suspended solids changes along the treatment line.
For PI, the concentrations show a marked peak in the morning (from 06:00 to 09:00), where both the median values and the dispersion reach their maximum. This behaviour is consistent with the COD profiles. In these hours, the variability is particularly pronounced due to the higher presence of coarse material in the wastewater. Depending on whether such large particles end up in the filter during analysis or not, the measured TSS and VSS values can vary significantly, leading to the observed high dispersion. Outside the morning peak, the values remain lower and more stable.
In PE, concentrations are significantly reduced compared to the influent, reflecting the sedimentation efficiency of PC. A diurnal trend is still visible, with higher values at night and in the morning (from 06:00 to 09:00), and lower concentrations during midday and afternoon. However, the variability is less pronounced than in the influent, indicating a more homogeneous effluent quality.
SE exhibits the lowest concentrations of TSS and VSS, consistently below 25 mg/L. Taking into account the different scales on the y-axis across the panels, it is evident that the variability is limited compared to the upstream points, showing the damping role of biological treatment and the final clarification on suspended solid concentrations, which attenuate the diurnal fluctuations observed in the influent.

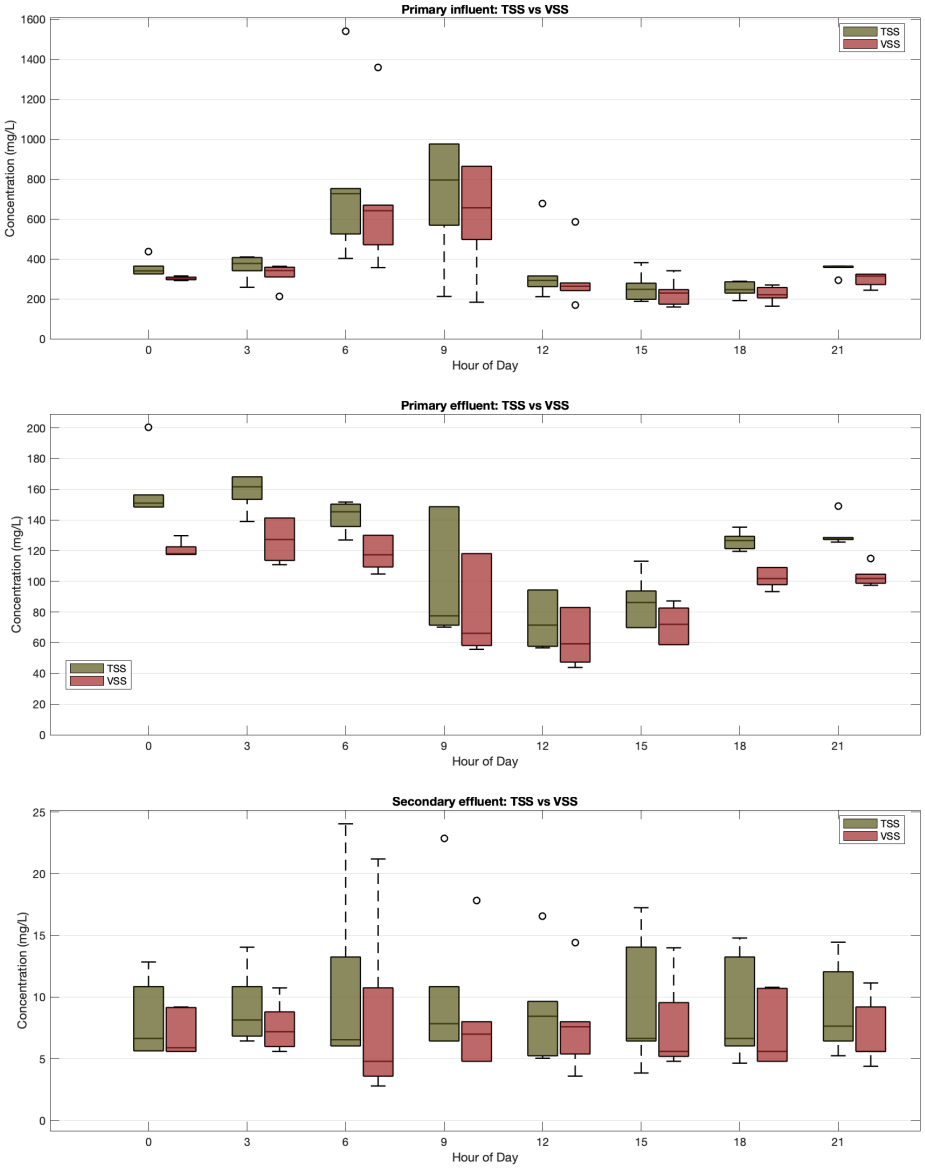


Figure 4.3: Hourly distribution of flow-weighted TSS and VSS concentrations across the three stages of the plant. Box plots refer to six sampling days values.

4.2 Chemical oxygen demand fractionation

COD fractionation was carried out to make the collected data more readable, comparable, and suitable for model calibration.
Figure 4.4 illustrates the diurnal variation of COD fractions in PI (top) and PE (bottom). The fractions are divided into soluble (S_i, S_s) and particulate (X_i, X_s), with further distinction between biodegradable and non-biodegradable components, as described in Section 2.1.1. In both plots, the total COD profile for the respective sampling point is shown as a dashed blue line, with its corresponding unit displayed on the secondary y-axis. All values represent the average of the six sampling days, highlighting the typical diurnal pattern observed across the monitoring period.
In PI, particulate COD accounts for the largest share, with X_s and X_i together accounting for more than half of total COD. After sedimentation in the PC, a clear reduction in particulate fractions can be observed. Both X_s and X_i decrease significantly, in favour of the relative growth of S_i and S_s. This confirms the well established role of the PC in removing settleable solids while leaving soluble COD fractions almost unaffected. This is further supported by an analysis of the absolute values from which the shown percentages were derived, revealing that the greatest reduction occurs in inert particulate matter, while the soluble fractions remain nearly unchanged.
The reduction in particulate COD becomes especially clear when the influent peak observed in PI between 06:00 and 09:00 is compared with the corresponding period in PE, approximately 4.5 hours later (09:00–12:00), accounting for the average hydraulic residence time of the system. During this period, this portion of COD decreases dramatically in the effluent, as already seen by analysing Fig. 4.1, but is further highlighted through this fractionation. This trend is associated with the plant's operational strategy, in which pumping is intensified during these hours to take advantage of low influent flow conditions and to empty the pumping chambers in preparation for the subsequent daily peak. Consequently, accumulated settled solids are directed towards the clarifier, resulting in a more pronounced removal of particles.
The particulate COD peak observed at PI between 06:00 and 09:00 also corresponds to the absolute maximum of total COD, indicating that the overall organic load entering the plant during these hours is particularly high. However, when analysing total COD at PE, it appears that, aside from the particulate fraction settled in the primary clarifier, the most substantial organic load reaching the effluent occurs later in the day, corresponding to the influent recorded at PI between 21:00 and 00:00, and observed at PE approximately 4.5 hours later, between 00:00 and 03:00. A secondary, smaller peak in total COD around midnight at PI may also suggest the presence of a load, possibly dominated by soluble organic matter.
These results highlight that the performance of the PC is not constant throughout the day but strongly influenced by diurnal variations in flow and plant operation.

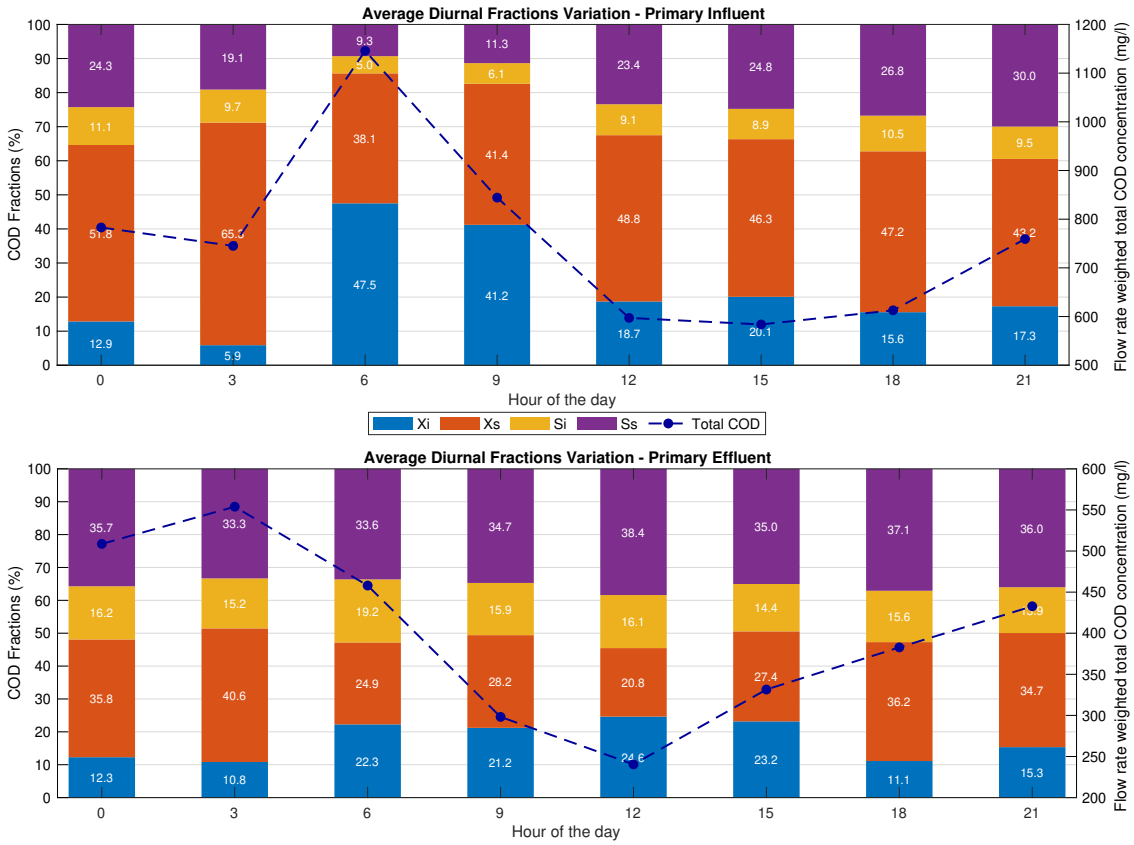


Figure 4.4: Hourly variation of total COD and COD fractions in the primary influent and primary effluent. The presented data are an average of the results of the six days of sampling.

4.3 Nitrogen conversion profiling results

The first five panels in Fig. 4.5 report ammonium (NH_4^+-N) and nitrate (NO_3^--N) concentrations across six sequential zones of the activated sludge line, each panel corresponding to a different sampling time/day as exposed in section 3.3.5. In all cases, NH_4^+-N decreases along the reactor train. Conversely, NO_3^--N increases progressively and reaches its highest values in zone 6.
This behaviour is consistent with the process configuration adopted at Viikinmäki WWTP, which operates according to a denitrification–nitrification scheme. The first three zones are typically maintained under anoxic conditions to promote denitrification, whereas the subsequent three are aerated to enable nitrification, see Fig. 3.5. Zones 2 and 3, however, are designed as swing zones, where aeration can be switched on or off depending on process requirements and online NH_4-N measurements. During the present sampling campaign, the HSY staff confirmed that both swing zones were operated without aeration. Nevertheless, a slight increase in NO_3^--N concentration was occasionally observed in zone 3, possibly due to minor superficial oxygen backflow or residual aeration extending into this area. Such conditions could initiate nitrification earlier than intended and may explain the deviation from the theoretical trend, suggesting that zone 3 acts as a transitional region where the balance between denitrification and nitrification is particularly sensitive to operational control.
The last plot reports nitrite (NO_2^--N) concentrations across the six zones of the activated sludge process. In all profiles, nitrite levels remain very low, generally below $200\,\mu g\,L^{-1}$, with only minor fluctuations along the treatment line. In most cases, values are close to the detection limit and show no consistent accumulation in either the anoxic or oxic zones. This behaviour is consistent with the expected dynamics of a stable DN–N process as nitrite is an intermediate of both nitrification and denitrification, but it is normally kept at low concentrations because both ammonia-oxidizing and nitrite-oxidizing bacteria are active under the given operational conditions. The slight peaks occasionally observed are transient and can be attributed to momentary imbalances.

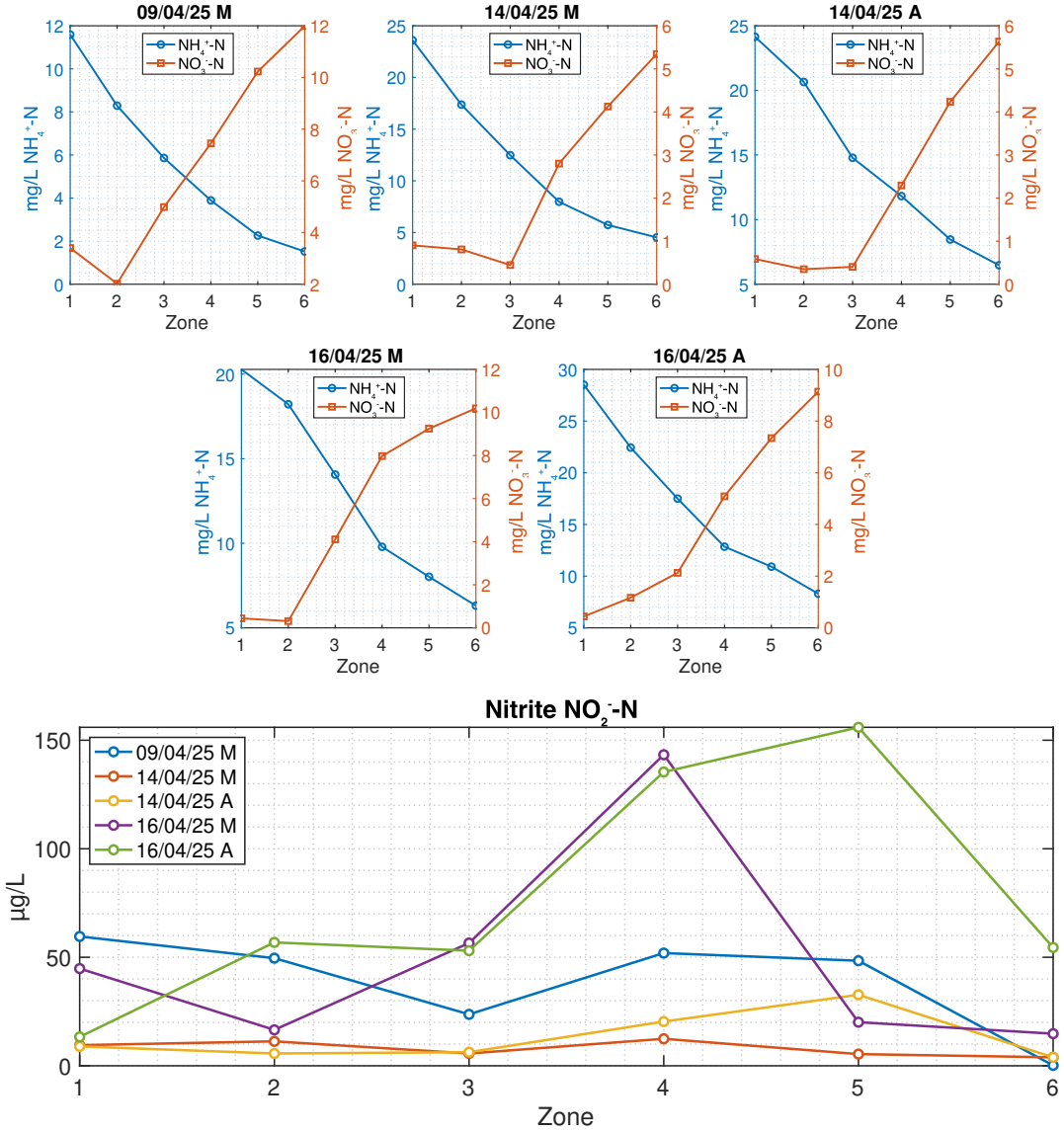


Figure 4.5: Profiles of ammonium (NH_4^+-N), nitrate (NO_3^--N), and nitrite (NO_2^--N) across six zones of the activated sludge unit. Panels correspond to different sampling dates/times (M = morning, A = afternoon).

4.4 Modelling results

4.4.1 Calibration results

Figure 4.6 reports the results of the model calibration. Providing PI data and tuning configurable parameters (as described in the Methods section 3.4), the simulated outputs were generated and compared with the measurements observed at PC effluent.
In all plots, discrete points (circles and triangles) represent observed values, whereas continuous lines denote model predictions. The left-hand column shows total COD, filtered COD and filtered flocculated COD, while the right-hand column illustrates TSS and VSS. Each row corresponds to a different configuration of the primary sedimentation model.

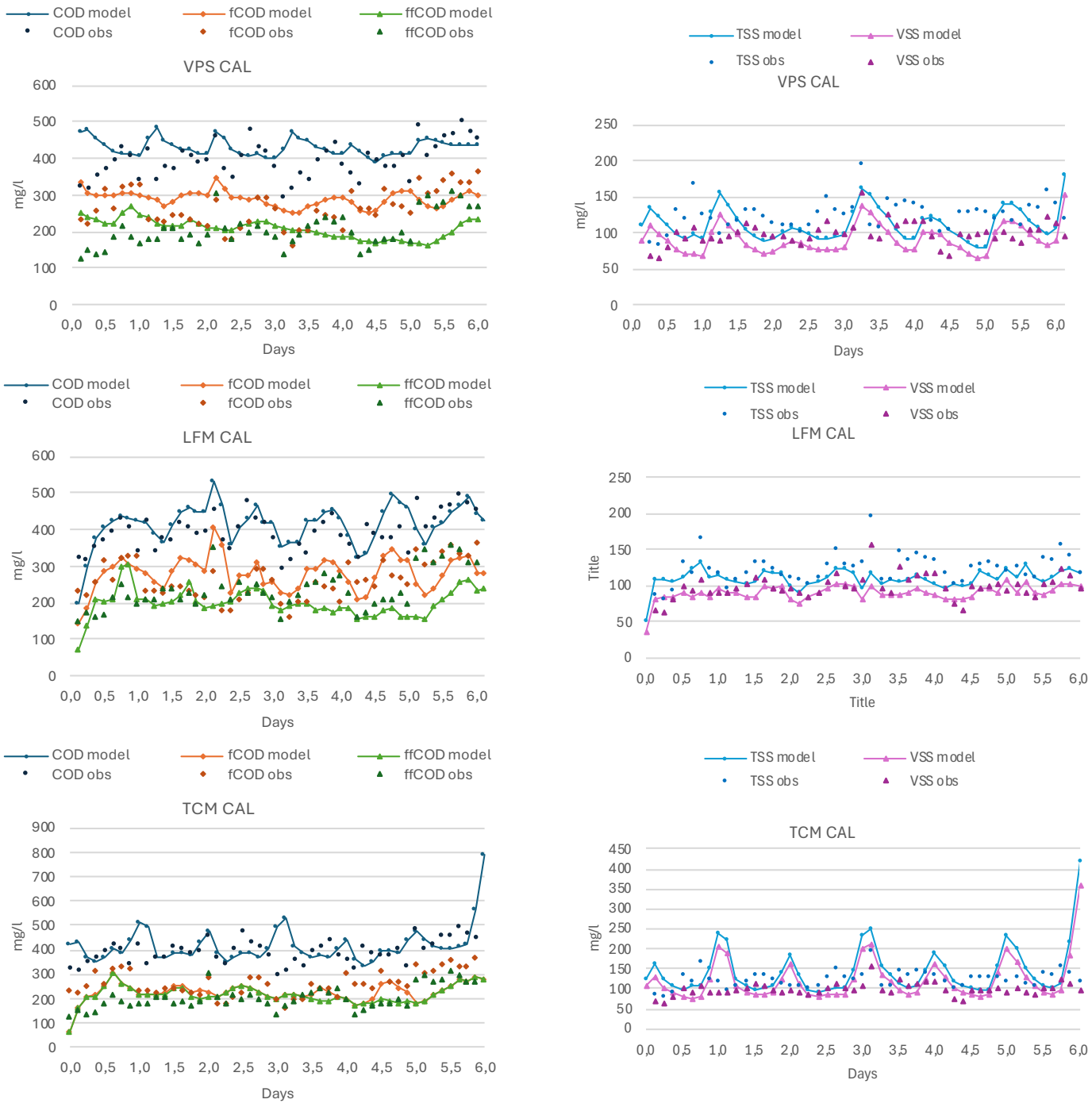


Figure 4.6: Calibration results of the primary clarifier model: comparison between observed data (obs) and simulated outputs (model) for COD, fCOD, ffCOD, TSS, and VSS under three model configurations (VPS, LFM, TCM).

4.4.2 Validation results

The validation results are presented in Fig. 4.7, following the same layout used for the calibration results. The left column displays total COD and its fractions, while the right column reports TSS and VSS.
The input and output data used for model validation originate from the sampling campaign conducted by Petäjä (2025), which spans five sampling days but includes total COD measurements only for the last day. This limitation reduces the robustness of the comparison, but nevertheless provides a meaningful basis for evaluating model performance.

4.4.3 Model evaluation (goodness of fit)

The Mean Absolute Percentage Error (MAPE), defined in equation 3.5, was used to quantitatively assess the model performance. Lower values indicate better agreement between model predictions and the experimental data. MAPE was calculated for both calibration and validation phases to evaluate the model's fitting accuracy and its ability to generalise to independent data.
A more detailed discussion of these results is provided in Section 5.2.

	VPS				
	COD	fCOD	ffCOD	TSS	VSS
Calibration	13.5%	21.3%	17.8%	23.9%	23.3%
Validation	35.3%	14.2%	13.3%	23.8%	18.7%

	LFM				
	COD	fCOD	ffCOD	TSS	VSS
Calibration	10.8%	21.7%	20.1%	13.4%	12.6%
Validation	28.8%	14.8%	15.7%	19.3%	19.8%

	TCM				
	COD	fCOD	ffCOD	TSS	VSS
Calibration	14.4%	18.2%	20.2%	37.9%	40.4%
Validation	29.4%	15.3%	17.3%	38.4%	33.3%

Table 4.1: MAPE (%) of COD, fCOD, ffCOD, TSS, and VSS for VPS, LFM, and TCM models during calibration and validation.

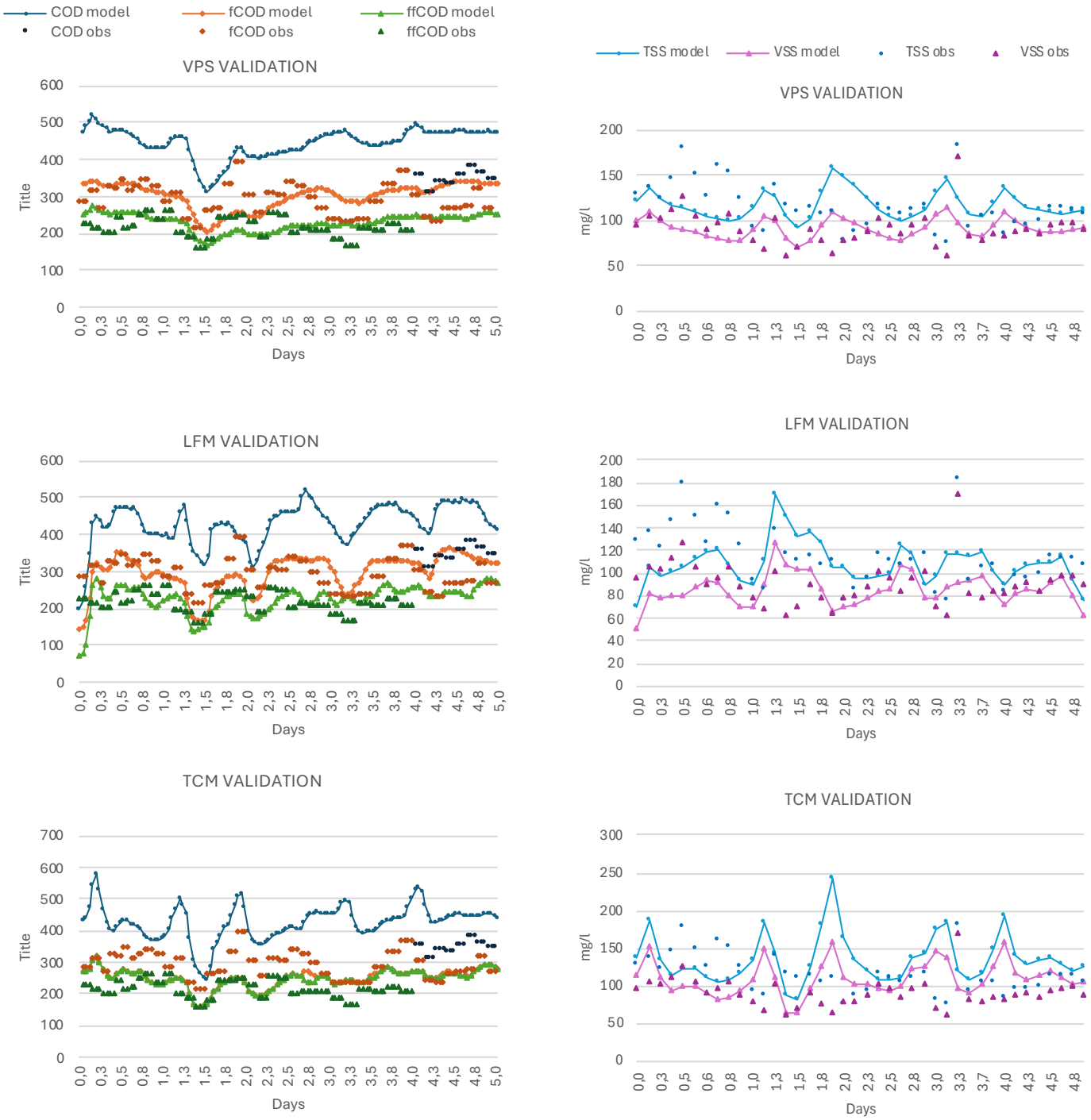


Figure 4.7: Validation results of the primary clarifier model: comparison Petäjä (2025) data (obs) and simulated outputs (model) for COD, fCOD, ffCOD, TSS, and VSS under three model configurations (VPS, LFM, TCM).

Chapter 5

Discussion

5.1 Comparison with previous sampling campaign results and HSY data

Overview of reference datasets

The results obtained during the present sampling campaign were compared with two reference datasets: the Viikinmäki environmental monitoring campaign carried out by HSY, and the previous sampling campaign performed by Petäjä (2025) within the same framework and following the same laboratory analysis standards. It should be noted, however, that Petäjä (2025) campaign was conducted in June 2024, approximately nine months earlier, and therefore under different seasonal conditions. For this reason, a more meaningful comparison can be made by considering the relative differences with the corresponding HSY data, or by analysing the relative reduction of a parameter along the treatment process, rather than directly comparing absolute values.
In general, the total COD measured in this study was consistent with that reported by HSY, confirming the reliability of the analytical methods applied and the robustness of the results across different datasets.
When compared with the previous campaign, the present results showed similar trends in both the PI and PE, supporting the reproducibility of the fractionation while also indicating that the particular climatic conditions during March–April 2025 did not lead to substantially different influent characteristics compared to June 2024.

The HSY monitoring campaign provides only a subset of the parameters analysed in this study, reported as daily averages on selected weekdays (see Fig. 3.2). Specifically, HSY employs flow-proportional automatic samplers, ensuring that the 24-hour composite samples are directly representative of the loads. For consistency, these values were compared with flow-weighted results of the present campaign. The HSY dataset includes:

- Primary influent (PI): TSS, COD, BOD_7
- Primary effluent (PE): TSS, COD, BOD_7
- Secondary effluent (SE): TSS

Chemical Oxygen Demand

Fig. 5.1 compares COD values from the present campaign and those reported by HSY. On average, the present measurements are higher by about 22% in the PI and 17% in the PE compared to HSY data. Although most of these deviations fall within the 20% uncertainty range declared by HSY for COD measurements (indicated by the vertical bars in the figure), the trend systematically points to an overestimation in the present dataset. This bias may arise from differences in sampling strategy and analytical methodology.
To further support the hypothesis of a systemic discrepancy between the two methods used for COD analysis, Table 5.1 presents data collected during the previous sampling campaign (Petäjä, 2025). In that campaign, the COD values showed an average overestimation of 18% for PI, 2% for PE, and 80% for SE when compared with corresponding HSY control data. This might be explained by the fact that concentrations and loads are much lower at this stage and small fluctuations result in large relative differences, making the comparison less meaningful.

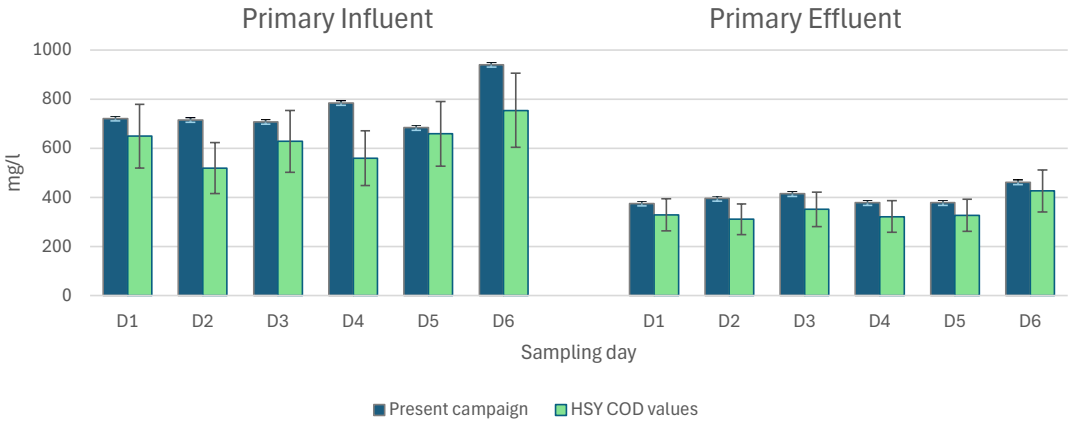


Figure 5.1: Comparison of COD values between the present campaign and the HSY monitoring campaign for the primary influent and primary effluent. For HSY values, the vertical bar represents the declared measurement uncertainty (20%), while for the present campaign it corresponds to an analytical uncertainty of 10 mg/L.

Table 5.1: Comparison of COD values among different sources. Values from the present campaign and Petäjä (2025) are reported as flow-weighted averages. All values expressed in mgCOD/L.

Stage of the plant	**Present campaign**	**HSY (reported)**	**Petäjä (2025)**	**HSY (Petäjä campaign)**
Primary influent	759	628	725	612
Primary effluent	401	345	353	346
Secondary effluent	92	n.a.	92	51

n.a. = not available measurement.

Total Suspended Solids

Comparison of TSS values (Fig. 5.2) shows that, for both PI and PE, the results of the present campaign are generally close to those reported by HSY, despite a slight tendency toward overestimation. In contrast, larger discrepancies are observed in the SE. Since the concentrations and corresponding loads are much lower at this stage, small absolute variations translate into larger relative differences, making these data less significant and less reliable for direct comparison.
From Table 5.2, the values appear to be in reasonable agreement with HSY and Petäjä (2025) data for both the PI and effluent. It is noteworthy that PE TSS is the only parameter showing an underestimation in both the present campaign and Petäjä (2025) data when compared to the corresponding HSY values. The removal efficiency of TSS by the PC in the present campaign showed a removal efficiency of 71%, Petäjä (2025) reported a similar value of 70%. The corresponding HSY datasets yielded a efficiency of 67% in both cases. Taking into account the analytical uncertainties and possible inaccuracies, these results appear consistent and suggest a reasonable reliability of the measurements.

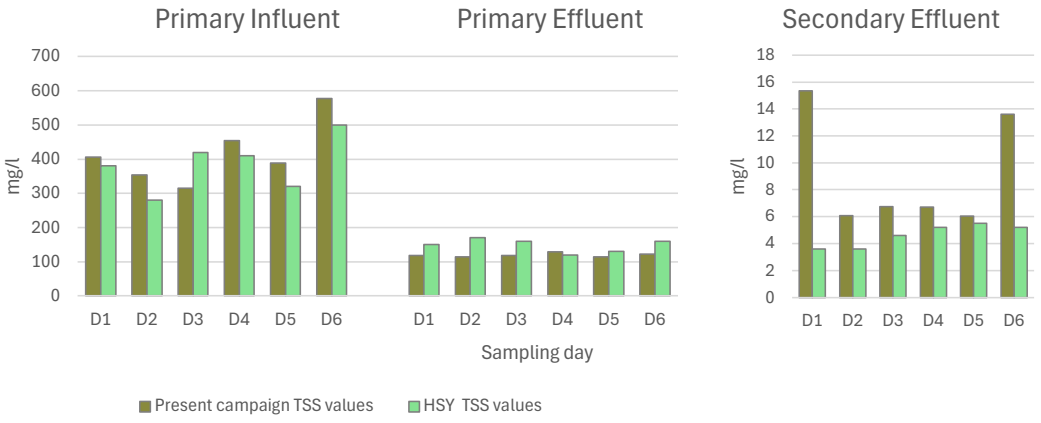


Figure 5.2: Comparison of TSS values between the present campaign and the HSY monitoring campaign for primary influent, primary effluent, and secondary effluent. Six sampling days are reported.

Table 5.2: Results of TSS values from different sources. Values from the present campaign and Petäjä (2025) are reported as flow-weighted averages. All values expressed in mgTSS/L.

Stage of the plant	Present campaign	HSY (reported)	Petäjä (2025)	HSY (Petäjä campaign)
Primary influent	416	385	370	390
Primary effluent	120	128	110	130
Secondary effluent	9	4.6	6.5	5.3

Biochemical Oxygen Demand

Figure 5.3, shows the comparison carried out between the BOD_7 values obtained in the present sampling campaign and the reference data provided by HSY.
During the campaign, additional BOD_7 measurements were possible only on the last day (D6), due to the limited availability of the necessary equipment. For this reason, the PE values are reported only for D6, while the PI values are available for all days.
For PI, the average BOD_7 values are overestimated by 20% with respect to the corresponding HSY data, with almost all values higher than those of HSY. As shown in Table 5.3, Petäjä (2025) findings indicate an average positive discrepancy of approximately 30% in the BOD_7 of the PI compared to the corresponding HSY dataset, suggesting the presence of a systematic bias that can arise from overestimation in campaign data or underestimation in HSY monitoring results.

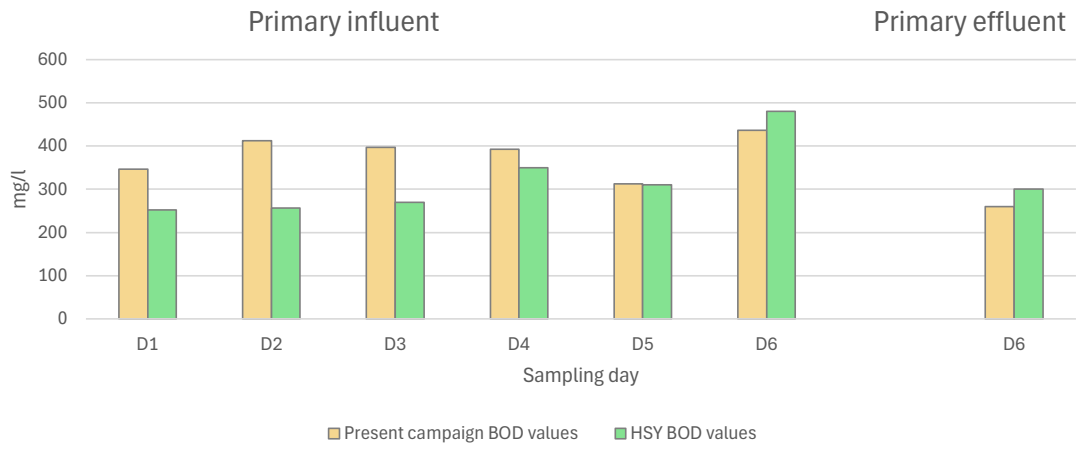


Figure 5.3: Comparison between BOD_7 values obtained in the present sampling campaign and HSY monitoring campaign.

Table 5.3: Comparison of BOD values among different sources. Values from the present campaign and Petäjä (2025) are reported as flow-weighted averages.(All values expressed in mgBOD/L).

Stage of the plant	Present campaign	HSY (reported)	Petäjä (2025)	HSY (Petäjä campaign)
Primary influent	383	320	413	289
Primary effluent	260*	300	n.a.	170
Secondary effluent	8.2	n.a.	n.a.	10.7

n.a. = not available measurement.
* Value refers to a single measurement taken on the last sampling day, unlike the other values which represent flow-weighted averages over multiple days.

Chemical oxygen demand fractionation results

Finally, the average COD fractions obtained in the two sampling campaigns are compared (Table 5.4). A more detailed comparison between the PI and effluent has already been provided above. In both stages of the process, the results appear fairly similar. On the one hand, this strengthens the reliability of the measurements, as comparable outcomes were obtained despite being carried out at different times and by different operators. On the other hand, it also indicates that the peculiar climatic conditions during March–April 2025, characterized by an unusually mild winter in the Helsinki region with little to no snow accumulation and, consequently, no significant snow melt, did not result in sampling conditions substantially different from those of June 2024, as was originally planned and expected.

It should be noted, however, that this type of tabular representation, while having the clear advantage of condensing the results into a few comparable figures, inevitably leads to a loss of information regarding the variability over time. By averaging hourly data, the intrinsic dynamics of the fractions are flattened into a less informative picture.

Nevertheless, including the standard deviation in this table, provides additional insight into the variability and robustness of the measured data. As can be observed, the standard deviations are particularly high for the particulate fractions. This behaviour is consistent with their formulation. As discussed in Section 2.1.1, the value of X_i inherits the uncertainties of the other fractions, since it is obtained by subtraction from the total COD. Similarly, the variability of X_s partly arises from the calculation of the biodegradable fraction based on BOD measurements, which, due to equipment limitations, were performed less frequently and with fewer replicates than COD. In addition to these methodological aspects, the natural hourly and daily fluctuations in the influent wastewater composition further contribute to the observed variability.

Stage of the plant	Source	Xi	Xs	Si	Ss
Primary influent	Present campaign	22 ± 10	48 ± 11	9 ± 3	21 ± 6
	Petäjä (2025)	21	52	6	21
Primary effluent	Present campaign	18 ± 11	31 ± 13	16 ± 5	35 ± 9
	Petäjä (2025)	14	33	13	40

Table 5.4: Average COD fractions and corresponding standard deviations obtained in the present sampling campaign, compared with the results from Petäjä (2025). All values are expressed as percentages and refer to the primary influent (PI) and primary effluent (PE).

5.2 Modelling

The modelling results highlight both the potential and limitations of applying simplified PC models within a data-limited experimental context. The calibration process, necessarily carried out in a semi-quantitative manner due to the absence of automatic fitting tools in SUMO version used for this work (see Section 3.4), introduced a certain degree of subjectivity, mitigated by comparing averages, peaks, and trends between observed and simulated data.
From the analysis of Fig. 4.6 and 4.7 and Table 4.1, it can be seen that VPS, although the simplest approach and the least demanding from a computational point of view,provides a reasonable, robust, and easy-to-interpret prediction of PE. Its formulation respects the main physical constraints of the system and allows for a clear overall representation of the separation process. However, it is not suitable for representing more complex dynamics, such as the degradation of organic matter or differential behaviour of individual fractions.
The comparison between calibration and validation confirmed the good predictive capacity of the Layered Flux Model, which emerged as the most balanced approach. Its performance suggests that introducing a limited degree of complexity, such as settling parameters and size characteristics, can significantly improve the description of COD fractions and solids removal without compromising model robustness. However, the model could be further refined by incorporating laboratory-derived settling parameters specifically measured for the Viikinmäki WWTP, which would allow a more accurate and site-specific characterization of the sedimentation behaviour. Alternatively, an approach similar to that adopted by Polorigni et al. (2021) could be implemented, in which several settling velocity groups are defined and associated with distinct particulate fractions, thereby capturing the heterogeneity of the suspended solids and their different settling dynamics. SUMO also offers different options to support this process, including empirical correlations that use routine laboratory tests, such as the sludge volume index (SVI), to estimate the Vesilind parameters, as well as an Excel-based tool that can automatically analyse laboratory settling curves to determine the maximum Vesilind settling velocity and the hindered settling parameter.
In contrast, the Three Compartment Model, although conceptually more detailed and

mechanistic, did not provide a clear advantage over LFM under the conditions of this study. Its higher data requirements, particularly for estimating compartment volumes and for capturing reaction-driven processes such as flocculation and polymer dosing, exceeded the scope of the available monitoring data. Finally, the observed behaviour of the colloidal COD fractions revealed a further limitation of the TCM model configuration. The negligible difference between fCOD and ffCOD (corresponding to the colloidal fraction) in the simulations, in contrast with the observed data, indicates that the model tended to overestimate reaction rates occurring in the clarifier. This is at odds with the operational design of PCs, which are intended to minimise biological reactions and direct such processes to the activated sludge stage. Future improvements to this modelling configuration should include, in addition to measurements to determine the detailed sedimentation rate and volume of the sludge blanket, particular attention to the colloidal fraction and its separation dynamics. However, it is worth mentioning that LFM is the most computationally demanding configuration, as it involves ten reactors compared to three in TCM and none in VPS.

5.2.1 Related Studies on Primary Clarifier Modelling

Having discussed the results obtained within the SUMO environment, it is worth momentarily setting aside this specific framework, albeit highly valuable for digital twin development, to briefly explore the current state of the art in primary clarifier modelling. This allows us to consider how other studies have approached similar challenges and what modelling strategies or calibration practices have been proposed in the literature.
In this regard, Polorigni et al. (2021), when developing a a primary clarifier model, also highlighted, in agreement with the present analysis, the critical importance of accurately estimating and representing the different settling velocities of suspended particles. Equally essential, according to their findings, is the use of an appropriate COD fractionations, particularly for the particulate, both biodegradable and unbiodegradable as well as for inorganic settleable solids. Moreover, Polorigni et al. (2021) emphasized that explicitly including the primary clarifier model within a whole-plant modelling framework is fundamental to avoid misestimations of particulate fractions, which could otherwise compromise the reliability of the overall simulation results.
Similarly to the present study, Gernaey et al. (2001) developed a primary clarifier model with the explicit aim of making it compatible and integrable with the ASM1 framework and its state variables (X_i, X_s, S_i, S_s). However, only total and filtered COD data were available, the latter being used as an approximation of the soluble fraction, which allowed the particulate portion to be estimated. The initial model, based on the Takacs clarifier model (Takács et al., 1991), produced unsatisfactory results, as the soluble fraction remained nearly unchanged, an issue also observed in this study with the simpler VPS model before increasing the volume of the pre-aeration tank. To address this limitation, Gernaey et al. (2001) assumed a residence time sufficient to ensure partial mixing. Consequently, Gernaey et al. (2001) adopted an approach similar to LFM, in which settling velocity equations govern mass balances across ten layers. The soluble components were satisfactorily described thanks to the inclusion of ammonification reactions, a flocculation model, and soluble residence time. Conversely, the estimation of settling parameters

proved particularly challenging due to the limited availability of experimental data, which ultimately affected the accuracy of the particulate fraction modelling.

Chapter 6

Conclusions

This thesis presents the research conducted to support the development of a digital twin of the Viikinmäki wastewater treatment plant, within the framework of the DIGICARBA project. Viikinmäki is the main treatment plant serving the Helsinki region and receives a combination of domestic, industrial, and stormwater inflows. The digital twin aims to enhance process understanding and optimization, with the ultimate goal of reducing greenhouse gas emissions and improving overall plant performance.

To support the calibration of the model underlying the digital twin, a seven-day sampling campaign was carried out between March and April. The campaign provided insight into the daily variability of total COD, filtered COD, filtered flocculated COD, BOD_7, TSS, and VSS in different stages of the plant. The COD fractionation analysis provides a clear picture of the organic load and its main components at the influent and effluent of the primary clarifier over a 24-hour period, revealing the diurnal trends and removal behaviour of particulate and soluble fractions.

In addition, five sets of nitrogen conversion profiling were performed to obtain quantitative information on the dynamics of nitrite, nitrate, and ammonium along the six zones of the activated sludge process. These profiles also contributed to the calibration of the digital twin model. The results indicate that the activated sludge process was well established and stable throughout the monitored period, suggesting consistent biological performance. However, the suspected backflow of oxygen detected in the upper layer of zone three should be further investigated, as it may locally affect the efficiency of nitrogen removal.

Although slightly revised compared with the previous campaign, the adopted methodology proved reliable and easily replicable. Overall, the results were consistent with those from the HSY environmental monitoring campaign, except for a slight overestimation of COD and BOD_7 compared with HSY routine analysis data, a deviation also reported during the earlier characterization campaign conducted under summer conditions.

In parallel to influent characterization, the data collected during this study were used to calibrate a prototype model of the Viikinmäki primary clarifier, implemented using the three different configurations available in the SUMO software. The models were validated using data from the previous characterization campaign, assessed through the Mean Absolute Percentage Error (MAPE), and compared in terms of performance and

applicability. Based on these promising results and the insights gained from the literature review, suitable modelling strategies were identified as promising approaches for extending the digital twin model to include the primary clarifier unit.
Overall, these outcomes highlight the breadth of information that can be extracted from such a sampling campaign. Beyond supporting the calibration of the digital twin and the prototyping of a primary clarifier model, the collected data provided a broader understanding of the plant's functioning, operational strategies, limitations, and strengths.

Bibliography

Sfs-en iso 5815-1:2019, 2019. Water quality. Determination of biochemical oxygen demand after n days (BODn). Part 1: Dilution and seeding method with allylthiourea addition (ISO 5815-1:2019) Available at: https://sales.sfs.fi/fi/index/tuotteet/SFS/CENISO/ID2/5/813325.html.stx.

Lobna Amin. doctoral project on vivianite recovery at wwtp viikinmäki with insa toulouse and siaap, 2024.

APHA 4500-NO_3^-. 4500-NO_3^- NITROGEN (NITRATE), pages 4–120–4–121. 2005. doi: 10.2105/SMWW.2882.089. URL https://www.standardmethods.org/doi/abs/10.2105/SMWW.2882.089.

Giulia Bachis, Thibaud Maruéjouls, Sovanna Tik, Youri Amerlinck, Henryk Melcer, Ingmar Nopens, Paul Lessard, and Peter A. Vanrolleghem. Modelling and characterization of primary settlers in view of whole plant and resource recovery modelling. Water Science and Technology, 72:2251–2261, 2015.

Guanghao Chen, George A. Ekama, Marinus C. M. van Loosdrecht, and Damir Brdjanovic, editors. Biological Wastewater Treatment: Principles, Modelling and Design. IWA Publishing, London, 2020. ISBN 9781789060362. doi: 10.2166/9781789060362.

Dynamita. West software and dynamita wiki, 2022a. URL https://wiki.dynamita.com/en/home.

Dynamita. Separation models. https://wiki.dynamita.com/en/PU_spec_models#three-compartment-separation-model, 2022b. Accessed: June 2025.

European Union. Directive (eu) 2024/3019 of the european parliament and of the council of 27 november 2024 concerning urban wastewater treatment (recast). Official Journal of the European Union, L 2024/3019, 2024. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024L3019.

Krist Gernaey, PA Vanrolleghem, and P Lessard. Modeling of a reactive primary clarifier. Water science and technology, 43(7):73–81, 2001.

Henri Haimi. Data-derived soft sensors in biological wastewater treatment – with application of multivariate statistical methods. Doctoral thesis (article-based), Aalto University, School of Engineering, 2016. URL https://urn.fi/URN:ISBN:978-952-60-6650-9.

Mogens Henze, Willi Gujer, Takashi Mino, and Mark C. M. van Loosdrecht, editors. Activated Sludge Models: ASM1, ASM2, ASM2d and ASM3. Scientific and Technical Report No. 9. IWA Publishing, London, UK, 2000. ISBN 1-900222-50-7.

HSY. Viikinmäki wastewater treatment plant. Technical report, Helsinki Region Environmental Services Joint Municipal Authority, 2017. URL https://www.hsy.fi/en/.

Jose A. Jimenez, Enrique J. La Motta, and Denny S. Parker. Kinetics of removal of particulate chemical oxygen demand in the activated-sludge process. Water Environment Research, 77:437–446, 2005.

Bruce Johnson, Cheng Yang, Keaton Lesnik, Joshua Registe, Tom Johnson, and Adrienne Menniti. Full-scale soft-sensor implementations enable wrrf hybrid digital-twins. Technical report, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), 2023.

Anmol Kiran. Data analysis and pre-processing for digital twin development, predictive modeling of missing variables at viikinmäki wastewater treatment plant. 2025.

Daniel Mamais, David Jenkins, and Paul Prrr. A rapid physical-chemical method for the determination of readily biodegradable soluble COD in municipal wastewater. Water Research, 27(1):195–197, 1993. ISSN 0043-1354. doi: https://doi.org/10.1016/0043-1354(93)90211-Y. URL https://www.sciencedirect.com/science/article/pii/004313549390211Y.

Eero Petäjä. Influent fractionation for viikinmäki wastewater treatment plant, 2025.

C.L. Polorigni, D.S. Ikumi, and G.A. Ekama. Primary sedimentation modelling with characterized setting velocity groups. Water Research, 189:116621, 2021.

PUB, Singapore's National Water Agency. Digitalising Water – Sharing Singapore's Experience. IWA Publishing, London, UK, 2020. ISBN 9781789061405. URL https://iwaponline.com/ebooks/book/806/Digitalising-Water-Sharing-Singapore-s-Experience.

Leiv Rieger, Sylvie Gillot, Günter Langergraber, Takayuki Ohtsuki, Andrew Shaw, Imre Takacs, and Stefan Winkler. Guidelines for using activated sludge models. IWA publishing, 2012.

P. J. Roeleveld and M. C. van Loosdrecht. Experience with guidelines for wastewater characterisation in the netherlands. Water Science and Technology: A Journal of the International Association on Water Pollution Research, 45(6):77–87, 2002.

SFS 3029. Sfs 3029: Wastewater — vocabulary. Technical report, Suomen Standardisoimisliitto SFS, 1976. URL https://sales.sfs.fi/fi/index/tuotteet/SFS/SFS/ID2/3/1019.html.stx.

M. Sieranen. doctoral project on generating data of ghg emissions from finnish centralized wwtps in cis-fin project with uoulu., 2024.

Suomen Standardoimisliitto SFS. Kemiallisen hapen kulutuksen (cod cr) määrittäminen suljetulla putkimenetelmällä: Vertailu standardimenetelmään. Standard, Suomen ympäristökeskus, 1986. URL https://helda.helsinki.fi/items/ff802b2a-5f51-4769-804f-fd78221410cf.

Suomen Standardoimisliitto SFS. Sfs-en 872: Water quality – determination of suspended solids – gravimetric method using glass-fibre filters. Standard, CEN, 2005. URL https://sales.sfs.fi/fi/index/tuotteet/SFS/CEN/ID2/8/11190.html.stx.

Imre Takács, Gilles G Patry, and Daniel Nolasco. A dynamic model of the clarification-thickening process. Water research, 25(10):1263–1271, 1991.

Elena Torfs, Niels Nicolaï, Saba Daneshgar, John B. Copp, Henri Haimi, David Ikumi, Bruce Johnson, Benedek B. Plosz, Spencer Snowling, Lloyd R. Townley, Borja Valverde-Pérez, Peter A. Vanrolleghem, Luca Vezzaro, and Ingmar Nopens. The transition of wrrf models to digital twin applications. Water Science and Technology, 85: 2840–2853, 2022.

Aninka Urho, Anna Kuokkanen, Jenni Raatikainen, Hanna Riihinen, Perttu Saarinen, and Maria Valtari. Wastewater treatment in the helsinki metropolitan area 2024 – viikinmäki and blominmäki wastewater treatment plants. Technical Report 2/2025, Helsinki Region Environmental Services Joint Municipal Authority, Helsinki, 2025. Paperback ISBN: 978-952-7146-95-8; PDF ISBN: 978-952-7146-94-1; Online ISBN: 978-952-7146-93-4. ISSN (print): 1798-6087; ISSN (online): 1798-6095. Language: Finnish.

B Valverde-Pérez, B Johnson, C Wärff, D Lumley, E Torfs, I Nopens, and L Townley. Operational digital twins in the urban water sector: case studies, 2021.

Ai-Jie Wang, Hewen Li, Zhejun He, Yu Tao, Hongcheng Wang, Min Yang, Dragan Savic, Glen T Daigger, and Nanqi Ren. Digital twins for wastewater treatment: A technical review. Engineering, 36:21–35, 2024.

Cheng Yang. Wrf 5121 final report overview: Long-term evaluation of hybrid digital twins at three full-scale wrrfs. In WRF 5121 Final Report Overview: Long-Term Evaluation of Hybrid Digital Twins at Three Full-Scale WRRFs. Water Environment Federation, 2024.