$$\frac{dA}{dt} = r_1(A+L)\left(1 - \frac{A+L}{K+L}\right) g_1(I_t)\, g_2(T_t)\, g_3(H_t)$$

Modelling Diatom Growth Using Ordinary Differential Equations

A Data-Driven Approach to Biomass Yield Prediction in Collaboration with Swedish Algae Factory

Master's Thesis in Engineering Mathematics and Computational Science

Simon Berggren

Department of Mathematical Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

MASTER'S THESIS 2026

Modelling Diatom Growth Using Ordinary Differential Equations

A Data-Driven Approach to Biomass Yield Prediction in Collaboration with Swedish Algae Factory

Simon Berggren

Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026

Modelling Diatom Growth Using Ordinary Differential Equations
A Data-Driven Approach to Biomass Yield Prediction in Collaboration with Swedish Algae Factory
Simon Berggren

Supervisor: Marija Cvijovic, Department of Mathematical Sciences
Examiner: Philip Gerlee, Department of Mathematical Sciences

Master's Thesis 2026
Department of Mathematical Sciences
Division of Applied Mathematics and Statistics
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Final equation with the goal of describing the diatom growth at Swedish Algae Factory.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2026

Modelling Diatom Growth Using Ordinary Differential Equations
A Data-Driven Approach to Biomass Yield Prediction in Collaboration with Swedish Algae Factory
Simon Berggren
Department of Mathematical Sciences
Chalmers University of Technology

Abstract

This thesis explores how cultivation data from Swedish Algae Factory can be used to identify key environmental factors affecting diatom growth and how these insights can inform improvements to the cultivation process. By formulating and tuning an ordinary differential equation (ODE) model based on environmental variables such as light intensity, temperature, and pH-levels, the work aims to capture the pace of diatom growth. However, due to limited output resolution, the model's predictive capabilities were constrained and model validation was not feasible. With that constraint, the model highlight areas where additional or more frequent measurements could improve understanding and modelling of the cultivation.

Keywords: Diatom, ODE, Algae, Growth, Temperature, Light, LUX, pH, Nutrient, Modelling.

Acknowledgements

I would like to thank Marija Cvijovic at the department of Mathematical Sciences at Chalmers. Moreover, I would like to thank Swedish Algae Factory for their contribution to the work. Moreover, everyone who has contributed in another way, you are helpful, thanks!

Simon Berggren, Gothenburg, May 21, 2026

Contents

List of Figures

1
Introduction

Swedish Algae Factory, founded in 2016, focuses on cultivating a specific sort of diatom to extract a silicon based material called Algica. Silicon based materials from diatoms have demonstrated significant potential in various industrial sectors, including but not limited to personal care and development of solar panels [23]. As of January 2025, the company possesses a cultivation area of roughly 3700 m^2, and they are aiming to both expand the company and make the process of growing diatoms more effective. To optimize the growth of diatom species, several variables such as temperature, light intensity, and nutrient levels must be taken into account.

In this context, the data analysis perspective offers a unique opportunity to advance the understanding of how these diatoms respond to environmental changes and how to optimize their growth. By systematically correlating algae growth with these diverse factors, Swedish Algae Factory hopes to refine their cultivation processes, identify crucial predictors of algae growth, and improve overall yield. Therefore, they have contributed with numerical data points from their cultivation for this research. From an academic viewpoint, this project also resonates strongly with ongoing research in computational biology, particularly in modelling time-series data and tackling issues of integrating information gathered at different resolutions.

As a consequence, this collaboration serves a dual purpose. On one hand, it addresses a practical industry need for optimizing production through data-driven methods, while on the other, it provides a test bed for exploring how complex biological systems can be modelled to balance growth and survival. Additionally, this thesis draws inspiration from the modelling approaches presented in *Mathematical Biology I: An Introduction* by J.D. Murray [15], which is part of the curriculum at Chalmers University of Technology. This connection highlights how academic theory and literature can be utilised for practical applications in Swedish industry.

1.1 Purpose and Research Questions

The overarching purpose of this project is to investigate how existing cultivation data from Swedish Algae Factory can be leveraged to both identify crucial predictors of algae growth and propose data-driven improvements to the cultivation process. More specifically, the idea is to explore and analyse the data given by Swedish Algae Factory with the intention to identify the features that most strongly correlate with variations in diatom growth. Furthermore, it is to evaluate if there are important parameters that are not currently being measured at Swedish Algae Factory that affect the growth of the diatoms. With this in mind, the following research questions are to be evaluated in this report:

- **Key Parameters:** From the dataset given by Swedish Algae Factory, which features emerge as the strongest predictors for diatom growth, and how do these features interact with one another?

- **Model Integration:** How can data of varying resolutions be integrated effectively into a unified model?

- **Data Gaps:** Is it possible that other features currently not measured or higher resolution data would significantly enhance the model's predictive capabilities and provide deeper insights?

- **Practical Recommendations:** How can these findings be translated into actionable recommendations for improving cultivation efficiency at Swedish Algae Factory?

To fulfil the purpose, the remainder of this thesis mostly follows a traditional set-up. Chapter 2 outlines the theoretical background and Chapter 3 presents the methodology for the research. However, Chapter 4 named *Procedure and Results* is intentionally made two-sided. It both showcases the result and motivates why the choices in the modelling process were made as such. This is to mimic the iterative approach to modelling, which is being presented more detailed in the methodology chapter. Finally, Chapter 5 provides conclusions and practical recommendations from the results.

2
Theory

In the theory section all theory which will be used for the paper is presented. It will include main parameters which affects the growth rate of diatoms found in literature and an explanation for each one of them. Additionally, it will include how modelling in biology is done using ordinary differential equations.

2.1 Growth Parameters

Most photoautotrophic organisms, such as diatoms, relies on two key processes for growth and survival. Firstly it is photosynthesis, which stores energy from sunlight, and secondly cellular respiration, which releases that energy for use. Photosynthesis is naturally occurring only when there is light available, meanwhile cellular respiration is occurring all the time and is consuming stored energy [8]. For a bioorganism to grow there are different growth parameters which needs to be present and which are contributing to a different degree to the growth of the organism. When it comes the growth of diatoms there are five different parameters which are contributing the most to the growth: water, nutrients, pH, light intensity, and temperature [1].

2.1.1 Nutrients and Water

Diatoms play a central role in global carbon fixation and are key drivers in the cycling of several nutrients, such as carbon, nitrogen, and phosphorus, as well as silicon and iron [18]. To sustain their relatively high growth rates and large size, diatoms require conditions rich in nutrients. This is due to their low surface to volume ratios compared to smaller phytoplankton, which are better adapted to low nutrient environments. When nutrients are abundant, diatoms often dominate the phytoplankton community if there is competition [18].

Water is essential not only as a physical medium but also because it directly supports physiological processes in algae. Depending on the species, the algae grows optimally in different saline levels of the water. Since evaporation over time increases salinity in water, it is important in cultivation to find a method to keep salinity levels even. Still, many algae are resilient and can tolerate a range of salinities [1].

2.1.2 Light

For photosynthesis to occur there needs to be some light source for the algae. Moreover, the photosynthesis of algae are dependent on different factors of the light such as intensity, periodicity, and quality [1]. The amount of light, the frequencies, and the periodicity which is required for optimal growth is different depending on the species of algae. So the following section will present how theory regarding light and photosynthesis works in general. Regarding light intensity, the relationship between light intensity and rate of photosynthesis in algae follows a pattern similar to a downward-opening convex curve. At low light intensities, photosynthesis is typically proportional to light available, increasing the light intensity is approximately linearly increasing the photosynthesis rate. However, once light intensity reaches a threshold the algae becomes light-saturated. At this stage, the rate of photosynthesis no longer increases with additional light. Instead, if the light intensity continues to increase beyond the threshold, the photosynthesis may begin to decline due to deactivation of key proteins within the algae [4].

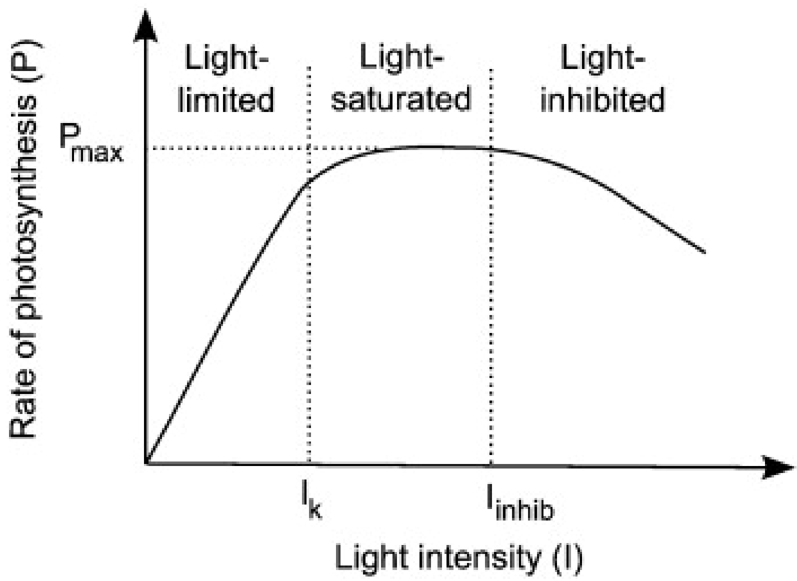

Figure 2.1: *Image showing the relationship between the rate of photosynthesis (P) and light intensity (I) [4].*

While light is essential for photosynthesis, continuous exposure to light reduces the efficiency of photosynthesis. Algae, like most organisms, depend on a cycle of light and darkness. During the light phase, photosynthesis occurs, but in the dark phase, the cells perform essential processes such as respiration and repair. Including a regular dark period is therefore necessary for healthy and efficient growth [14]. The optimal duration of light exposure for diatoms is not fixed. One on hand, it appears to depend on the species. On the other, experiments indicate that the high adaptability of diatoms makes them adapt to the current light–dark cycle. For example, experiments alternating between 12:12 and 16:8 light–dark cycles have shown that over time diatoms are adapting their pigment levels to their environment. However, continuous light without interruption can lead to reduced pigment accumulation and physiological stress. These findings highlight that maintaining a natural day–night rhythm is likely more important for diatom health and productivity than simply maximizing the number of light hours [17].

2.1.3 Temperature

Just like with light, temperature influences photosynthetic rate and can be incorporated into algae growth models in different ways. A common approach is to assume that light and temperature are uncoupled and effect growth independently. In these uncoupled models, the photosynthetic rate is modelled as the product of two separate multiplicative terms, one for light and one for temperature [4]. For example, the Arrhenius equation, presented in equation 2.1, can be used to describe how temperature influences the rate of growth directly. This equation calculates a growth rate k [s^{-1}] which is dependent on the constants A, E_a, and R. A is a scaling constant in the inverse time unit s^{-1}, E_a is the activation energy for the given chemical event and R is the universal gas constant. Furthermore, the Arrhenius equation is dependent on the variable T which represents the temperature given in kelvin [7].

$$k = Ae^{-\frac{E_a}{RT}} \tag{2.1}$$

However, the Arrhenius equation has its limitations. It was originally developed for describing chemical reactions, and when describing temperature's effect on growth rate it fails on high temperatures. In reality, high temperatures can lead to reduced efficiency of photosynthesis. Therefore, models that only use the Arrhenius equation has a tendency to overestimate the growth rate [4].

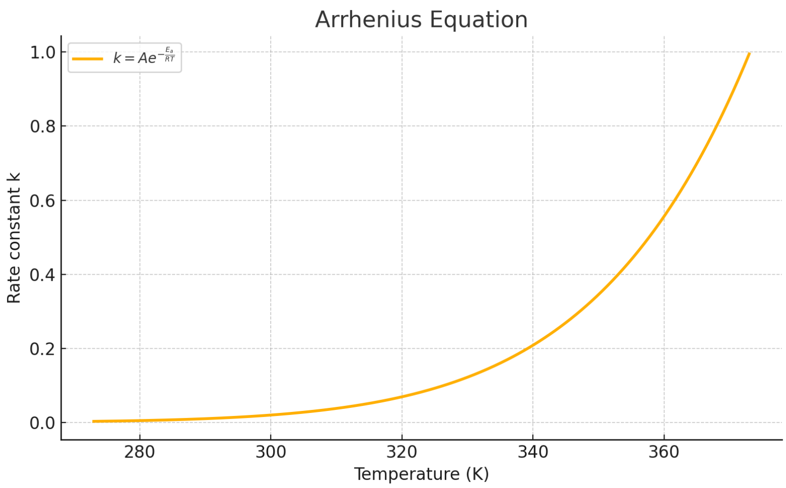

Figure 2.2: *An example plot over the Arrhenius equation.*

One way to address this is simply to include how diatom growth is responding to higher temperatures. This can be done in multiple ways, and there is no single approach which is the correct one. With that mentioned, the main thing to catch is that after some optimal temperature further increase in temperature should result in a decreased growth rate [13]. Some models incorporate coupled effects, where temperature not only directly affects growth, but also modifies how the organism responds to light. In these models, parameters that were previously constant such as the optimal light intensity can vary with temperature. This reflects the fact that the efficiency of light capture and energy absorption within the algae using photosynthetic rate is in itself dependent on the temperature [6].

While a coupled model often provide a more realistic description of how light and temperature interact in algae, they also bring a more complex behaviour of the model. Coupled models do often require fitting of several parameters from experimental data, which increases the risk of overfitting [6].

2.1.4 pH levels

pH levels also affects the growth rate, i.e the amount of biomass produced per time unit, of algae. Similarly to how light was described to affect the growth rate in section 2.1.2, pH levels from low to high do affect the growth rate similarly as a bell curve function. For every algal species it exists an optimum pH level in which it thrives. The further away from that level the pH diverges, the lower the growth rate [s^{-1}] will become. This optimal pH for growth is often somewhere in between pH 6-7, but the level depends on the specie [12].

There has been shown for two different types of diatoms, namely *Thalassiosira pseudonana* and *Thalassiosira oceanica*, that they are not being affected negatively by pH levels between 7.0 and 8.8 [5]. Moreover, another paper which researched the effect different pH-levels had on the growth rate of the diatom *Skeletonema costatum* found that it was approximately unchanged between 6.5–8.5. Furthermore, it decreased significantly at levels above 9.4 [22].

2.2 Modelling Biological Systems

Mathematical modelling of biological systems is the act of translating biological processes into numbers and equations. A well known example of this is the Malthusian population growth model from the late 18th century [11]. This is an early example of an ODE model over population growth, which is the main methodology utilised throughout this paper. Moreover, when discussing the growth of algae and diatoms, both logistic growth, which accounts for environmental limitations, and Michaelis-Menten kinetics, which captures saturation dynamics, are used for modelling of their growth.

2.2.1 Logistic Growth

Logistic growth is a population behaviour which is describing the growth of a population in an environment where there is a limiting factor, such as a set amount of resources or lack of space, which determines how large a population can become. This is used in a variety of field, including biology. When modelling a population which follows logistic growth it is often preferable to work with the change of growth over time, or in other words, the derivative of the population with respect to time. The expression for the derivative for a population which is growing logistically then becomes,

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right). \tag{2.2}$$

Observing this equation we find the following terms. N is the population which we are trying to describe. r is called the growth rate, which regulates the growth rate of the population and has the unit of s^-1. The term K is the carrying capacity, in some literature it can be written as N_{max} since it is the largest value which N can become before $\frac{dN}{dt}$ becomes negative.

This formulation of the logistic equation is continuous in nature, as it models population growth using a differential equation. That is, the change in population is smooth, assuming that the population N can take on any real positive value and changes in infinitesimally small increments. This is often a reasonable approximation in biological or chemical systems where populations are large and change gradually. For example, when describing biomass in industrial settings which often is measured in volumes or mass [15].

2.2.2 Michaelis-Menten Kinetics

Michaelis–Menten kinetics is originally a concept in enzyme kinetics, describing how the rate of enzymatic reactions depends on the concentration of the substrate. It also often appears in description of biological systems using ODE:s. The equation,

$$v = V_{max}\frac{S}{K_m + S}, \tag{2.3}$$

originally relates a reaction velocity v to a substrate concentration S, where V_{max} is the maximum reaction rate and K_m is the Michaelis constant, which represents the substrate concentration at which the reaction rate is half of V_{max}. This equation behaviour can be seen in the exemplary plot [20]. In the context of modelling bio-

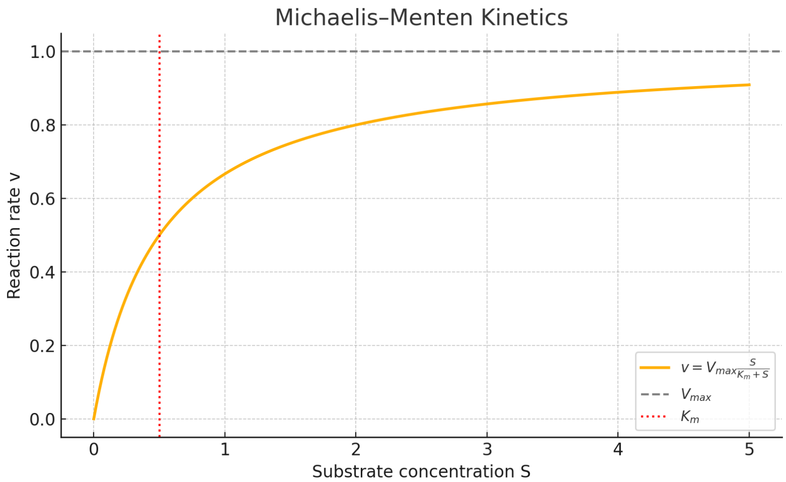

Figure 2.3: *Example plot viusalising the behaviour of Michaellis-Mentis Kinetics. Here $V_{max} = 1.0$ and $K_m = 0.5$.*

logical growth, such as algae cultivation, Michaelis–Menten kinetics often appears in modified forms to describe uptake of nutrients, sunlight, or other input parameters. It serves as a practical and interpretable equation in systems where detailed mechanistic data may be lacking. Because of this, it has become a standard component in models across many areas of biological and environmental modelling [20].

3
Method

In this section we will present the methodology of the project. As commonly needed in project including datasets, the work began with manipulating the dataset to a form which could be utilized through the whole project. Thereafter, due to the nature of this project, there was an iterative method with creating a model, using parameter estimation to fit it to the data, and thereafter evaluate the models usability. Finally, one of the main goal with this paper is to present actions which Swedish Algae Factory can implement to increment diatom yield. These considerations are primarily brought up in the discussion chapter, but are briefly introduced here as part of the overall methodology.

3.1 Data Cleaning

The first step was to clean the data received from Swedish Algae Factory. In the data there were mainly three different issues which had to be addressed.

- Firstly, missing data values needed to be addressed.
- Secondly, finding and handling so called frozen data values, i.e. values received from sensors which occasionally stops measuring new data and gives repetitive values, had to be addressed.
- Thirdly, identify gaps were it could be possible to interpolate missing data.

All of these issues basically come down to the same problem, which is finding a balance between data resolution and data accuracy. When deciding what resolution to use, the purpose of the data is the most important factor. In this case, where the goal is to model the growth of algae over weekly periods, a daily resolution for parameters such as sunlight, temperature, and salinity is not enough to produce an accurate model. On the other hand, a resolution of one measurement per second is more detailed than necessary. A time resolution somewhere between minutes and hours is therefore a reasonable compromise. Fortunately, the dataset contains measurements of important parameters such as pH levels and light intensity, measured in LUX, at a high enough resolution. Because of this, there is no need to interpolate between data points, which could otherwise lead to inaccurate values and ultimately to inaccurate model output.

3.1.1 Yield Data

As stated in the previous section, the resolution of the data in general were extensive. Which in this case implies that the time resolution is at least hourly. However, this is only true for the input parameters such as LUX, temperature, and pH levels. In the data there are two different output parameters, either yield or contamination. The time resolution on these parameters are only documented when harvest occurs at Swedish Algae Factory. Which effectively do result in a time resolution of once every third week. As a consequence, vital information of the algae growth behaviour is missing in the data.

3.1.2 The Cultivation Area

The diatoms is being produced inside Swedish Algae Factory's cultivation area in Kungshamn, Sweden. Details about the measurements and other metrics can not be mentioned due to non disclosure agreements brought up in the section *Company Confidential Details*. However, for the interested reader, one can observe figure 3.1 which showcase an image taken from within the cultivation area. The figure is showing layers of cultivations boards which each of them have diatoms growing on them. The harvesting system is half-automatic and the diatoms are being harvested once every third week.

Figure 3.1: *Image from Swedish Algae Factory's website of their cultivation area.*

3.1.3 Company Confidential Details

Due to non disclosure agreements between Swedish Algae Factory and the writer of this thesis, certain details can not be presented directly. The details which can not be presented in this paper can be divided into two categories. Firstly, the precise algae species which is being cultivated at the factory can not be disclosed. Secondly, numeric values which has been obtained through their sensors can not either be disclosed. This will limit some of the discussion and conclusion which can be made further on. As a consequence, the main focus will be on the development and generic results that the final model produce rather than specific details which could be connected to the input parameters such as numeric values from sensors.

3.2 Iterative Model Generation

The goal was to use data from Swedish Algae Factory to find an ODE model which describes the growth of the diatom species which are being cultivated at their factory. This since ODE models are widely used in biological systems due to their interpretability and ability to capture dynamic processes [15]. With this in mind, the approach was to work iteratively using the following methodology,

- Use the existing literature as guidance when developing the current model.
- Use either data or theory to fit the model parameters.
- Evaluate the model's accuracy and perform steady state analysis.
- Go back to the first step and either build upon the last model or generate a new one.

Since the cultivation area is limited in the factory and the diatoms are growing organisms the biomass should follow some sort of logistic population growth. With this as a ground truth, one can then build upon the model by adding multiplicative terms which corrects the growth rate based on input data, such as LUX or temperature. For example, let us call a generic logistic function of the diatoms $f(A)$ where A is the current diatom biomass. Thereafter, one can build upon that function using a function g for how another parameter γ is effecting the growth. As a result, one can find a function,

$$\frac{dA}{dt} = f(A)g_1(\gamma_1)g_2(\gamma_2)...g_n(\gamma_n) \tag{3.1}$$

under the assumption that g_i is independent of g_j where $i \neq j$. Occasionally there could be a function g which are dependent on more than one parameter. For two parameters it would look as follows $g_i(\gamma_{i_1}, \gamma_{i_2})$.

3.2.1 Parameter Estimation

During the iterative model generation, two different methods for parameter estimation were utilised. The first and primary idea was to use the julia package *PEtab.jl* to fit Ordinary Differential Equation models to the data [16]. However, during model development, the usage of this methodology for deciding parameters faced an issue. After some iterations, the problem formulation had more parameters to estimate compared to the available output data points. This was a concern since the estimation of parameters became an underdetermined problem. As a consequence, the *PEtab.jl* package were used for determining the parameters in section 4.1 and 4.2. Thereafter, the parameters were adjusted manually to generate a reasonable fit to the output data. Additionally, if there were theory backing up some estimations of the parameters, then this was also utilised.

3.2.2 Numerical Solution of ODE Models

Once parameter estimation was completed, the ODE models were solved numerically using the forward Euler method [3]. Although it is one of the simpler numerical methods available, it was sufficient for the data resolutions used in this work. The method works simply by advancing the solution one time step at a time according to

$$y(t+h) = y(t) + y'(t) \cdot h,$$

where h is the step size. To reflect the harvesting cycle at Swedish Algae Factory, A was reset to zero every third week. The forward Euler method were implemented in Python for generating the results which can be seen in Procedure and Results.

4

Procedure and Results

This chapter presents each model developed during the project, along with the motivation for its design and inclusion. The first one is a pure logistical ground model, representing $f(A)$ presented in equation 3.1. Thereafter each section will discuss how to build upon this model by including parametric functions with the form of $g_i(\lambda_i)$ as multiplicative terms, such as in equation 3.1. Throughout the chapter, the unit of A is in dm^3, which represents the volume of the potential harvest in the cultivation area.

4.1 Logistic Ground Model

As stated in the section *Iterative Model Generation*, the base assumption is that the diatoms are following a logistical growth model and that the growth is influenced by environmental parameters. Therefore, the original approach was to generate a ground model $f(A)$ which is growing logistically as in function 2.2. However, the logistical function is describing the growth of an population. The main goal is actually to describe how yield evolves, rather than modelling the diatoms themselves.

This is achieved by shifting the logistic function downward by a term L which is larger than 0. Doing that will result in the following function for the growth rate,

$$\frac{dA}{dt} = r(A + L)\left(1 - \frac{A + L}{K + L}\right). \tag{4.1}$$

This model is motivated by the fact that when harvest occurs, not all algae in the cultivation are being harvested. Some of the algae are being left behind so that they can produce more yield. Taking one step back to the original logistic growth equation 2.2, one can see that the term N must be strictly larger than zero for the population to be able to grow. This comes naturally from the fact that a non existent population can not reproduce itself. Beneficially, the final goal is to model how much yield which has been produced, which is different compared to the total amount of biomass. One solution to this is the substitution of N in equation 2.2 with $A+L$, where the term A is the amount of potential yield to be harvested at a specific time. The term L can be interpreted as the biomass which is being spared after harvest so that new biomass can be generated. The sum $A+L$ is then equal to the total biomass N. Henceforth, the notation N will be dropped and $A+L$ will be used in its place, exemplified in equation 4.1. The other substitution done is from K to $K+L$, which changes both the interpretation of what K represents and of the equation. Instead of symbolising the carrying capacity, K hereinafter will denote the maximum yield which the cultivation can produce before harvest. Furthermore, the term $K + L$ is now representing the max capacity of the cultivation in terms of pure biomass.

In simple terms, the model represented in equation 4.1 is a shifted version of a normal logistical model over population growth, since the model maintains the same behaviour. Importantly, the model 4.1 is only defined when $-L \leq A \leq K$, similarly to the original logistical equation 2.2, which follows from the fact that the biomass $A+L$ can not be negative and that the cultivation areas maximum capacity is $K+L$. When A is close to $-L$ the growth term will approach 0. Similarly, the growth term will approach 0 when A is approaching K. Additionally, the closer A is to $\frac{K+(-L)}{2}$ the larger $\frac{dA}{dt}$ will become, which is of importance when searching for the maximum value of $\frac{dA}{dt}$.

At this stage the parameter estimation was deferred, this since the model was not able to describe the growth of the diatoms in a representable manner. With that mentioned, the figure below is showcasing how A is behaving when only this function is used for describing the growth.

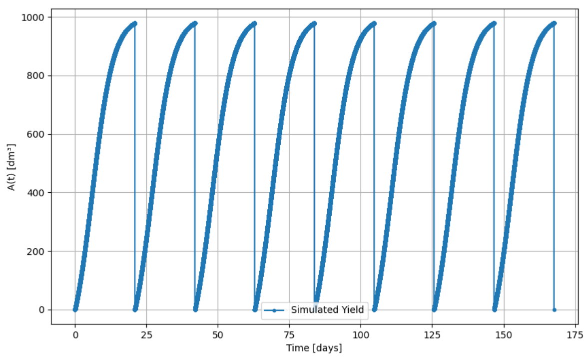

Figure 4.1: *Example showing how the model is behaving only using equation 2.2 for simulation using* $r = 0.27$, $L = 200$, *and* $K = 1000$.

In this example, note that harvesting occurs every third week which sets $A = 0$ which is done to mimic the harvest process at Swedish Algae Factory. The event of harvesting is manually set in the code for A. However, the parameters of r, L, and K were manually set and should in this scenario not be considered as potential values for the final model.

4.2 Light Term

The next step is to include the influence of light on cultivation. The approach was to use a model which indicates how light influence the rate of photosynthesis in the model [10]. This is because the rate of photosynthesis is highly coupled to the rate of growth. Here it is important to note that this model, mathematically identical to the Michaelis-Menten kinetics brought up in section 2.2.2, is one of a handful different models which is presented in the literature review *Modeling the effects of light and temperature on algae growth* made by Béchet, Shilton, and Guieysse [4]. With that mentioned, here is the model of photosynthetic rate brought up in the paper,

$$P(I_t) = P_m \frac{I_t}{I_K + I_t} - R_X. \tag{4.2}$$

Here the photosynthetic rate at time t, P, is being decided by P_m which is the maximum value P can become. The photosynthetic rate has units of $\left[\frac{\text{mg}\, O_2}{\text{g}_{\text{biomass}}\, \text{h}}\right]$. I_t is simply the light intensity at time t and I_K is a half constant. Both of these are measured in photosynthetic photon flux density (PPFD) which has the units of $\left[\frac{\mu\, mol}{m^2 s}\right]$. Lastly, the last term R_X is called the respiration term and has the same units as P and P_m. It indicates how much energy from the incoming light, or from stored chemical energy, which is used for respiration processes.

Considering equation 4.2 as a multiplicative term for the growth of the diatoms, then the original logistic equation 2.2 simply becomes the following,

$$\frac{dA}{dt} = r(A + L)\left(1 - \frac{A + L}{K + L}\right)\left(P_m \frac{I_t}{I_K + I_t} - R_X\right). \tag{4.3}$$

By introducing a new growth term $r_1 = r\, P_m$ and a new respiration term $R_1 = \frac{R_X}{P_m}$ one will receive a new equation $g_1(I_t) = \left(\frac{I_t}{I_K + I_t} - R_1\right)$. Here one can note that some parameters lose their original physical interpretation. However, the functionality with the equation will still remain the same with one less constant to think about, which is advantageous in our case. Therefore, the equation will then become,

$$\frac{dA}{dt} = r_1(A + L)\left(1 - \frac{A + L}{K + L}\right)\left(\frac{I_t}{I_K + I_t} - R_1\right). \tag{4.4}$$

While the original formulation in Equation 4.2 is expressed in terms of PPFD, the available data from Swedish Algae Factory was instead recorded in LUX. Although these units are not directly equivalent, LUX was used under the assumption that the behaviour of the model is not being affected by alternating between these units. Thus, the model parameters I_K has the unit of LUX and R_1 is becoming non-dimensional. The parameter values used to generate plot 4.2 were first generated using the *PEtab* library, then altered manually to a degree, since two of the data points seemed to be outliers.

4.2.1 Practical Interpretation

The last parenthesis in equation 4.4 is built up by two positive constants I_K and R_1 and additionally a non-negative parameter I_t. When no light is available this results in $I_t = 0$, which in extension yields $P(t) = -R_X$. Since this term is negative, and all

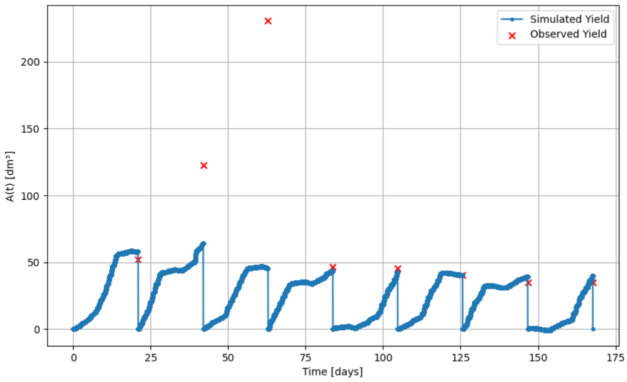

Figure 4.2: *Example simulation using equation 4.4 with* $r_1 = 0.27$*,* $L = 200$*,* $K = 1000$*,* $I_k = 30000$*, and* $R_1 = 0.015$*. The parameters* r_1*,* L*, and* K*, were set manually as constants. Thereafter,* I_k *and* R_1 *were generated using the PEtab library, then altered manually to a degree, since two of the data points used for fitting seemed to be outliers.*

the other terms is positive, this results in $\frac{dA}{dt}$ to be negative. This is mimicking the de-growth which comes from respirational processes which always occurs including periods of no light.

4.3 Temperature Term

When including a temperature term it can be incorporated as either uncoupled or coupled to the light term presented in the section above [4]. Furthermore, it could also be coupled to other terms affecting growth such as pH levels [12]. The approach used is largely up to preference. For instance, using a coupled model allows generating more complex behaviour which mimics biological growth in a more accurate manner. However, for this approach to be executed well one should have a good perception of growth processes. A more intuitive and straightforward approach is, as mentioned in the section *Iterative Model Generation*, to use uncoupled terms rather than coupled ones.

In the theory section *Temperature* it was discussed that the Arrhenius equation 2.1 could be used to describe a multiplicative term of how temperature affects growth of algae with the downside of not being able to mimic the behaviour at higher temperatures [7]. However, it has been shown that a curve similar to a bell-curve is a good approximation both at higher and lower temperatures for how temperature affects the growth rate of algae. For example, one could use the following statement based equation as an good approximation [9],

$$g_2(T_t) = \begin{cases} \exp\left[-K_A^T(T_A - T_t)^2\right] & \text{for } T_t < T_A \\ 1 & \text{for } T_A \leq T_t \leq T_B \\ \exp\left[-K_B^T(T_t - T_B)^2\right] & \text{for } T_t > T_B. \end{cases} \quad (4.5)$$

Equation 4.5 is using four species specific constants. K_A^T and K_B^T alter the derivative of the function near its lower and upper ends. T_A and T_B are the temperatures which in between there are optimal growth rate. An example of this multiplicative term can be seen in figure 4.3. Observe how a lower value of the K_B^T term compared to the K_A^T term makes the right side of the function not as steep as the left side.

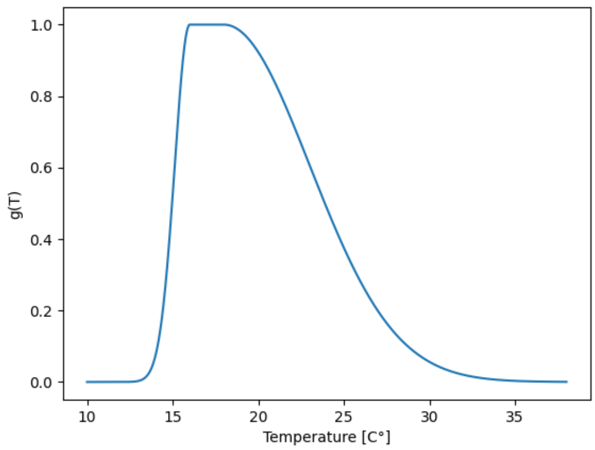

Figure 4.3: *An example of equation 4.5 using* $T_A = 16.0$*,* $T_B = 18.0$*,* $K_A^T = 0.65$*, and* $K_B^T = 0.02$*.*

The advantage of using this equation compared to an Arrhenius equation is that it captures the growth behaviour for large temperatures. With this in mind, including equation 4.5 as a multiplicative term in equation 4.4, and additionally, calling $g_1(I_t) = (\frac{I_t}{I_K + I_t} - R_1)$, one will receive the following,

$$\frac{dA}{dt} = r_1(A + L)\left(1 - \frac{A + L}{K + L}\right) g_1(I_t)\, g_2(T_t). \quad (4.6)$$

When equation 4.6 is simulated using the forward Euler method in Python with data from the Swedish Algae Factory, one will receive the plot in figure 4.4. As mentioned in the section *Parameter Estimation*, equation 4.6 has more parameters compared to the number of output data points, which makes a parameter estimation difficult. Consequently, an appropriate way to estimate the parameters is to formulate an informed hypothesis based on existing literature on diatoms about behaviour in different temperatures, and thereafter set the parameters so that the behaviour of the function mimics the theory. However, this approach could generate information about the diatom species used in production. Therefore, the parameters have been sporadically chosen in such a way that it gives an interesting result, without revealing information about the diatom species at Swedish Algae Factory. Keeping the other parameters similar as to earlier experiments, and since $g_2(T_t) \leq 1$, one should expect the generated yield to be lesser or as big as the same function without that term. Which can be seen is true when comparing figure 4.4 to figure 4.2.

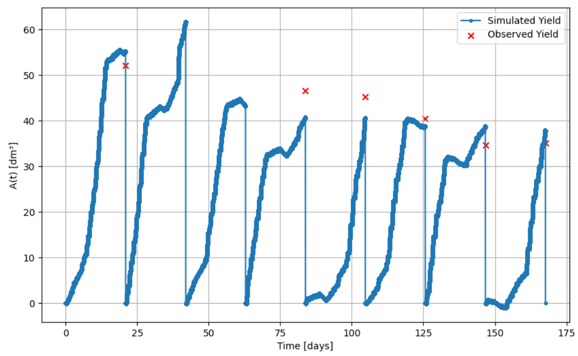

Figure 4.4: *Example simulation using equation 4.6 with* $r_1 = 0.27$*,* $L = 200$*,* $K = 1000$*,* $I_k = 30000$*,* $R_1 = 0.015$*,* $T_A = 16.0$*,* $T_B = 18.0$*,* $K_A^T = 0.65$*, and* $K_B^T = 0.02$*. Here,* r_1*,* L*,* K*,* I_k*, and* R_1 *is being copied from the section 4.2. The other parameter values are example values which describe a general but arbitrary behaviour for algae growth [7].*

An interesting dynamic can be observed between the terms $g_1(I_t)$ and $g_2(T_t)$. Since the term $g_1(I_t)$ can be both positive and negative and the overarching goal of the model is to maximizing yield. Then, if $g_1(I_t)$ is negative, there is an incentive to alter the temperature T_t such that $g_2(T_t)$ is close to zero. This would result in minimizing de-growth in periods of no light. Hence, according to the model this would be the most effective. However, as mentioned in section 4.2, it does not take into consideration the necessary respiratory processes which occurs in the diatoms during the night period. Alter the temperature during the night such that $g_2(T_t) \approx 0$ might have negative consequences on the growth which is not included in the model at hand.

4.3.1 pH Term

Based on the theory section regarding *pH levels*, there seems to be an indication that the optimal pH-level for diatoms is somewhere between 6 - 8.5. Furthermore, and without revealing specific information about the diatoms used, data from the cultivation reveal that the pH level was above 7 at least 99.8% of the time. Additionally, several data points of the measured pH level which were below 7 were significantly lower and only lasted for a short time, suggesting that they could be due to measurement errors.

As a result it seems reasonable to use a model which only assumes that the growth rate of the diatoms is not being affected by pH levels below 8.5. Thereafter, one can

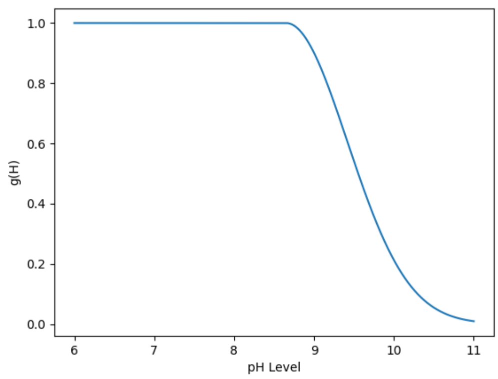

Figure 4.5: *Example image of equation 4.7 when* $H_A = 8.65$ *and* $K^H = 0.85$*.*

reasonably use a similar approach as made for the temperature term in equation 4.5. This results in the following equation,

$$g_3(H_t) = \begin{cases} 1 & \text{for } H_t \leq H_A \\ \exp\left[-K^H (H_t - H_A)^2\right] & \text{for } H_t > H_A. \end{cases} \tag{4.7}$$

In the equation above, H_A and K^H are constants and H_t is the pH level at time t. A reasonable approach based on the theory section would then be to set H_A to somewhere between 8.5 - 8.8 depending on the species. This is because it showed for three different diatoms that they were not effected negatively for pH levels up to this point, given that the level was not below 7. Image 4.5 is an example of how $g_3(H_t)$ could look like. Including pH levels in the simulation using equation 4.7 as an multiplicative term to equation 4.6 gives no major difference when using the parameter values of $H_A = 8.65$ and $K^H = 0.85$.

4.3.2 Nutrients Term

Currently, there is no available data on nutrient levels for the cultivation area. However, literature study is showing that apart from light, temperature, and pH levels, nutrients seem to be one of the most common limiting factors for algae growth. Therefore, a proposal is made for how limitation of nutrients could be incorporated in future models. One common approach is to state that the growth term depends on each nutrient following Michaelis-Menten kinetics [21]. Simply put, one could include a multiplicative term as the following for nitrogen,

$$g_4(N_t) = \frac{N_t}{K_N + N_t}. \tag{4.8}$$

This follow the same logic as presented in equation 2.3. The constant K_N represents the nitrogen concentration at which the reaction rate is half of optimal, while N_t is the nitrogen rate at time t. Following the same principle, one could generate $g_5(P_t)$ and $g_6(C_t)$ for phosphor and carbon. These terms would allow the model to account for possible nutrient saturation effects once such data becomes available.

4.4 Parameter Analysis

The final equation including the original logistic term and how light, temperature, and pH levels affect the growth rate came to look as follows.

$$\frac{dA}{dt} = r_1(A+L)\left(1 - \frac{A+L}{K+L}\right) g_1(I_t)\, g_2(T_t)\, g_3(H_t), \tag{4.9}$$

where $g_1(I_T)$, $g_2(T_t)$, and $g_3(H_t)$ are the following,

$$g_1(I_t) = \left(\frac{I_t}{I_K + I_t} - R_1\right),$$

$$g_2(T_t) = \begin{cases} \exp\left[-K_A^T(T_A - T_t)^2\right] & \text{for } T_t < T_A \\ 1 & \text{for } T_A \le T_t \le T_B \\ \exp\left[-K_B^T(T_t - T_B)^2\right] & \text{for } T_t > T_B, \end{cases}$$

$$g_3(H_t) = \begin{cases} 1 & \text{for } H_t \le H_A \\ \exp\left[-K^H(H_t - H_A)^2\right] & \text{for } H_t > H_A. \end{cases}$$

Observing this equation with the intention to find out which terms that have the biggest impact on the growth rate $\frac{dA}{dt}$ one can quickly find an interesting aspect. $g_1(I_T)$, $g_2(T_t)$, and $g_3(H_t)$ all have a maximum value which is lesser than or equal to one. Analysing equation 4.9 further, one do find that the parameters which has the largest effect on harvest are r_1, K, L, I_K, and R_1. The parameters K and L are both used to describe the scale of the cultivation area, when the cultivation increases, so do these. Additionally, it is intuitive that a larger cultivation area generate more yield.

A more interesting aspect comes when looking at the constants I_K, R_1 and r_1. R_1 is connected to respiratory processes and is the only constant making it possible for loss of biomass. The smaller R_1, the less energy is consumed for those processes. In extension, a lower value of this parameter would generate more biomass over time. In reality, R_1 is not constant but are coupled with other environmental factors such as light and temperature [19]. For example, it has been found that respiration decreases with lower temperature [2]. However, as mentioned in the section *Light*, respiration is also necessary for the algae to grow correctly. Therefore, inhibit the respiration process during the nighttime, with the intention to maximize growth, could as well have a negative impact on the growth process.

5
Conclusion

The purpose of this work was to investigate how cultivation data from Swedish Algae Factory could be used to identify key predictors of diatom growth using ODE modelling as the main approach. The analysis from chapter 4 highlights some conclusions which will be brought up in the section *Model Conclusion*. Moreover, limitations related to data resolution, particularly the low frequency on the output measurements, are discussed in the section *Data Limitation*.

5.1 Model Conclusion

The final model can be viewed in equation 4.9. Since there is no obvious method to use for validation of the model, it is not obvious what the model can be used for. In general, one can use the theoretical statements used as motivation for the models in the chapter *Procedure and Results* as springboard for which measurements to be done or not. For example, in the section *Parameter Analysis* one could conclude that even though the model is telling us that the optimal thing would be to regulate the temperature to minimize $\frac{dA}{dt}$ during night. However, this would probably inhibit necessary respiratory processes to occur and this effect on the growth is not captured by the model.

Additionally, one of the main ideas brought up during the theory which was not included in the final equation was the dependence of day and night cycles. Diatoms seems to grow optimally when there is repeated day and night cycles. Following the models approach, one would optimize growth by having the most light available at all time. However, this is not true, as could be found in the section Light.

5.2 Data Limitation

One of the issues with this project, also mentioned in the methodology section, was the lack of yield data. Since the yield, which directly represents the gain of biomass, only was measured after harvest this resulted in a resolution of the growth once every third week. This has resulted in an uncertainty of how accurate the model is in terms of diatom growth behaviour, specially in between harvests. To capture a more detailed behaviour of the growth there is a need of higher resolution on the output data than once every third week.

For example, in the paper *Simulating pH effects in an algal-growth hydrodynamics model* [9] the data on the change of biomass is being measured once every 24 hours. This makes it possible to more accurately capture daily behaviour of the algae growth, resulting in a more precise model. If there is a wish to capture how the growth is being affected by day and night cycles one would beneficially have an even higher resolution such as every 6 hours. The required detail of resolution depends on the goal of the model which is being generated. For a visual example, one can look at figure seven in the paper mentioned.

During the course of this work, the output resolution has been sufficient for the goal of getting an ODE model which could have the potential to predict how much yield the cultivation will generate in between harvest. However, the data has not been sufficient for creating a test set to verify the model against. Rather, the output data have been sufficient to do manual parameter tuning. Would the intention with the data be to do a more computational heavy method, such as machine learning methods, the amount of data points simply are insufficient. The intended application of the output data should be a key consideration for Swedish Algae Factory when determining the appropriate frequency for measuring biomass production during cultivation.

Bibliography

[1] Shamshad Ahmad, Kashifa Iqbal, Richa Kothari, Har Mohan Singh, Ahmet Sari, and V.V. Tyagi. A critical overview of upstream cultivation and downstream processing of algae-based biofuels: Opportunity, technological barriers and future perspective. *Journal of Biotechnology*, 351:74–98, 2022.

[2] Owen K Atkin and Mark G Tjoelker. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends in plant science*, 8(7):343–351, 2003.

[3] B N Biswas, Somnath Chatterjee, SP Mukherjee, and Subhradeep Pal. A discussion on euler method: A review. *Electronic Journal of Mathematical Analysis and Applications*, 1(2):2090–2792, 2013.

[4] Quentin Béchet, Andy Shilton, and Benoit Guieysse. Modeling the effects of light and temperature on algae growth: State of the art and critical assessment for productivity prediction during outdoor cultivation. *Biotechnology Advances*, 31(8):1648–1663, 2013.

[5] Edward G Durbin. Effects of ph on the growth and carbon uptake of marine phytoplankton. *Marine Ecology-Progress Series*, 109:83–83, 1994.

[6] Jiale Fan, Futian Li, Siyu Hu, Kunshan Gao, and Juntian Xu. Larger diatoms are more sensitive to temperature changes and prone to succumb to warming stress. *Limnology and Oceanography*, 68(11):2512–2528, 2023.

[7] Joel C. Goldman and Edward J. Carpenter. A kinetic approach to the effect of temperature on algal growth. *Limnology and Oceanography*, 19(5):756–766, 1974.

[8] Kasper Hancke and Ronnie N Glud. Temperature effects on respiration and photosynthesis in three diatom-dominated benthic communities. *Aquatic Microbial Ecology*, 37(3):265–281, 2004.

[9] Scott C James, Vijayasarathi Janardhanam, and David T Hanson. Simulating ph effects in an algal-growth hydrodynamics model. *Journal of Phycology*, 49(3):608–615, 2013.

[10] You-Chul Jeon, Chul-Woong Cho, and Yeoung-Sang Yun. Measurement of microalgal photosynthetic activity depending on light intensity and quality. *Biochemical Engineering Journal*, 27(2):127–131, 2005.

[11] Thomas Malthus. An essay on the principle of population. In *British Politics and the Environment in the Long Nineteenth Century*, pages 77–84. Routledge, 2023.

[12] Aloice W Mayo. Effects of temperature and ph on the kinetic growth of unialga chlorella vulgaris cultures containing bacteria. *Water Environment Research*, 69(1):64–72, 1997.

[13] David JS Montagnes and Montagnes Franklin. Effect of temperature on diatom volume, growth rate, and carbon and nitrogen content: reconsidering some paradigms. *Limnology and oceanography*, 46(8):2008–2018, 2001.

[14] Norio Murata, Shunichi Takahashi, Yoshitaka Nishiyama, and Suleyman I Allakhverdiev. Photoinhibition of photosystem ii under environmental stress. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1767(6):414–421, 2007.

[15] James D Murray. *Mathematical biology: I. An introduction*, volume 17. Springer Science & Business Media, 2007.

[16] Sebastian Persson, Fabian Fröhlich, Stephan Grein, Torkel Loman, Damiano Ognissanti, Viktor Hasselgren, Jan Hasenauer, and Marija Cvijovic. Petab.jl: advancing the efficiency and utility of dynamic modelling. *Bioinformatics*, 41(9):btaf497, 2025.

[17] Maria Ragni and Maurizio Ribera d'Alcalà. Circadian variability in the photobiology of phaeodactylum tricornutum: pigment content. *Journal of Plankton Research*, 29(2):141–156, 2007.

[18] Géraldine Sarthou, Klaas R. Timmermans, Stéphane Blain, and Paul Tréguer. Growth physiology and fate of diatoms in the ocean: a review. *Journal of Sea Research*, 53(1–2):25–42, 2005.

[19] Eleonora Sforza, Diana Simionato, Giorgio Mario Giacometti, Alberto Bertucco, and Tomas Morosinotto. Adjusted light and dark cycles can optimize photosynthetic efficiency in algae growing in photobioreactors. *PloS one*, 7(6):e38975, 2012.

[20] Bharath Srinivasan. A guide to the michaelis–menten equation: steady state and beyond. *The FEBS journal*, 289(20):6086–6098, 2022.

[21] Chayarat Tantanasarit, Andrew J. Englande, and Sandhya Babel. Nitrogen, phosphorus and silicon uptake kinetics by marine diatom chaetoceros calcitrans under high nutrient concentrations. *Journal of Experimental Marine Biology and Ecology*, 446:67–75, 2013.

[22] MORTEN TARALDSVIK and SVERRE M. MYKLESTAD. The effect of ph on growth rate, biochemical composition and extracellular carbohydrate production of the marine diatom skeletonema costatum. *European Journal of Phycology*, 35(2):189–194, 2000.

[23] Engel G Vrieling, Theo PM Beelen, Rutger A van Santen, and Winfried WC Gieskes. Diatom silicon biomineralization as an inspirational source of new approaches to silica production. In *Progress in industrial microbiology*, volume 35, pages 39–51. Elsevier, 1999.

A

Appendix

A.1 Data

This part will include information about the dataset that were produced by the Swedish Algae Factory. In short, the main thing presented are the different names of the column-data which were received.

Table A.1: Translation of data column names from Swedish to English

Original Column Name	**English Translation**
Datum	Date
Volym systemtank - A	Volume system tank - A
pH - A	pH - A
Antal aktiva banor - A	Active lanes - A
Tillfört CO2 - A	CO2 supplied - A
Konduktivitet - A	Conductivity - A
Temp - A	Temperature - A
Skördsump - A	Harvest sump - A
Volym systemtank - B	Volume system tank - B
pH - B	pH - B
Antal aktiva banor - B	Active lanes - B
Tillfört CO2 - B	CO2 supplied - B
Konduktivitet - B	Conductivity - B
Temp - B	Temperature - B
Skördsump - B	Harvest sump - B

Table A.2: Translation of data column names from Swedish to English

Original Column Name	**English Translation**
Datum	Date
CO2 - Gustav	CO2 - Gustav
Medeltemp 24h - Gustav	Average temp 24h - Gustav
Temp summa - Gustav	Temp sum - Gustav
LUX 1h - Gustav	LUX 1h - Gustav
kWh per dag - Gustav	kWh per day - Gustav
Ljussumma - Gustav	Light sum - Gustav
Veckodag	Weekday
vecka	Week
Ljus %	Light %
Ljus i watt	Light in watts
Beräknad korrigerad kWh per dygn	Estimated corrected kWh per day
Mätt ljuseffekt kWh per dag	Measured light effect kWh per day
Tot Ljus kWh	Total light kWh
Ljussumma Sol LUX	Light sum Solar LUX
CO2	CO2
Temp Vmedel	Mean temp V
Temp Vmax	Max temp V
Temp Vmin	Min temp V
TEMP SUMMA	TEMP SUM
Skörd (L pelletvolym)	Harvest (L pellet volume)
Diatom skörd	Diatom harvest
Glidande medel diatom 3	Moving average diatom 3
Glidande medel diatom 5	Moving average diatom 5
Kontamination	Contamination
Skörd ml	Harvest ml
Microbiol	Microbiological
Förväntad Algica 15% Yeild	Expected Algica 15% Yield

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

AVANCEZ
1829

CHALMERS
UNIVERSITY OF TECHNOLOGY