UNIVERSIDADE DE BRASILIA

Programmable photobioreactors for studies of microalgae growth dynamics

Author:
Ícaro Lorran Lopes Costa

Supervisor:
Luiz Roncaratti

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Institute of Physics

January 2, 2023

"Death is caused by swallowing small amounts of saliva over a long period of time."

George Carlin

UNIVERSIDADE DE BRASILIA

Abstract

Institute of Physics

Master of Science

Programmable photobioreactors for studies of microalgae growth dynamics

by Ícaro Lorran Lopes Costa

Most of the free energy available on the surface of the Earth comes from sunlight, and photosynthesis uses this natural resource to sustain almost all known forms of life. Such a process took billions of years to evolve and became the primary energy transformation process in our planet. Microalgae stand out among the many photosynthetic organisms because of their relative simplicity and variety. In this work, a network of computer-controlled photobioreactors for precise studies of the microalgae growth dynamics was developed and tested. These reactors can be used to characterize a given species or strain of microalgae as a function of the physical parameters that condition the development of the culture. They can also be used to study the microalgae response to stress conditions, adaptation to environmental changes, and the search for optimized culture parameters.

A fotossíntese é um processo que permite que os organismos vivos produzam energia usando a luz solar como fonte primária de energia. O processo depende do comprimento de onda dos fótons, que devem cair dentro da região de radiação fotossinteticamente ativa (PAR) do espectro solar entre 400 e 700 nanômetros, correspondendo aproximadamente à faixa de luz visível ao olho humano.

As microalgas são organismos fotossintéticos simples encontrados em vários habitats em todo o mundo, com uma estimativa de 30.000 a 70.000 espécies naturais. Eles são atraentes para uso em aplicações tecnológicas, especialmente se forem geneticamente manipuláveis, mas há desafios significativos para a produção economicamente viável de biocombustíveis a partir de microalgas. A evolução laboratorial adaptativa (ALE) é um método para melhorar o desempenho de uma linhagem de microalgas por meio do estudo de fenômenos evolutivos e adaptativos em um ambiente controlado.

A dinâmica de crescimento de uma cepa de microalga pode ser difícil de medir em ambientes naturais devido a flutuações nos parâmetros físicos, químicos e biológicos, mas esses parâmetros podem ser controlados em ambientes artificiais como fotobiorreatores (PBRs) para entender melhor suas influências individuais na dinâmica de crescimento. Os PBRs são dispositivos usados para conter e controlar culturas de microalgas e podem variar em complexidade, desde dispositivos analógicos simples até dispositivos mecatrônicos totalmente automatizados. Os PBRs podem ser usados para experimentos de longo prazo para estudar a dinâmica de crescimento de microalgas e otimizar a eficiência e a produtividade usando algoritmos de otimização como algoritmos evolutivos, que são adequados para problemas de "caixa preta" onde pouco se sabe sobre o comportamento analítico da função que está sendo otimizada.

Um sistema de controle foi desenvolvido para uma matriz de PBRs programáveis para ser usado como um aparato experimental para estudar a dinâmica de crescimento de microalgas. O sistema permite interação entre o usuário e cada reator, operação autônoma de acordo com regras pré-definidas e execução de algoritmos de otimização para explorar parâmetros físicos que influenciam o desenvolvimento da cultura. A matriz PBR consiste em sete PBRs idênticos desenvolvidos e construídos no Laboratório de Fotobiorreatores do Instituto de Física da Universidade de Brasília.

Um algoritmo genético foi implementado usando os PBRs como indivíduos com genes que representam frequências de intensidade de LED e uma função de aptidão baseada na taxa de crescimento do meio algal ou eficiência, definida como a quantidade de algas produzidas por quantidade de energia investida. A matriz PBR programável permite o estudo da dinâmica de crescimento de microalgas sob várias condições e a otimização do desempenho da PBR por meio do uso de algoritmos evolutivos. O sistema de controle e os algoritmos de otimização desenvolvidos neste estudo apresentam potencial para melhorar a eficiência e produtividade do cultivo de microalgas em PBRs.

Acknowledgements

My gratitude goes to Dr. Roncaratti and Dr. Gabriela, whose eagerness to teach and intermittent dedication to this project and many before it allowed me to complete it.

Contents

List of Figures

List of Tables

Dedicated to my Roomba, who's been cleaning the floor throughout the production of this manuscript

Chapter 1

Introduction

1.1 Photosynthesis

Most of the free energy available on the surface of the Earth comes from sunlight. Photosynthesis [1] uses this natural resource to sustain almost all known forms of life. Such process took billions of years to evolve within photosynthetic organisms [2] and became the primary energy transformation process in our planet [3, 4]. The conversion from light to usable energy begins with its absorption by highly structured protein complexes that contain chlorophylls and other auxiliary pigments [5]. Light absorption generates excitations that diffuse through the photosynthetic apparatus, eventually reaching the reaction centers where charge separation occurs. The resulting free electrons are moved through a transport chain, generating free Gibbs energy in the form of a proton gradient across the thylakoid membrane [6]. Finally, this energy synthesizes adenosine triphosphate (ATP), an organic compound that provides energy for various processes in living cells.

Light energy acts on photosynthesis in a similar way to electrical energy that makes a pump capable of accumulating water in an elevated reservoir [1] (see Fig. 1.1). In the case of photosynthesis, the accumulated chemical energy is maintained due to the metastability of the "organic matter+oxygen" system. In organisms, this intrinsically unstable system does not react spontaneously but in a controlled way through special channels controlled by catalytic enzymes that cause the accumulated energy to be released in small, gradual steps rather than all at once. The release of the accumulated energy at once would be equivalent to direct combustion, which, even not causing the organism to ignite, would, in addition to wasting energy, cause local damage. *Respiration* is the process that gradually releases the energy accumulated by photosynthesis in small amounts suitable for the organism's different physical and chemical functions. These small amounts of energy are transported by high-energy phosphate bonds, the most common form of which is the ATP molecule. This molecule releases about 10 Kcal/mol after being transformed into adenosine diphosphate (ADP) and inorganic phosphate (Pi) via hydrolysis. As we will see below, this phosphate bond energy is less than 10% of the energy accumulated in the elementary process of photosynthesis. In other words, it is as if photosynthesis delivered 100 bills and breath spent that amount in 10 coins.

Both the reactants and products of photosynthesis consist of the same number of atoms, C, H, and O. Each of these atoms is composed of a positive nucleus and surrounding electrons. When two atoms approach each other, the electrical interactions cause a rearrangement of the electrons. A repulsion between them means that the two electronic systems are more stable when the two nuclei are farther apart. The energy released or consumed in a chemical reaction is determined by the change in stability of the systems involved. Photosynthesis starts with a very stable arrangement of C, H, and O atoms in the CO_2 and H_2O molecules and ends with a much

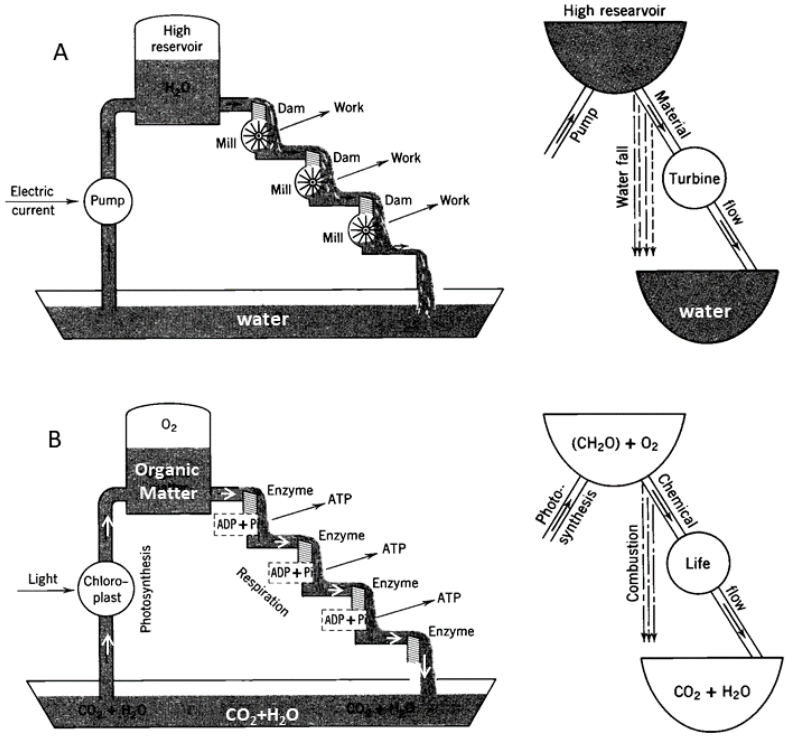

FIGURE 1.1: A - Pump and water wheels. B - Photosynthesis and respiration. Adapted from Ref. [1].

less stable arrangement of the same nuclei and electrons in the CH*system*.$_2$O+O_2. The main reason for this change in stability is that the oxygen atoms are weakly attracted to each other. The O_2 molecule is a relatively weak arrangement; its two oxygen atoms tend to bond more easily to carbon and hydrogen atoms.

The energy released when two atoms (or groups of atoms) A and B bond together is called binding energy. It is negative because when a bond is formed, the energy of the A+B system decreases.

As a first approximation, the total energy of a molecule can be considered the sum of the energies of its bonds, disregarding the cross interactions between atoms that do not form bonds. For example, the energy of the molecule H_2O can be considered the sum of the energy of two O-H bonds, whose energy is -110 Kcal/mol, totaling -220 Kcal/mol. The cross-interaction between the two H atoms is weak and can be neglected. The total binding energy of the CO_2 molecule, -380 Kcal/mol, can similarly be considered the sum of four C-O bonds, -95 Kcal/mol since each oxygen is linked to carbon by a double bond. In this case, the cross-interaction between the two oxygen atoms is also weak and can be neglected. The energy of each O-O bond in the O_2 molecule is only -58 Kcal/mol, totaling -116 Kcal/mol. Therefore, every time an oxygen molecule is consumed (in respiration or combustion), a considerable amount of energy is released due to the new, more stable arrangement of the oxygen atoms with the carbon and hydrogen atoms. Such a process is the primary explanation for the high energy accumulation in photosynthesis, where O-O bonds (-58 Kcal/mol) are formed, and O-H bonds (-110 Kcal/mol) are destroyed. Other bond changes also occur in photosynthesis, such as the change from O-H bonds to C-H bonds, but these contribute considerably less to the resulting change in energy than

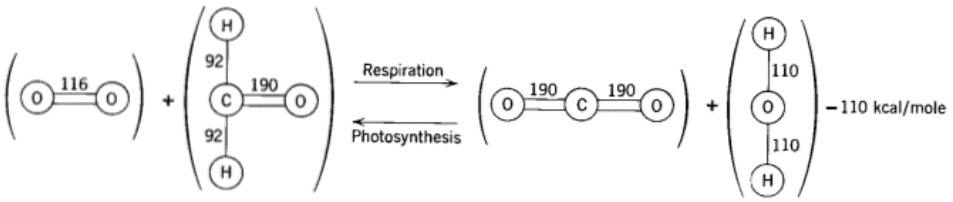

FIGURE 1.2: Bond energies and overall energy of photosynthesis and respiration. Adapted from Ref. [1].

changes from O-O bonds to C-O and O-H bonds. From this analysis, based on the conservation of bond energy, it results that the energy released by the combustion of the CH_2O molecule (basic unit of carbohydrates) is approximately 110 Kcal/mol (see Fig. 1.2).

Many free oxygen molecules surround all organic matter (carbohydrates) on Earth. Respiration enzymes control tiny gates that allow organic molecules, one by one, to encounter oxygen.

Unlike artificially constructed materials for solar conversion, natural photosynthesis is remarkably robust and adapted to continuous changes in environmental conditions due to various regulatory processes. In particular, photosynthetic organisms often experience sudden changes in the intensity of incident light over periods ranging from seconds to months, and the organism may receive an amount of light that exceeds its capacity of use in assimilating carbon dioxide. This excess energy can cause photo-oxidative damage to the photosynthetic apparatus due to the formation of reactive oxygen species (ROS) [7, 8]. In chloroplasts, there are two main pathways for the formation of ROS [9]: (i) electron transfer to molecular oxygen on the acceptor side of photosystem I (PSI) or II (PSII), leading to the formation of the superoxide radical (O_2^-) and, in subsequent reactions, hydrogen peroxide (H_2O_2) or hydroxyl radical ($OH\cdot$); and (ii) energy transfer from triplet chlorophyll ($^3Chl^*$) to molecular oxygen, which leads to the formation of singlet oxygen ($^1O_2^*$). In PSII, carotenoids play a crucial role in deactivating $^3Chl^*$ and $^1O_2^*$ and reducing ROS formation through thermal dissipation of excess light energy (*non-photochemical quenching*, NPQ) at the level of $^1Chl^*$. Beta-carotene is responsible for the deactivation of $^1O_2^*$ in the PSII reaction center, while xanthophylls are involved in the NPQ of excitations in the antenna complexes. These are examples of mechanisms employed by photosynthetic organisms to protect themselves from the adverse effects of high light intensity whose by-product are organic pigments.

In other words, photosynthetic organisms are remarkable examples of biological nanomachines that have dominated the process of harvesting sunlight over billions of years of evolution. Understanding such mechanisms is essential for realizing promising scientific and technological achievements.

1.2 Spectral irradiance

The Photosynthetically Active Radiation (PAR) refers to the portion of the solar spectrum lying between 400 and 700 nanometers, and roughly matches the visible light range to the human eye. It is a zone that photosynthetic organisms put to use in photosynthesis. Photons with longer wavelengths lack the necessary energy to trigger it, while photons at shorter wavelengths are filtered out by the ozone layer because

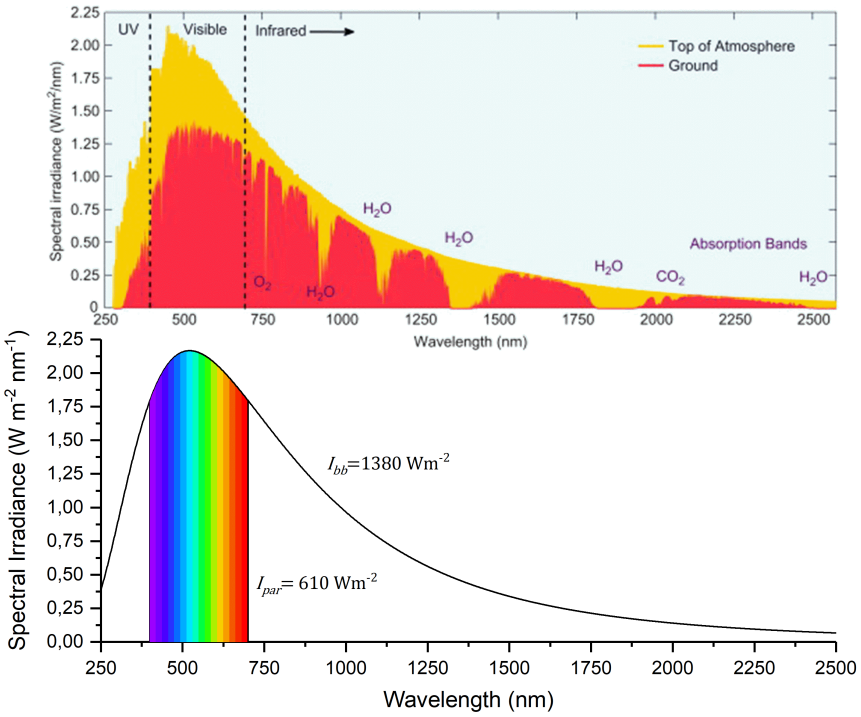

FIGURE 1.3: Solar Spectrum as a function of wavelength. Notice that the 400 to 700 nanometer range lies at the peak of the spectrum. Image credit: Handbook of Energy Volume I [11]

they are more energetic. Another reason that shorter wavelengths are undesirable for photosynthesis is that they can harm cells and tissues.

The irradiance of a given light source is the radiant energy emitted per unit time per unit of area. The SI unit of irradiance is the watt per square metre ($\mathrm{W\,m^{-2}}$). Irradiance is often called intensity, but this term is avoided in radiometry where such usage leads to confusion with radiant intensity. In astrophysics, irradiance is called radiant flux. Spectral irradiance is the irradiance of a surface per unit wavelength and is commonly measured in watts per square metre per nanometer ($\mathrm{W\,m^{-2}\,nm^{-1}}$) [10, p. 161].

Defining $S(\lambda)$ as the spectral irradiance of a given light source, its irradiance I is given by

$$I = \int S(\lambda) d\lambda. \tag{1.1}$$

In a naturally occurring setting, the light source comes from our Sun, whose spectral irradiance obeys Planck's function for the blackbody radiation curve to an approximate degree when atmospheric absorption is ignored [10, p. 72]. Such curve can be seen on figure 1.3.

1.3 Microalgae

Microalgae stand out among the many photosynthetic organisms because of their relative simplicity and variety. The biodiversity of microalgae is enormous. A conservative estimate puts somewhere between 30,000 and 70,000 natural species [13]. They are ubiquitous in nature [14], and their biodiversity is a valuable source of resources for humanity [15, 16]. Over the last decades, thousands of new components have been obtained from microalgae biomass, such as carotenoids, fatty acids,

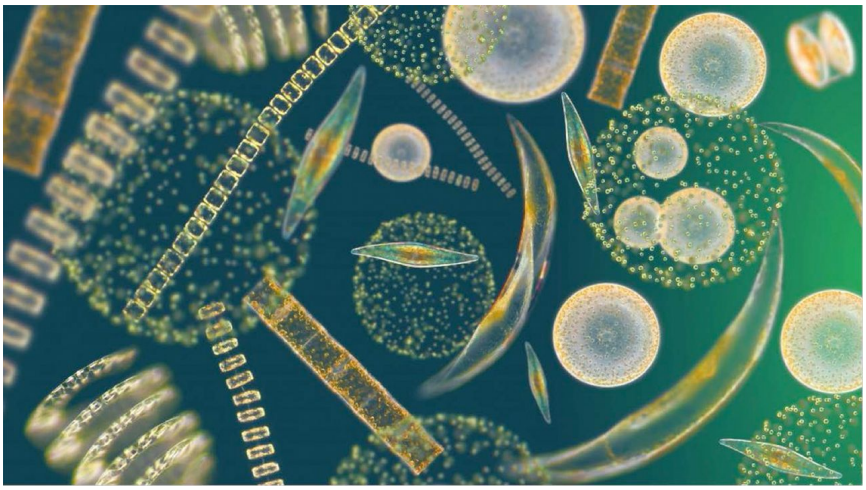

FIGURE 1.4: Example of microalgae. Image credit: Dr. Richard R Kirby [12]

antioxidants, toxins, enzymes, polymers, peptides, and steroids [17]. Furthermore, depending on the techniques used, microalgae can be converted into various types of fuels that are promising alternatives to fossil fuels and other known sources of biofuels such as sugarcane and corn [18–21]. Several companies and government agencies fund efforts to reduce the required capital and operating costs to make microalgae fuel production viable [22–24].

Microalgae-derived biomass has considerable advantages over traditional raw materials: high productivity, usually 10 to 100 times higher than planted crops; highly efficient carbon capture; grown in salt water, brackish water, and even wastewater and sewage. Moreover, microalgae can be harvested continuously throughout the year and integrated into a continuous production chain like in a traditional oil refinery. In addition, both cultivation and processing can occur at the same location, which favors sequential and integrated production, thus reducing logistical costs. However, there are still significant technological challenges for an economically viable production of biofuels derived from microalgae, such as the development and characterization of strains, crop protection, improvements in lipid harvesting and extraction methods, and efficient cultivation methods.

The diversity of characteristics and habitats make microalgae extremely attractive for technological applications, especially if the candidate lineage is accessible to genetic manipulation. Microalgae have numerous advantages over plants for producing recombinant proteins through transgenesis. They progress to large-scale production in a matter of weeks, which is a short time compared to months or years in the case of higher plants such as maize or tobacco. Microalgae are unicellular organisms, so there should be less variation in the accumulation of recombinant proteins, making the process more uniform. They can be grown in closed systems, thus reducing the risk of contamination in the production system and increasing the protection of the ecosystem close to potential contamination by genetically modified organisms, a recurring problem in vegetable crops in open or semi-open environments.

Another valuable technique for improving the performance of a given microalgae strain is adaptive laboratory evolution (ALE). This methodology allows the analysis of evolutionary and adaptive phenomena in a controlled environment. During microbial ALE, an organism is cultured under clearly defined conditions for long periods, allowing for improved phenotypes to be selected. Its basic principles go back

to the work of Anton van Leeuwenhoek, Louis Pasteur, Robert Koch, and Charles Darwin, with their discoveries about microorganisms, germs theory, and the importance of natural and artificial selection in the evolution and reproduction of organisms. Despite being carried out for a long time, the advent of new genetic sequencing technologies and modern control and automation techniques in recent decades makes ALE an even more promising methodology in biotechnology.

A fundamental feature of a given microalgae strain is its growth dynamics. In natural environments (oceans, rivers, lakes, and ponds), physical (light, temperature), chemical (molecules and ions dissolved in water), and biological parameters (viruses, bacteria, fungi, and predator-prey interactions) can fluctuate over time, making it difficult to measure the role of each of these parameters in growth dynamics. In artificial environments (photobioreactors), these parameters can be controlled and their influences individualized.

1.4 Brazilian scenario

Brazil has a vast coastal area, with more than 10,000 km, approximately 12% of the planet's freshwater, and receives an average solar irradiance higher than the portion received by most other countries. In addition, the country is home to the richest flora in the world (40,989 species, 18,932 of which are endemic), with more than 3400 species of algae cataloged [25]. Currently, several Brazilian institutions study and maintain algae collections, five of which have more than a hundred [26] lineages. Embrapa and Petrobras lead initiatives that characterize and domesticate native microalgae strains. While Petrobras focuses on using marine microalgae for biofuel production, Embrapa focuses on the characterization of continental microalgae and its uses in bio-products [27].

1.5 Photobioreactors

Photobioreactors (PBR) are devices used to contain and control a microalgae culture. Although several models have already been presented, the rational design of PBR still under development. A typical PBR is composed of a sample (microalgae) immersed in a culture medium under a constant mixture of air with the help of an air compressor. The sample is kept alight with photons of wavelengths between 400 and 700 nanometers, which defines the photosynthetically active region of the spectrum.

There are two types of PBR. The first type is called a refinery PBR, consisting of large reactors whose objective is to produce as much biomass as possible. To this end, they are designed to keep the crop in optimal conditions continuously and over long periods. This type of PBR design requires knowledge of the interactions between the sample and environmental parameters, and its operation requires expensive and specific installations. The second type is a benchtop PBR, whose objective is to study the growth dynamics of the sample and its dependence on environmental parameters. These reactors can be used to characterize a given species or strain of microalgae as a function of the parameters that condition the development of the sample, e.g., temperature, culture medium, dissolved gases, and light (intensity, spectrum, and modulation). They can also be used to study the sample's response to stress conditions, adaptation to environmental changes, and the search for optimized conditions. This type of equipment is used in several universities and research centers worldwide, each custom-made. There are also benchtop PBRs sold

FIGURE 1.5: Examples of various types of bioreactors. [28] [29] [30] [31]

by companies that, in general, are very expensive and, because they are designed as *user friendly*, have few configuration and operation options, presenting a questionable cost benefit. From a technological point of view, a PBR can be designed as anything from a simple, utterly analog device to a fully automated mechatronic device.

1.6 Autonomous PBRs

Ecology is the study of the relationships between living organisms and their environment. In artificial systems, like PBRs, the environment is defined by the reactor settings. This closed system has both organic (microalgae) and mechatronic parts (PBR with its sensors and actuators), where sensors monitors conditions and signals when changes occur, and actuator receives a signal and performs an action. Such a system can be viewed as a cyborg (portmanteau of cybernetic and organism) since the involved organic part may present enhanced abilities due to the integration with the mechatronic parts that relies on some sort of feedback. Cybernetics is a wide-ranging field concerned with regulatory and purposive systems. One of the most well known definitions is that of Norbert Wiener who characterised cybernetics as concerned with "control and communication in the animal and the machine" [32]. A programmable reactor open several possibilities for scientific studies and technological applications, in particular, stand out the use of PBRs as autonomous devices able to perform long run experiments accurately, to study the microalgae growth dynamics in changing environments and to perform optimization strategies in order to improve efficiency and productivity.

An optimization algorithm will generally use the historical input data, also known as training data, to approximate an optimal function by consecutively tweaking the control parameters in order to maximize a score associated with its performance. When little is known about the analytical behavior of such a function, we say that

the function is a "black box". Optimization algorithms that better perform with black boxes are stochastic. One class of such algorithms is evolutionary algorithms, which are inspired by the process of natural selection. The case of PBRs can fit better within this class as the interaction between their sample, and the device itself is complex enough to be regarded as a black box problem.

1.7 Objective and outline

The main objective of this work is to develop a control system for an array of programmable benchtop PBRs in order to obtain an experimental apparatus for studies of microalgae growth dynamics. Such a system should be able to: i) provide the interaction between the user and each reactor; ii) allow reactors to work autonomously according to predefined rules; and iii) execute optimization algorithms in order to explore the physical parameters that condition the development of the culture.

The methodology employed in constructing the PBR array and an overview of the algorithms developed for this project is discussed in the next chapter. Subsequently, we report on the outcome of our work in the "results" section. Lastly, a thorough analysis of the results, critique of the methodology, and proposal of the following steps are found in the "conclusion."

Chapter 2

Methodology

2.1 Microalgae

Microalgae from the genus *Chlorella* are eukaryotic, unicellular, free living photosynthetic organisms with a diameter ranging from 1 to 20 μm that can be spherical or ellipsoid [33]. They can be found in freshwater and marine environments, in soils and in the air, and can also colonize several invertebrate phyla such as Protozoa, Porifera, Platyhelminthes, Coelenterata, sponges or ascomycetes fungi in lichen [34]. In particular, *Chlorella sorokiniana* has been indicated as a promising species for biotechnology processes and applications [35]. Several studies have demonstrated that this species has a robust ability to produce biomass, proteins, sugars and high oil yields under a variety of conditions and can also be genetically manipulated [35]. Moreover, they can acclimatize to artificial conditions including those involving the use of LED light, high temperatures and low-cost urea, ammonia and nitrate based fertilizers [36, 37].

The microalgae *Chlorella sorokiniana* Embrapa LBA39, a Tropical strain isolated from the Brazilian Savanna (Cerrado) [38], was used in this work. This strain has been deposited in the Collection of Microorganism and Microalgae for Agroenergy and Biorefineries of the Brazilian Agricultural Research Corporation - Embrapa (Brasília-DF) and preserved through protocols described by [39]. Cultures were stored in liquid BG11 medium and kept in two liters PBR under constant temperature of 25 °C, with photoperiod of 12 h/12 h of neutral white (5000K) LED light.

2.2 Photobioreactors

Seven identical PBRs, developed and built in the Laboratory of Photobioreactors of the Institute of Physics of the University of Brasília, were used in this work. Each PBR is an independent mechatronic device capable of carrying out a series of actions automatically in order to growth microalgae in controlled conditions. They can be used alone or as part of a network. In the following, the system is described described in terms of its hardware and software components.

2.2.1 Hardware

The central part of the reactor is a borosilicate glass tube 20 cm high and 3 cm in diameter, with a total volume of approximately 140 ml (figures 2.1A and 2.3). Of this total, 100 ml is used for microalgae culture. The upper part of the tube is hermetically closed with a lid (2.1B) that has four tubes for inlet and outlet of air and liquid (culture medium and micro algae), a thermometer, and a pH sensor. The bottom of the tube is fitted to the holder attached at the base of the reactor (figures 2.1C and

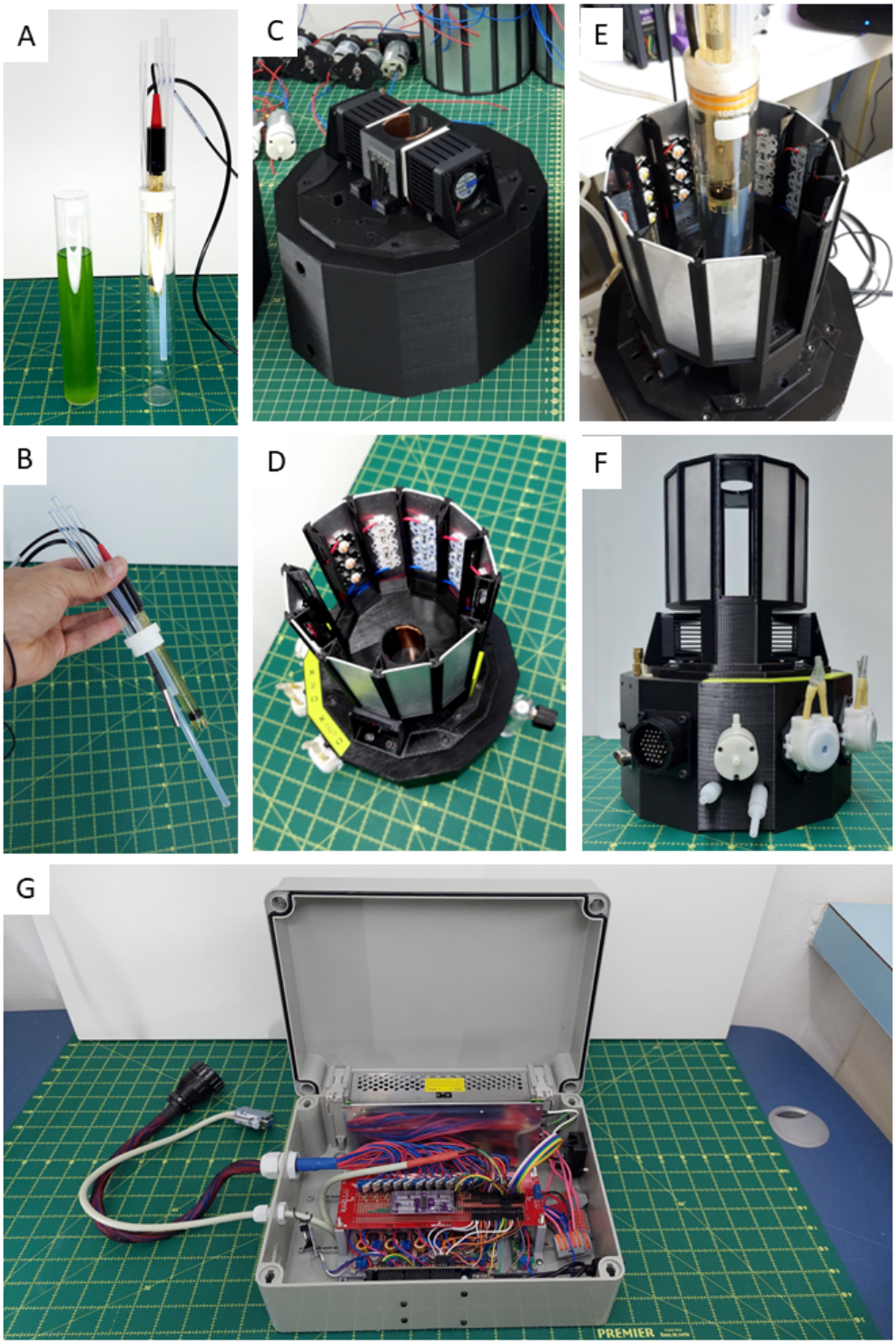

FIGURE 2.1

FIGURE 2.2

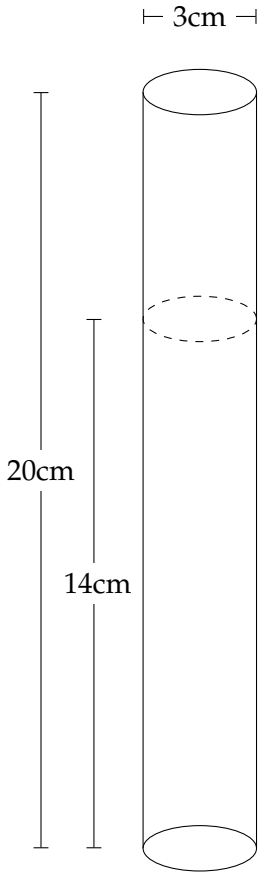

FIGURE 2.3: Diagram with the dimensions of the borosilicate glass tube

2.1E). This holder has an opening with a light sensor (TSL2591 Light-to-Digital Converter) and three LEDs with wavelengths in blue 470 nm, red 680 nm, and infrared 840 nm. The light scattering readings measured by the sensor for each wavelength can be correlated with the microalgae density and the chlorophyll concentration in the culture. In addition, this base has also two thermoelectric modules (TEC1-12706, 4cm x 4cm heat sinks and fans) attached on its sides to control the temperature of the glass tube (figures 2.1C and 2.2F) and a magnetic stirrer at its bottom to keep the microalgae culture homogeneous inside the tube during cultivation. This base is part of the reactor structure (figure 2.1F), which also contains an air compressor for the aeration of the culture, two peristaltic pumps for dilution and density control, electrical connectors, and the lighting panel. The compressor and peristaltic pumps are connected to the glass tube cap through silicone hoses. The lighting panel consists of 10 LED channels (figures 2.1D, and 2.2D), each having a different color in the visible region of the spectrum (440, 470, 495, 530, 595, 634, 660, 684, neutral white, and full spectrum). The intensity of each channel can be individually controlled; thus, a lighting spectrum with arbitrary shape and modulation can be programmed and used.

All these sensors and actuators are powered by an individual power supply of 200 W and correlated through a microprocessor board (Arduino MEGA2560) installed inside a control box (figure 2.1G). The control box is connected to a computer with a Linux system (Raspberry pi 3B).

Finally, each computer can be connected to an ethernet local area network, wired or wirelessly, and controlled individually or together from a computer with control

software. All data is stored and visualized in real-time (more details in sections 2.2.2 and 2.3).

The entire system, hardware, software, and mechanical parts, was designed and built by our group over the last five years. Some essential features of the system are (i) it makes use of free software and hardware; (ii) electronic components are low cost and easily accessible; and (iii) the mechanical parts were designed in CAD and built with a 3D printer.

2.2.2 Software

All photobioreactors are controlled by a microcontroller board (Arduino Mega 2560) running a custom firmware designed in our lab called Spectrum. This firmware allows communication to an external computer by spawning a shell via the Arduino's serial bus, which in turn can either be manipulated by a user or a bot. The LVP library is responsible for interpreting all the serial's incoming instructions according to its syntax, which listens for either a 'set' and 'get' commands to operate on one or multiple parameters or user-defined commands.

A parameter can be either an actuator or a sensor. Actuators can be divided into engines, such as the air and water pumps, coolers and the magnetic stirrer, the Peltier devices, or each one of the LEDs in the LED array. In contrast, the sensors get data from temperature; red, blue, and infrared light; CO2 levels; and the pH of the sample.

The commands can instruct the Arduino to execute specific pre-programmed tasks such as printing the PBR sensor and actuator state into the shell, automatically diluting the culture to a preset concentration of algae, switching the LEDs on and off over a 12h period, or automatically control the CO2 levels in the air mixture.

2.3 Programming and control

The serial protocol allows only a single transmitter to communicate to a receiver simultaneously. Connecting multiple PBR devices in parallel to a computer through their Arduino USB port, such as in a USB hub, may raise the possibility that more than one device attempts to send a message simultaneously; This generates a jumbled input that the receiver cannot decode, this effect is known as bus contention (see figure 2.4)[40]. This effect makes a robust long term network through a USB hub unfeasible to be established.

To avoid such problems, each serial can be connected to an intermediate computer acting as a middle-man, relaying the data over a more robust channel against bus contention, such as Ethernet or Wi-Fi. In this case, the data is transferred over HTTP, and the intermediate computer acts as a server. Another reason to implement this protocol is that it allows for easy scalability: The PBR devices could be sparsely placed in several geographical locations and be readily available for external researchers to retrieve data and submit custom experiments.

By adopting the HTTP structure, the network of our experimental apparatus 2.5 consists of a "main computer" attached to an array of several PBRs. Each contains an Arduino as a control unit and a Raspberry Pi as a computer unit and HTTP relay.

Since each PBR has a computer attached to them, it can also behave as a standalone HTTP server. However, when controlled by the "main computer," they are operated manually by a user or through a Python bot. If the bot is chosen, the user

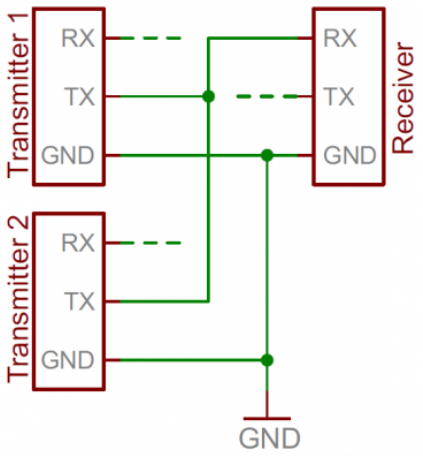

FIGURE 2.4: Example of topology that could cause bus contention [40]. Two transmitters (TX end) connecting to a single receiver (RX) may try to send a message at the same time, creating a superposed code that the receiver can't understand.

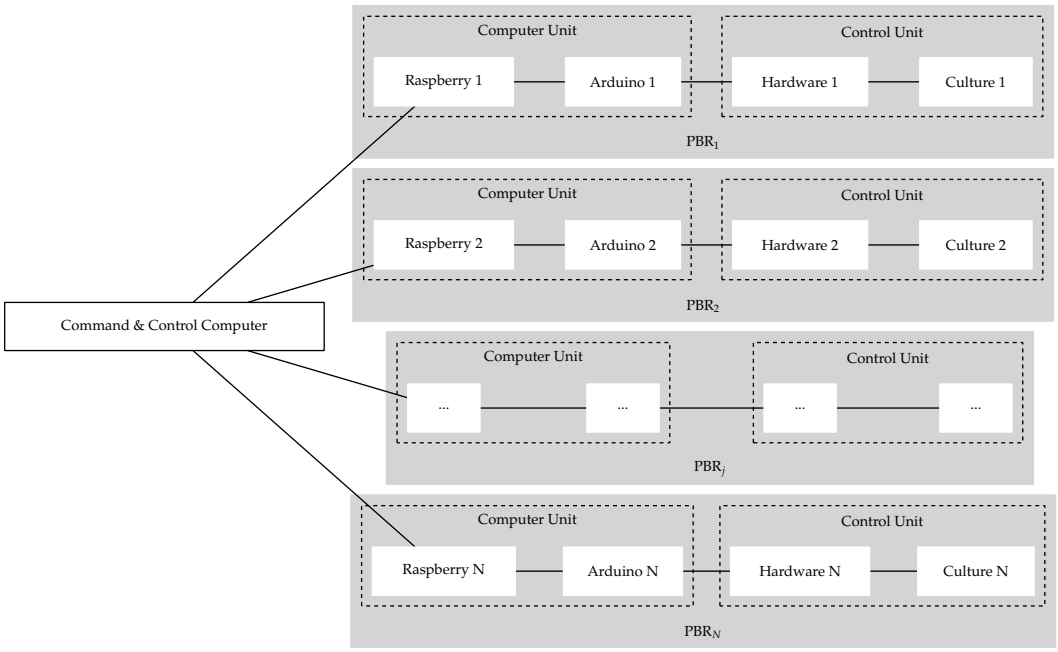

FIGURE 2.5: Network structure of the experimental apparatus.

can decide between letting the PBRs advance independently or optimizing a parameter with a built-in genetic algorithm.

Also, to account for possible power outages, the Python script controlling the Serial communication was installed as a systemd service ([41]), so the connection would be re-established automatically after a restart or system crash. While the state of the PBRs is constantly cached in the "main computer" for quick recovery when the system is brought back.

2.3.1 Optimization Algorithms

An optimization algorithm is a process capable of solving a minimization or maximization problem numerically. A plethora of areas ranging from Artificial Intelligence to Physics and Biology make use of these methods. The most common types are deterministic, designed to solve convex functions, i.e., functions that only present a single optimal point within the desired search space. There are many variants such as Gradient Descent, Nelder-Mead, or Newton-Raphson's method.

In the context of experiments, the function to be optimized can be either the "growth rate", defined as the amount of cells being generated per unit time over a volume element; or the efficiency, which is the ratio between the growth rate with the power invested into the system from the light panels. The unknown behavior of the growth rate function embodies a black box optimization problem.

In the case of functions presenting multiple optimal points, we use stochastic techniques. The implementation of randomness allows these algorithms to find global optima at the cost of becoming non-deterministic. The class of stochastic algorithms that we will focus on are the Evolutionary Algorithms, which receive this name for being inspired by the process of evolution of the species.

The Genetic Algorithm (abbreviated to GA) is a subclass of Evolutionary Algorithms. Although there are many possible implementations, our choice considers the search space embedded in a binary vector space of given dimension n ($\{0,1\}^n$). Within this context, an "individual" is a candidate for a solution to the optimization problem whose genotype is represented by a point in the aforementioned binary space (see figure 2.1). A collection of such individuals form a "population," which undergoes two types of operations along each iteration of the algorithm:

Firstly, a "crossover" operation selects pairs of individuals randomly from a population with a probability distribution based on the function's response to their genotype. Then, each pair exchange their corresponding vector components by selecting an index randomly and switching all components after that index with the components of the other individual (figure 2.2). After that, a "mutation" operation is applied over the resulting pair of individuals by flipping some of their vector components at random over a minuscule probability. A pseudo algorithm is presented later in the appendix A; see 1.

It may also be of interest to preserve the optimal solution across all iterations of the GA. In this case, we say that the selection process is "elitist" and will replace the individual with the worst performance with an intact version of the best one before the crossover step. When elitism is performed, there is a guarantee that the optima will evolve monotonically at each iteration, but at a cost that the population might get "trapped" at the vicinities of the elite individual, being unable to explore different solutions.

$$G = \begin{array}{c} \\ I_1 \\ I_2 \\ I_3 \\ I_4 \\ I_5 \\ I_6 \\ I_7 \end{array} \begin{array}{c} \overbrace{}^{x_1} \quad \overbrace{}^{x_2} \quad \overbrace{}^{x_3} \quad \overbrace{}^{x_4} \\ \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

2.1: Example of binary population matrix (G) tracking the state of an iteration of a genetic algorithm. Each individual is represented as a row while groups of columns are used to make a binary representation of the optimization parameters (x_1 to x_4).

$$\begin{array}{c} \\ I_a \\ I_b \\ \\ \bar{I}_a \\ \bar{I}_b \end{array} \begin{array}{c} \overbrace{}^{\text{cut}} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \\ \overbrace{} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \end{array}$$

2.2: An example of crossover operation carried at only two individuals. A random cut is made at the fourth column and values up to this column are switched between them, generating two more individuals.

Chapter 3

Results

The primary feature of the constructed reactors is that they are equipped with a programmable light panel that can cast light in the visible region. Other than the light panel, the reactors can be programmed to work autonomously and execute optimization algorithms. They can also monitor the density and growth rate of the inserted culture in real time through an optical sensor. With so many features and adjustable parameters, validation and calibration proved to be necessary.

In this chapter, it will be presented some of the results that allowed for validating the functioning and usefulness of the system.

3.1 Illumination panel

3.1.1 Modelling of the Spectral Irradiance of the LED Panel

The reactors' light panels are composed of ten color channels having different spectra. The channels are made of three identical 1.5 W LEDs, and the spectral irradiances of each channel were measured using a spectrometer (CCS200 from Thorlabs) and a light power meter (PM160T from Thorlabs). Since they were provided by the spectrometer in arbitrary units, they were converted to $\mathrm{W\,m^{-2}\,nm^{-1}}$ through normalization performed with the readings from the light power meter. The spectral irradiances obtained in this way are shown in Fig. 3.1 (symbols, scatters). In the panels of this figure, it can be seen that the spectral irradiance of each channel has asymmetric bell-shaped profiles that can be fitted with a function given by [42]

$$F(\lambda) = fL(\lambda) + (1-f)G(\lambda), \tag{3.1}$$

where

$$G(\lambda) = \frac{A}{\gamma(\lambda)}\sqrt{\frac{4\ln 2}{\pi}}\exp\left(-4\ln 2\left(\frac{\lambda-\lambda_0}{\gamma(\lambda)}\right)^2\right), \tag{3.2}$$

$$L(\lambda) = \frac{2A}{\pi\gamma(\lambda)}\frac{1}{1+4\left(\frac{\lambda-\lambda_0}{\gamma(\lambda)}\right)^2} \tag{3.3}$$

and f is the fraction of Lorentzian character contributing to the net line shape. For example, when $f = 1$, the shape is a pure Lorentzian and when $f = 0$, the shape is a pure Gaussian. The sigmoid function

$$\gamma(\lambda) = \frac{2\gamma_0}{1+\exp\left(a(\lambda-\lambda_0)\right)} \tag{3.4}$$

is used to take into account the asymmetry of the spectral irradiances. The parameter a is a measure of asymmetry. Negative values skew the spectrum towards higher

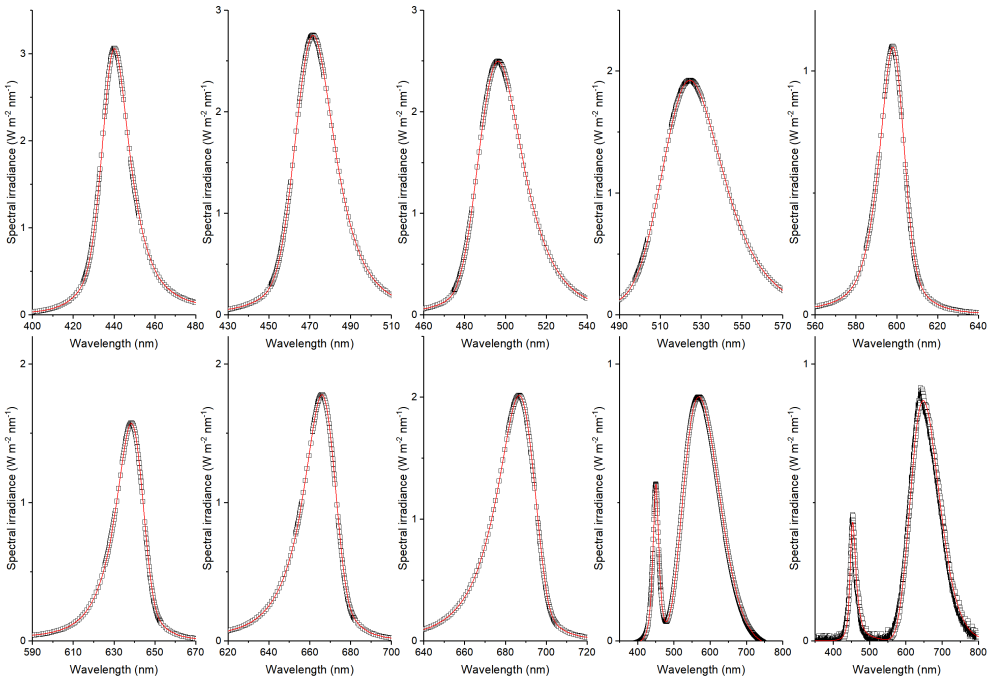

FIGURE 3.1: Curves fitting of the spectral irradiances for each channel of the Illumination panel according to function 3.1.

Wavelength (nm)	f	A (W m^{-2})	γ_0 (nm)	a (10^{-2}nm^{-1})	λ_0 (nm)	f	A (W m^{-2})	γ_0 (nm)	a (10^{-2}nm^{-1})	λ_0 (nm)	Irradiance (W m^{-2})
440	0.7	74.4	17.2	-3.6	441.0	–	–	–	–	–	72.2
470	0.6	88.3	24.1	−2.4	472.6	–	–	–	–	–	85.0
495	0.4	86.4	27.1	−2.5	497.8	–	–	–	–	–	82.2
530	0.4	89.6	36.4	−1.4	526.0	–	–	–	–	–	85.6
595	0.6	22.4	14.9	3.0	597.5	–	–	–	–	–	21.7
634	0.6	35.8	17.1	5.0	637.5	–	–	–	–	–	34.4
660	0.6	47.3	20.1	4.8	664.6	–	–	–	–	–	45.2
684	0.4	60.0	23.5	4.2	685.0	–	–	–	–	–	57.4
White	0.9	18.7	20.3	−1.1	450.7	−0.6	99.2	116.6	−0.4	575.4	115.1
Full	0.8	12.7	19.6	−4.1	454.8	−0.0	96.0	94.7	−0.5	652.3	103.5

TABLE 3.1: Parameters used for curve fit presented in figure

wavelengths, while positive values skew it towards lower wavelengths. When a is zero, equation 3.4 reduces to γ_0 and the resulting spectral irradiance is a standard symmetric Gaussian or Lorentzian profile with a constant width γ_0.

In total, five parameters were used to generate each curve fit in figure 3.1. However, the case for white and full-spectrum LEDs required two F functions to address their bimodality. The parameters for each curve, obtained using the Levenberg-Marquardt algorithm, are show in table 3.1. The obtained spectral irradiances for each channel are shown in figure 3.2.

Limited only by the power that the LED panel can provide, any given spectrum in the PAR region can be approximated with a linear combination of such curves. In mathematical language, let C denote the set of LEDs available and denote the function used to fit a LED of wavelength $j \in C$ as F_j. Also, define the real function S as

$$S = \sum_{j \in C} c_j F_j. \tag{3.5}$$

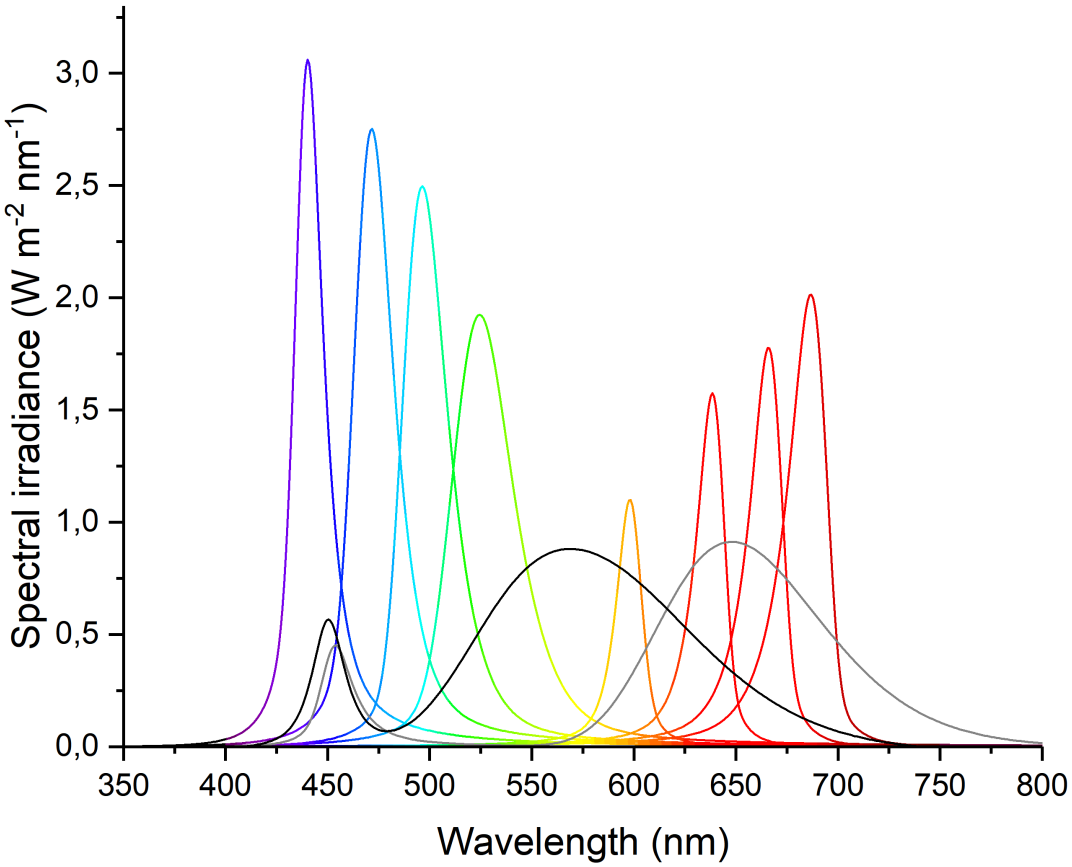

FIGURE 3.2: Spectral irradiances for each channel of the Illumination panel over the PAR region.

A proper approximation of an arbitrary spectrum S is one that satisfies the condition

$$S' = \text{argmin}_{X \in U} d\left(X, S\right). \tag{3.6}$$

Where the set U represents all real functions that can be written in the form of 3.5 (see figure 3.3) and d is a distance defined over the real function space. Furthermore, the set U must be constrained by the physical significance of the scalars c_j: As they are related to the PWM intensity of each LED, or rather, the fraction of power that each of them must provide, they cannot be negative and cannot exceed unit. This condition can be summarized in mathematical notation as follows:

$$U = \left\{ X = \sum_{j \in C} c_j F_j \middle| \sum_{j \in C} \sqrt{c_j}^2 < 1 \right\} \tag{3.7}$$

A particular example of an element of the set U is the "neutral" spectrum, which remains fairly flat within the visible region as it is illustrated in figure 3.4.

3.2 Density

The densities of the microalgae culture inside each reactor are obtained from the readings of the optical sensor at the 840 nm wavelength. The sensor reading at this wavelength is then related to the microalgae density through a calibration curve obtained with direct microalgae counting with a Neubauer chamber and microscope.

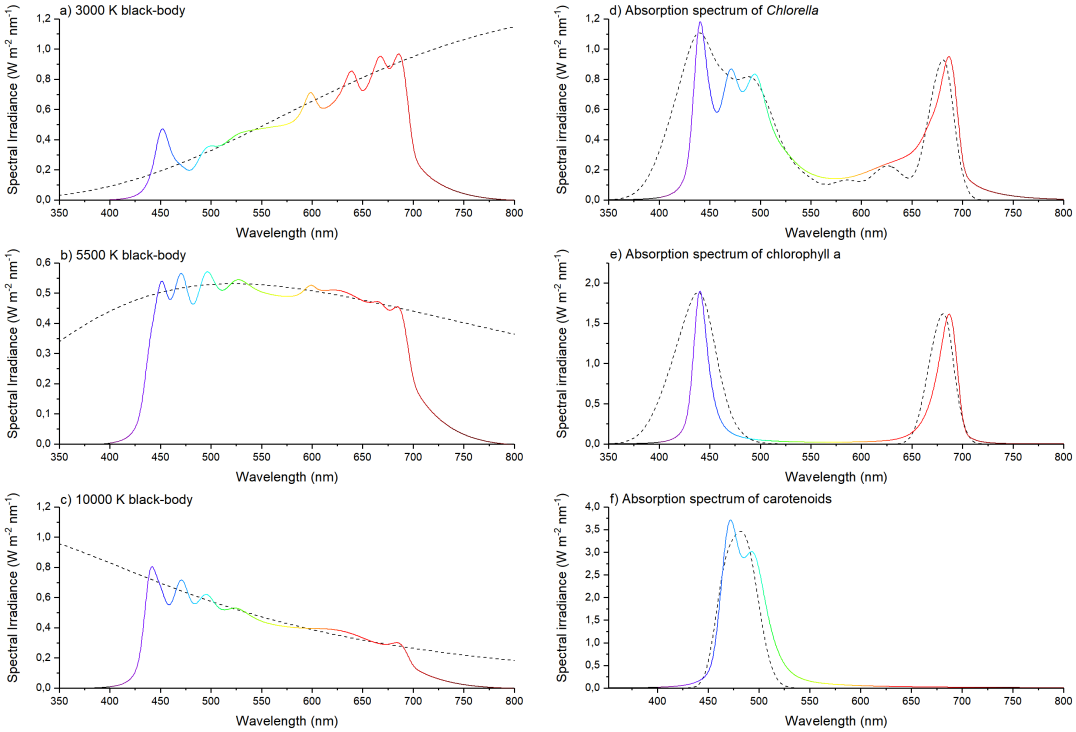

FIGURE 3.3: Some examples of spectral irradiances that can be produced by the illumination panels by the amplitude modulation of the curves shown in figure 3.2. The dotted lines represent the target spectra while the continuous lines are the actual outputs of the panels.

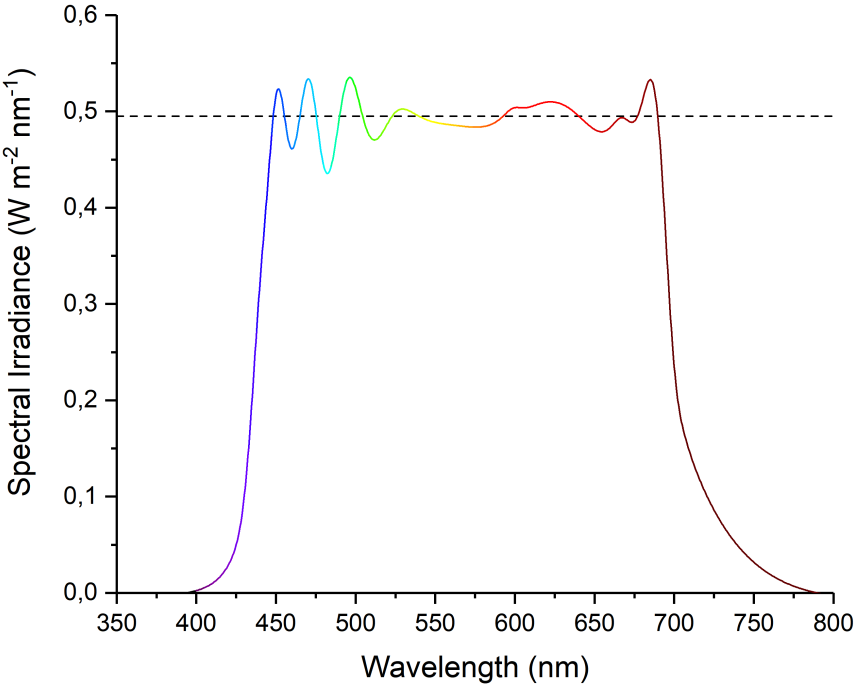

FIGURE 3.4: Neutral white spectra generated by the panel. Here is shown the curve with the maximum irradiance possible (130 W m^{-2}).

Reactor	a_1	t_1	a_2	t_2
1	71.76	354.32	8.00	5506.99
2	68.10	341.86	7.56	6009.17
3	69.41	312.08	7.85	4884.57
4	68.82	332.10	7.95	5353.53
5	72.15	325.08	7.47	4989.12
6	68.10	292.42	7.37	4406.68
7	70.45	355.07	7.84	5570.97

TABLE 3.2: Parameters obtained for the calibration curve 3.8 of each reactor

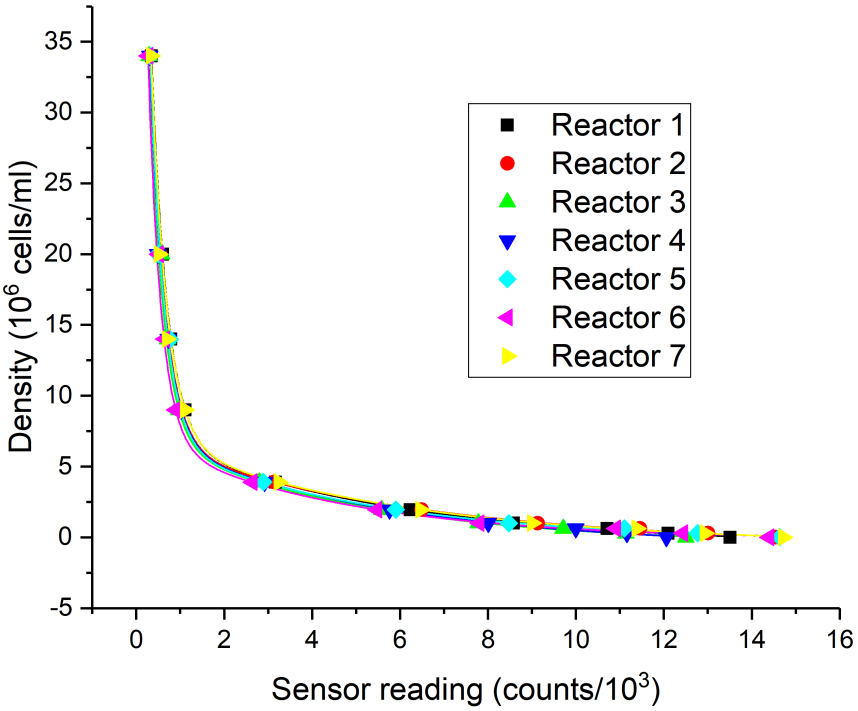

FIGURE 3.5: Calibration curves for microalgae density of each reactor.

The calibration curve of each reactor is shown in figure 3.5. They where obtained using ten different values of densities (between 0 and 35×10^6 cells/ml) and the corresponding sensor readings. Each measured data is fitted by a sum of two exponential curves, that is,

$$\rho(x) = a_1 \exp\left(-\frac{x}{t_1}\right) + a_2 \exp\left(-\frac{x}{t_2}\right). \tag{3.8}$$

Where ρ is the density and x the sensor readings at 840 nm and lies in the interval $(0, \infty)$. The best-fit parameters are given in table 3.2.

3.3 Growth dynamics

As a first approach to study the time evolution of the population of the algae sample, we subjected them under a square pulse of neutral white light of a period of 12 hours with the intent to simulate a day night cycle as it is illustrated in the in figure 3.6 A. The response to this cycle is tracked by the density time series in the plot of

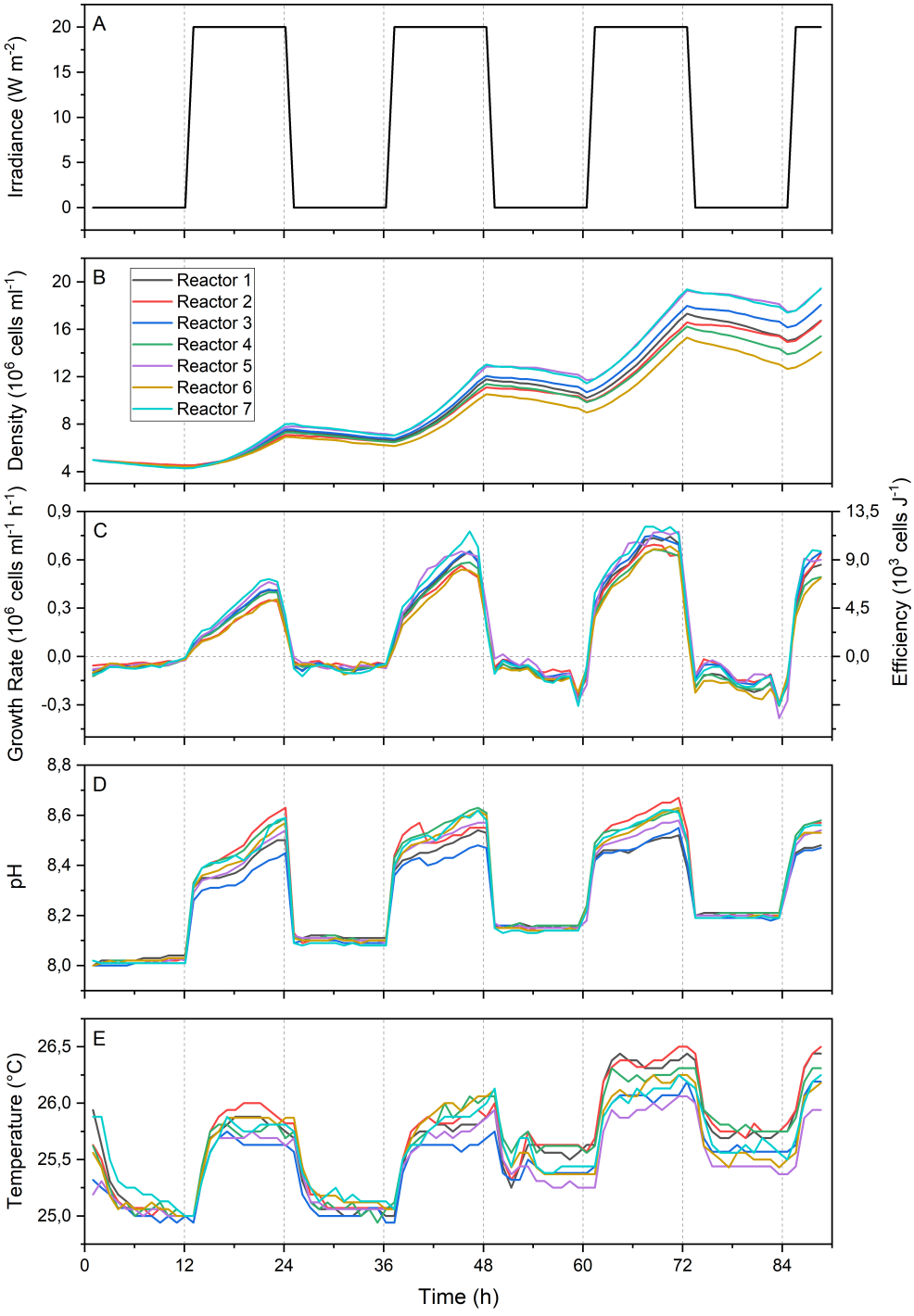

FIGURE 3.6: Growth dynamics of the culture under 12/12 light cycle of neutral white light. A) 12/12 light cycle. In panels B, C, D and E are shown microalgae densities, growth rates, pH and temperatures inside each reactor over time.

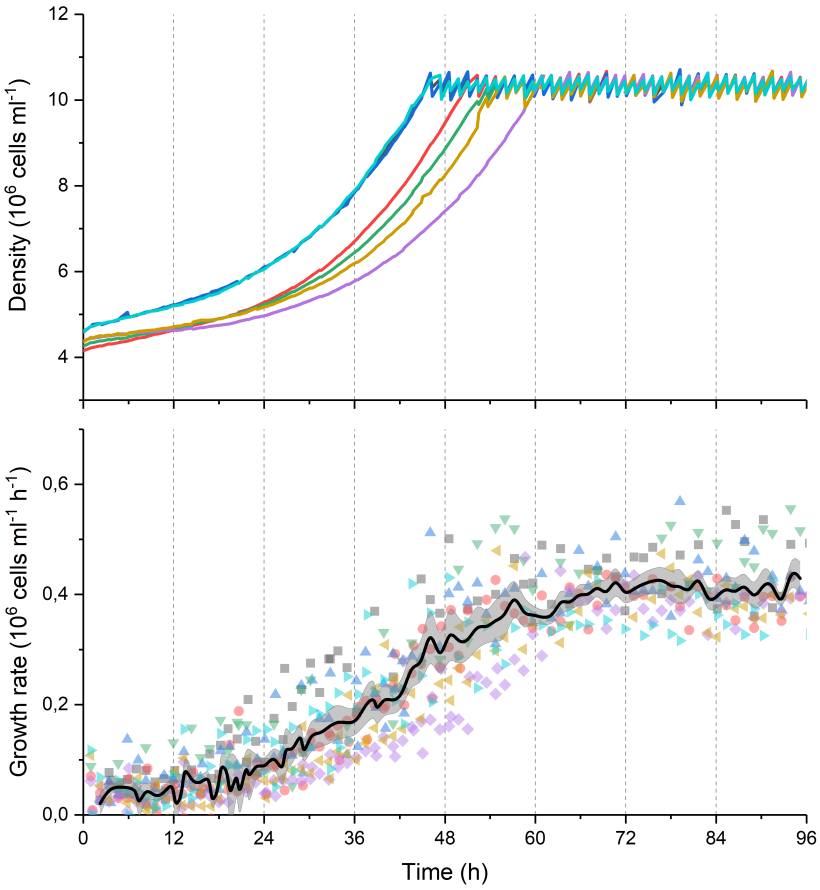

FIGURE 3.7: Upper panel: Reactors reaching a target density and keeping it for several hours. Lower panel: The corresponding growth rate for each reactor (symbols) and the average value (continuous black line). The shaded band are limited by the corresponding standard error values.

figure 3.6 B, in which it is noticed that growth occurs in periods where the light is at its maximum and the converse happens when it is turned off, as expected from a photosynthesizing organism.

A quantity of major interest is the growth rate, which can be obtained as the derivative of the density curve. As the plot C in figure 3.6 shows, the growth rate makes it clear to detect changes in the culture's behavior. A co-variable with similar significance is the pH of the culture's medium, which highly correlates with the growth rate, as it can be seen on the plot in figure 3.6 D.

The reactors also have a mode of operation in which they'll seek a predefined density and keep it at a quasi-static regimen. Figure 3.7 demonstrates the time evolution of the reactors in this mode, starting off from a density that is far below the target. Also in this case the reactors were set with 20 $W\,m^{-2}$ of neutral white light and the cultures have growth freely until they reach the predefined density. Whenever the density exceeds the target, the reactor dilutes the culture with more medium and discards some of the resulting mixture to keep the volume of culture constant. This procedure will repeat until the density reaches the target or a point below it. Figure 3.7 also shows the growth rates of each reactor and its average value over time.

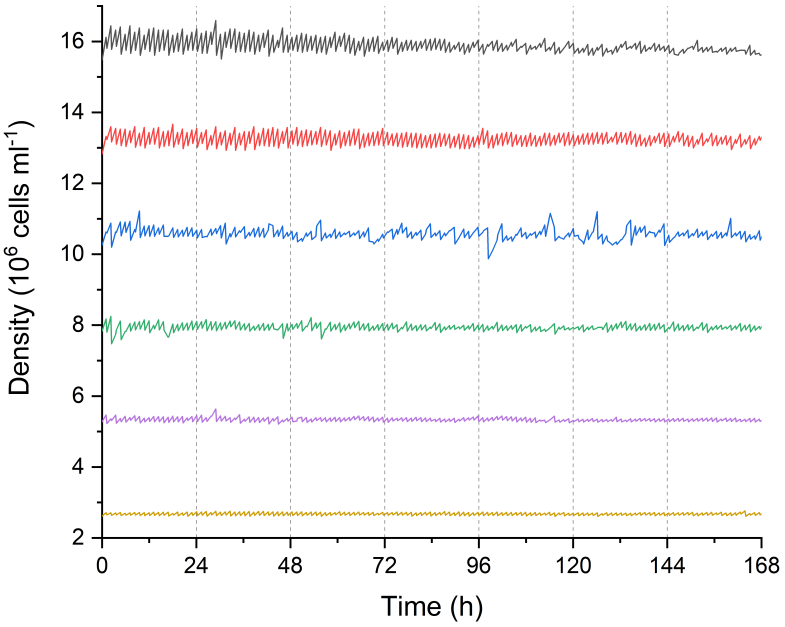

FIGURE 3.8: Densities kept constant by the reactors over the course of a week.

The constant density mode of operation can be executed in any range achievable by the PBRs. Some examples are demonstrated in figure 3.8, where the targets were kept constant over the course of a week.

In order to study how the microalgae respond to different incident irradiance, the reactors were programmed to vary the irradiance of the neutral white light over time accordingly with the pattern shown in the panel D of figure 3.9. For each irradiance value, in the increasing and in the decreasing part of the irradiance curve over time, the growth rate and the pH in each reactor were measured and the values are shown in panels A and B of figure 3.9. The efficiencies, defined as

$$E = c\frac{k}{I}, \tag{3.9}$$

where, k is the growth rate and I the irradiance, are shown in panel C. The constant $c = 0,3$ ml m^2 is the product of the reactor volume (100 mL) and the illuminated area (0,003 m^2), see figure 2.3. In this way, the efficiency can be represented in cells/Joules.

3.4 Analysis of Convergence of the Genetic Algorithm

Due to the stochastic nature of the genetic algorithm, there is no deterministic way to tell how long it would take for it to converge. However, we can perform a statistical analysis in a simplified version of it to estimate the time of convergence. The appendix A provides an estimate by only taking into account selection over a constant fitness function. For a population of size N with l genes and G_0^k ones on gene k, the average hamming distance $\bar{h}_{t+1}$ is given by equation A.9:

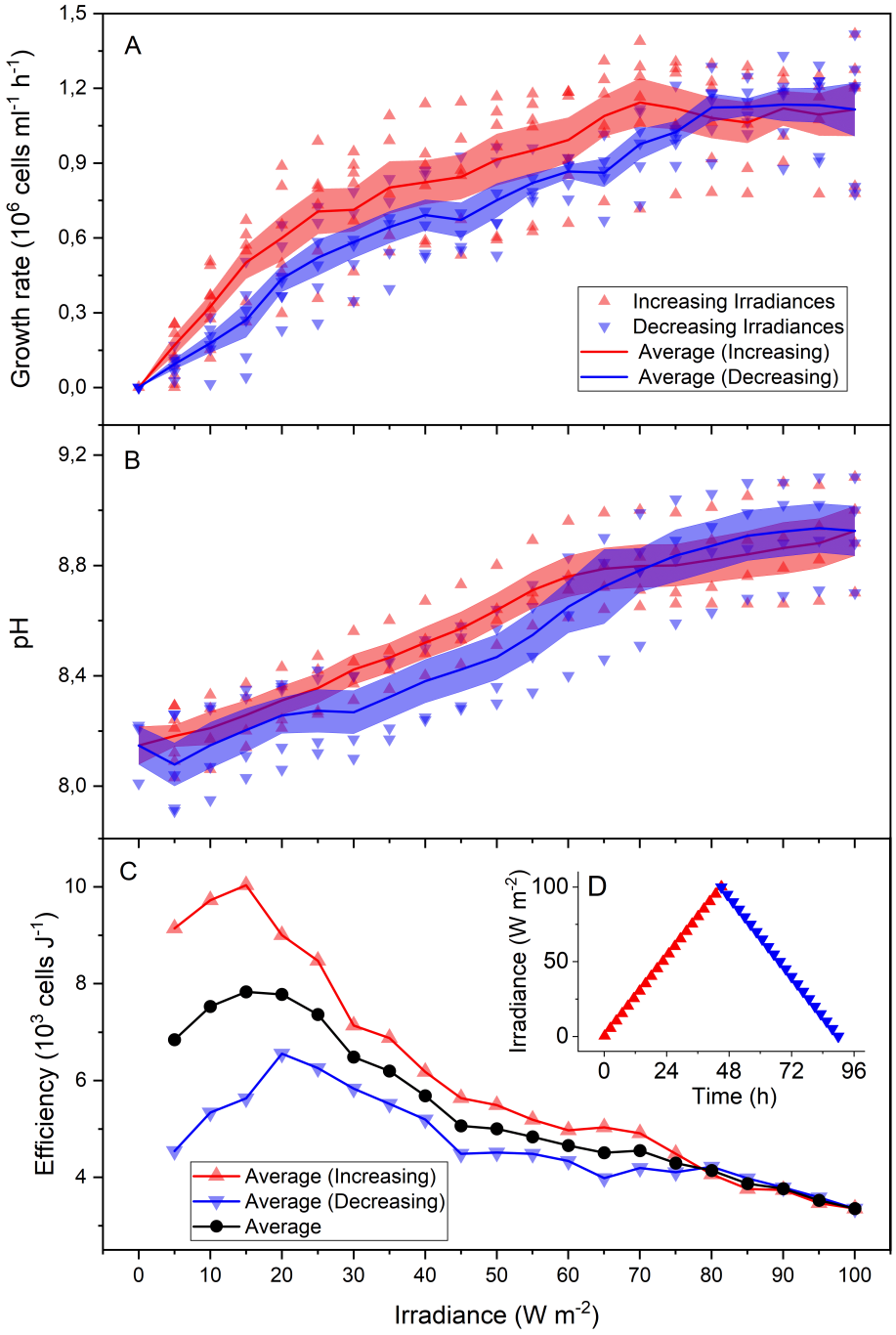

FIGURE 3.9: Growth rate, pH and efficiency as function of neutral white light irradiance values. The inset D shows how the irradiance values were varied over time. In panel A and B, symbols represent the measurements of each reactor and continuous lines the average values. Shaded bands are limited by the corresponding standard error values.

$$\bar{h}_{t+1} = \frac{1}{l}\sum_{k=1}^{l}\left(G_0^k\left(1-\frac{1}{N}\right)^t\left(N - G_0^k\left(1-\frac{1}{N}\right)\right)\right)$$

We will be using this measure as an indication of homogeneity of the population. An average hamming distance of zero means that the population is completely homogeneous, and the algorithm has completely converged. In a realistic scenario, this measure won't monotonically converge to zero due to mutation, but it can be used to estimate when it is a good time to stop the iterations.

By running the genetic algorithm repeatedly over a fixed random initial population, we can check the validity of equation A.10 with numerical results. To align with the case concerning this project, we ran a simulation with a population of size eight with 64 genes over 500 epochs. In figure 3.10, we can see that the decay is nearly what equation A.9 describes. When we include mutation, however, the hamming distance no longer converges to zero. Instead, it oscillates with low amplitude, as in figure 3.11. Whether looking at the numerical ensemble or at the theoretical estimate, the algorithm seems to converge within 20 iterations.

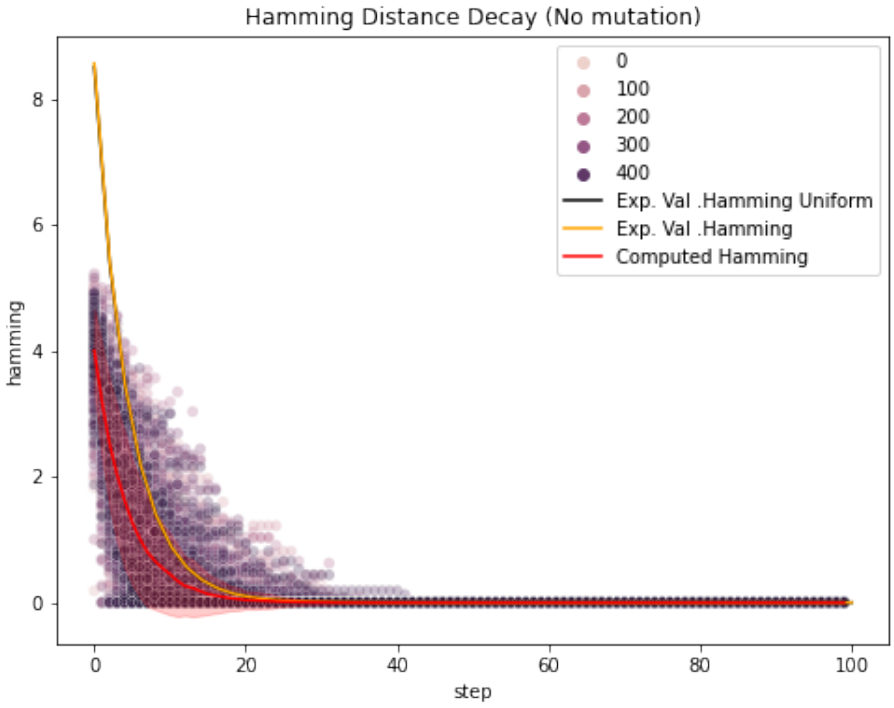

FIGURE 3.10: Hamming distance decay over time for a population of 8 individuals and 64 genes. The points in the scatter plot were obtained by running the algorithm 500 times with no mutation.

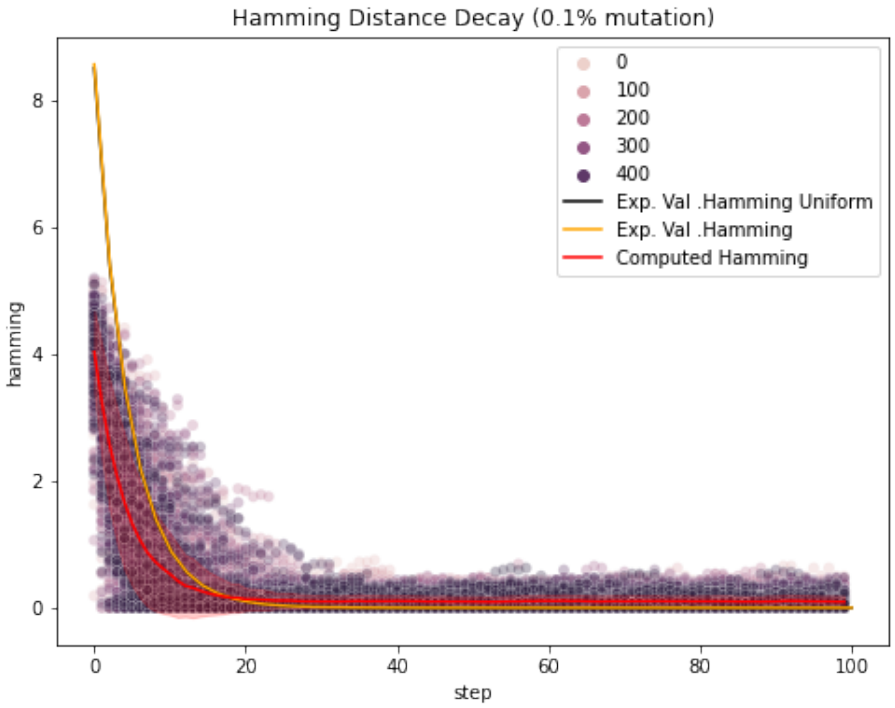

FIGURE 3.11: Hamming distance decay over time for a population of 8 individuals and 64 genes. The points in the scatter plot were obtained by running the algorithm 500 times with a mutation rate of 0.001.

3.5 Genetic Algorithm Optimization

The optimization procedure involved the use of only irradiance of the neutral white light as the dependent parameter and the growth rate as the quantity to be optimized. The mutation rate was set to be 0.1 and the population started homogeneously with an irradiance of $6\,\mathrm{W\,m^{-2}}$. We also keep track of the efficiency and irradiance as the experiment develop over the course of 4.5 days (figure 3.12).

Initially, the system rapidly reaches an apparent maximum within the first 12 hours, with an irradiance of $15\,\mathrm{W\,m^{-2}}$. Soon after this point, the performance of the system slowly degrades over time, even though it is still set to maximize the growth rate.

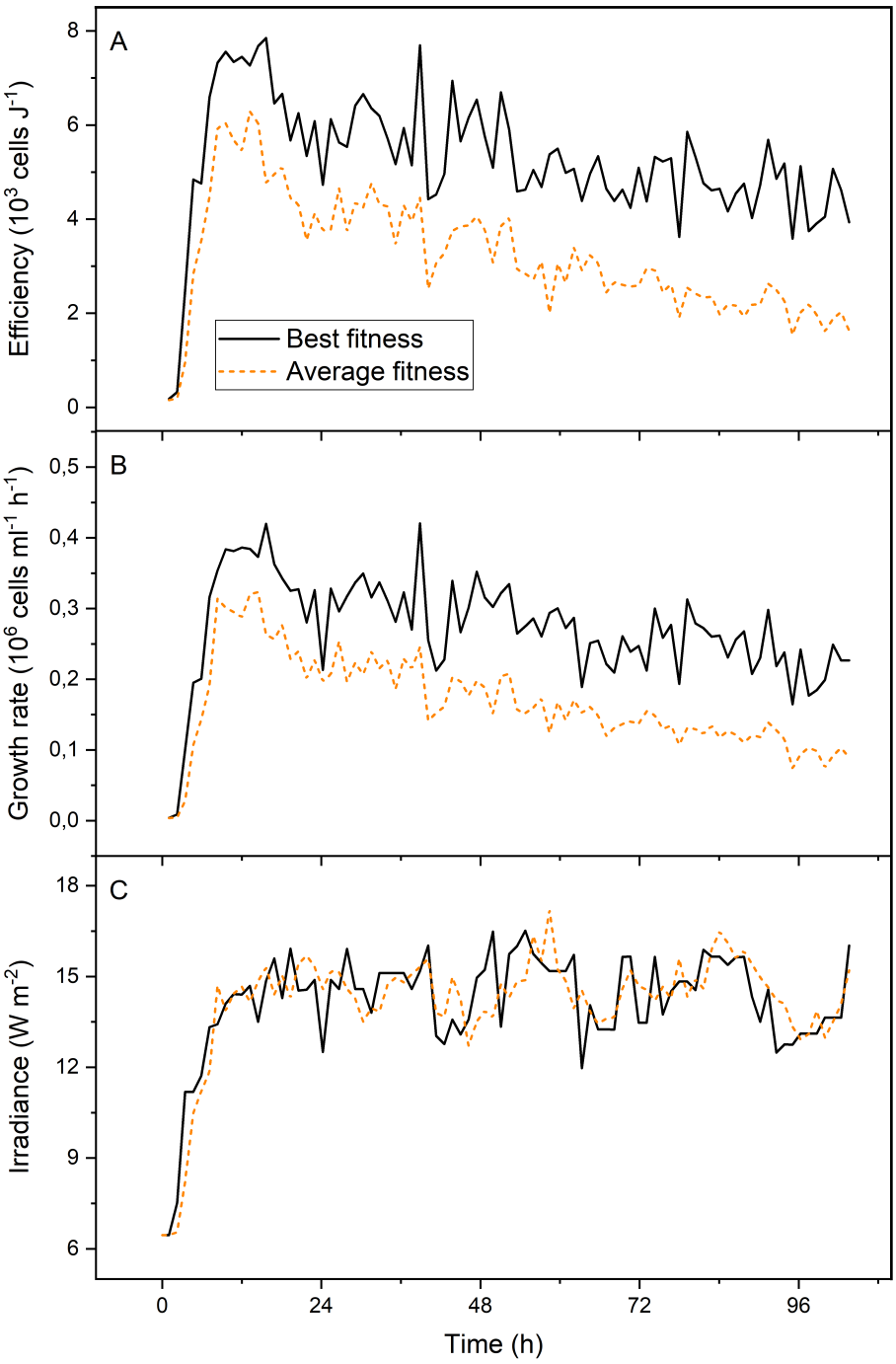

FIGURE 3.12: Time evolution of the genetic algorithm running with only white light. The dashed lines represent the average over all bioreactors, while the black solid line is the maximum at each instant.

Chapter 4

Discussion

The programmable illumination panel is one of the main features of the reactors; it allows for a virtual study of culture under the influence of various light sources while remaining in a controlled environment as illustrated in 3.3. These units under a computer-controlled array provide a framework to run numerous autonomous experiments simultaneously or to concur towards a single optimization task.

The stability of the system can last up to the order of a whole week. It is also equipped with some features to avoid disruptions or start from where it left off. Each bioreactor can also be controlled individually so that a specific unit can be debugged even during the execution of an experiment.

However, to allow for reliable readings of the culture's density, the reactor's optical sensor readings cannot be saturated to the point where it is almost zero, as it would imply on a high reading error, which is related to the derivative of the curves in figure 3.5. Indeed, whenever the optical sensor has completely reached zero counts, it would be impossible to deduce the culture's density from it as all incoming light would be absorbed by the culture. Conversely, counts that are too excessive will saturate the sensor; also causing reading errors and potentially damaging it. For such reasons, the ideal strategy is to keep the density in a quasi-constant regimen around 4 to 12 counts/10^3.

Figures 3.6, 3.7 and 3.8 show that the system is stable and accurate, being able to detect important details of microalgae growth dynamics. Figure 3.9 shows that the system is able to detect not only the effect of the incident irradiance, but also the effect of its change over time. In the particular presented study of figure 3.9, what seems to happen is that irradiances higher than 70 $W\,m^{-2}$ cause an irreversible photodamage to the *Chlorella sorokiniana* and this may be the cause for the lost of efficiency in the second half of the experiment.

An unforeseen complication to the optimization task was the apparent time dependence of the growth rate, as show in figure 3.9. Since the genetic algorithm assumes that there are no time related changes in the environment, it couldn't possibly perform well, as the fitness function would continuously change. A consequence of that is seen in figure 3.12, where it is clearly noticed that the optimal solution degrades over time.

However, even with such drawbacks, the experiment presented in section 3.5 still seems to point in the right direction: The irradiance of the first maximum coincides with the increasing and decreasing averages of figure 3.9 C, suggesting that even with time dependence, the algorithm is still trying to keep up with the optimal point, which happens to be around 15$W\,m^{-2}$, while the efficiency travels the plot of figure 3.9 C maintaining a constant irradiance.

This leads us to believe that further tinkering with the system could yield more revealing answers. To account for time dependence, a possible alternative would be

to create a variant of the genetic algorithm or find another one in which the parameters are only allowed to increase at a slow pace.

Chapter 5

Conclusion

We managed to develop a stable framework capable to manage an array of photobioreactors in a local area network, allowing for the parallel execution of replicates, experiments running with different parameter or collectively under the control of an intelligent system. Several tests were carried out to validate the operation and accuracy of the system. In particular, we implemented a genetic algorithm on such framework in order to optimize the growth efficiency of the microalgae *Chlorella sorokiniana*. We can conclude that the presented system works well, constituting a new tool capable of revealing details of the microalgae growth dynamics that are impossible to be detected with conventional reactors. Next, we plan to use this apparatus to characterize the growth dynamics of different strains of microalgae and to study how this dynamics depends on the spectra of the incident light.

Appendix A

Mathematical Methods

A.1 Genetic Algorithm

A genetic algorithm is an optimization method inspired by the process of the evolution of the species. Although there are many possible implementations, we will focus over the following case (A Python implementation is available on GitHub [43]):

Definition 1 *For* $\mathcal{U} := \{0,1\}^l$, *given* $f : \mathcal{U} \to \mathbb{R}$, *find* $u \in \mathcal{U} | f(u) = \max_{x \in \mathcal{U}} f(x)$

Algorithm 1 Genetic Algorithm

$N \leftarrow$ Population size (even number)
$l \leftarrow$ Number of alleles
$K \leftarrow$ Number of steps
$mp \leftarrow$ Probability of mutation
$G \in \mathcal{U}^N \leftarrow$ Random binary matrix $N \times l$
$f \leftarrow$ Optimization Function
for $k \leftarrow 1..K$ **do**
    $F \leftarrow N$ dimensional vector to store output of f
    for $n \leftarrow 1..N$ **do** $F_n \leftarrow f(G_n)$
    $P \leftarrow \text{softmax}(F)$
    selected $\leftarrow$ Choose pairs from $1..\frac{N}{2}$ with probability$P$
    $G \leftarrow \text{crossover}(\textit{selected})$
    $G \leftarrow \text{mutation}(G, mp)$
return G

In short, the algorithm spans a binary matrix of N rows and l columns representing a collection of the genomes of N individuals from $\mathcal{U}$, all with l genes. For each iteration, a call to the optimization function f is made for each individual, which is then turned into a probability by the softmax function. From this probability, pairs of individuals are selected and cross-combined at random splits until we can repopulate the matrix. For the case of an even number of individuals there will be $\frac{N}{2}$ pairs whereas for an odd case, we will need $\frac{N}{2} + 1$ pairs and drop the last child.

After these steps, a mutation transformation is applied over the matrix G such that each element has a chance of flipping its state (0 to 1 or 1 to 0). The probability of mutation is given by mp.

A.1.1 Estimated Conditions for Convergence

Since the genetic algorithm relies on stochastic processes, we cannot define a deterministic stopping time. Instead, we will say that the algorithm converges when the

population is homogeneous and track the expected value of this homogeneity over time. As [44] recommends, we will drop the mutation and crossover operators to ease the analysis and measure the homogeneity of the population as the average of the hamming distances between each pair of the individuals' genomes.

For a given gene j, the hamming distance (h) between individuals a and b is defined as follows:

$$h(G_{aj}, G_{bj}) = \begin{cases} 1, & \text{if } G_{aj} \neq G_{bj} \\ 0, & \text{otherwise} \end{cases} \tag{A.1}$$

.

So, for the entire population matrix G, the average hamming distance is

$$\bar{h}(G) = \frac{1}{2l} \sum_{k=1}^{l} \sum_{i,j}^{N} h(G_{ik}, G_{jk}) \tag{A.2}$$

.

As [44] mentions, the case for a constant optimization function f is of particular interest since the algorithm can do no better than a random guess for the selection process. That is, the probability distribution for selection is uniform. Under this assumption, the probability that a given column in the matrix G will have k individuals with ones given its current counts (X_t) is given by a binomial probability:

$$P\left(X_{t+1} = k | X_t, ..., X_0\right) = \binom{N}{k} \left(\frac{X_t}{N}\right)^k \left(1 - \frac{X_t}{N}\right)^{N-k} \tag{A.3}$$

.

To calculate the average hamming distance given an initial population matrix, we will focus on a specific column in the matrix G and analyze the sum of hamming distances (H_t) over the populations instead. As the counts of ones are known (X_t), we can directly compute the sum as follows:

$$H_t = X_t \left(N - X_t\right) \tag{A.4}$$

The expected value of H_t given an initial count X_0 is represented by

$$\mathbb{E}\left[H_t | X_0\right]$$

From equation A.3, we can compute $\mathbb{E}\left[X_{t+1} | X_t, ..., X_0\right]$ and $\mathbb{E}\left[X_{t+1}^2 | X_t, ..., X_0\right]$:

$$\begin{aligned} \mathbb{E}\left[X_{t+1} | X_t, ..., X_0\right] &= X_t \\ \mathbb{E}\left[X_{t+1}^2 | X_t, ..., X_0\right] &= X_t^2 + X_t \left(1 - \frac{X_t}{N}\right) \end{aligned} \tag{A.5}$$

.

By defining a σ-algebra generated by the collection of random variables $X_t, ..., X_0$ as $\mathcal{F}_t$, we can make use of the following property to calculate the expected values in function of the first step (t=0) as explained by [45]:

Theorem 1 *Given $\mathcal{F}_a$ and $\mathcal{F}_b$ σ-algebras such that $\mathcal{F}_a \subset \mathcal{F}_b$, and an observable X, then*

$$\mathbb{E}\left[X | \mathcal{F}_a\right] = \mathbb{E}\left[\mathbb{E}\left[X | \mathcal{F}_b\right] | \mathcal{F}_a\right]$$

Since the σ-algebra generated by X_0, denoted by $\mathcal{F}_0$, is contained in the σ-algebra generated by the collection $X_t, ..., X_0$, we can use theorem 1 to compute $\mathbb{E}\left[X_{t+1}|\mathcal{F}_0\right]$ and $\mathbb{E}\left[X_{t+1}^2|\mathcal{F}_0\right]$:

$$\begin{aligned}\mathbb{E}\left[X_{t+1}|\mathcal{F}_0\right] =& \mathbb{E}\left[\mathbb{E}\left[X_{t+1}|\mathcal{F}_t\right]|\mathcal{F}_0\right] \\ =& \mathbb{E}\left[\mathbb{E}\left[X_{t+1}|X_t, ..., X_0\right]|\mathcal{F}_0\right] \\ =& \mathbb{E}\left[X_t|X_0\right] \\ =& \mathbb{E}\left[X_0|X_0\right] \\ =& X_0\end{aligned} \tag{A.6}$$

$$\begin{aligned}\mathbb{E}\left[X_{t+1}^2|\mathcal{F}_0\right] =& \mathbb{E}\left[\mathbb{E}\left[X_{t+1}^2|\mathcal{F}_t\right]|\mathcal{F}_0\right] \\ =& \mathbb{E}\left[\mathbb{E}\left[X_{t+1}^2|X_t, ..., X_0\right]|\mathcal{F}_0\right] \\ =& \mathbb{E}\left[X_t^2 + X_t\left(1 - \frac{X_t}{N}\right)|\mathcal{F}_0\right] \\ =& \mathbb{E}\left[X_t|\mathcal{F}_0\right] + \left(1 - \frac{1}{N}\right)\mathbb{E}\left[X_t^2|\mathcal{F}_0\right] \\ (Induction) \implies & \\ =& \left(1 - \frac{1}{N}\right)^{t+1}\mathbb{E}\left[X_0^2|\mathcal{F}_0\right] + \mathbb{E}\left[X_0|\mathcal{F}_0\right]\sum_{u=0}^{t}\left(1 - \frac{1}{N}\right)^u \\ =& \left(1 - \frac{1}{N}\right)^{t+1}X_0^2 + NX_0\left(1 - \left(1 - \frac{1}{N}\right)^t\right)\end{aligned} \tag{A.7}$$

.

Applying the results of equations A.6 and A.7 to compute the expected value of H_t, we have:

$$\mathbb{E}\left[H_{t+1}|\mathcal{F}_t\right] = \mathbb{E}\left[X_{t+1}(N - X_{t+1})|\mathcal{F}_0\right] = X_0\left(1 - \frac{1}{N}\right)^t\left(N - X_0\left(1 - \frac{1}{N}\right)\right) \tag{A.8}$$

.

Notice that, since the term $\left(1 - \frac{1}{N}\right)$ is smaller than one, as the number of steps grow, we have that the hamming distance tends to zero, resulting in uniform alleles which are either filled with ones or zeros. By plugging this result back into our desired metric for the average of the hamming distance of the population (equation A.2), we conclude that

$$\bar{h}_{t+1} = \frac{1}{l}\sum_{k=1}^{l}\left(G_0^k\left(1 - \frac{1}{N}\right)^t\left(N - G_0^k\left(1 - \frac{1}{N}\right)\right)\right) \tag{A.9}$$

.

Where G_t^k represents the counts of ones for allele k on step t. For the case where all alleles start off with the same proportion G_0, the equation A.9 can be further simplified to

$$\bar{h}_{t+1} = \frac{1}{N}\left(G_0\left(1 - \frac{1}{N}\right)^t\left(N - G_0\left(1 - \frac{1}{N}\right)\right)\right) \tag{A.10}$$

.

Glossary

bot A program which emulates a human performing a specific task. 13

firmware The software responsible to provide low-level control a specific harware. 13

fitness function A function that is optimized by an Evolutionary Algorithm. 24

gene A unit of heredity holding information about the characteristics of an offspring. For genetic algorithms, this could be an index on a binary array. 24

genomes A set of genes. 33

HTTP HyperText Transfer Protocol. The protocol used to transfer web pages. 13

PWM Pulse Width Modulation. A way to send a signal between a HIGH and a LOW by turning the emitter on and off over a fast rate.. 19

script A computer program. This term is mostly used when referring to dynamic programming languages like Python. 15

serial A communication protocol to send data sequentially over a computer bus. 13

server A computer responsible to provide data and other resources to other computers. 13

shell Command line interface used to interact with a computer. 13

Bibliography

[1] E. Rabinowitch and R. Govindjee, *Photosynthesis* (Wiley, 1969), ISBN: 9780471704232, https://books.google.com.br/books?id=pXJef3jnZ-sC (cited on pages 1–3).

[2] B. E. Schirrmeister, J. M. de Vos, A. Antonelli, and H. C. Bagheri, "Evolution of multicellularity coincided with increased diversification of cyanobacteria and the great oxidation event", Proceedings of the National Academy of Sciences **110**, 1791–1796 (2013) (cited on page 1).

[3] N. Nelson and A. Ben-Shem, "The complex architecture of oxygenic photosynthesis", Nature Reviews Molecular Cell Biology **5**, 971–982 (2004) (cited on page 1).

[4] N. Nelson, "Photosystems and global effects of oxygenic photosynthesis", Biochimica et Biophysica Acta (BBA)-Bioenergetics **1807**, 856–863 (2011) (cited on page 1).

[5] N. Nelson and C. F. Yocum, "Structure and function of photosystems i and ii", Annu. Rev. Plant Biol. **57**, 521–565 (2006) (cited on page 1).

[6] R. E. Blankenship, *Molecular mechanisms of photosynthesis* (John Wiley & Sons, 2014) (cited on page 1).

[7] P. Müller, X.-P. Li, and K. K. Niyogi, "Non-photochemical quenching. a response to excess light energy", Plant physiology **125**, 1558–1566 (2001) (cited on page 3).

[8] Z. Li, S. Wakao, B. B. Fischer, and K. K. Niyogi, "Sensing and responding to excess light", Annual review of plant biology **60**, 239–260 (2009) (cited on page 3).

[9] P. Jahns and A. R. Holzwarth, "The role of the xanthophyll cycle and of lutein in photoprotection of photosystem ii", Biochimica et Biophysica Acta (BBA)-Bioenergetics **1817**, 182–193 (2012) (cited on page 3).

[10] B. W. Carroll and D. A. Ostlie, *An introduction to modern astrophysics*, en, 2nd edition (Pearson Education, Philadelphia, PA, July 2006) (cited on page 4).

[11] "Section 10 - solar", in *Handbook of energy*, edited by C. J. Cleveland and C. Morris (Elsevier, Amsterdam, 2013), pages 405–450, ISBN: 978-0-08-046405-3, https://doi.org/10.1016/B978-0-08-046405-3.00010-3, https://www.sciencedirect.com/science/article/pii/B9780080464053000103 (cited on page 4).

[12] R. Horne, "Seeking sailors to help measure phytoplankton populations", Proceedings of the National Academy of Sciences **110**, 7107–7107 (2013) 10.1073/pnas.1306732110, https://www.pnas.org/doi/abs/10.1073/pnas.1306732110 (cited on page 5).

[13] M. D. Guiry, "How many species of algae are there?", Journal of phycology **48**, 1057–1063 (2012) (cited on page 4).

[14] D. G. Boyce, M. R. Lewis, and B. Worm, "Global phytoplankton decline over the past century", Nature **466**, 591–596 (2010) (cited on page 4).

[15] E. E. Luiten, I. Akkerman, A. Koulman, P. Kamermans, H. Reith, M. J. Barbosa, D. Sipkema, and R. H. Wijffels, "Realizing the promises of marine biotechnology", Biomolecular Engineering **20**, 429–439 (2003) (cited on page 4).

[16] O. Pulz and W. Gross, "Valuable products from biotechnology of microalgae", Applied microbiology and biotechnology **65**, 635–648 (2004) (cited on page 4).

[17] K. H. Cardozo, T. Guaratini, M. P. Barros, V. R. Falcão, A. P. Tonon, N. P. Lopes, S. Campos, M. A. Torres, A. O. Souza, P. Colepicolo, et al., "Metabolites from algae with economical impact", Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology **146**, 60–78 (2007) (cited on page 5).

[18] H. C. Greenwell, L. Laurens, R. Shields, R. Lovitt, and K. Flynn, "Placing microalgae on the biofuels priority list: a review of the technological challenges", Journal of the royal society interface **7**, 703–726 (2010) (cited on page 5).

[19] S. A. Scott, M. P. Davey, J. S. Dennis, I. Horst, C. J. Howe, D. J. Lea-Smith, and A. G. Smith, "Biodiesel from algae: challenges and prospects", Current opinion in biotechnology **21**, 277–286 (2010) (cited on page 5).

[20] T. M. Mata, A. A. Martins, and N. S. Caetano, "Microalgae for biodiesel production and other applications: a review", Renewable and sustainable energy reviews **14**, 217–232 (2010) (cited on page 5).

[21] S. S. Oncel, "Microalgae for a macroenergy world", Renewable and Sustainable Energy Reviews **26**, 241–264 (2013) (cited on page 5).

[22] C. J. Unkefer, R. T. Sayre, J. K. Magnuson, D. B. Anderson, I. Baxter, I. K. Blaby, J. K. Brown, M. Carleton, R. A. Cattolico, T. Dale, et al., "Review of the algal biology program within the national alliance for advanced biofuels and bioproducts", Algal research **22**, 187–215 (2017) (cited on page 5).

[23] P. J. Lammers, M. Huesemann, W. Boeing, D. B. Anderson, R. G. Arnold, X. Bai, M. Bhole, Y. Brhanavan, L. Brown, J. Brown, et al., "Review of the cultivation program within the national alliance for advanced biofuels and bioproducts", Algal research **22**, 166–186 (2017) (cited on page 5).

[24] B. L. Marrone, R. E. Lacey, D. B. Anderson, J. Bonner, J. Coons, T. Dale, C. M. Downes, S. Fernando, C. Fuller, B. Goodall, et al., "Review of the harvesting and extraction program within the national alliance for advanced biofuels and bioproducts", Algal research **33**, 470–485 (2018) (cited on page 5).

[25] R. Forzza, J. Baumgratz, C. Bicudo, D. Canhos, A. Jr, M. Nadruz, A. Costa, D. Costa, M. Hopkins, P. Leitman, L. Lohmann, E. Nic Lughadha, L. Maia, G. Martinelli, M. Menezes, M. Morim, A. Peixoto, J. R. Pirani, J. Prado, and D. Zappi, "New brazilian floristic list highlights conservation challenges", BioScience **62**, 39–45 (2012) 10.1525/bio.2012.62.1.8 (cited on page 6).

[26] S. I. Hadi, H. Santana, P. P. Brunale, T. G. Gomes, M. D. Oliveira, A. Matthiensen, M. E. Oliveira, F. C. Silva, and B. S. Brasil, "Dna barcoding green microalgae isolated from neotropical inland waters", PloS one **11**, e0149284 (2016) (cited on page 6).

[27] B. Brasil, F. Silva, and F. Siqueira, "Microalgae biorefineries: the brazilian scenario in perspective", New Biotechnology **39**, 90–98 (2017) 10.1016/j.nbt.2016.04.007, https://doi.org/10.1016/j.nbt.2016.04.007 (cited on page 6).

[28] S. Huo, Z. Wang, S. Zhu, Q. Shu, L. Qin, W. Zhou, P. Feng, F. Zhu, W. Qi, and R. Dong, "Biomass accumulation of chlorella zofingiensis g1 cultures grown outdoors in photobioreactors", Frontiers in Energy Research **6**, 10.3389/fenrg.2018.00049 (2018) 10.3389/fenrg.2018.00049 (cited on page 7).

[29] *Algae photobioreactor vizualization for uniwersytet gdański*, https://www.deviantart.com/silentcenter/art/Photobioreactors-299621647 (cited on page 7).

[30] *Alpha omega electronics*, https://www.alphaomega-electronics.com/en/home/5961-large-scale-photobioreactors-25-l-or-100-l-volume.html (cited on page 7).

[31] *Plant cell photobioreactors market size, share, development by 2024*, 2019, https://www.openpr.com/news/1887062/plant-cell-photobioreactors-market-size-share-development (cited on page 7).

[32] N. Wiener, *Cybernetics: or control and communication in the animal and the machine* (Technology Press, 1948), https://books.google.com.br/books?id=B7XPoAEACAAJ (cited on page 7).

[33] C. Bicudo and M. Menezes, "Gêneros de algas de águas continentais no brasil - terceira edição", cited By 1 (2017) (cited on page 9).

[34] J. Wehr, R. Sheath, and J. Kociolek, *Preface*, cited By 53 (2015), pages xv–xvi, 10.1016/B978-0-12-385876-4.09998-9 (cited on page 9).

[35] J. Olivares, "National alliance for advanced biofuels and bioproducts synopsis (naabb) final report", National Alliance for Advanced Biofuels and Bioproducts Synopsis (NAABB) Final Report, cited By 37 (2014) (cited on page 9).

[36] S. Cazzaniga, L. Dall'Osto, J. Szaub, L. Scibilia, M. Ballottari, S. Purton, and R. Bassi, "Domestication of the green alga chlorella sorokiniana: reduction of antenna size improves light-use efficiency in a photobioreactor", Biotechnology for Biofuels **7**, cited By 102, 10.1186/s13068-014-0157-z (2014) 10.1186/s13068-014-0157-z (cited on page 9).

[37] D. M. Ribeiro, L. F. Roncaratti, G. C. Possa, L. C. Garcia, L. J. Cançado, T. C. R. Williams, and B. dos Santos Alves Figueiredo Brasil, "A low-cost approach for chlorella sorokiniana production through combined use of urea, ammonia and nitrate based fertilizers", Bioresource Technology Reports **9**, 100354, ISSN: 2589-014X (2020) https://doi.org/10.1016/j.biteb.2019.100354, https://www.sciencedirect.com/science/article/pii/S2589014X19302440 (cited on page 9).

[38] S. Hadi, H. Santana, P. Brunale, T. Gomes, M. Oliveira, A. Matthiensen, M. Oliveira, F. Silva, and B. Brasil, "Dna barcoding green microalgae isolated from neotropical inland waters", PLoS ONE **11**, cited By 46, 10.1371/journal.pone.0149284 (2016) 10.1371/journal.pone.0149284 (cited on page 9).

[39] M. Fernandes, L. Calsing, R. Nascimento, H. Santana, P. Morais, G. de Capdeville, and B. Brasil, "Customized cryopreservation protocols for chlorophytes based on cell morphology", Algal Research **38**, cited By 4, 10.1016/j.algal.2018.101402 (2019) 10.1016/j.algal.2018.101402 (cited on page 9).

[40] Jimblom, *Serial communication*, https://learn.sparkfun.com/tutorials/serial-communication (cited on pages 13–14).

[41] *Systemd*, https://github.com/systemd/systemd (cited on page 15).

[42] A. L. Stancik and E. B. Brauns, "A simple asymmetric lineshape for fitting infrared absorption spectra", Vibrational Spectroscopy **47**, 66–69 (2008) 10.1016/j.vibspec.2008.02.009, https://doi.org/10.1016/j.vibspec.2008.02.009 (cited on page 17).

[43] Ícaro Lorran Lopes Costa, *GAPY - Genetic Algorithm for Python*, https://github.com/icarosadero/gapy (cited on page 33).

[44] S. J. Louis and G. J. Rawlins, "Predicting convergence time for genetic algorithms", Foundations of Genetic Algorithms **2**, 141–161 (1993) (cited on page 34).

[45] *Stochastic processes*, https://www.cs.yale.edu/homes/aspnes/pinewiki/Martingales.html (cited on page 34).