ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
ENGENHARIA QUÍMICA

LARA FADEL BEARZI

COMPUTER-AIDED PRODUCT DESIGN FOR COCOA BUTTER ALTERNATIVES

SÃO PAULO
2024

LARA FADEL BEARZI

COMPUTER-AIDED PRODUCT DESIGN FOR COCOA BUTTER ALTERNATIVES

Revised Version

Thesis presented to the Chemical Engineering Department at the Escola Politécnica da USP, to obtain the degree of Master of Science.

Supervisor: Prof. Dr. Moisés Teles dos Santos

SÃO PAULO
2024

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____de _____de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catalogação-na-publicação

Bearzi, Lara
COMPUTER-AIDED PRODUCT DESIGN FOR COCOA BUTTER ALTERNATIVES / L. Bearzi -- versão corr. -- São Paulo, 2024.
85 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia Química.

1.CAPD 2.Otimização 3.Chocolate I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia Química II.t.

Nome: BEARZI, Lara Fadel

Título: Computer-Aided Product Design for Cocoa Butter Alternatives

Dissertação apresentada ao Departamento de Engenharia Química da Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Aprovado em:

Banca Examinadora

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Prof. Dr. _____

Instituição: _____

Julgamento: _____

Agradecimentos

Ao meu orientador, Prof. Dr. Moisés Teles dos Santos, pela paciência, confiança e ensinamentos nesses últimos anos. Agradeço a disponibilização do VOSTAT, que tornou possível todo esse projeto.

Aos meus pais, Cristiane e Gustavo, por todo o suporte emocional e financeiro. Agradeço por terem me ensinado que, com muito trabalho e dedicação, conseguimos chegar em qualquer lugar.

À minha irmã, Marina, por todo apoio durante todos os momentos da minha vida.

Aos demais membros da minha família, Marlene, Manir, Sirleine, Karen, Sophia, Manir Jr. e João Bosco pelos momentos de descontração e alegria.

Ao meu parceiro dos últimos anos, Daniel, por todo o companheirismo, paciência e suporte.

Aos meus amigos, agradeço o apoio incondicional.

A USP e aos funcionários do Departamento de Engenharia Química da USP.

Agradeço, também, à CAPES/PROEX pelo auxílio financeiro para a participação no V Congresso de Óleos e Gorduras.

A Deus, pela saúde e força nos momentos de dificuldades.

Resumo

BEARZI, Lara Fadel. Computer-Aided Product Design for Cocoa Butter Alternatives. 2023. Dissertação (Mestrado em Engenharia Química) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2023.

A busca por substitutos à manteiga de cacau em produtos de chocolate é uma atual realidade, dados os riscos existentes à produção do cacau, seu preço instável, e a disponibilidade de outras matérias-primas mais baratas. Além disso, o mercado para as indústrias químicas vem se tornando cada vez mais competitivo, exigindo que as empresas foquem na satisfação do cliente, tanto em qualidade do produto quanto em tempo para lançamento. Para auxiliar no desenvolvimento de novos produtos, a metodologia *Computer-Aided Product Design* utiliza conceitos de Engenharia Reversa usando simulações computacionais e modelos de predição de propriedades físico-químicas. Visando proporcionar uma busca otimizada de substitutos à manteiga de cacau, o presente trabalho trata da criação de uma ferramenta de otimização em Python, acoplada a um modelo de predição do cálculo de Conteúdo de Gordura Sólida (CGS), cujo objetivo é encontrar uma mistura que possa vir a substituir a manteiga de cacau em produtos de chocolate. O CGS foi escolhido, pois quantifica uma propriedade fundamental em chocolates: sólido a temperatura ambiente, e líquido à temperatura corporal. As predições do CGS foram feitas pela solução de um problema termodinâmico de equilíbrio sólido-líquido. A etapa de Engenharia Reversa construída nesse trabalho, utilizou um Algoritmo Genético (AG), e para isso, foi definida uma função objetivo baseada no CGS das temperaturas de 5, 25, e 35ºC. Dentro dos testes de óleos individuais e misturas binárias, foi feita uma análise de sensibilidade para otimizar os parâmetros quantitativos, como o tamanho da população, taxa de crossover e mutação. Para as misturas binárias, com tamanho de população de 100 indivíduos, taxa de crossover de 90%, taxa de mutação de 5%, e tempo de simulação de 12 horas, foi possível obter uma função objetivo 15% acima da obtida com óleos individuais, ou seja, foi possível encontrar uma solução mais bem adaptada para substituir a manteiga de cacau, o que demonstra que a metodologia pode vir a ser utilizada como guia em processos de desenvolvimento de novos produtos. Como trabalhos futuros recomenda-se unificar as ferramentas utilizadas na otimização, o que reduziria consideravelmente o tempo de simulação, realizar uma análise de sensibilidade dos parâmetros qualitativos do Algoritmo Genético, como os métodos de seleção e substituição, e buscar soluções formadas por misturas ternárias de óleos e gorduras.

Palavras-chave: Computer-Aided Product Design, Chocolate, Algoritmo Genético, Otimização, Conteúdo de Gordura Sólida.

Abstract

BEARZI, Lara Fadel. Computer-Aided Product Design for Cocoa Butter Alternatives. 2023. Dissertação (Mestrado em Engenharia Química) – Escola Politécnica da Universidade de São Paulo, São Paulo, 2023.

The search for cocoa butter substitutes in chocolate products is a current reality, given existing risks to cocoa production, its unstable prices, and the availability of cheaper raw material. Besides that, market is becoming more competitive for Chemical industries, demanding companies to focus on customer satisfaction, both in product quality and time-to-market. To fasten new product development, Computer-Aided Product Design methodology uses Reverse Engineering concepts, through computational simulations and property prediction models. Aiming at an optimized search for cocoa butter substitutes, the current work creates an optimization tool in Python, coupled with a pre-existing prediction model for Solid Fat Content (SFC) estimation, whose objective is to find a possible substitute for cocoa butter in chocolate products. SFC was chosen, because it quantifies a fundamental property in chocolates: solid at room temperature, and liquid at body temperature. SFC prediction was calculated by solving a thermodynamic solid-liquid equilibrium problem. The reverse engineering step built in this work used Genetic Algorithm (GA) and, thus, an objective function was defined based on SFC values for 5, 25, and 35ºC temperatures. Within individual and binary mixture tests, a sensitivity analysis was performed to optimize quantitative parameters, such as population size, crossover and mutation rates. For binary mixtures, population size of 100, crossover rate of 90%, mutation rate of 5%, and simulation of 12 hours, provided results 15% better than the one obtained with single vegetable oil simulation, i.e., it was possible to find a better adapted solution to the current problem, which demonstrates that the methodology can be used as a guide in new product development processes. As future work, it is recommended a unification of the tools, which would decrease considerably simulation time, a sensitivity analysis on Genetic Algorithm's qualitative parameters, such as selection and replacement methods, and the search for ternary mixtures of oil and fats, besides single and binary mixtures.

Keywords: Computer-Aided Product Design, Chocolate, Genetic Algorithm, Optimization, Solid Fat Content.

Summary

List of figures

List of tables

1. Introduction

1.1. Market Context

According to market research done by Fortune Business Insights (2022) cocoa and chocolate global market size was valued at $47 billion in 2021 and was expected to reach $68 billion by 2029. They also see cocoa as an emerging healthy ingredient and a greater demand for high-quality cocoa derivatives. A greater demand for cocoa substitutes and equivalents is also seen, mainly driven by the increasing price of cocoa and the higher availability of its substitutes. Many industries are already offering cocoa butter equivalent blends.

From historical data provided by the International Cocoa Organization, 2022, in Figure 1, we can see a significant oscillation of the cocoa beans' prices throughout the years.

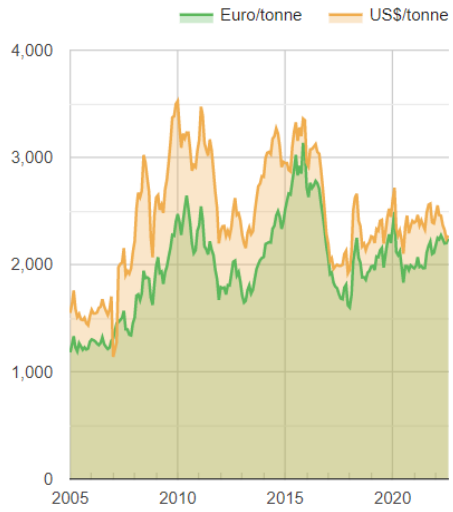

Figure 1: Cocoa Beans Monthly Price Oscillation. Source: (INTERNATIONAL COCOA ORGANIZATION, 2022)

In Brazil, there was an increasing cocoa production until 1989, when witch's broom disease (WBD) outbroke (DIAS, 2022). In Figure 2 we see that between the 60's and the 80's, Brazil had an increasing production and exported a part of it. After the start of this disease, production decreased and has not yet come back to the same levels as before. According to Teixeira, Thomazella, and Pereira (2015), WBD is one of the most severe problems that could affect the cocoa crop and can result in a production loss up to 90%. Even though this disease has not yet reached any country

in Africa, which is the biggest cocoa producer, WBD represents a real threat to the world chocolate industry (TEIXEIRA; THOMAZELLA; PEREIRA, 2015).

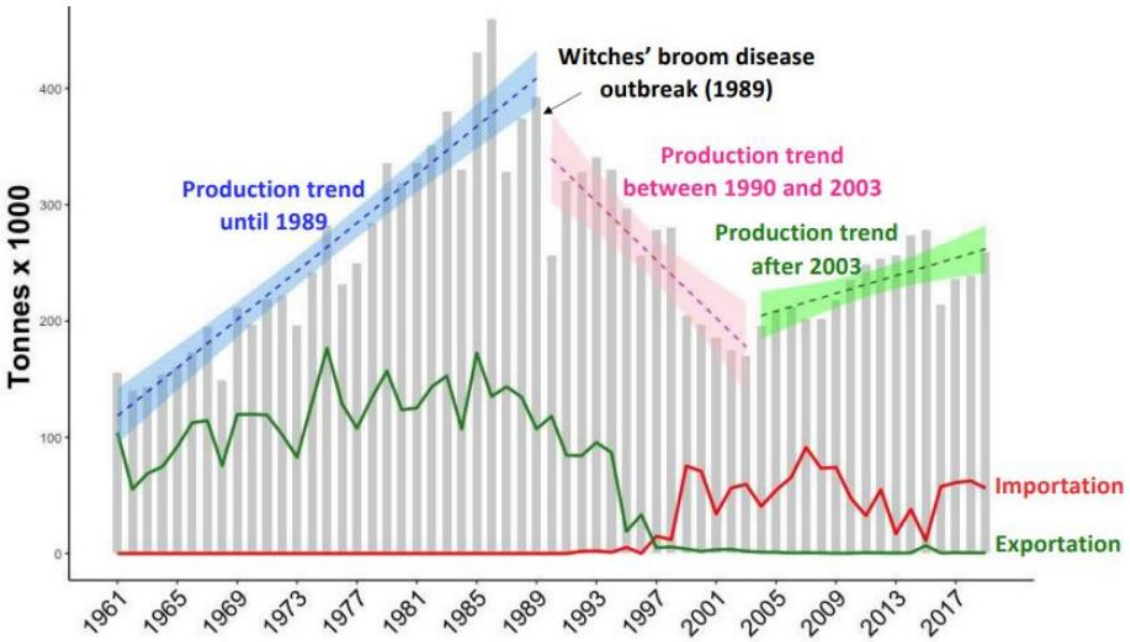

Figure 2: Cocoa production, importation, and exportation in Brazil throughout the years.
Source: Dias, 2022

With an increasing cocoa demand, a reduction in cocoa production in some countries, an oscillating price, and the fact that other vegetable oils and fats are cheaper (JAHURUL et al., 2013), the use of alternative oils and fats to replace cocoa butter in chocolate formulations, which is already a current reality, is becoming even more desirable.

There are different kinds of cocoa butter alternatives: Cocoa Butter Equivalents (CBEs), Cocoa Butter Replacer (CBRs) and Cocoa Butter Substitutes (CBSs) (LIPP; ANKLAM, 1998). Main differences and some examples of each of them are shown in Table 1.

Table 1: Summary for Cocoa Butter Alternatives

	Cocoa Butter Equivalent (CBE)	Cocoa Butter Replacer (CBR)	Cocoa Butter Substitute (CBS)
Type	non-lauric vegetable fat	non-lauric vegetable fat	lauric fat
Properties	Similar physical and chemical properties to CB	Similar fatty acid distribution, different TAG composition	Some physical similarities, very different chemically
Mixing	Can be mixed with CB in any amount	Partially compatible to CB	Suitable for 100% substitution only
Main Fatty Acids	Palmitic, Stearic, and Oleic acids	Elaidic, Stearic, Palmitic, Linoleic acids	Lauric, Myristic Acid
Main TAGs	POP, POS, SOS	PEE, SEE	LLL, LLM, LMM
Examples	Palm oil, Illipé Butter, Shea Butter, Kokum Butter, Sal Fat	Soya oil, Rape Seed oil, Cotton oil, Ground Nut oil, Palm olein	Coconut oil, Palm Kernel Oil

Sources: (LIPP; ANKLAM, 1998) (BUDIANTO; KUSMARDINI, 2021).
L: Lauric acid (C12:0), M: Myristic acid (C14:0), P: palmitic acid (C16:0), S: Stearic acid (C18:0), O: Oleic acid (C18:1), E: Elaidic acid (C18:1 trans-9), L: Linoleic acid (C18:2)

The need to develop innovative products is spread over several industries besides the food and chocolate industry. Nowadays, companies are facing increased market competition, mainly because of globalization and the spread of information. Smith and Ierapepritou, (2010), made a study to evaluate the trends in chemical industries regarding product development. It was discovered that global competitiveness and demand for product variety were the main topics that affected new product strategies for these companies, as stated in Figure 3. Thus, aiming at a good performance, companies ought to seek competitive advantage against their competitors.

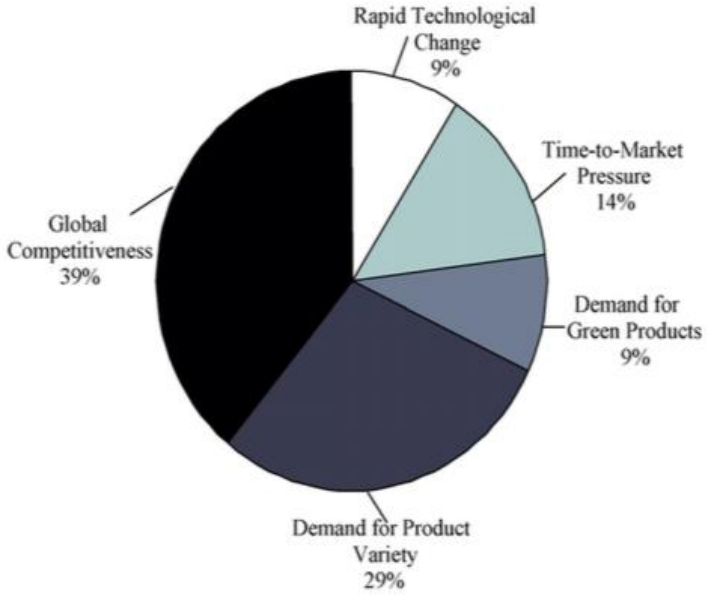

Figure 3: Trends affecting new product development strategies. Source: (SMITH; IERAPEPRITOU, 2010)

This context requires that companies shorten their time-to-market for chemical products to keep up with the new tendencies that emerge and consider more consumer acceptance in the product development process (ARRIETA-ESCOBAR et al, 2019a; ZHANG et al., 2020). However, there is no point in launching a new product fast if it won't sell as expected. According to Yang et al. (2003), the expectations from consumers are more diverse and rapidly changing, which results in the reduction of the product's life cycle (SEIDER et al., 2009). According to Terzi et al. (2010), companies need to not only reduce time spent in product development and its costs, but also improve the final quality of the product.

One strategy that demonstrates the importance of customer's acceptance is Mass Customization, which is a strategy focused on producing personalized products and services (DAVIS, 1989; PINE II; VICTOR; BOYNTON, 1993). Some studies have shown that Mass Customization is an important strategy regarding companies' competitive advantage and have shown successful case studies in different sectors, such as the food industry and electronics (FOGLIATTO; DA SILVEIRA; BORENSTEIN, 2012).

Market research within research and development employees working in different chemical industries in Brazil showed that the usual time to launch a new

product is around 3 years (verbal information)[1], proving that long time-to-market is still a reality in industries nowadays. One issue, that emerged mostly after the pandemic, is shortage of raw material. This means that the company must quickly substitute a material for another, due to unavailability of the original one. Another necessity, mainly in the food industry, is the adaptation of products that are trend in other countries, which results in ingredient substitution, according to regional availability and market necessity.

1.2. Product Design

According to Hill (2009), chemical product development is normally done by scientific hypothesis, trial-and-error, and, sometimes, can be accelerated by methods such as high-throughput experimentation (HTE) or experimental design (CONTE; GANI; NG, 2011). While in the HTE approach many experiments are performed in parallel in small quantities, the experimental design relies on statistical analysis to reduce the number of experiments. In both cases, there is a reduction in the total amount of materials required to perform the experiments.

The above-mentioned techniques result in an adequate product, however, since it is not possible to test all alternatives, a better product was probably missed (ZHANG et al., 2018). Also, these traditional methods are often costly and take a long time to be performed. (VENKATASUBRAMANIAN; CHAN; CARUTHERS, 1994).

Another aspect is that, with these approaches being highly dependent on previous knowledge, makes it even more complex to search for optimal properties without systematic selection tools (CHURI; ACHENIE, 1996).

Given the resourceful and prolonged current product development, opposed to the fast-changing market, a more robust and accurate framework for Product Development must be adopted to guarantee the companies' success (COSTA; MOGGRIDGE; SARAIVA, 2006). In this context, arises Product Design, a systematic framework to minimize experiments and launch new products faster and more efficiently (COSTA; MOGGRIDGE; SARAIVA, 2006; HILL, 2009). It also represents a shift from focus on purity to performance (CONTE et al., 2012).

[1] Interviews performed by Beatriz Mazzini and Lara Bearzi with R&D professionals from different private companies.

According to the framework developed by Moggridge and Cussler (2000) the starting point of the Chemical Product Design is the identification of the customer's needs, then product ideas are generated, with the most promising ones being selected for the development of a process for manufacturing it. Most of the different proposed Chemical Product Design Frameworks are market-pull (COSTA; MOGGRIDGE; SARAIVA, 2006), which means that the products are developed to meet an already existing market demand.

The customers' needs for a specific product are qualitative and subjective; and ought to be modeled by quantitative parameters, the performance indices, which depend on the product's composition, structure and usage conditions (COSTA; MOGGRIDGE; SARAIVA, 2006).

According to Terzi et al. (2010), an increased focus on product design is essential for the success of the current industry. Previous investments on just in time, quality management and maintenance aimed at improving cost, quality, and time to market, are not enough anymore for a company to succeed. Nowadays, the focus must be on innovation to guarantee customer satisfaction. Figure 4 shows this shift of importance from the Manufacturing to Design and End of Life phases (TERZI et al., 2010).

The Product Design concept has been growing over the past few years and is even considered by some authors to be a possible third paradigm of Chemical Engineering, after Unit Operations, which was developed in the 1920s and 1930s, and Transport Phenomena, developed in the late 1950s (HILL, 2009).

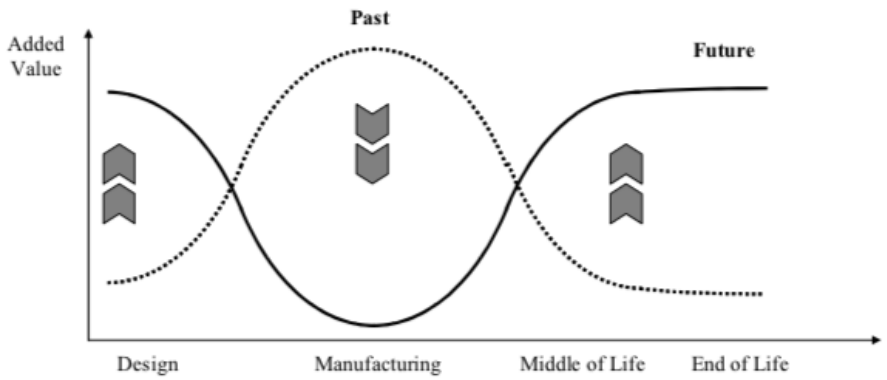

Figure 4: Added value by lifecycle stage comparison between past and future trends in companies. Source: (TERZI et al., 2010)

1.3. Computer-Aided Product Design

Computer-Aided Product Design, also referred to as CAPD or CAMD (Computer-aided molecular design), is the Product Design methodology assisted by computational tools. According to Harper et al. (1999) , a CAMD main goal is to find compounds that match previously defined properties. Harper and Gani (2000) defined the CAMD methodology in a three-step procedure: problem definition (pre-design step), molecules or product search (design step) and results analysis (post-design step). In the first step, it is important to identify the problem, define the compounds which will take part in the search and the desired properties and properties' constraints. In the second step, the CAMD design, possible molecules are generated and evaluated based on property constraints. Some approaches here can be taken to avoid combinatorial explosion, such as the multi-level approach or an evolutionary algorithm. The last step, post-design, is used to analyze results with more robust property predictions methods or experiments.

Figure 5 shows a 5-step framework for Computer-Aided Product Design. It starts with defining customers' needs, which may involve market research; then, this qualitative information is translated into quantitative properties and constraints. Depending on the case study, a database containing the property's values for different substances could be enough. For more complex properties or more innovative materials, property prediction models are necessary and, if mixtures are being considered, mixing models should also be included. The CAPD step is the optimization one and, according to Harper et al. (1999) can be classified as a database search, Generate and Test Method, or Mathematical Programming and Genetic Algorithm Method, which will be further explained on the Literature Review Section. The last step is the Validation, which may include experiments or more rigorous property calculations.

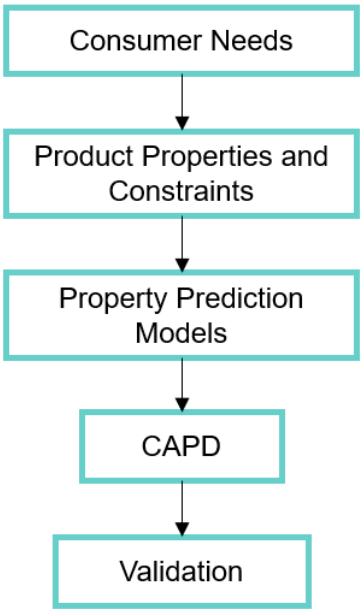

Figure 5: Computer-Aided Product Design Generic Framework

2. Objectives

Based on the previous discussed need to improve search methods for innovative products, the general objective of this work is to develop a computational tool to aid the design of fat-based products. With this goal, the following specific objectives can be outlined:

- To model the problem of cocoa butter replacers design with the conceptual methodologies of chemical product design.
- To develop, implement and test a search method based on evolutionary methods.
- To integrate the search method with the computational tools previously developed for solid fat content prediction.
- To apply the methodology to identify optimal mixtures of fats and oils able to replace cocoa butter.

3. Literature Review

3.1. Product Design

3.1.1. General Formulation

According to Austin; Sahinidis; Trahan (2016), chemical product development is based on trial-and-error, limited by the number of chemicals, time and financial resources. The set of possible solutions (molecules/mixtures) – the “design space”— is too large to allow an exhaustive test and an approve/disapprove approach. Also, aiming at a sustainable growth, product development must balance expenditure of resources, environmental impact, product performance and cost (GANI, 2004). Product Design arises in this context, to speed up new product launches, bring a customer-centric methodology and propose a more sustainable product development process.

In Chemical Product Design, the desired behavior of the products is known, and the goal is to identify a composition of the final product that matches such behavior. (GANI, 2004). After the product search and validation, there is a next step of finding out if it’s possible to manufacture it in a profitable manner with positive environmental impact (GANI, 2004). Moggridge and Cussler (2001) suggested a four-step product design process: needs definition, generation of ideas, selection of best ideas and product manufacturing. According to Gani (2004) there are two possible types of chemical product design problems: molecular and mixture design, and the applicability of the solution is highly dependent on reliable target property models (GANI, 2004).

Product Design has been differently classified among researchers in the literature. Ng, Gani and Dam-Johansen, (2007) divided Product Design approaches into three types: trial-and-error experiments approach, where experiments are planned based on previous knowledge; model-based approach, where mathematical models are used for properties estimation; and integrated experiment-modeling approach. According to Conte, Gani, and Ng (2011), the integrated approach is the best one, as it requires less time and resources when compared to approach 1, and leads to more accurate outputs, when compared to approach 2.

Zhang et al. (2018) listed five different Product Design approaches: experiment based, database search, heuristic-rule based, model-based computer-aided, and integrated model-based computer-aided techniques combined with experiments.

According to Zhang et al (2018), all methods can have a good product as output, however, it is not feasible to test all alternatives, and a better product might have been missed.

3.1.2. Experiment-guided product design

Being highly dependent on experiments, strategies as to reduce the number of experiments have emerged in the Product Design field. One of them is the Experimental Design. As stated by Montgomery (2012), statistical design of experiments is the process of planning experiments and have reliable conclusions using statistical methods. Experimental Design has three principles: randomization, replication, and blocking.

Factorial design is one Experimental Design strategy, in which all possible combinations of the levels of the variables/factors and the individual contribution of each of them to the response are analyzed (DEAN; VOSS, 1999). Factors vary together, which saves time and resources (MONTGOMERY, 2012). Montgomery shows an example of experimental design applied to a robust product design regarding choice of battery material that would result in a long life, in a wide range of different temperatures. This is a two-factor factorial experiment, in which observations are selected at random, and the Analysis of Variance method, also known as ANOVA, can be used to interpret the results (MONTGOMERY, 2012)

The response surface methodology (RSM) can also be used, in which the response of interest is represented graphically (MONTGOMERY, 2012). Several studies used this approach. Particularly in the oils and fats area, studies using RSM can be found in the works of Ahn et al. (2008), Santos et al. (2014), Vitolo; Ract; Guebara (2017), Sivakanthan; Jayasooriya; and Madhujith (2019) and Boroujeni et al. (2020).

3.2. Search Methods

3.2.1. Introduction

Optimization methods are vastly used in the Chemical Engineering field to improve product quality, and reduce costs and environmental impact (GUT, 2021). The optimization process normally needs an objective function, which indicates the quality of the results, and process and product constraints applied to the search space. There are different optimization methods, and classifications.

Mathematical programming is a systematic approach used for optimizing an objective function, considering a set of constraints (HUANG; LAI; CHENG, 2009). A generic mathematical programming problem can be stated as: (MASHWANI; HAIDER; BELHAOUARI, 2021)

$$
\begin{aligned}
&minimize\ F(x) = f_1(x),\ f_2(x), \dots, f_m(x), \\
&g_i(x) \leq 0, \qquad i = 1, 2, \dots, p \\
&h_i(x) = 0, \qquad j = 1, 2, \dots, q \\
&x_l^i \leq\ x^i \leq\ x_u^i, \qquad i = 1, 2, \dots N
\end{aligned}
$$

Where F(x) is the function with m objectives, g are the inequality constraints, h are the equality constraints, and $x_l^i \leq\ x^i \leq\ x_u^i$ are constraints over x.

Mathematical programming problems are usually deterministic, in which the same starting point leads to the same solution. Linear programming (LP) involves solutions for problems with linear and continuous objective function and restrictions. Linear problems are convex, which means that the optimal solution is always the global one. Simplex is an example of methodology to solve LP problems. Non-linear programming (NLP) includes non-linear equations, which result in non-convexities, i.e., multiple local minimum solutions. Some examples of solution methods are successive linear programming (SLP), sequential quadratic programming (SQP), and generalized reduced gradient (GRG). Discrete optimization includes variables assuming only specific values, different from continuous restrictions. Binary variables are one common restriction, in which the result can only be yes/no (1/0). Mixed integer linear programming (MILP), Mixed integer non-linear programming (MINLP) and Integer

programming (IP) are examples of discrete optimization problems. Examples of methods to solve these equations are Branch and Bound, and Decomposition based.

Decomposition-based is a very common solution approach for computer-aided chemical product design, in which the number of feasible mixtures decreases in each level (CONTE; GANI; NG, 2011). It divides the problem into smaller subproblems, which can be: (KARUNANITHI; ACHENIE; GANI, 2005)

1. Generation of feasible molecular structures
2. Evaluation of pure component properties constraints
3. Evaluation of mixture properties constraints
4. Evaluation of material miscibility among each other
5. Evaluation of process model constraints (if applicable)

There are numerous examples of mathematical programming and deterministic optimization methods applied to computer-aided product design. Some few examples were compiled and are listed in Table 2.

Table 2: Mathematical Programming Methods Applied to Chemical Product Design

Case Study	Problem Formulation	Solution Approach	References
Paint Formulation and Insect Repellent Lotion	MINLP	Decomposition-based	(CONTE; GANI; NG, 2011)
Gasoline and Lubricant Blends	MINLP	Decomposition-based	(YUNUS et al., 2014)
Solvent Blend	MINLP	Decomposition-based	(CIGNITTI; ZHANG; GANI, 2015)
Gasoline and Diesel Blends	MINLP	Decomposition-based and two-step solution	(KALAKUL et al., 2018)
Gasoline and Jet-Fuel Blends	MINLP	Two-steps solution	(ZHANG et al., 2018)
Cosmetics Design	MILP	CPLEX	(ARRIETA-ESCOBAR et al., 2019)

Complexity increases with non-linearity and the presence of discrete variables, which makes it harder to reach global optimum. Normally, the real-world objective functions indeed are non-linear, non-continuous, multidimensional, which increases complexity and computational cost (MASHWANI; HAIDER; BELHAOUARI, 2021).

As an alternative to exact methods with increasing computational costs, heuristic techniques have become more popular, resulting in reasonably good local optimal solution (HUANG; LAI; CHENG, 2009). Heuristics are procedures for solving complex combinatorial optimization problems using intuitive approaches, which requires a new solution for each problem (ÓLAFSSON, 2006). Metaheuristics were also developed to deal with complex optimization problems, but they provide a more generic approach when compared to heuristics and require less work than developing a specialized heuristic for each application (ÓLAFSSON, 2006). Most of the Metaheuristic algorithms are associated with behaviors found in nature, such as the evolution process (genetic algorithm), or the self-organization of ant colonies (ant colony optimization) (BELSLEY, 2009). Metaheuristic algorithms can be classified into single solution and population based. Single solution use candidate solution and improve it by using local search, which can result in local optima solutions (KATOCH; CHAUHAN; KUMAR, 2021). Examples of common single solutions are simulated annealing and tabu search. Population-based, on the other hand, use multiple candidate solution in the search process. Some examples being genetic algorithm and ant colony optimization.

In stochastic optimization, random variables are usually associated with the search process: the same starting point does not necessarily result in the same output. Heuristic and metaheuristic have some examples of stochastic methods, that obtain acceptable solutions easily and quickly, but not necessarily optimal (LIN, 1975; ZANAKIS; EVANS, 1981).

Meta-Heuristic methods are also seen in computer aided product design in the chemical engineering field, although less frequently than deterministic methods, as listed in Table 3.

Table 3: Some meta-Heuristic Optimization Methods Applied to Chemical Product Design

Case Study	Solution Approach	References
Polymer Design	Genetic Algorithm	Venkatasubramanian, Chan, Caruthers (1994)
Refrigerant and Solvent Design	Simulated Annealing	Ourique and Telles (1998)
Molybdenum Catalyst	Tabu search	Lin et al. (2005)
Blanked Wash Mixture, Chlorinated Paraffins, Solvents	Genetic Algorithm	Heintz et al. (2014)
Cosmetics Design	Efficient Ant Colony Algorithm (EACO)	Gebreslassie and Diwekar (2015)

3.2.2. Genetic Algorithm

3.2.2.1. Introduction

Genetic Algorithms are based on Darwinian models of natural selection and evolution, in which a new population is generated iteratively through biological-inspired operators, passing on the fittest characteristics (VENKATASUBRAMANIAN; CHAN; CARUTHERS, 1994). The main terminologies used in this methodology are described in Table 4.

Table 4: Genetic Algorithm Main Concepts and Terminologies. Adapted from (ALHIJAWI; AWAJAN, 2023)

Concept	Definition
Population/Generation	List of possible solutions, composed by several individuals.
Individual	One possible solution, also named chromosome. It represents a single solution to the problem.
Initialization	First generated population in the GA algorithm, usually chosen at random.
Evaluation	Every individual is evaluated using an objective function that will attribute to the individual a performance (fitness). This fitness will rank the individual in the population.
Selection	This step selects the individuals that will compose the parents pool and generate offsprings, by going through genetic operators (crossover and mutation).
Crossover	Recombination process involving two individuals, responsible for generating offsprings.
Mutation	Random modification of an individual's part.
Replacement	Process of substituting the previous population for a new one, including the offsprings generated by genetic operators.
Stop Criteria	Criteria for the algorithm to stop.

Each generation has a certain number of chromosomes that represent possible solutions, and that are composed of different genes/alleles (ALHIJAWI; AWAJAN, 2023). Each individual in the population has a fitness score attributed to it, that represents how close they are to a certain solution (KOZENY, 2015), previously defined according to the problem the genetic algorithm is trying to solve.

In Figure 6, there is a generic framework for Genetic Algorithm. First, we have the random initial population, which is one of the stochastic step in this method. Then, there's an iteration involving a fitness score and genetic operators, until a certain termination criterion is reached.

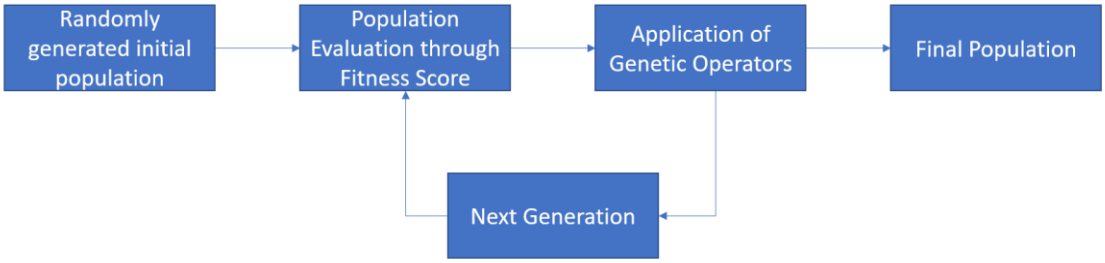

Figure 6: Generic Genetic Algorithm Framework

Evolutionary algorithms (EA), including the genetic algorithm, have variable parameters to be chosen upon when building the optimization solution. For example, the main components of an EA are the genetic operators, which includes selection and variation operators (EIBEN; SMIT, 2011), and there are different methods to perform these modifications. Thus, the choice of these parameters is very important to guarantee a good performance of the algorithm.

Algorithm design is the parameter choices needed to specify an algorithm and solve a specific problem. Within algorithm design, there is parameter tuning, which is a methodology to study how parameters affect final performance and find specific ones that optimize algorithm performance (EIBEN; SMIT, 2011).

Eiben et al. (2011) divides the possible parameters for EA into qualitative and quantitative, and within each qualitative choice, more quantitative parameters can emerge, which are called sub-parameters. Also, qualitative parameters define the main structure of the algorithm, and different quantitative values create instances of the algorithm.

The generic algorithm has two main downsides. The first is that the optimal solution is not guaranteed, as it is for other stochastic methods. The second is that there are many variable parameters in the GA, quantitative and qualitative, that have a big impact on the optimization performance.

Depending on the qualitative and quantitative parameters chosen to run the algorithm, one can encounter premature convergence. Convergence in GA means that there's a reduction in diversity in population's genetic material, which reduces the number of possible distinct offspring and reduces the potential search space of recombination (BAKER, 1989).

3.2.2.2. Qualitative Parameters

There are three main types of qualitative parameters in genetic algorithm: selection and variation methods, and termination criteria.

In each iteration of the GA, genetic operators process the previous population by replacing it by a new one (KATOCH; CHAUHAN; KUMAR, 2021), through crossover, mutation, and selection operations (KOZENY, 2015). In the selection step, individuals are selected based on its fitness value; in the crossover step, a random part of the chromosome is changed between another one to create off-springs; and in the mutation step, a random part of the chromosome will be changed (KATOCH; CHAUHAN; KUMAR, 2021). This process is repeated until the termination criteria is reached.

The selection step, also known as reproduction operator, determines if an individual will go through the reproduction process or not (KATOCH; CHAUHAN; KUMAR, 2021). Selection processes can be divided into fitness proportionate and ordinal selection (BURKE; KENDALL, 2005). Examples of fitness proportionate selection are roulette-wheel selection, stochastic universal selection, and elitist selection. Examples of ordinal selection are tournament and truncation selection. These two types are explained in Table 5.

Table 5: Different Selection Techniques for Genetic Algorithms. Adapted from (KATOCH; CHAUHAN; KUMAR, 2021)

Method	Description	Pros & Cons
Roulette-wheel (ALHIJAWI; AWAJAN, 2023)	A selection probability is calculated for each individual, proportional to its objective function, then individuals with highest probability are selected as parents	Individuals with lower performance are less probable to be selected. Risk of premature convergence to a local optimum (JEBARI; MADIAFI, 2013)
Stochastic Universal (ALHIJAWI; AWAJAN, 2023)	Population is ordered according to performance. A single random number is selected (P_1), and the parents are selected starting with P_1 and spaced by 1/M, where M is the number of parents to be selected.	Reduced risk of premature convergence (JEBARI; MADIAFI, 2013)
Elitist	The first M individuals with best performance are selected. (ALHIJAWI; AWAJAN, 2023)	Best chromosomes are kept in the population
Tournament	S individuals are randomly chosen from the population, and the fittest one wins the tournament. This tournament is repeated M times to select M parents (ALHIJAWI; AWAJAN, 2023)	Loss of diversity when tournament size is large (KATOCH; CHAUHAN; KUMAR, 2021)
Truncation	The X chromosomes with highest fitness value are chosen for reproduction, and are reproduced 1/X times (HUANG; LAI; CHENG, 2009)	Simple, more used in very large populations (JEBARI; MADIAFI, 2013)

Crossing-over is the exchange of alleles between two different individuals (HOLLAND, 1992). One generic crossover process proceeds in three steps (HOLLAND, 1992):

1. Two individuals A and B are selected at random from the population.
2. A number *x* is selected at random.
3. Two new structures (offsprings) derive from A and B, by exchanging the elements to the right of position x.

By varying *x* or the number of *xs*, there are different possible crossover techniques. In the single point crossover, there's only one crossover point for the allele swap; in the two-point or k-point crossover, on the other hand, two or more random crossover points are selected to perform the segments swapping (KATOCH; CHAUHAN; KUMAR, 2021). Crossover does not introduce new alleles to the population, it creates new combinations of existing alleles (BAKER, 1989).

Another genetic operator is mutation. In the mutation, a random segment of the chromosome is replaced or modified to generate a new structure; normally, it is associated with a low probability of happening (HOLLAND, 1992). Whereas the main role of crossover is to generate new structures, mutation ensures that no gene is permanently gone from the population and avoid local optima (HOLLAND, 1992).

After performing the genetic operators and generating the new offsprings, there are different ways to replace the new generation. Some examples are listed in Table 6.

Table 6: Different Replacement Techniques, with methodology, pros and cons (ALHIJAWI; AWAJAN, 2023; BURKE; KENDALL, 2005)

Method	Description	Pros & Cons
Delete-All	Current generation (i.e. parents) are replaced by the new offsprings generated	Children may improve population fitness; no new parameters are added to the optimization process. Parents with good fitness value are deleted from the population
Steady State	N individuals from the previous population are deleted and replaced by N new ones	Two new parameters are added to the optimization: N value and how to choose the parents to be deleted (the best ones, the worst ones or at random)
Elitism	Selects the fittest individuals from the parents and the new offspring	Parents with good fitness can be maintained in the next generation
Delete n-last	Worst n individuals of the current population are replaced by n individuals from the new offspring group.	Parents with good fitness are maintained in the next generation

To end the genetic algorithm, a termination criterion also must be chosen. Termination criteria can be based on number of iterations (generations) or a minimum value for the objective function.

3.2.2.3. Quantitative Parameters

Regardless of qualitative parameters chosen, there are 3 main quantitative values to be chosen: population size, mutation rate, and crossover rate. Population size is the number of individuals that compose each generation, that will have its fitness function calculated and will go through genetic operators. Mutation and crossover rate are probability operators that define if individuals will go through those processes.

Other quantitative sub-parameters can be necessary depending on qualitative parameters' choice. For example, elitism selection requires the number of individuals that will be chosen for reproduction; and termination criteria depends on the number of iterations or the minimum objective function value, which adds another parameter necessity. All these parameters must be chosen for each problem.

3.2.3. Computer-Aided Product Design

3.2.3.1. Concept

Computer-Aided Product Design (CAPD) is one of the ten technologies associated with the Industry 4.0 concept (DALENOGARE et al., 2018) and it is essentially the Product Design assisted by computational tools. It is the reverse problem of property prediction, meaning that given a set of previously specified properties, the CAPD uses optimization-based tools to find the molecules or mixtures that achieve these properties (GANI, 2004). CAPD requires target properties that depend on the function of the designed material, and on the available property prediction methods (HARPER; GANI, 2000). Property prediction methods are extremely important for CAPD methods as they ensure that product development is not strictly dependent on experimental data (CIGNITTI; ZHANG; GANI, 2015).

The CAPD method can be divided into computer-aided molecular design (CAMD) or computer-aided mixture/blend design (CAM^{b}D).

In the CAMD, a molecule needs to be found, given a set of building blocks and a target property. According to Gani (2004), the main steps of a CAMD are: generate feasible structures, estimate desired properties, and choose those that satisfy properties constraints. There are different methods to solve this type of problem, such as the generate and test approach, mathematical programming methods, or hybrid methods (GANI, 2004).

In the generate and test approach, one way to avoid combinatorial explosion is to use the multilevel methodology, developed by (HARPER et al., 1999). This approach consists of four levels, and, in each of them, there is a generation and screening step. The results from each level are used as input for the next one, and the most complex steps are performed at the higher levels, where the candidates are already the most promising ones. Levels 1 and 2 use macroscopic representation of molecules, while levels 3 and 4 employ microscopic representations. Macroscopic representation of molecules is composed of first and second order group building blocks, and microscopic representation includes the atomic representation of the molecules, in level 3, and a three-dimensional representation, in level 4.

Mathematical programming techniques solve the product design problems with numerical solvers, modeling each problem as MILP or MINLP models.

Hybrid techniques integrate the two previous methods, by using them in the region where they perform best (GANI et al., 2016).

In the CAM^{b}D, the potential molecular structures are known, but the identity and the fraction of each one in the final mixture are unknown. Some application examples on computer-aided mixture/blend design are solvent mixture, polymer formulations, oil blend, pesticides and paint formulations (GANI, 2004). One challenge of the mixture design, according to Gani (2004), is the need for appropriate property prediction models for mixtures, given that most formulation design problems involve non-ideal mixtures and its property models have not been developed yet, or require complex implementations in a CAM^{b}D framework.

3.2.3.2. Property Prediction Methods

Property prediction methods and models are extremely important for the success of a computer-aided product design. According to Gani and O'Connell (2001), these methods have three different roles in chemical product design: service role, where property values are provided; service plus advice role, where, besides the property values, they are used in evaluating the feasibility of the design; and integration role, where in addition to advice and service roles, the property methods contribute to the solution of the problem.

One possible classification of properties, according to Kontogeordis and Gani (2004), is between **primary** (single value properties, calculated from molecular structure information only, such as critical temperature), **secondary** (need the molecular structure information, and other properties to be calculated, such as entropy of fusion, calculated using the melting temperature and melting enthalpy) and **functional** (properties dependent on intensive variables, such as temperature and pressure, besides the molecular structural information. One example of such property is the viscosity, which depend on the temperature). Primary properties are calculated using group contribution methods; secondary properties are functions of primary properties; and functional properties are functions of primary properties, temperature and pressure (ZHANG et al., 2018).

Property prediction methods, on the other hand, can be classified as **theoretical** (e.g. Molecular Modeling) and **semi-empirical and empirical** (correlated functions based on experimental data) (GANI, 2019). Most common property models are semi-empirical, where properties are estimated from regressed values for a set of model parameters (GANI, 2019). One example of a semi-empirical method is the Group-Contribution (GC), also known as additive method, in which the molecular structure is composed of several groups (building blocks), and the final property of the molecule is the sum of each group's contribution (GANI, 2019). In the GC-Method, it is assumed that the contribution value of a given group is the same in all the compounds where it appears, and that the property value of the compound is a sum of the partial contributions of all the groups in its representation of the molecular structure (GANI, 2019). Also, in GC methods, molecular groups are represented as 1st order, 2nd order, and 3rd order, where an increase in order increases the precision of the prediction (CIGNITTI; ZHANG; GANI, 2015).

Group-Contribution methods can be useful when there is not a strong need for accuracy, such as product design itself where the promising candidates generated will be verified by experiments or more rigorous models later (GANI, 2019).

3.2.3.3. Property Prediction methods on fats and oils

Property prediction methods specific for fats and oils have been widely researched and tested in the literature. Table 7 and Table 8 show different prediction methods for viscosity and surface tension, respectively, in fatty based products. Viscosity and surface tension are some of the key properties needed to study the deodorization process of fats and oils, which is an important step in the edible oil industry (DÍAZ-TOVAR; GANI; SARUP, 2011).

Table 7: Different Prediction Methods for viscosity of fatty acids, TAGs, and vegetable oils.

Method	Sources	Inputs	Chemical species
Modified Andrade + Kay's Rule	(NOUREDDINI; TEOH; CLEMENTS, 1992) (1); (AZIAN et al., 2001) (2); (VALERI; MEIRELLES, 1997) (3)	Temperature and experimental correlation constants	FA (1) TAG (2) Vegetable Oil (3)
Group Contribution	(RABELO et al., 2000)	Temperature, number of carbons and number of double bonds	FA
Ceriani Equation	(CERIANI et al., 2007); (CERIANI; GONÇALVES; COUTINHO, 2011)	Temperature, compound chemical formula	FA
Fragment-based methodology	(ZONG; RAMANATHAN; CHEN, 2010)	Fragment composition	FA
Viscosity from Density	(ESTEBAN et al., 2012); (RODENBUSH; HSIEH; VISWANATH, 1999)	Density	Vegetable Oils
Viscosity from IV and SV	(TOSCANO et al., 2012)	Iodine and saponification	Vegetable Oils

		number, and temperature	
Fatty Acid Viscosity	(FERRER et al., 2017)	Temperature, melting temperature and constants	FA

TAG: triacylglycerol. FA: fatty acid

Table 8: Different Prediction Methods for surface tension of fatty acids, and vegetable oils.

Method	Sources	Inputs	Chemical Species
Artificial Neural Networks and Multiple Linear Regression	(MELO-ESPINOSA et al., 2014)	Fatty Acid composition	Vegetable Oils
Group Contribution	(DÍAZ-TOVAR; GANI; SARUP, 2011)	Molecular structure (functional groups)	Fatty Acids
Temperature Correlation	(CHUMPITAZ; COUTINHO; MEIRELLES, 1999)	Temperature, Experimental data	Fatty Acids

3.2.3.4. Applications

The CAPD methods are used in the pre-design step of a chemical product development, to help choose a set of possible products (ACHENIE; GANI; VENKATASUBRAMANIAN, 2003).

There are many different frameworks for developing a new chemical product, depending on the type of the final product desired and the ingredients used in the process. It is improbable that a single ingredient satisfies the needs of a product, so a mixture of different ingredients is normally sought (CONTE; GANI; NG, 2011).

Zhang et al. (2018) proposed a 7-step framework, that has the following steps: 1. identification of product attributes, 2. conversion of product attributes to property constraints, 3. identification of possible ingredients, 4. generation of a basic set of ingredient-chemicals, 5. Mixed Integer Linear/Nonlinear Programming (MILP/MINLP)

model formulation, 6. optimization model solution, using optimization platforms such as GAMS, 7. verification, and experimental iteration.

Considering the large database of compounds and non-linear constraints, normally a large combinatorial problem emerges. A way of managing this complexity is to use a decomposition-based approach (KARUNANITHI; ACHENIE; GANI, 2005), in which the search method, a MINLP algorithm, is divided into several sub-problems that are easier to solve. As previously mentioned, there are several authors that used the decomposition-based approach for the CAPD solution.

Cignitti, Zhang and Gani (2015) applied the CAM^{b}D framework for the development of solvents, using a MINLP formulation, solved through a decomposed optimization approach, where the optimum product could be a single compound, or a mixture/blend.

Zhang et al. (2018) developed a MINLP formulation for the design of surrogate fuels. Target properties included viscosity and density, which were predicted using a group-contribution method, and phase equilibrium-based properties, which were calculated through a UNIFAC model. The UNIFAC model, used to calculate the activity coefficient values of properties such as Reid vapor pressure, made the optimization model highly nonlinear and hard to solve directly. Thus, a decomposition-based algorithm was used to solve the MINLP problem.

Kalakul et al. (2018) also used a decomposition-based approach for fuel development but also included a two-step solution approach in the CAM^{b}D resolution. In this methodology, the constraints of the MINLP model are grouped into two parts: all constraints except UNIFAC model, and the UNIFAC model itself. Based on this grouping, the first group of constraints is solved with a fixed activity coefficient. Then, on step-2, the UNIFAC equations are solved with the ingredients' compositions calculated in step-1, and a new set of activity coefficients are obtained. Step-1 is repeated with the new activity coefficient values, and this iteration is repeated until a specified error is satisfied.

Conte, Gani, and Ng (2011) developed a framework for liquid formulated products, applying CAPD to case studies such as a paint formulation and an insect repellent. A formulation consists of an active ingredient (AI), which is the ingredient responsible for the main function of the product; additives, which enhance the desired properties; and a solvent mixture, responsible for dissolving the AI and other additives.

The main difference between the blend and liquid formulation product design is the existence of one new step for each category of ingredient in the latter.

Mattei, Kontogeorgis, and Gani (2014) applied a computer-aided product design framework for surfactant selection and the development of emulsions, with UV sunscreen and hand wash case studies. Their developed framework is shown in Figure 7. An emulsion is a type of formulation that consists of previously immiscible liquids with an emulsifier (also known as surfactants) that stabilizes the mixture. As in the liquid formulation, it can be also decomposed in active ingredients, additives, and solvent mixture.

The advisor of this MSc work is one of the authors of a computer-aided molecular design tool called IBSS (HEINTZ et al., 2014). The tool was design to cope with biomass-sourced molecules and it has been used in several cases, including blanket wash, chlorinated paraffins substitutes, and solvent design for active compounds extraction (HEINTZ et al., 2014). The search method used by the authors is a Genetic Algorithm.

In Table 9, there are different CAPD examples from the literature, including the above-mentioned ones.

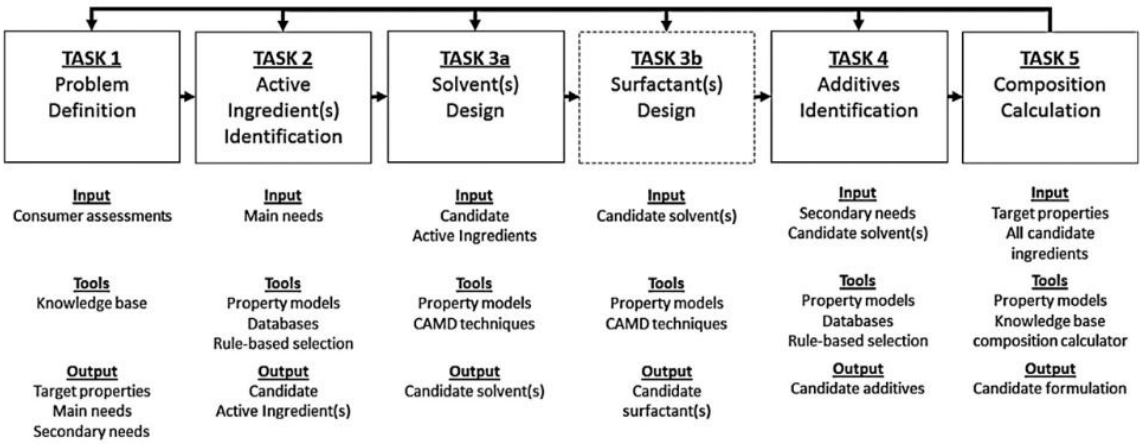

Figure 7: Computer Aided Mixture-emulsion design framework. Source: (MATTEI; KONTOGEORGIS; GANI, 2014)

Table 9: Summary of Some Computer-aided product design models in the open literature

Case Study	Solution approach	References
Paint Formulation and Insect Repellent Lotion	Decomposition-based	(CONTE; GANI; NG, 2011)
Biomass-based molecules	Genetic Algorithm	(HEINTZ et al., 2014)
Surfactant Design	Not informed	(MATTEI; KONTOGEORGIS; GANI, 2014)
Gasoline and Lubricant Blends	Decomposition-based	(YUNUS et al., 2014)
Solvent Blend	Decomposition-based	(CIGNITTI; ZHANG; GANI, 2015)
Gasoline and Diesel Blends	Decomposition-based and two-step solution	(KALAKUL et al., 2018)
Gasoline and Jet-Fuel Blends	Two-steps solution	(ZHANG et al., 2018)
Cosmetics Design	CPLEX	(ARRIETA-ESCOBAR et al., 2019)

3.2.3.5. Product Design using fats and oils

Cocoa butter is extracted from the cocoa bean, the seed from the Theobroma cocoa tree (DEPOORTERE, 2010). Cocoa butter consists of 97% of triacylglycerols (TAGs) and 3% of free fatty acids, mono- and diacylglycerols and phospholipids (SMITH, 2001). Cocoa beans are composed of of 85% cotyledon (nib) and 15% shell; the nibs contain around 55% of fat (JAHURUL et al., 2013).

A triacylglycerol (TAG) is formed by three fatty acids attached to a glycerol backbone. In cocoa butter, the main fatty acids are palmitic (P, 20 to 26%), stearic (S, 29 to 38%), oleic (O, 29 to 38%), linoleic (L, 2 to 4%), and arachidic (A, 1%) acids (DEPOORTERE, 2010). Cocoa Butter TAGs are mainly POP (palmitic-oleic-palmitic),

POS (palmitic-oleic-stearic) and SOS (stearic-oleic-stearic) (DEPOORTERE, 2010). The amount of each fatty acid and TAG vary according to the origin of the Cocoa Butter (SMITH, 2001). A higher number of unsaturated fatty acids is associated with a lower melting point (JAHURUL et al., 2013).

In chocolate formulations, cocoa butter is responsible for the dispersion of other components and for the physical behavior of chocolate, such as its typical melting profile, as it is the continuous phase (LIPP; ANKLAM, 1998).

The melting of cocoa butter occurs between 27 and 33ºC, which makes it ideal for confectionary applications (JAHURUL et al., 2013). The steepness of its melting profile has an impact on flavor release: the flavor is released as the cocoa butter melts; if the fat has a sharp melting profile, the flavor is released in a short space of time, yielding an intense flavor (SMITH, 2001). The brittleness is another important property of the cocoa butter, responsible for the "snap" when the chocolate breaks (LIPP; ANKLAM, 1998).

Cocoa butter can crystalize into different polymorphic forms, such as α, β, β', γ, with melting points of 17, 35-37, 26, 23ºC, respectively and only β crystal is used for chocolate production, because of its melting point (JAHURUL et al., 2013).

Thus, according to the temperature, the triacylglycerol mixtures that compose fats and oils are distributed among solid and liquid phases. This solid-liquid equilibrium determines the physical and sensorial properties of many fat-based products. As the temperature increases, the TAGs molecules are more likely to be in the liquid state.

The melting profile of a fat is important to product development, as it influences the application of the fat (TELES DOS SANTOS; GERBAUD; LE ROUX, 2012). Fats and oils don't have sharp melting points, different from pure compounds (FASINA et al., 2008). This happens, because they are composed of mixtures of TAGs that gradually melt before becoming completely liquid, as the temperature increases (O'BRIEN, 1988).

Solid Fat Content (SFC) is used to describe food properties, and its behavior in different conditions (AUGUSTO et al., 2012). However, the search for new mixtures of vegetable oils and fats based on its melting profiles is usually guided by heuristics, prior knowledge, and experimental data; thermodynamic and computational modeling are still not common (TELES DOS SANTOS; GERBAUD; ROUX, 2013). Teles dos Santos; Gerbaud; Le Roux, (2014) showed how modeling and simulation can aid

Product Design of vegetable oils and fats to match a desired melting profile, which can be helpful in the first steps of computer-aided mixture/blend design, to avoid combinatorial explosion in the experimental steps.

Block, Figueiredo and Gomide (1997) developed a Neural Network designed to formulate ternary mixture fats from based on a desired solid profile. Results showed that it is possible to use neural networks to automate fat formulation, however, this is highly dependent on experimental data and would probably not work when dealing with new fats and oils others than those used to train the neural network. Thus, the prediction capability for Solid Fat Content is very limited.

Fasina et al. (2008) developed a linear equation to correlate melting characteristics of a vegetable oil with the amount of monounsaturated or polyunsaturated fatty acids. Soares et al. (2009) calculated multiple regression coefficients over experimental data related to mixtures of palm stearin and palm olein, for different properties, such as softening and melting points, consistency, and solid fat content.

Augusto et al. (2012) evaluated different sigmoidal functions (Gompertz model, Power decay model, and the Logistic model, shown in equations 8, 9 and 10, respectively) to model SFC curves as a function of temperature for animal and vegetable oils, which resulted in R^2 higher than 0.96. The data for this modeling was obtained from the literature, calculated by either nuclear magnetic resonance (NMR) or differential scanning calorimeter (DSC). Farmani, (2015) also used the sigmoidal Gompertz model to fit SFC data as a function of temperature and saturated fatty acids in chemically interesterified vegetable oils.

$$SFC(\%) = a * \ e^{-e^{(b-c\,T)}} \tag{8}$$

$$SFC(\%) = \frac{a}{1+(b*t)^c} \tag{9}$$

$$SFC(\%) = \frac{a}{1+b*e^{c\,T}} \tag{10}$$

Silva, Barrera-Arellano and Ribeiro (2022) calculated Pearson correlation coefficients between trans fatty acid, saturated fatty acid, monounsaturated fatty acids

and polyunsaturated fatty acids content and some physical properties, such as adhesiveness, hardness, spreadability and solid fat content.

As stated above, several studies have used sigmoidal functions and empirical models for modeling SFC curves for oils and fats. However, equations and curves fitted to data from specific experiments cannot be generalized (PEREIRA; MEIRELLES; MAXIMO, 2020). This is where thermodynamic solid-liquid equilibrium models can provide a more accurate and general prediction model for the SFC.

Concerning rigorous thermodynamics models to the solid-liquid phase behavior of vegetable oils and fats, composed of dozens of different TAGs, Teles dos Santos; Gerbaud; Roux (2013) developed a Solid-Liquid Equilibrium model with an algorithm to minimize Gibbs free energy. Once the phase equilibria is solved, the solid fat content of a given fats or oil can be estimated in a given temperature. This method was validated both with experimental data from the literature (TELES DOS SANTOS; GERBAUD; ROUX, 2013 (TELES DOS SANTOS; GERBAUD; LE ROUX, 2014 (TELES DOS SANTOS; MORGAVI; LE ROUX, 2018), and with experiments done by the authors (TELES DOS SANTOS; GERBAUD; LE ROUX, 2012) (TELES DOS SANTOS et al., 2016). The validations performed are shown in Table 10, with the respective absolute errors and oils/fats used for validation. This is the method used in this work.

Table 10: Validation of Teles dos Santos model for SFC and melting profile values

Oils and fats used for validation	Property evaluated	Experimental data source	Average Absolute Error	References
Palm, Peanut and Grapeseed Oil	Phase Transition Temperature	Both Literature and Experiments	Between -0.72 ^{o}C and -1.29 ^{o}C	(TELES DOS SANTOS; GERBAUD; LE ROUX, 2012)
Canola oil, fully hydrogenated palm oil stearin, Palm oil stearin, Cottonseed oil, Milkfat and corn oil	SFC	Experimental data from literature	3.33% (without CI) 4.13% (after CI)	(TELES DOS SANTOS; GERBAUD; ROUX, 2013)
Palm oil, Sunflower oil, Palm kernel oil	SFC	Experimental data from literature	Between 5.2% and 6.3% before CI	(TELES DOS SANTOS;

			4.2% after CI	GERBAUD; LE ROUX, 2014)
Palm stearin, Canola oil, fully hydrogenated soybean oil	Melting Profile and SFC	Experiments	Not informed	(TELES DOS SANTOS et al., 2016)
Tucumã pulp and kernel oils, Rubber seed oil, Passion fruit oil, Buriti pulp oil, Ucuúba butter, Pracaxi seeds oil, Cupuassu fat and Palm stearin	SFC and Melting Point	Experimental data from literature	4.5ºC (MP) 3.8% (SFC)	(TELES DOS SANTOS; MORGAVI; LE ROUX, 2018)

Pereira et al. (2019) described the melting behavior of fats through a solid-liquid equilibrium model and used UNIFAC and UNIQUAC models to calculate the compounds' activity coefficients for both liquid and solid phases. This modeling was validated with predictive curves of two fats. Different from the Teles dos Santos *et al.* modeling (2013, 2014, 2016, 2018), Pereira's work considers the non-ideal behavior of both liquid and solid phases, whereas Teles dos Santos considers only the solid phase as non-ideal.

Pereira, Meirelles and Maximo (2020) gathered all thermodynamic modeling approaches for Solid Fat Content estimation for triacylglycerols and compared them in Table 11.

Table 11: Summary of different thermodynamic approaches for Solid Fat Content in Vegetable oils and fats estimation

Modeling Approach	Activity coefficient model	References
Experimental values or Group-Contribution method for pure compounds, solid-liquid equilibrium for mixtures	Equation 11	(WON, 1993)
Number of carbons for pure compounds, solid-liquid equilibrium for mixtures	Margules equations	(WESDORP, 1990) (HJORTH et al., 2015)
Group-Contribution method for pure compounds, solid-liquid equilibrium for mixtures	Margules equations	(TELES DOS SANTOS et al., 2016); (TELES DOS SANTOS; GERBAUD; LE ROUX, 2013, 2014); (TELES DOS SANTOS; MORGAVI; LE ROUX, 2018).
Number of carbons for pure compounds, solid-liquid equilibrium for mixtures	Predictive UNIQUAC	(PEREIRA et al., 2019)

Source: Adapted from Pereira, Meirelles, and Maximo (2020)

A diversity of chemical products has attracted the interest in the Product Design area, such as fuels additives, pharmaceuticals, polymers, ionic liquids and solvents, as previously shown in section 3.1.3. However, up to our knowledge, there is a lack of integrated computational-experimental works in the early steps of guiding product design in the fats and oils industry. This work aims to fill this gap, coupling the model previously developed by Teles dos Santos et al. with a genetic algorithm-based search model developed in this work.

3.2.4. Cocoa Butter Equivalent Problem

Cocoa is mainly consumed as chocolate or cocoa powder, including beverages and cakes (AFOAKWA, 2016). The consumption of chocolate products has significantly increased worldwide. (LOULLIS; PINAKOULAKI, 2018) highlighted that to overcome the challenges of the increasing demand for cocoa, the variable cocoa supply and its rising price, it is necessary to use other fats as substitutes. The research and use of alternative fats as cocoa butter substitutes is a traditional practice in the food industry. The European Chocolate Directive (Directive 2000/36/EC) allows the addition of up to 5% of vegetable fats (CBE) other than cocoa butter in chocolate. Also, it defines that only the following 6 vegetable fats may be added to chocolate products, besides cocoa butter: shea butter, palm-oil, sal fat, kokum butter, illipe fat and mango kernel fat. Although other solutions have been proposed to cope with this problem, such as the use of oleogels (LI; LIU, 2019) and interesterification reaction between different oils and fats (ABIGOR et al., 2003), the diversity of underexploited components and the possibility to mix them can still offer new solutions for such an old problem. Thus, given the large diversity of oils and fats, a product design combinatorial problem is set and computational tools can aid in this search procedure, prior to the experimental step.

4. Methodology

Methodologies used in this project are shown in Figure 8, and described in the sections 4.1, 4.2, and 4.3.

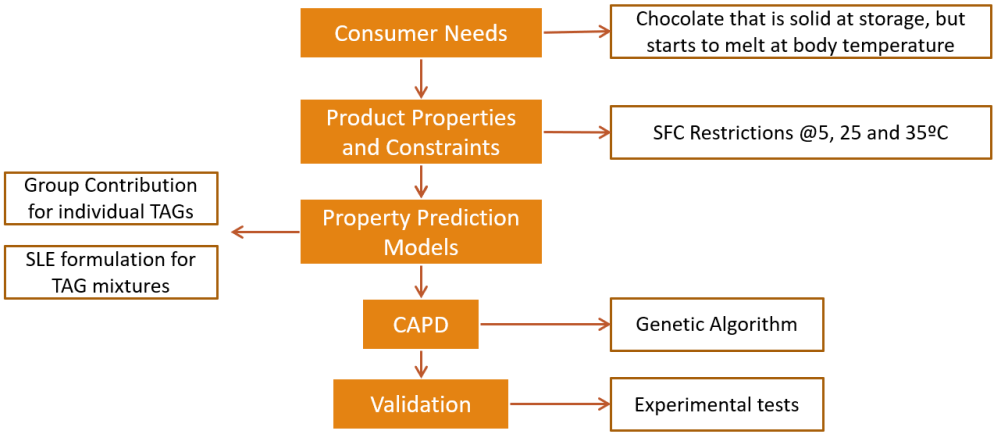

Figure 8: Methodologies used in each step of the CAPD.

4.1. Property Models

A set of desired physical chemical properties can be chosen in the cocoa butter equivalents design. One of the most important is the Solid Fat Content, as the solid percentage of TAGs determines physico chemical properties and sensory attributes of many products (Teles dos Santos et al; 2013). Thus, this work uses the SFC as the target property.

The SFC prediction model used in this work is divided in three steps: generation of TAGs, solution of the solid-liquid equilibrium model, and computation of the SFC for a given temperature., This methodology was developed in a previous work (TELES DOS SANTOS; GERBAUD; ROUX, 2013), and was used as a starting point for the current methodology. Also, it is important to mention that the modeling and simulation was already validated with experimental data from literature, as stated in Teles dos Santos et al. (2013).

Triacylglycerol (TAG) composition is predicted by combinatorial assembly of fatty acids composition of a given fat/oil in the glycerol structure. In this work, the random distribution of fatty acids in the glycerol (no preferential position) is assumed. Thus, only fatty acid mass distribution is necessary as input to generate all possible

TAGs that account for 95% of the total mass, as described in previous work (TELES DOS SANTOS; GERBAUD; ROUX, 2013). A FORTRAN 90 code embedded in a main framework called VOSTAT copes with this task, along with the calculation of thermodynamic interaction parameters, melting temperature, and melting enthalpy which are used in the next steps.

The prediction of SFC in each temperature can be viewed as a thermodynamic solid-liquid equilibrium (SLE) problem. This problem is solved by minimizing the Gibbs Free Energy function and finding the distribution between solid and liquid TAG composition (the second order condition for phase equilibrium), as stated in equation 12. This optimization problem is nonlinear and aims at minimizing the Gibbs free energy function (G), subject to linear material balance constraints (equations 13 and 14), where nc and np are the number of different TAGs and the number of phases in the mixture, respectively; n_i^j and μ_i^j represent the number of mols and the chemical potential of TAG i in phase j, respectively and n_i is the total number of mols of TAG i.

$$min\ G(n) = \sum_{i=1}^{nc} \sum_{j=1}^{np} n_i^j \mu_i^j(n) = \sum_{j=1}^{np} n^j g^j \quad (12)$$

s.t.

$$n_i = \sum_{j=1}^{np} n_i^j \ , i = 1 \dots nc \quad (13)$$

$$0 \leq n_i^j \leq n_i \ , i = 1 \dots nc;\ j = 1 \dots np \quad (14)$$

The intensive Gibbs energy is the weighted sum of partial Gibbs energy of the components, as stated in equation 15. By definition, the partial Gibbs Energy is equal to its chemical potential. $\mu_{i,0}^j$ is the chemical potential of a pure TAG, and γ_i^j is the activity coefficient of TAG i.

$$g^j = \sum_{i=1}^{nc} x_j^j (\underline{g_i^j}) = \sum_{i=1}^{nc} x_j^j (\mu_i^j) = \sum_{i=1}^{nc} x_j^j (\mu_{i,0}^j + RTln\gamma_i^j x_i^j) \quad (15)$$

When j = liquid, $\mu_{i,0}^j = 0$ and $\gamma_i^j = 1$ (considered as ideal)

$$g^{liquid} = RT \sum_{i=1}^{nc} x_j^{liquid} (lnx_i^{liquid}) \quad (16)$$

When j = solid:

$$\mu_{i,0}^{j} = T\Delta H_{m,i}^{solid(j)}\left(\frac{1}{T} - \frac{1}{T_{m,i}^{solid(j)}}\right) \quad (17)$$

Where $T_{m,i}^{solid(j)}$ and $\Delta H_{m,i}^{solid(j)}$ are the melting temperature and melting enthalpy of pure TAG i in solid state j.

The activity coefficient for the non-ideal solid phases is calculated with 2-suffix Margules model, which is an Excess Gibbs Free Energy Model. This model was chosen due to the existence of an experimental database in triacylglycerols that allows computing the interaction parameters between a pair of TAGs i and j (Aij) (Wesdorp et al., 2005).

The melting temperatures and melting enthalpies necessary for the calculation can be found in the program via experimental data gathered from literature. When the data is not available, group contribution methods, such as that of (ZEBERG-MIKKELSEN; STENBY, 1999) are used to estimate these properties. For a complete description of the methods used, the reader must consult previous works (Teles dos Santos et al., 2013, 2014, 2016).

Once the number of moles of each TAG in each phase in equilibrium is determined, the SFC for a given temperature (SFC(T)) is computed according to Equation 18, where Mi is the molar mass of TAG i.

$$SFC\ (T) = \frac{\sum_{j=1}^{np=1}\sum_{i=1}^{nc}(n_i^j)M_i}{\sum_{j=1}^{np}\sum_{i=1}^{nc}(n_i^j)M_i} \quad (18)$$

The NLP optimization problem generated by the SLE model is then solved using a Generalized Reduced Gradient Method (CONOPT 3 solver). This optimization program was coupled with the VOSTAT framework, already described in the former sections. When a particular polymorphic form is evaluated, the number of mols of all TAGs in the other two is set to zero. We assume, as in Won (1993), that all the triglyceride constituent molecules exhibit the same polymorphic behavior in one phase. Despite β being the most common solid crystal for chocolate usage (JAHURUL et al., 2013), and the extensive experimental validations on the VOSTAT tool, it was

important to have a confidence interval on the results. Thus, all results of this work refer to the β' solid state, as it represents an intermediate result between α and β polymorphs.

For further details, a previous work of the authors can be consulted (TELES DOS SANTOS; GERBAUD; LE ROUX, 2012).

4.2. Integration of Computational Tools

The VOSTAT software previously developed by Teles dos Santos, *et al.* (2013) predicts the whole melting curve for a given fat/oil mixture based on the Solid Fat Content predictions, as shown in section 4.1. The inputs and outputs of the VOSTAT are shown in Figure 9; temperature interval and step are needed, along with fatty acid composition and minimal fraction, which stands for the percentage of TAGs considered in the SFC estimation. For these calculations, three different tools are involved: Fortran, GAMS and Matlab, illustrated in Figure 10.

The program starts manually with the Product Designer choosing the desired vegetable oils with its composition in the mixture, chosen from a pre-existing vegetable oil database that already contains fatty acid composition. This selection is done in the .exe file main program, written in Fortran. This program also includes TAG composition prediction, melting temperature, and melting enthalpy databases and predictions, when necessary. After the Fortran calculations are performed, GAMS is then called via batch files to solve the NLP optimization problem using a CONOPT 3 solver. The GAMS program reads a set a txt files, containing data generated by Fortran, performs the optimization and saves the results in an Excel file. Then, Matlab is called to plot the SFC by temperature charts.

One of the steps of the current work is to automate this search, by adding another layer of optimization (a genetic Algorithm, in Python) on top of the SFC calculation, as illustrated in Figure 11. A program in Python was added and, at the beginning, generates a random list of vegetable oils to send via txt file to Fortran and GAMS. After the SFC is calculated with the VOSTAT for all vegetable oils in the list, the program goes back to Python, which via a Genetic Algorithm, selects the best options, creates new ones, and goes back to the Fortran. This iteration needs to have a termination criterion, which will be further explained in section 4.3.

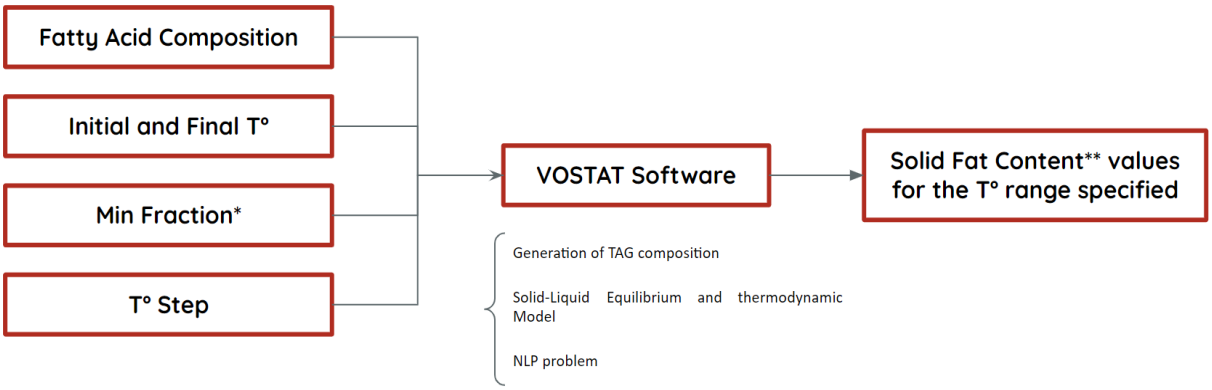

Figure 9: Inputs and Outputs of VOSTAT Software

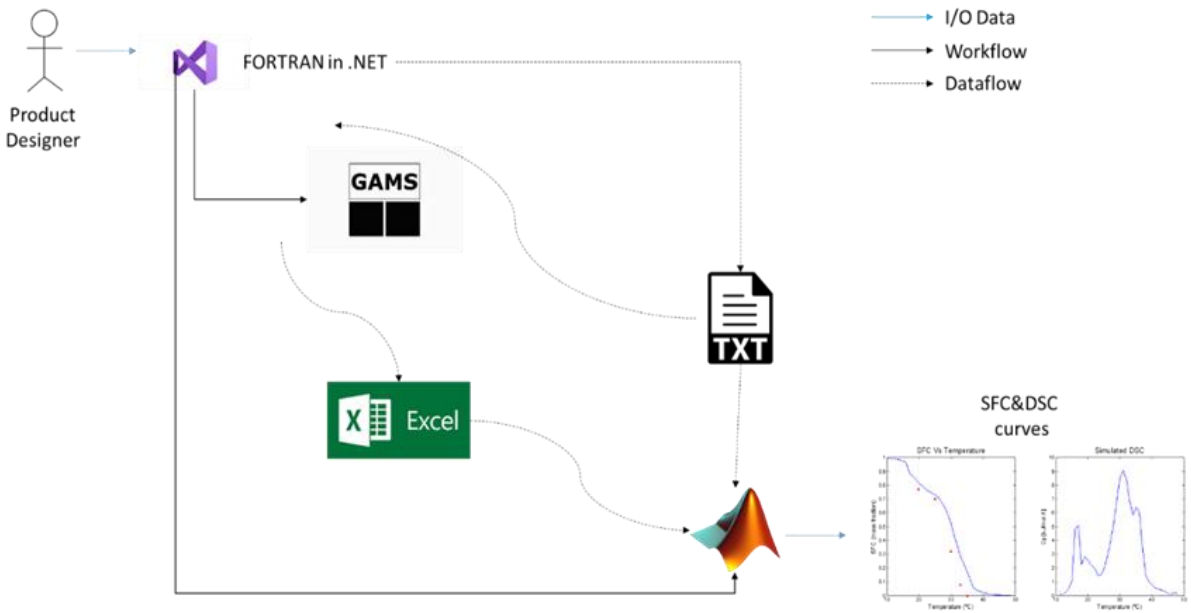

Figure 10: Initial SFC calculation Framework

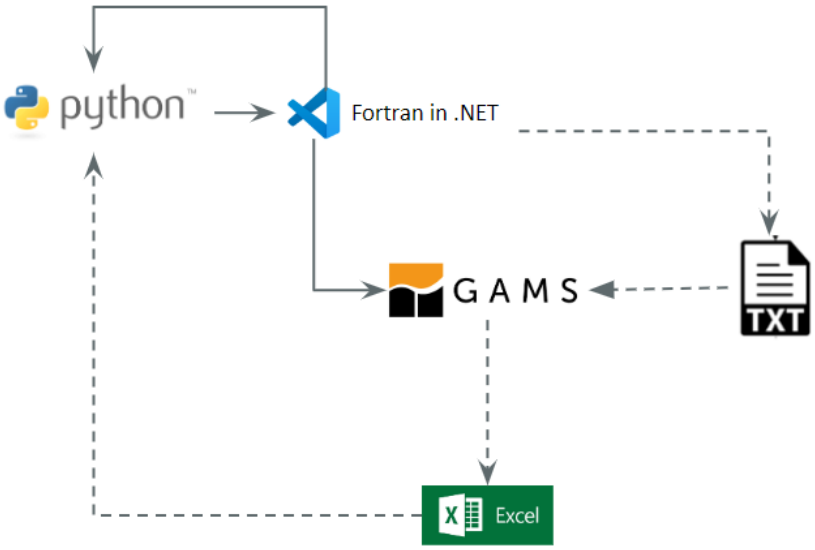

Figure 11: Modified tool by a Python framework developed in this work

4.3. Search Algorithm

In this work, a second layer of optimization was added on top of the SFC thermodynamic optimization that already existed in the program. This extra optimization was developed using a Genetic Algorithm. A Meta-Heuristic algorithm was chosen, due to the high complexity of the SFC model and the high number of possible mixtures. The Genetic Algorithm was chosen specifically, because it is simpler to implement than other Meta-Heuristic algorithms and because there are examples in the literature that show this can be a good option for chemical optimization problems.

In this project, qualitative parameters were chosen based on previous studies and remained fixed throughout the tests; a small sensitivity analysis was performed on quantitative parameters and sub parameters. Parameters used and varied in this simulation are described in Table 12. Crossover method is shown in Figure 12, and selection and replacement methods in Figure 13.

Table 12: Parameters applied to the simulation.

Parameter	Values for Single Oils Search	Values for Binary Mixtures Search
Population/Generation	10, 20, 30	30, 50, 100
Individual	One oil ID per individual	Two oil IDs and the percentage of the first one
Initialization	Random Individuals	Random Individuals
Evaluation	Fitness/Objective function = average of SFC values at three different temperatures	Same as Single Oils
Selection	NA	Elitism of 50%, 70%, 80% and 90%.
Crossover	NA	Single Point: exchange of oil ID and/or oil percentage.
Mutation	Random modification of oil ID	Random modification of oil ID and/or percentage
Replacement	Elitism	Elitism of 10%, 20%, 30% and 50%.
Stop Criteria	Number of generations (20)	Number of generations (50)

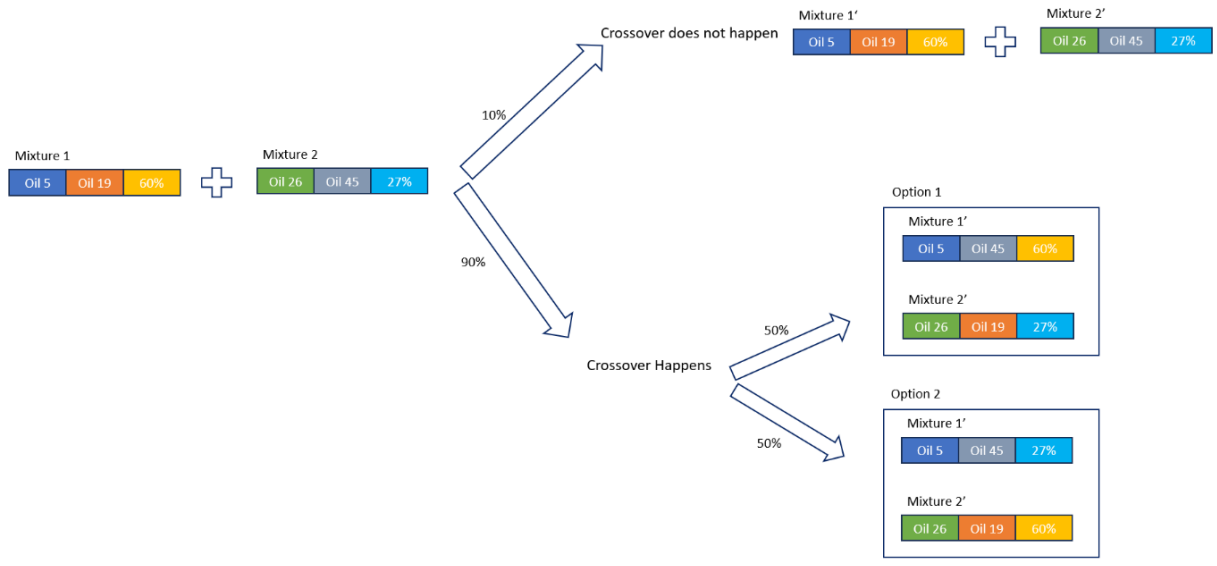

Figure 12: Example of the crossover method used in the proposed model.

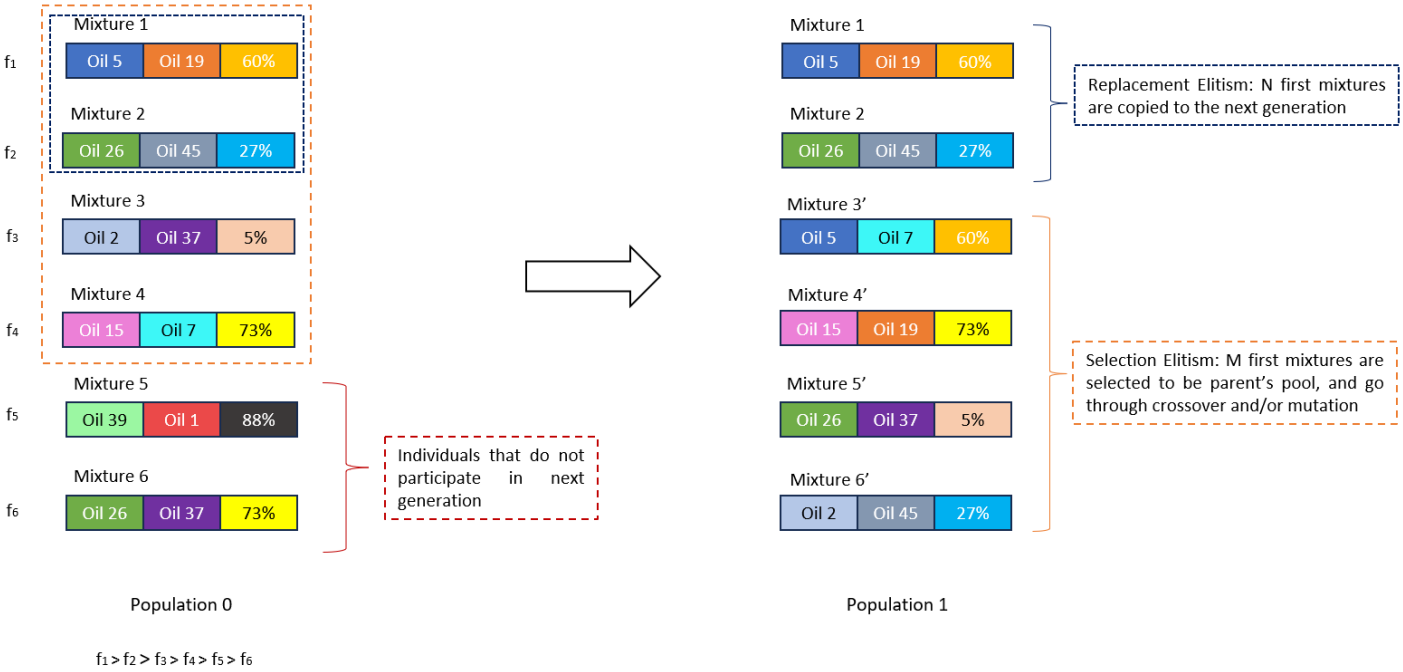

Figure 13: Selection and Replacement Methods used in the optimization.

The objective function created considers a desired solid fat content at three different temperatures: 5, 25 and 35ºC. These temperatures were chosen based on the current case study, which involves finding oils, fats and its mixtures that match melting properties of Cocoa Butter, that has a low and sharp melting point, ranging from 27 to 35ºC (JAHURUL et al., 2013), represented in Figure 14.

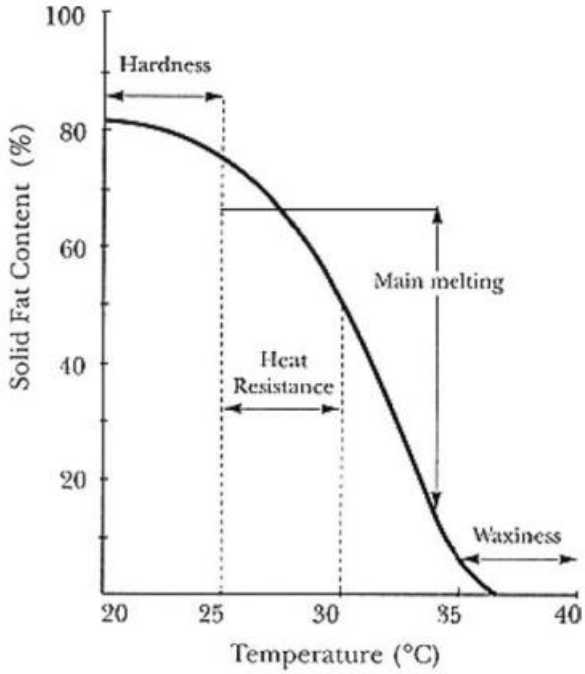

Figure 14: Solid Fat Content curve of Cocoa Butter. Source: (NILSON, 1986)

The three intermediate objective functions are represented in equations 19, 20 and 21, which represents the desired behavior of potential solutions:

- At 5ºC (fridge temperature), Cocoa Butter should contain at least 80% of its composition at solid state (constraint f(1), represented in equation 19).

- At 25ºC, which represents the shelf-temperature, it should be solid as well, with a solid fraction above 75% (constraint f(2), represented in equation 20); and

- At 35ºC it should be completed liquid, because it represents the body-temperature and, thus, the consumption temperature, and it should have solid fraction below 5% (constraint f(3), represented in equation 21).

$$f(1) = \begin{cases} 1, & SFC(@5^oC) \geq 0{,}80 \\ SFC/0{,}80, & SFC(@5^oC) < 0{,}80 \end{cases} \quad (19)$$

$$f(2) = \begin{cases} 1, & SFC(@25^oC) \geq 0{,}75 \\ SFC/0{,}75, & SFC(@25^oC) < 0{,}75 \end{cases} \quad (20)$$

$$f(3) = \begin{cases} 1, & SFC(@35^oC) \leq 0{,}05 \\ 0{,}05/SFC, & SFC(@35^oC) > 0{,}05 \end{cases} \quad (21)$$

$$f(objective) = \frac{f(1)+f(2)+f(3)}{3} \quad (22)$$

The SFC at these three temperatures are therefore calculated, and then a value between 0 and 1 was attributed to each one of them, with 1 being the closest to the desired constraint, and 0 the furthest. An average of the three values was calculated as the individual's fitness function (equation 22) and is used in the evaluation step.

After the fitness function was calculated, the population was ordered, and the genetic operators were applied.

For the single individuals, the elitism value chose the oils that would continue in the next population, and the remaining were changed for different ones. This process was repeated until maximum number of generations, which varied between 10 and 30.

For binary mixtures, the elitism number chose the parent pool, and they went through single point crossover of oil ID and proportion of the oils, and mutation. Then, worst individuals were deleted and replaced by generated offsprings. This iteration was repeated until the maximum number of generations, which varied between 10 and 50.

Figure 15 shows an overview of the Genetic Algorithm framework, and how it communicates to the VOSTAT software. The genetic algorithm aims to maximize the so-called performance (the arithmetic mean of the 3 objective functions).

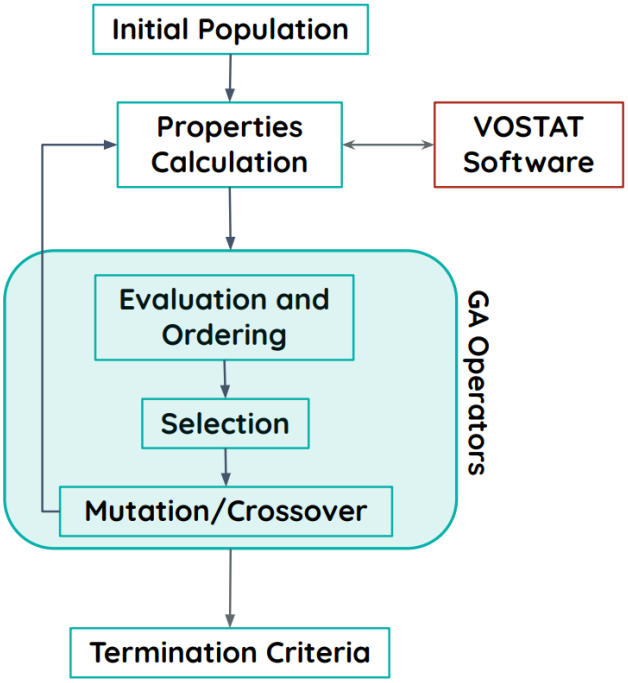

Figure 15: Overview of the Genetic Algorithm framework used in this work.

5. Results

5.1. Case Study: Cocoa Butter Alternatives

5.1.1. Introduction

The methodology was divided into single vegetable oils search and binary mixtures search. For both, some adjustments on the original VOSTAT Software were necessary, to adapt the program to run on another computer. As already mentioned, four main softwares were used in the entire optimization process: Fortran, Python, GAMS and Matlab, each of them with different licenses and access methods, which made the optimization process less efficient.

The objective function was as an average of three cocoa butter SFC constraints, as detailed in section 4.3 and the same was used for both use cases: single and binary mixtures.

Each intermediate function curve according to the SFC value is represented in Figure 16. For f(1) and f(2), the curve has a linear increasing shape until it reaches the desired value of 0.80 and 0.75, respectively, when the function is set to 1.0 (best value). For f(3), it is 1 at lower SFCs, and then it decreases as a function of 0.05/SFC.

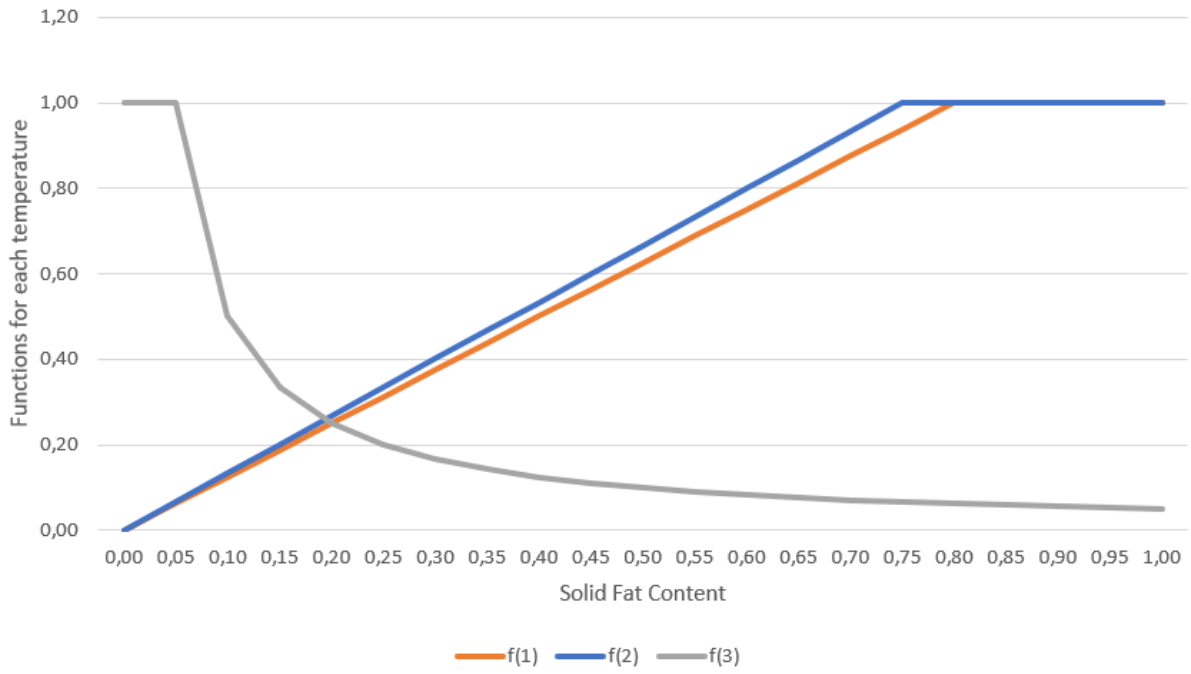

Figure 16: Variation of objective function in each temperature, according to solid fat content values.

A database of 45 oils and fats was used in the optimization process, presented in Table 13. In appendix A it is listed each oil's fatty acid composition. With 45 oils, a total of 98,010 combinations were possible in the binary mixture search space.

Table 13: Oil and Fats used in the Optimization Process

Oil ID	Oil Name
1	Açaí Oil
2	Argan Oil
3	Avocado Oil
4	Babaçu Oil
5	Brazilian Nut Oil
6	Buriti Pulp Oil
7	Canola Oil
8	Coconut Oil
9	Corn Oil
10	Cotton Seed Oil
11	Cupuaçu Oil
12	Andiroba Oil
13	Vaccenic-enriched Oil
14	Copaíba Oil
15	Fully Hydrogenated Soy Fat
16	Grapeseed Oil
17	Illipe Butter
18	Jatropa Butter
19	Kokum Kernel Fat
20	Licuri Oil
21	Linseed Oil
22	Macadamia Nut Oil
23	Macaúba Kernel Oil
24	Mango Oil
25	Mango Seed Butter
26	Microalgae Oil
27	Murumuru Butter
28	Olive Oil
29	Palm Oil
30	Palm Stearin
31	Passion Fruit Oil
32	Passion Fruit Seed Oil
33	Peanut Oil
34	Pequi Oil
35	Pomegranate Oil
36	Pracaxi Oil
37	Rice Brain Oil
38	Rubber Seed Oils

39	Sal Fat
40	Shea Butter
41	Soybean Oil
42	Sunflower Oil
43	Tucumã Pulp Oil
44	Tucumã Kernel Oil
45	Ucuúba Butter

5.1.2. Single Vegetable oils

As the single vegetable oil is the simplest application, two approaches were applied in this case: the exhaustive search, and the genetic algorithm search, which helped validate the methodology.

In the exhaustive search, the solid fat content values of a database previously built containing 45 oils and fats were predicted using the VOSTAT Software, methodology already explained in Section 4. Then, the objective function for each of them was calculated to rank the solutions according to the desired performance.

After calculating all SFC values, it was already possible to detect which were the best ones, by analyzing the objective function values. Tables 14 and 15 show the predicted SFC values for the 45 fats and oils in different temperatures. The color red represents low values of SFC, the color green indicates high values of SFC, and the color yellow represents intermediate values.

Table 14 shows SFC calculated for every fat/oil, where 100% mean a completely solid fat, and 0% a completely liquid oil. Table 15 contains the functions that compose the final objective at the 3 analyzed temperatures (5, 25 and 35^{o}C), and the final objective function, which is an average of those three. In Table 15 , the value 1 would be the product closest to the desired properties, and 0 the furthest one. In this case, best oils and fats for this specific case study would be 15, 20, 27, 45, 4, and 19, which are respectively fully hydrogenated Soy fat, Licuri oil, Murumuru butter, Ucuúba butter, Babassu oil, and Kokum kernel fat.

Table 14: Results from SFC calculation of the entire database

Temperature /Oil ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
5	11%	1%	0%	75%	1%	6%	0%	92%	0%	4%	34%	20%	38%	1%	100%	0%	78%	2%	100%	65%	0%	1%	48%	32%	48%	9%	100%	0%	42%	65%	0%	0%	1%	30%	0%	10%	6%	0%	45%	68%	0%	0%	17%	100%	100%
6	11%	1%	0%	70%	1%	6%	0%	87%	0%	4%	33%	19%	28%	1%	100%	0%	74%	1%	100%	59%	0%	0%	42%	32%	44%	9%	100%	0%	42%	64%	0%	0%	1%	30%	0%	9%	6%	0%	41%	62%	0%	0%	17%	97%	100%
7	11%	1%	0%	65%	1%	5%	0%	83%	0%	3%	32%	18%	10%	1%	100%	0%	69%	1%	100%	56%	0%	0%	35%	32%	43%	9%	100%	0%	41%	64%	0%	0%	0%	30%	0%	8%	5%	0%	40%	55%	0%	0%	16%	97%	100%
8	10%	1%	0%	62%	1%	5%	0%	79%	0%	3%	31%	16%	0%	0%	100%	0%	66%	1%	100%	52%	0%	0%	32%	32%	42%	8%	100%	0%	41%	62%	0%	0%	0%	29%	0%	7%	5%	0%	39%	51%	0%	0%	16%	96%	100%
9	9%	0%	0%	58%	0%	5%	0%	75%	0%	3%	31%	15%	0%	0%	100%	0%	65%	1%	100%	50%	0%	0%	29%	32%	41%	8%	100%	0%	40%	62%	0%	0%	0%	29%	0%	6%	4%	0%	38%	50%	0%	0%	15%	96%	100%
10	9%	0%	0%	55%	0%	5%	0%	72%	0%	2%	30%	14%	0%	0%	100%	0%	65%	0%	100%	48%	0%	0%	26%	32%	39%	8%	100%	0%	39%	61%	0%	0%	0%	28%	0%	5%	4%	0%	37%	49%	0%	0%	15%	94%	100%
11	8%	0%	0%	52%	0%	4%	0%	69%	0%	2%	28%	13%	0%	0%	100%	0%	64%	0%	69%	45%	0%	0%	23%	31%	38%	8%	100%	0%	38%	61%	0%	0%	0%	27%	0%	4%	3%	0%	36%	49%	0%	0%	14%	89%	100%
12	7%	0%	0%	49%	0%	4%	0%	66%	0%	2%	27%	11%	0%	0%	100%	0%	62%	0%	68%	43%	0%	0%	21%	31%	38%	8%	91%	0%	37%	60%	0%	0%	0%	26%	0%	4%	2%	0%	35%	48%	0%	0%	13%	86%	100%
13	6%	0%	0%	45%	0%	3%	0%	63%	0%	2%	26%	9%	0%	0%	100%	0%	61%	0%	68%	40%	0%	0%	19%	31%	37%	8%	87%	0%	36%	59%	0%	0%	0%	26%	0%	3%	2%	0%	34%	47%	0%	0%	12%	82%	100%
14	5%	0%	0%	43%	0%	2%	0%	60%	0%	2%	25%	7%	0%	0%	100%	0%	59%	0%	68%	36%	0%	0%	18%	31%	35%	8%	85%	0%	35%	59%	0%	0%	0%	24%	0%	2%	1%	0%	33%	47%	0%	0%	10%	78%	100%
15	4%	0%	0%	41%	0%	2%	0%	58%	0%	2%	25%	6%	0%	0%	100%	0%	56%	0%	68%	33%	0%	0%	16%	30%	35%	8%	84%	0%	33%	58%	0%	0%	0%	23%	0%	2%	0%	0%	32%	46%	0%	0%	9%	75%	100%
16	3%	0%	0%	39%	0%	1%	0%	54%	0%	2%	24%	5%	0%	0%	100%	0%	53%	0%	68%	32%	0%	0%	15%	30%	34%	8%	82%	0%	32%	57%	0%	0%	0%	21%	0%	2%	0%	0%	31%	45%	0%	0%	8%	73%	100%
17	2%	0%	0%	38%	0%	0%	0%	52%	0%	2%	23%	3%	0%	0%	100%	0%	52%	0%	67%	30%	0%	0%	14%	29%	33%	8%	80%	0%	30%	55%	0%	0%	0%	19%	0%	2%	0%	0%	30%	45%	0%	0%	7%	71%	100%
18	1%	0%	0%	36%	0%	0%	0%	51%	0%	2%	21%	1%	0%	0%	100%	0%	50%	0%	67%	28%	0%	0%	13%	28%	32%	8%	79%	0%	28%	53%	0%	0%	0%	17%	0%	1%	0%	0%	29%	44%	0%	0%	5%	69%	100%
19	1%	0%	0%	35%	0%	0%	0%	49%	0%	2%	20%	0%	0%	0%	100%	0%	47%	0%	67%	26%	0%	0%	12%	28%	31%	8%	78%	0%	25%	51%	0%	0%	0%	15%	0%	1%	0%	0%	28%	43%	0%	0%	4%	66%	100%
20	1%	0%	0%	33%	0%	0%	0%	47%	0%	2%	19%	0%	0%	0%	100%	0%	46%	0%	66%	25%	0%	0%	11%	27%	29%	8%	77%	0%	22%	49%	0%	0%	0%	13%	0%	1%	0%	0%	27%	42%	0%	0%	2%	65%	100%
21	1%	0%	0%	32%	0%	0%	0%	46%	0%	2%	18%	0%	0%	0%	100%	0%	44%	0%	66%	23%	0%	0%	10%	25%	28%	8%	76%	0%	20%	46%	0%	0%	0%	11%	0%	1%	0%	0%	26%	41%	0%	0%	2%	63%	100%
22	1%	0%	0%	31%	0%	0%	0%	44%	0%	2%	17%	0%	0%	0%	100%	0%	42%	0%	65%	22%	0%	0%	9%	24%	26%	8%	75%	0%	18%	42%	0%	0%	0%	9%	0%	1%	0%	0%	24%	40%	0%	0%	2%	62%	100%
23	1%	0%	0%	29%	0%	0%	0%	42%	0%	2%	16%	0%	0%	0%	100%	0%	40%	0%	64%	21%	0%	0%	8%	22%	24%	8%	74%	0%	15%	38%	0%	0%	0%	7%	0%	1%	0%	0%	23%	38%	0%	0%	2%	61%	100%
24	1%	0%	0%	28%	0%	0%	0%	41%	0%	1%	15%	0%	0%	0%	100%	0%	37%	0%	63%	19%	0%	0%	7%	21%	23%	8%	73%	0%	12%	35%	0%	0%	0%	6%	0%	1%	0%	0%	21%	36%	0%	0%	2%	59%	100%
25	1%	0%	0%	27%	0%	0%	0%	39%	0%	1%	14%	0%	0%	0%	100%	0%	33%	0%	62%	18%	0%	0%	6%	20%	22%	7%	72%	0%	11%	32%	0%	0%	0%	6%	0%	1%	0%	0%	20%	34%	0%	0%	2%	58%	100%
26	1%	0%	0%	25%	0%	0%	0%	37%	0%	1%	13%	0%	0%	0%	100%	0%	32%	0%	61%	16%	0%	0%	5%	19%	21%	7%	70%	0%	11%	28%	0%	0%	0%	6%	0%	1%	0%	0%	19%	32%	0%	0%	1%	57%	100%
27	0%	0%	0%	24%	0%	0%	0%	35%	0%	1%	11%	0%	0%	0%	100%	0%	31%	0%	59%	15%	0%	0%	4%	18%	19%	7%	69%	0%	11%	25%	0%	0%	0%	6%	0%	1%	0%	0%	17%	29%	0%	0%	1%	55%	99%
28	0%	0%	0%	22%	0%	0%	0%	33%	0%	1%	9%	0%	0%	0%	100%	0%	30%	0%	56%	13%	0%	0%	3%	16%	18%	7%	67%	0%	11%	25%	0%	0%	0%	6%	0%	1%	0%	0%	16%	28%	0%	0%	1%	54%	95%
29	0%	0%	0%	20%	0%	0%	0%	31%	0%	0%	7%	0%	0%	0%	100%	0%	28%	0%	53%	12%	0%	0%	3%	14%	15%	6%	66%	0%	10%	24%	0%	0%	0%	6%	0%	1%	0%	0%	14%	26%	0%	0%	1%	53%	93%
30	0%	0%	0%	17%	0%	0%	0%	29%	0%	0%	7%	0%	0%	0%	100%	0%	26%	0%	48%	10%	0%	0%	2%	11%	13%	6%	65%	0%	10%	24%	0%	0%	0%	5%	0%	1%	0%	0%	11%	24%	0%	0%	0%	51%	93%
31	0%	0%	0%	16%	0%	0%	0%	27%	0%	0%	7%	0%	0%	0%	100%	0%	23%	0%	43%	8%	0%	0%	1%	11%	11%	6%	63%	0%	9%	24%	0%	0%	0%	5%	0%	1%	0%	0%	9%	21%	0%	0%	0%	49%	92%
32	0%	0%	0%	14%	0%	0%	0%	23%	0%	0%	7%	0%	0%	0%	100%	0%	21%	0%	41%	7%	0%	0%	0%	10%	11%	5%	62%	0%	8%	23%	0%	0%	0%	4%	0%	1%	0%	0%	9%	18%	0%	0%	0%	47%	92%
33	0%	0%	0%	12%	0%	0%	0%	21%	0%	0%	6%	0%	0%	0%	100%	0%	21%	0%	38%	5%	0%	0%	0%	10%	11%	4%	60%	0%	8%	22%	0%	0%	0%	4%	0%	1%	0%	0%	9%	14%	0%	0%	0%	45%	92%
34	0%	0%	0%	9%	0%	0%	0%	18%	0%	0%	6%	0%	0%	0%	100%	0%	21%	0%	34%	4%	0%	0%	0%	9%	10%	4%	58%	0%	7%	21%	0%	0%	0%	3%	0%	1%	0%	0%	9%	14%	0%	0%	0%	42%	91%
35	0%	0%	0%	7%	0%	0%	0%	16%	0%	0%	5%	0%	0%	0%	100%	0%	20%	0%	29%	3%	0%	0%	0%	9%	10%	3%	55%	0%	7%	21%	0%	0%	0%	3%	0%	1%	0%	0%	8%	14%	0%	0%	0%	39%	91%

Table 15: Objective Function calculation for each vegetable oil.

Temperature/Oil ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	0.14	0.01	0.00	0.94	0.01	0.08	0.00	1.00	0.00	0.05	0.43	0.25	0.48	0.01	1.00
25	0.01	0.00	0.00	0.36	0.00	0.00	0.00	0.52	0.00	0.01	0.19	0.00	0.00	0.00	1.00
35	1.00	1.00	1.00	0.71	1.00	1.00	1.00	0.31	1.00	1.00	1.00	1.00	1.00	1.00	0.05
Final Objective	0.38	0.34	0.33	0.67	0.34	0.36	0.33	0.61	0.33	0.35	0.54	0.42	0.49	0.34	0.68

Temperature/Oil ID	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
5	0.00	0.98	0.03	1.00	0.81	0.00	0.01	0.60	0.40	0.60	0.11	1.00	0.00	0.53	0.81
25	0.00	0.44	0.00	0.83	0.24	0.00	0.00	0.08	0.27	0.29	0.09	0.96	0.00	0.15	0.43
35	1.00	0.25	1.00	0.17	1.00	1.00	1.00	1.00	0.56	0.50	1.00	0.09	1.00	0.71	0.24
Final Objective	0.33	0.56	0.34	0.67	0.68	0.33	0.34	0.56	0.41	0.46	0.40	0.68	0.33	0.46	0.49

Temperature/Oil ID	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
5	0.00	0.00	0.01	0.38	0.00	0.13	0.08	0.00	0.56	0.85	0.00	0.00	0.21	1.00	1.00
25	0.00	0.00	0.00	0.08	0.00	0.01	0.00	0.00	0.27	0.45	0.00	0.00	0.03	0.77	1.00
35	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.63	0.36	1.00	1.00	1.00	0.13	0.05
Final Objective	0.33	0.33	0.34	0.49	0.33	0.38	0.36	0.33	0.48	0.55	0.33	0.33	0.41	0.63	0.68

In parallel, a genetic algorithm was applied to the database, varying population size from 10 to 30, keeping elitism at 5, and varying the number of generations from 10 to 200. With 10 generations, it was already possible to obtain the same results as the exhaustive search methodology. Table 16 shows the top-5 oils and fats, according to the objective function. Table 17 shows how the best vegetable oil fitness function vary with the number of generations (termination criteria) and population size. After 10 generations, all metrics do not change anymore.

Table 16: Top 5 oils and fats, according to the objective function.

VO ID	VO Name	f(1)	f(2)	f(3)	Objective Function
45	Ucuúba Butter	1.000	1.000	0.055	0.685
20	Licuri Oil	0.813	0.240	1.000	0.684
27	Murumuru Butter	1.000	0.960	0.091	0.684
15	Fully hydrogenated Soy Fat	1.000	1.000	0.050	0.683
4	Babaçu Oil	0.938	0.360	0.714	0.671

Table 17: Best VO fitness function, keeping elitism at 5 and varying population size (10, 20, 30) and number of generations (5 to 20).

Number of generations/ Population size	5	10	15	20
10	0.685	0.685	0.685	0.685
20	0.684	0.685	0.685	0.685
30	0.685	0.685	0.685	0.685

Figure 17 shows the average performance of the population and elitism, varying with the number of generations from 1 until 200, using 10 as the population size and 5 as elitism.

In this first version of the Genetic Algorithm, the average performance of the entire population increases on average until a performance of 0.55, and the oscillates around this number. This value is mainly influenced by elitism part of the population, as the individuals for the following population are always chosen at random. Elitism reaches its maximum value at generation 32, as seen in Figure 17.

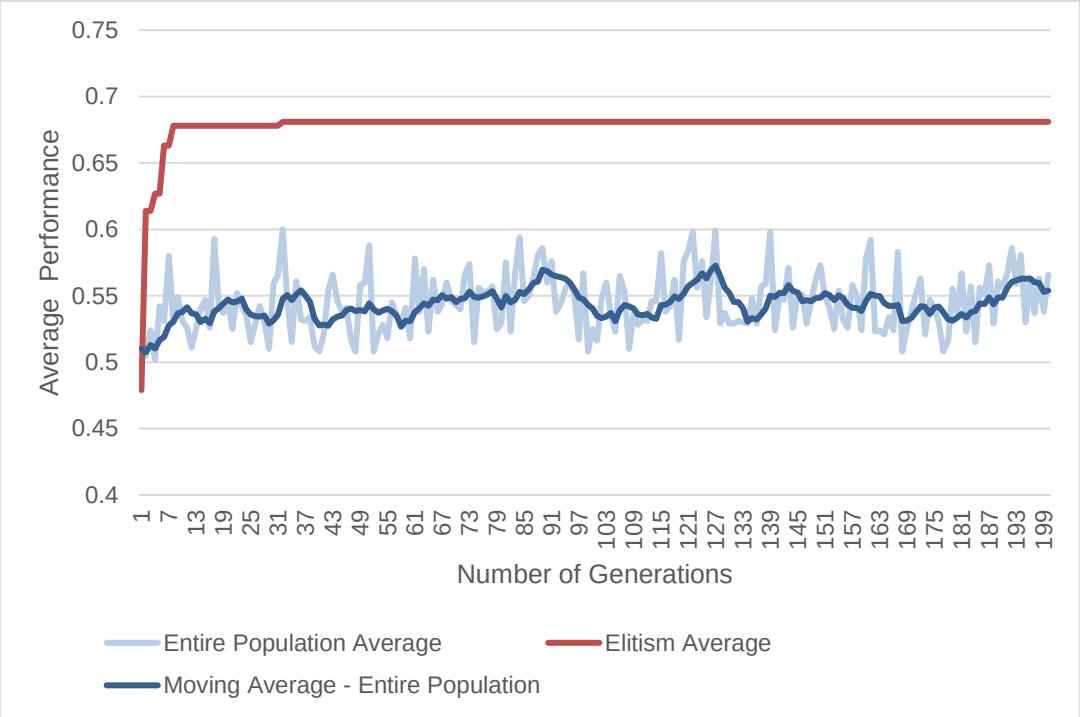

Figure 17: Average performance of the whole population and elitism according to the number of generations.

5.1.3. Binary Mixtures

In binary mixtures' case, the main challenge is to define the optimal quantitative and qualitative parameters for the Genetic Algorithm. A pseudo-code of this implementation is illustrated in Appendix B.

The tested structure considered Elitism as the Selection method and Elitism also as the Replacement technique. This means that N best mixtures were kept in the next population, and the M best individuals went through crossover and mutation processes, being the sum of M and N the number of individuals in each population. Different values were tested for population size, replacement elitism rate, crossover rate, and mutation rate. In each test, only the variable of interest varied, and the remaining ones were kept constant. Also, for the same population size, the same starting population was considered, as a way for the results to be more comparable.

5.1.3.1 Population Size Analysis

Population size was the first quantitative parameter varied. Quantitative Parameters' values are shown in Table 18, with population size varying between 30, 50, and 100. Overall, an early convergence was observed for populations with lower number of individuals, as seen in Figures 18, 19, and 20, resulting in final fitness function very close to best results obtained in single oil optimization for population size 30. For 50 and 100 individuals, best mixture's performance increased considerably, showing that a higher population size result in better mixtures. However, as shown in Table 18, increase in population size also results in higher computational cost, as the optimization process takes longer to run.

Table 18: Genetic Algorithm Quantitative Parameters for Population Size Variation

Population Size	Replacement Elitism	Replacement Elitism Rate	Mutation Rate	Crossover Rate	Number of Generations	Running Time (min)	Best Mixture ID	Best Mixture Performance
30	6	20%	5%	90%	50	140	26_30_45	0.697
50	10	20%	5%	90%	50	300	4_93_13	0.740
100	20	20%	5%	90%	50	720	13_27_8	0.787

26_30_45: 30% of Microalgae oil and 70% of Ucuúba butter;

4_93_13: 93% of Babassu oil and 7% of Vaccenic-enriched oil;

13_27_8: 27% of Vaccenic-enriched oil and 73% of Coconut oil.

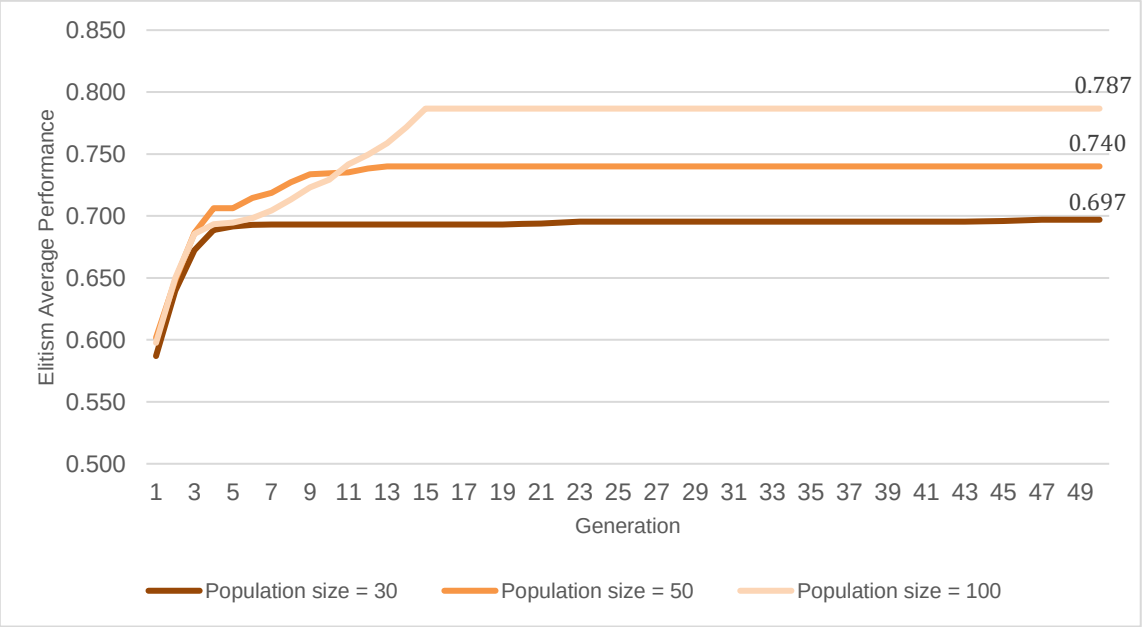

Figure 18: Elitism average performance for each population size tested, varying by generation.

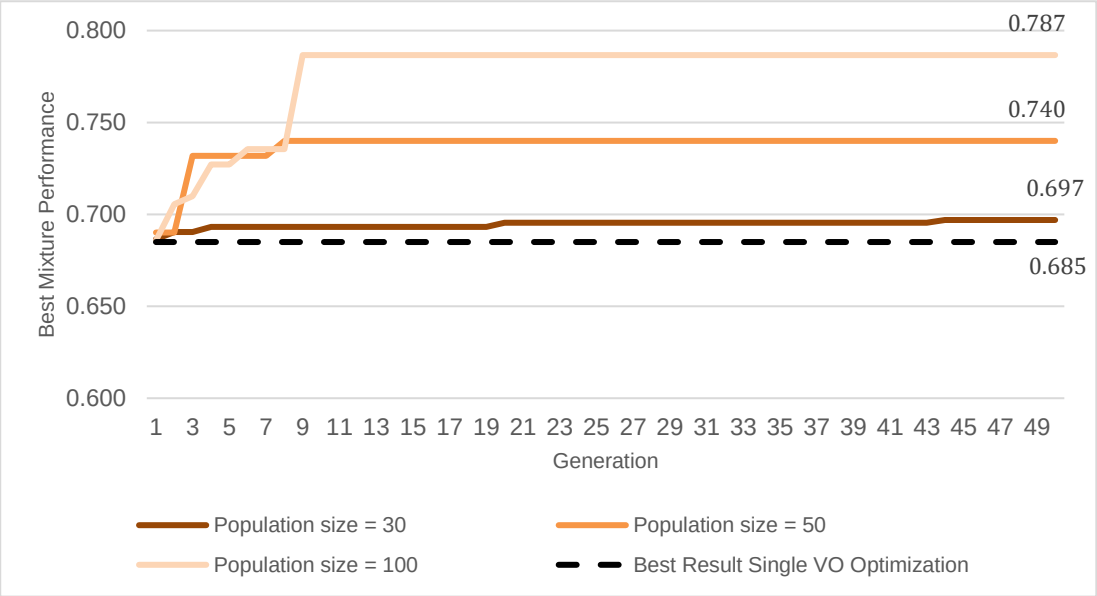

Figure 20: Best Mixture Performance for each population size tested, varying by generation.

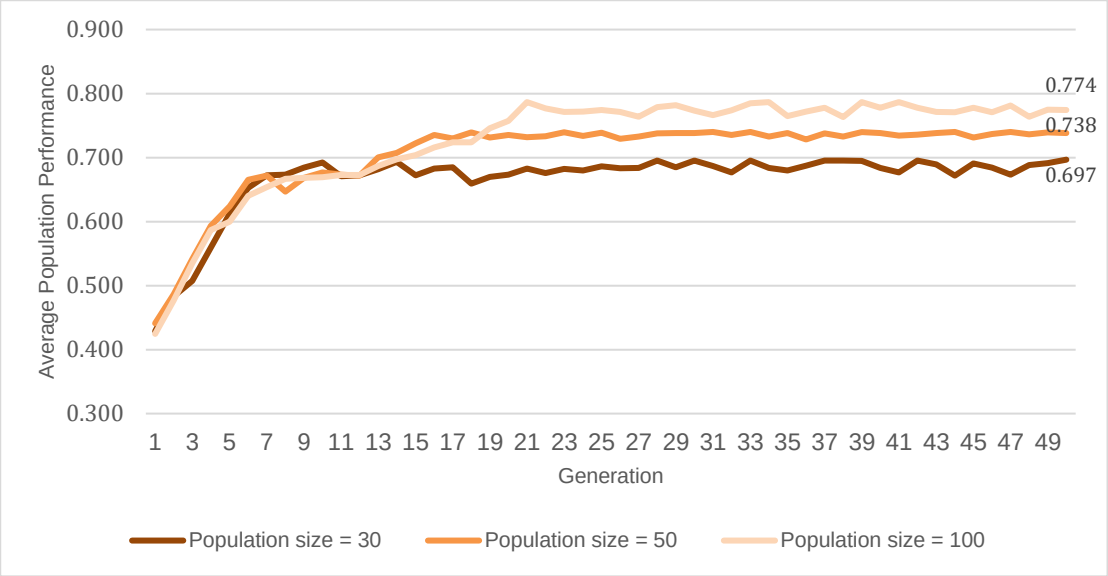

Figure 19: Entire Population Average Performance for each population size tested, varying by generation.

Average population performance tends to increase and converge to best mixture performance, as shown in Figure 19, because there is a convergence of all the individuals towards best mixtures. With a higher population size, convergence takes longer, and better results can be achieved.

Best mixtures obtained in binary optimization do not necessarily involve best oils and fats found in single optimization, as shown in Table 19. The best performance found using population size of 100, is composed of oils that are not in the top-5 individual performers, Vaccenic-enriched and Coconut Oil.

Table 19: Best Mixtures obtained in Binary Optimization, varying population size.

Best Mixture Obtained	Population size = 30	Population size = 50	Population size = 100
VO 1	Microalgae Oil	Babaçu Oil	Vaccenic-enriched Oil
VO 2	Ucuúba Butter	Vaccenic-enriched Oil	Coconut Oil
VO 1 Percentage	30%	93%	27%

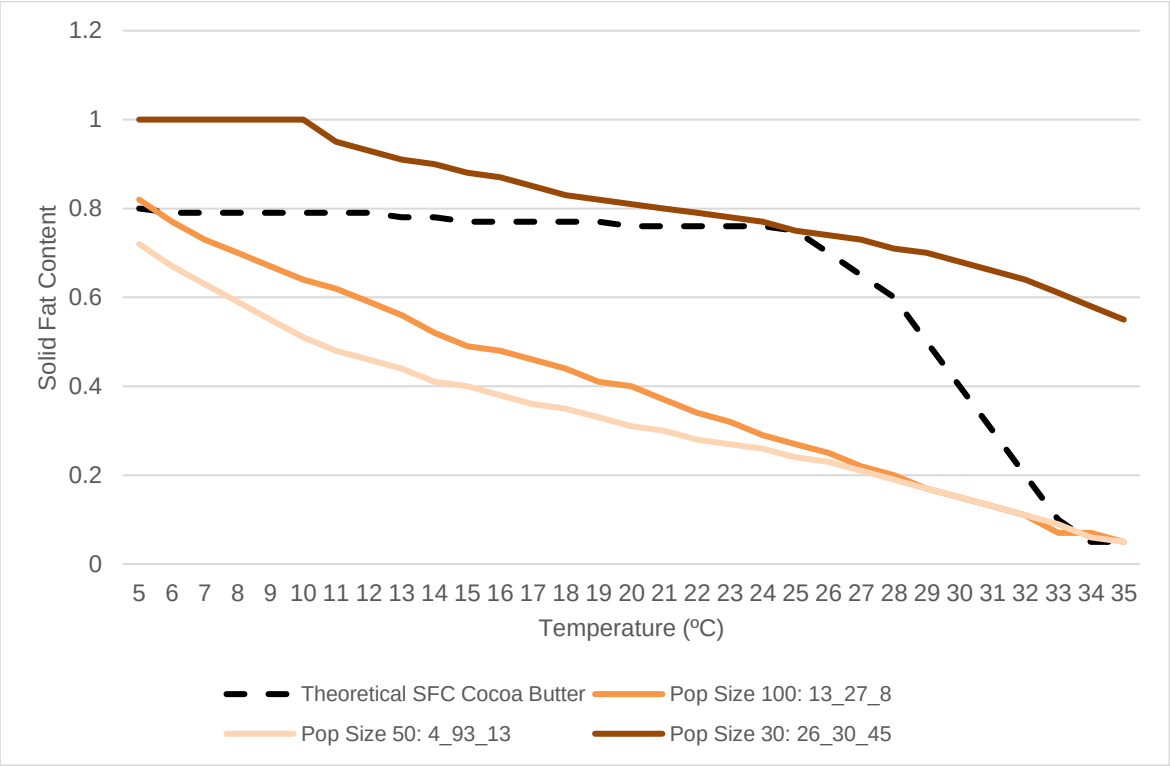

Figure 21: SFC curves from best mixtures obtained by each population size test. Theoretical SFC Cocoa Butter based on Figure 14.

In Figure 21, it is possible to compare the best mixtures obtained by each population size test. Population 30 is good in lower temperature, as it is solid, but it does not melt when reaches body temperature, which is an essential property of cocoa

butter. Populations 50 and 100, however, are adequate in very low or very high temperatures, but do not show the steep decrease between 25 and 35ºC.

Analyzing each mixture with the respective individual oils, Figure 22 shows that Ucuúba Butter has a higher melting point, which can be explained by higher amount of saturated fatty acids, such as Myristic acid. The microalgae oil, on the other hand, is composed of more than half of unsaturated fatty acids, which explains the lower melting point. The mixture shows a melting point closer to Ucuúba, as it represents 70% of the mixture.

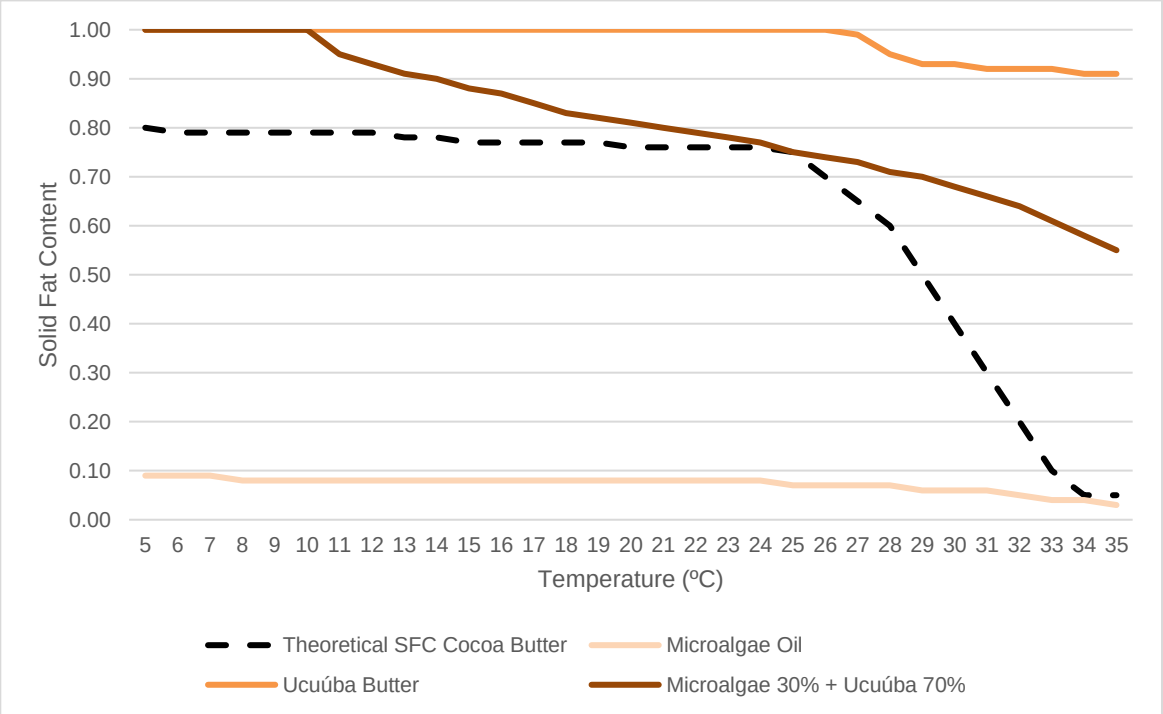

Figure 22: SFC curve for population with size 30, and its individual vegetable oils and fats.

Figure 23 shows the oils that compose the best result for simulation with 50 generations. Vaccenic-enriched oil is composed mostly of unsaturated fatty acids, and Babaçu oil mostly saturated, which explains the difference between the curves.

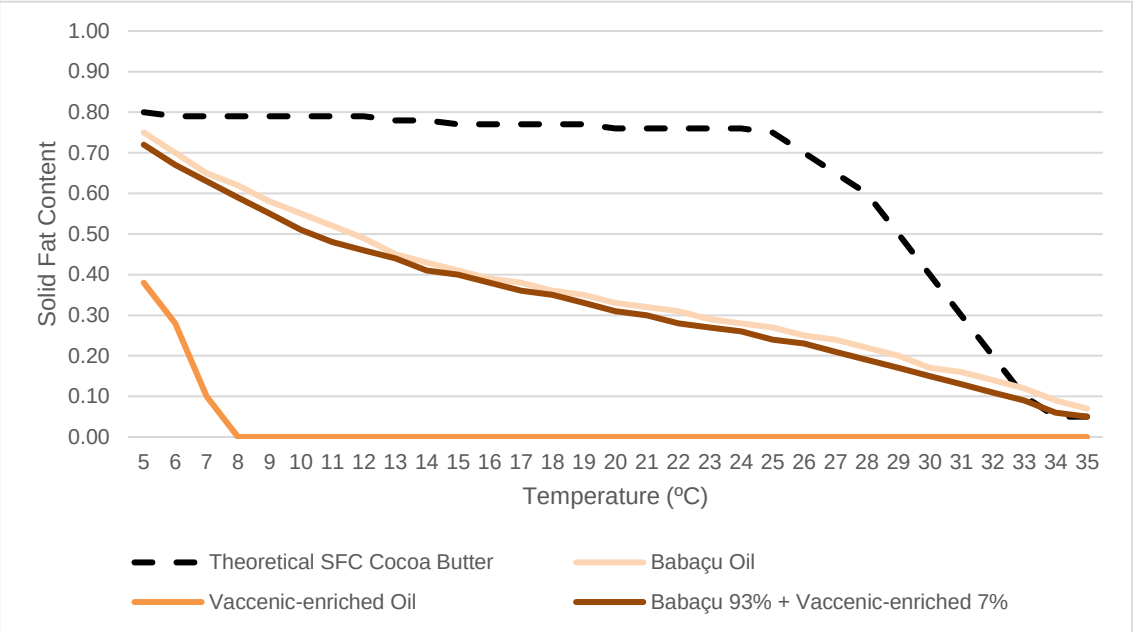

Figure 23: SFC curve for population with size 50, and its individual vegetable oils and fats.

In Figure 24, we also see Coconut oil with a higher melting point, because of a higher amount of saturated fatty acids. Even though this is the best result obtained, it is possible to the that the shape of the curve is very different from Cocoa Butter's.

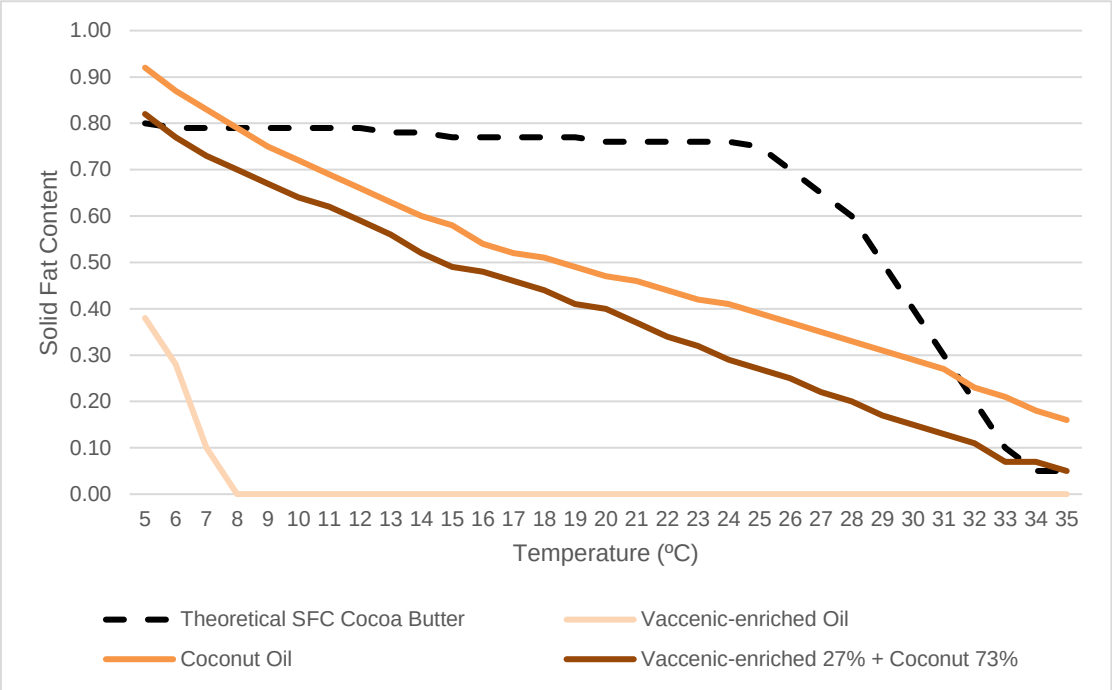

Figure 24: SFC curve for population with size 100, and its individual vegetable oils and fats.

Population size of 100 showed the best performances, and an even higher population size would probably result in better mixtures, but at the expense of computational cost. For the remainder of the sensitivity analysis, population size of 100 will be considered.

5.1.3.2 Replacement Elitism Rate Analysis

Replacement Elitism refers to the number of individuals that are copied to the next population without going through any modifications. This parameter was varied from 10%, 20%, 30%, and 50%.

On one hand, a higher elitism rate guarantees that good individuals will not be lost in following generations; on the other hand, it prevents that even more individuals go through crossover and mutation processes, which are the main responsible for introducing new genes and new combinations into the population.

In Table 20 and Figure 27, one can see that the best mixture found was using an elitism rate of 20%; both lower and higher elitism rates showed worse results. In

Figure 25, one can see that elitism rates of 10% and 30% both converge to same results after around 10 generations.

There's no clear trend as what happens to results with increase or decrease of replacement elitism rate. For example, the highest rate, 50%, has an intermediate result: 20% rate has a better result, and 10%/30% has a worse result. Figure 27 shows that all four tests are at least 9% better than results obtained with single vegetable oil optimization, which shows the potential of this methodology.

In Figure 26, we see overall population performance converging to best results, as also seen in previous simulations.

Table 20: Genetic Algorithm Quantitative Parameters for Replacement Elitism Rate Variation

Population Size	Replacement Elitism	Replacement Elitism Rate	Mutation Rate	Crossover Rate	Number of Generations	Running Time (min)	Best Mixture ID	Best Mixture Performance
100	10	10%	5%	90%	50	1110	20_70_8	0.744
100	20	20%	5%	90%	50	720	13_27_8	0.787
100	30	30%	5%	90%	50	480	8_31_20	0.744
100	50	50%	5%	90%	50	300	13_59_45	0.768

20_70_8: 70% of Licuri oil and 30% of Coconut oil;

13_27_8: 27% of Vaccenic-enriched oil and 73% of Coconut oil;

8_31_20: 31% of Coconut oil and 69% of Licuri oil;

13_59_45: 59% of Vaccenic-enriched oil and 41% of Ucuúba butter.

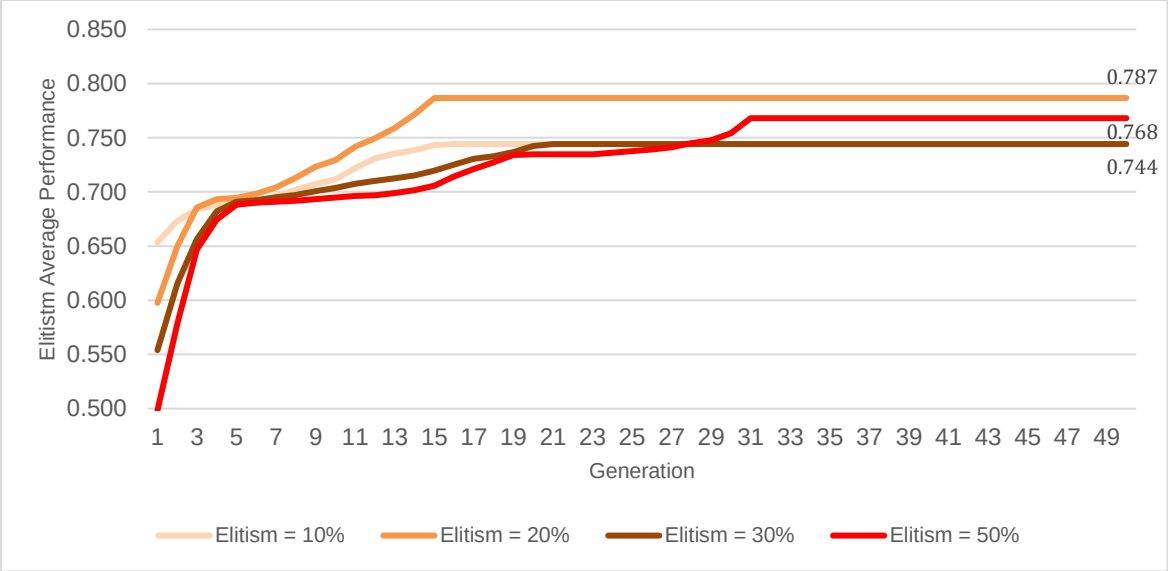

Figure 25: Elitism average performance for each Elitism Rate tested, varying by generation.

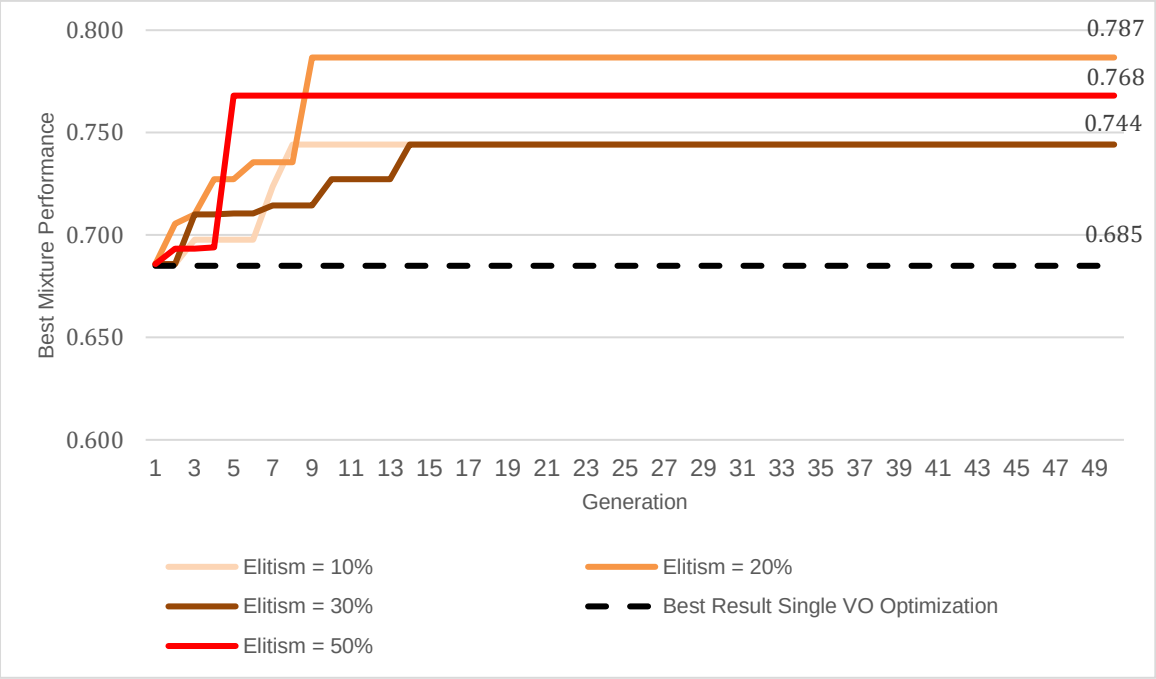

Figure 27: Best mixture performance for each Elitism Rate tested, varying by generation.

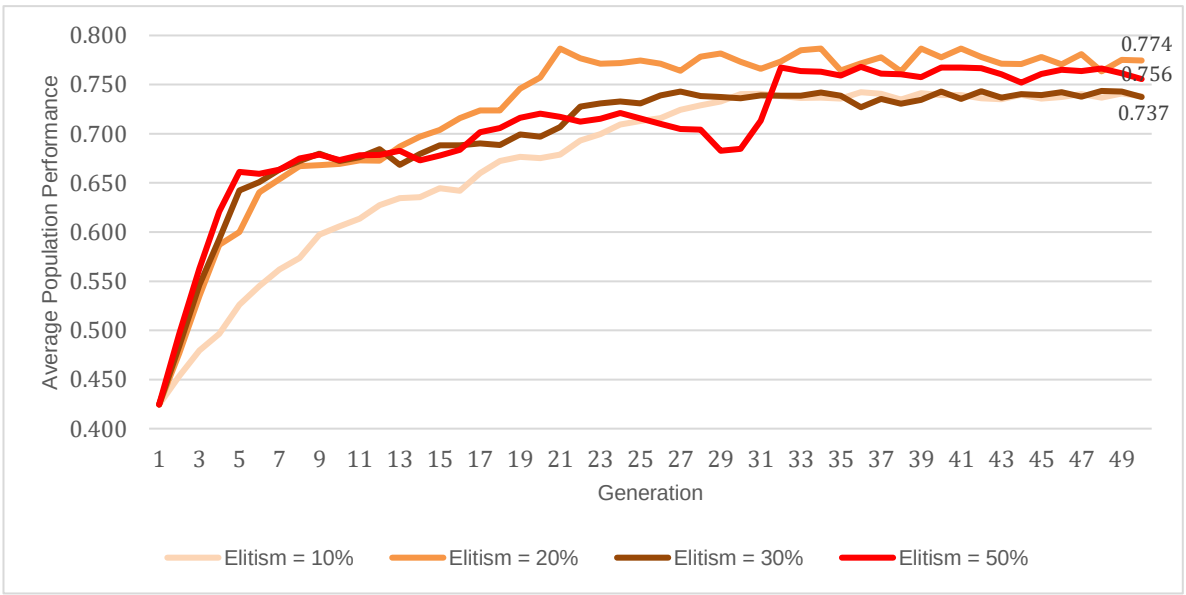

Figure 26: Entire Population Average Performance for each Elitism Rate tested, varying by generation.

When compared to population size optimization, in the elitism rate a new vegetable oil appeared among the best mixtures, the Licuri Oil, as it is seen in Table 21, and the best result (elitism rate of 20%) showed the same composition as best result in population size optimization.

Table 21: Best Mixtures obtained in Binary Optimization, varying elitism rate.

Best Mixture Obtained	Elitism Rate = 10%	Elitism Rate = 20%	Elitism Rate = 30%	Elitism Rate = 50%
VO 1	Licuri Oil	Vaccenic-enriched Oil	Coconut Oil	Vaccenic-enriched Oil
VO 2	Coconut Oil	Coconut Oil	Licuri Oil	Ucuúba Butter
VO 1 Percentage	70%	27%	31%	59%

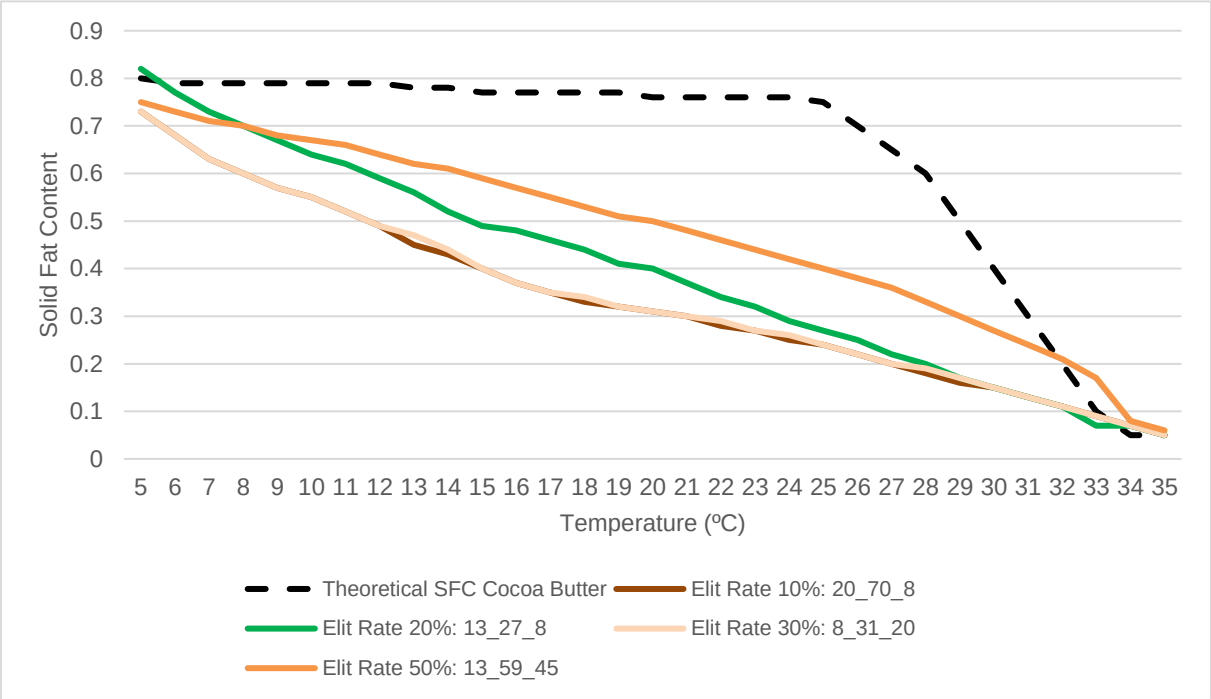

Figure 28: SFC curves from best mixtures obtained by each elitism rate test.

In Figure 28, it is possible to see that all elitism rate tests have a similar behavior: they match the requirements from temperatures 5 and 35ºC, but not at 25ºC and, thus, do not show the sharp melting point.

In Figure 30 and Figure 32, it is possible to see that Vaccenic-enriched Oil has a very low melting temperature, which is a direct result of its composition being mostly unsaturated fatty acids. In Figure 29 and Figure 31, one can see that the shape of the

Coconut and Licuri Oil curves are very similar, but at different levels, which can be explained by both having a high amount of saturated fatty acids, but Coconut oil's saturated composition being a bit above Licuri's.

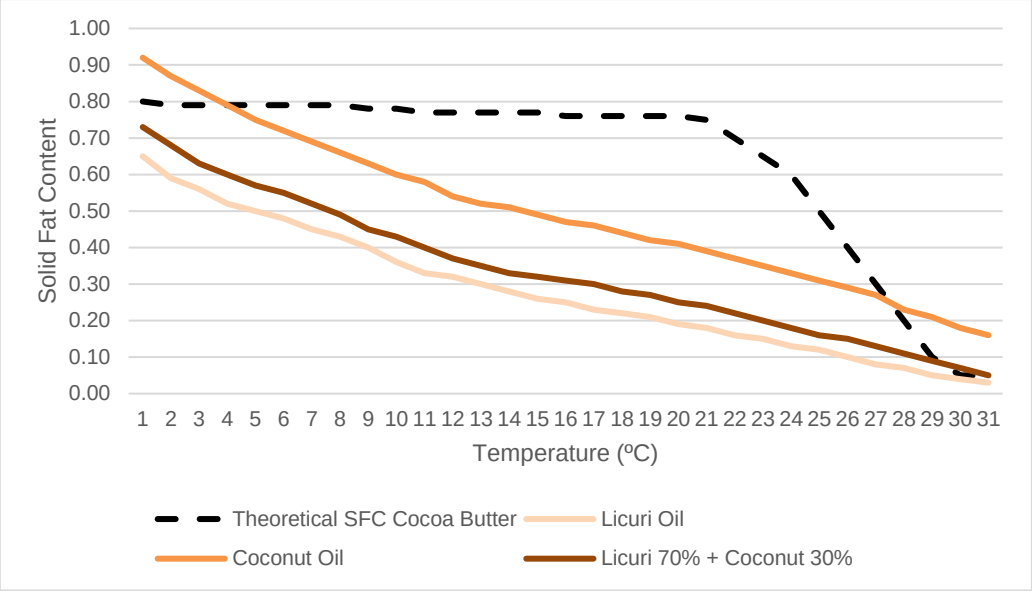

Figure 29: SFC curve for elitism rate of 10%, and its individual vegetable oils and fats.

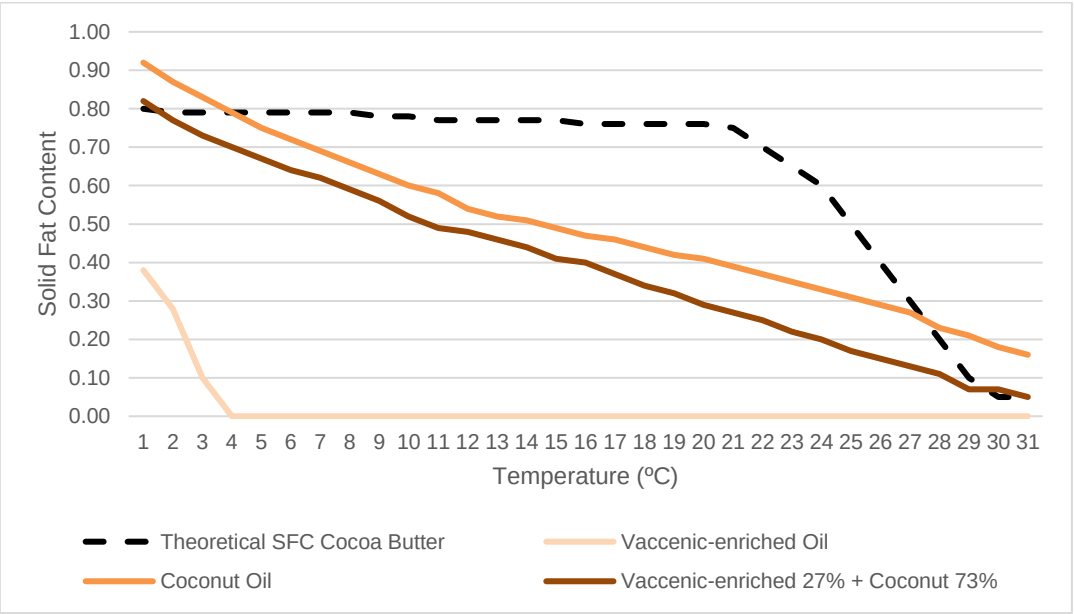

Figure 30: SFC curve for elitism rate of 20%, and its individual vegetable oils and fats.

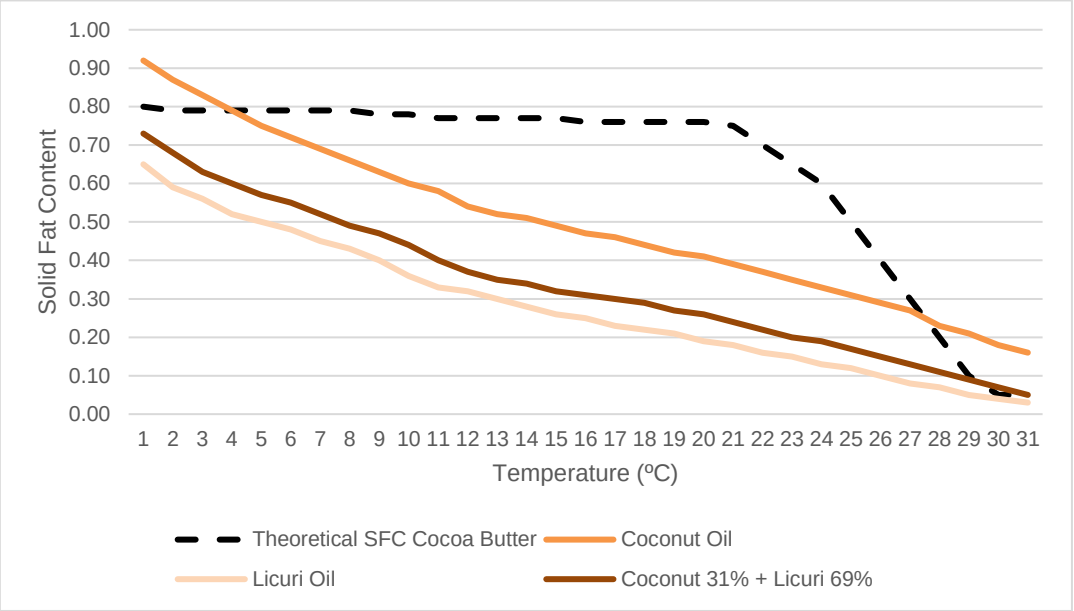

Figure 31: SFC curve for elitism rate of 30%, and its individual vegetable oils and fats.

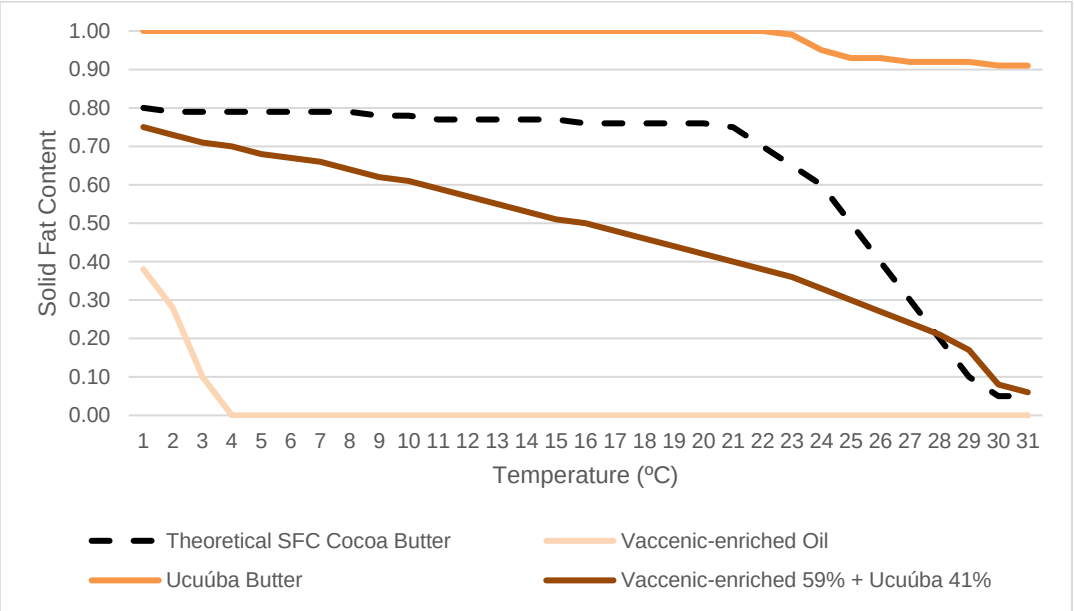

Figure 32: SFC curve for elitism rate of 50%, and its individual vegetable oils and fats.

5.1.3.3 Crossover Rate Analysis

Each pair of selected parents have a chance of going through crossover or not, and this percentage was the third parameter analyzed. Each pair of parents had a 50% chance of going through oil crossover, and a 50% chance of going through oil and percentage crossover, this 50%/50% ratio remained constant in all simulations.

Best results were obtained with a crossover rate of 90%, which also showed the second highest running time, as seen in Table 22. All the other simulations resulted practically in the same mixture: 70% of Licuri Oil and 30% of Coconut Oil, mixture already analyzed in Section 5.1.3.2.

In Figures 33, 34, and 35 it is possible to see that crossover rates of 70%, 80% and 100% converge to the same elitism average, best mixture performance, and population average. As it also happened to Replacement Elitism Rate, there's no clear trend regarding elitism rate: the best result is neither the lowest nor the highest rate.

Table 22: Genetic Algorithm Quantitative Parameters for Crossover Rate Variation

Population Size	Replacement Elitism	Elitism Rate	Mutation Rate	Crossover Rate	Number of Generations	Running Time (min)	Best Mixture ID	Best Mixture Performance
100	20	20%	5%	70%	50	480	20_71_8	0.744
100	20	20%	5%	80%	50	600	8_31_20	0.744
100	20	20%	5%	90%	50	720	13_27_8	0.787
100	20	20%	5%	100%	50	740	20_68_8	0.744

20_71_8: 71% of Licuri oil and 29% of Coconut oil;

8_31_20: 31% of Coconut oil and 69% of Licuri oil;

13_27_8: 27% of Vaccenic-enriched oil and 73% of Coconut oil;

20_68_8: 68% of Licuri oil and 32% of Coconut oil.

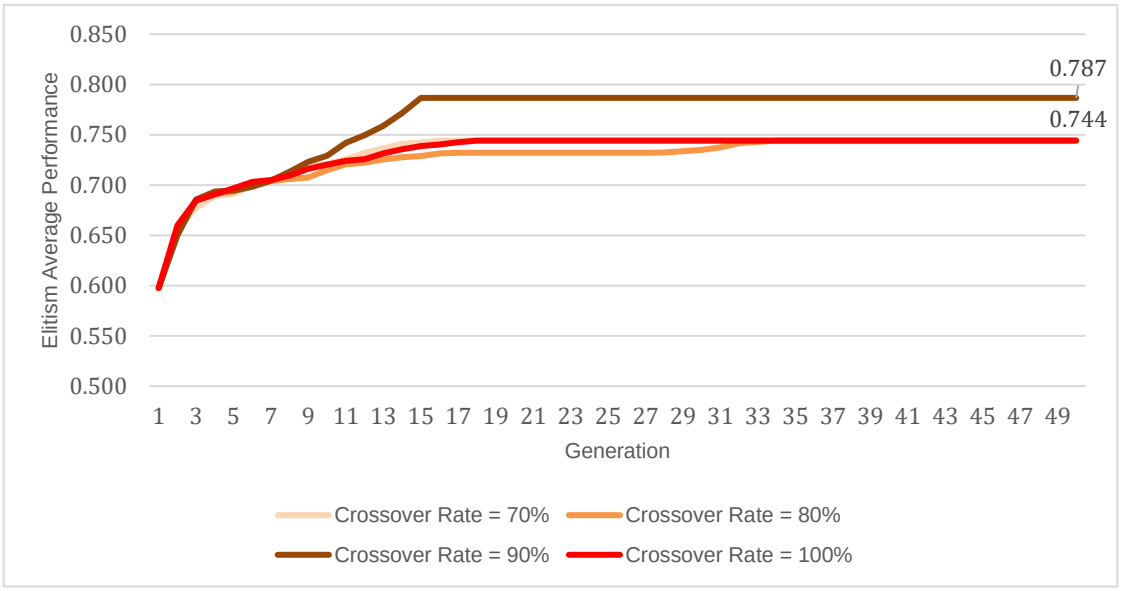

Figure 33: Elitism average performance for each Crossover Rate tested, varying by generation.

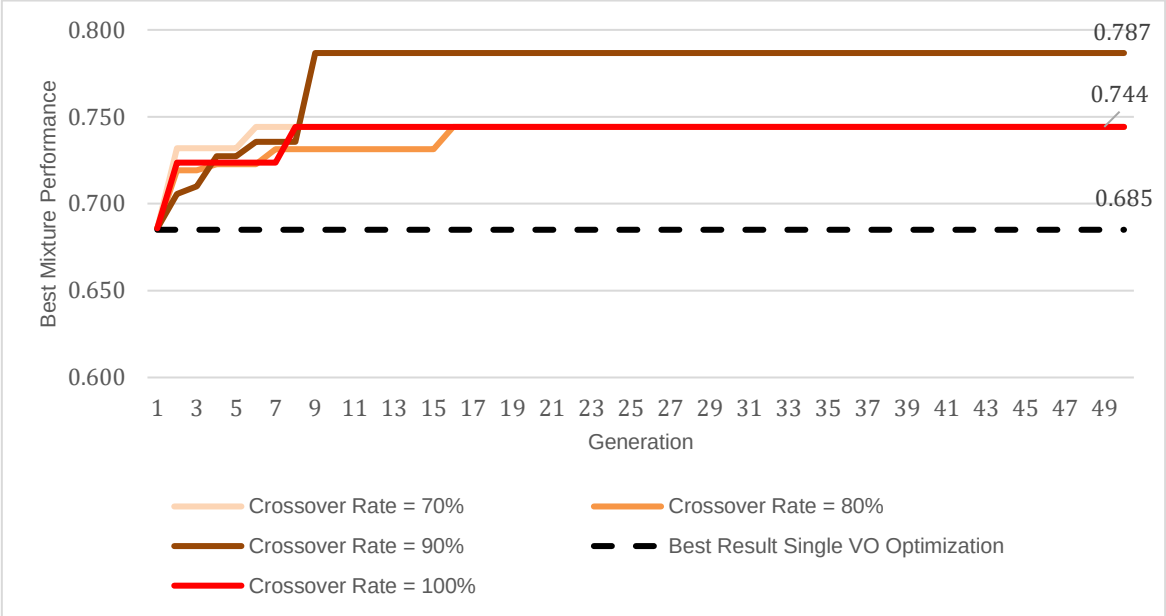

Figure 34: Best Mixture Performance for each Crossover Rate tested, varying by generation.

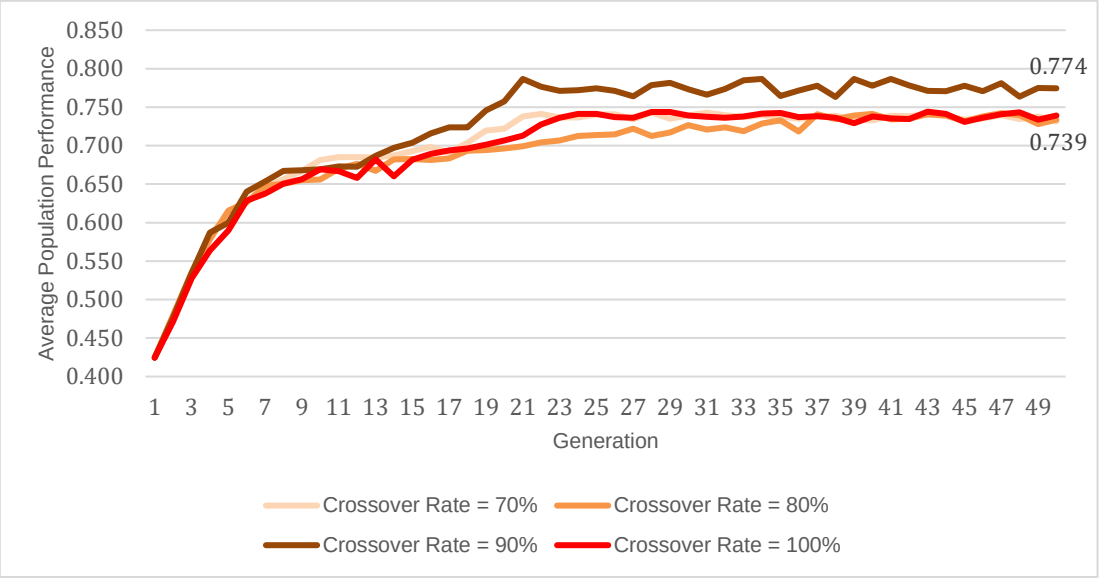

Figure 35: Entire Population Average Performance for each Crossover Rate tested, varying by generation.

In Figure 36 it is possible to see that best result from crossover rate of 90% is indeed closer to expected curve, however, it still does not show the steep decrease between 25 and 35ºC, as it also happened with other parameters analysis.

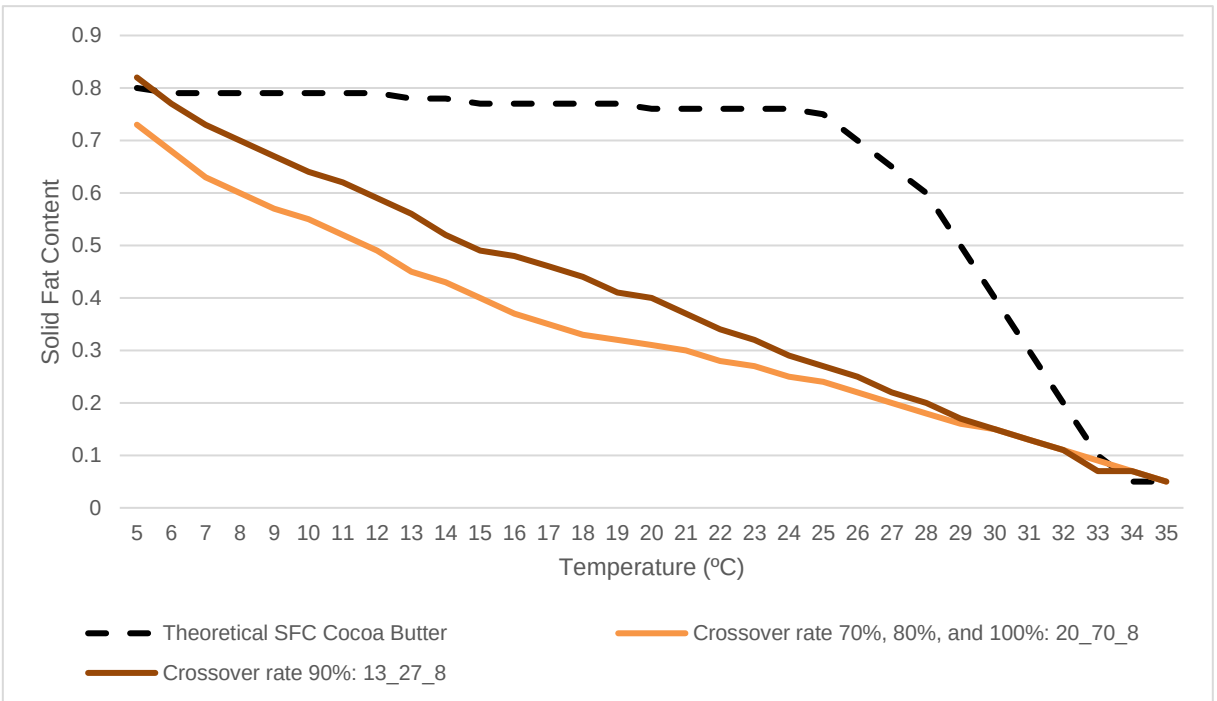

Figure 36: SFC curves from best mixtures obtained by each crossover rate test.

The analysis performed in previous sections 5.1.3.1 and 5.1.3.2 comparing best mixtures with its individual oils SFC will not be done for the current section, as the results are the same as presented before.

5.1.3.4 Mutation Rate Analysis

Mutation rate works very similarly to crossover rate. Each pair of parents selected have a change of going through mutation or not, being 50% of them going through oil mutation, and the other 50% through oil and percentage mutation. The mutation process is important, as it ensures that no gene is permanently removed from future generations.

As seen in Table 23 and Figure 38, mutation rates of 5%, 15%, and 25% resulted in the same best mixture. The running time between simulations are not 100% comparable, as results from previous simulations were available for usage, meaning that if a mixture has already been analyzed, the optimization process of finding its SFC curve does not have to be performed again. This explains the counterintuitive running time of a mutation rate of 15%, which should result in a higher running time than 2% and 5% mutation rate. Figure 37 shows that mutation rates of 5%, 15%, and 25% also resulted in the same elitism averages. In Figure 39 it is possible to see that average population varies in all mutation rates.

Table 23: Genetic Algorithm Quantitative Parameters for Mutation Rate Variation

Population Size	Replacement Elitism	Elitism Rate	Mutation Rate	Crossover Rate	Number of Generations	Running Time (min)	Best Mixture ID	Best Mixture Performance
100	20	20%	2%	90%	50	390	8_31_20	0.744
100	20	20%	5%	90%	50	720	13_27_8	0.787
100	20	20%	15%	90%	50	390	8_74_13	0.787
100	20	20%	25%	90%	50	780	8_74_13	0.787

8_31_20: 31% of Coconut oil and 69% of Licuri oil;

13_27_8: 27% of Vaccenic-enriched oil and 73% of Coconut oil;

8_74_13: 74% of Coconut oil and 26% of Vaccenic-enriched oil.

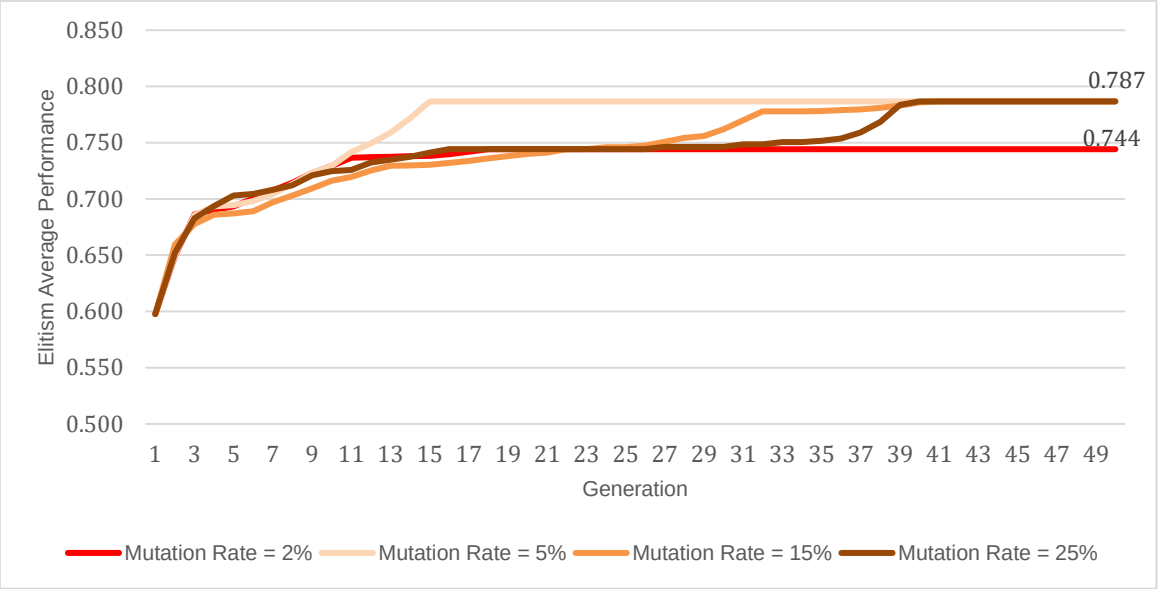

Figure 37: Elitism average performance for each Mutation Rate tested, varying by generation.

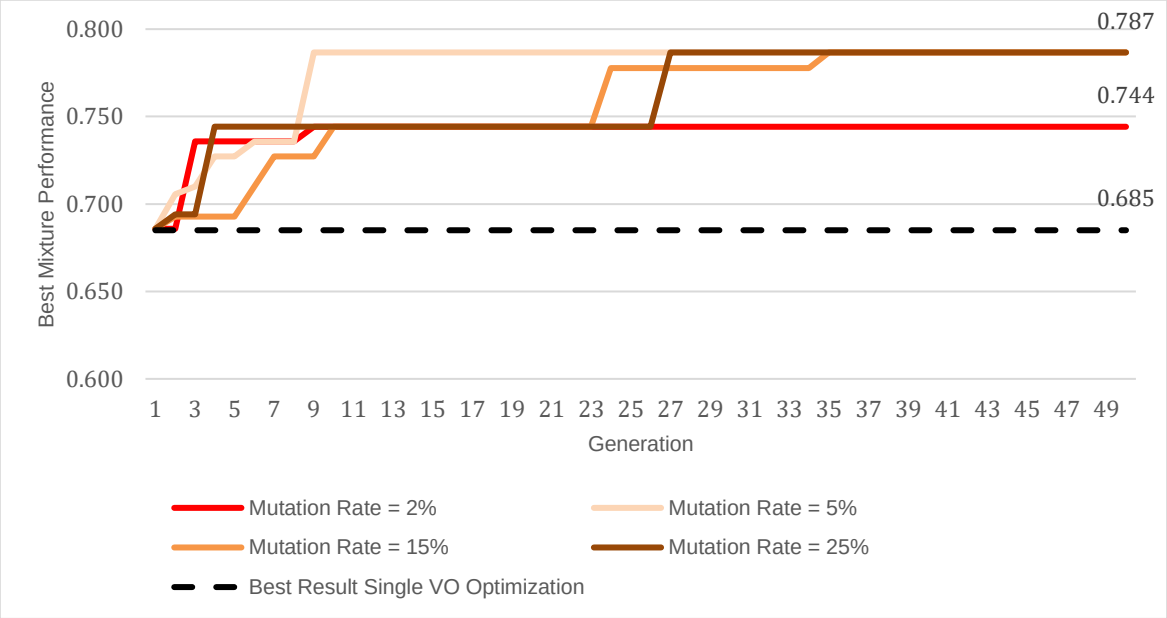

Figure 38: Best Mixture Performance for each Mutation Rate tested, varying by generation.

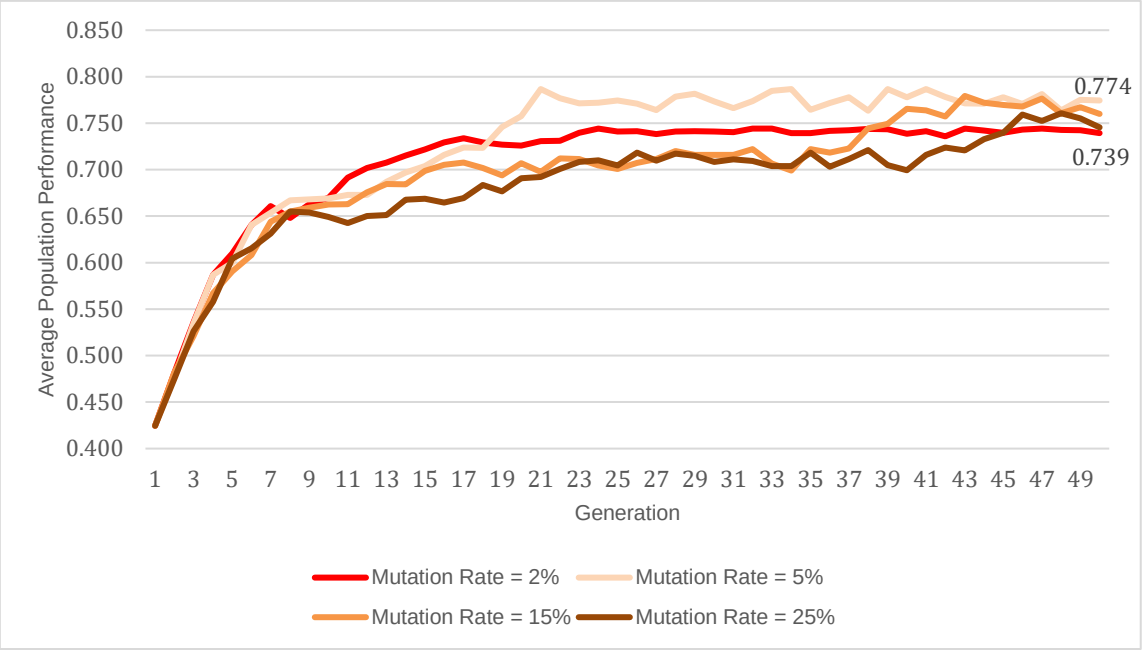

Figure 39: Entire Population Average Performance for each Mutation Rate tested, varying by generation.

Mutation rate seems to have a small impact on results, as 3 of the tests reached the same final mixture. So, the lowest one will be chosen, as it should result in the lowest running time.

6. Conclusions

Utilization of oils and fats alternatives to cocoa butter in chocolate products is a current reality. Increasing demand, together with higher prices and reduced in production in some countries, turns this substitution even more appealing to industries.

In parallel, Computer-Aided Product Design (CAPD) and Reverse Engineering concepts were developed to make product development process cheaper and faster, helping industries launch better products in a shorter period.

In the case study for cocoa butter, Solid Fat Content (SFC) was chosen as the main optimization property in CAPD Framework, because it represents the most important behavior of chocolate that needs to be reproduced: solid at fridge and shelf-temperature, and liquid at body/consumption temperature. The additional optimization layer was done using a Genetic Algorithm, because the SFC computation is itself a nonlinear optimization step.

Calculation of SFC for each simulation used a methodology previously developed, by minimizing the Gibbs Free Energy function. This requires a resolution of a NLP problem, using a CONOPT 3 solver.

The first challenge of this work was the tools' integration, since SFC optimization was done using Fortran, GAMs, and Matlab, and the GA was written in Python, a more recent programming language.

The main challenge of working specifically with GA is the number of modifiable parameters, which ideally would involve a thorough sensitivity analysis involving both qualitative and quantitative parameters. This analysis was performed only in quantitative parameters, maintaining qualitative choices fixed. Even with the optimization of GA parameters, best results are not guaranteed, and different simulations are not 100% comparable, as a result of the stochastic nature of this method.

In the case of single oils and fats, a simpler approach was performed, because only 45 simulations needed to be performed and compared to the genetic algorithm. In this case, Ucuúba butter, Licuri Oil, and Murumuru butter were the top 3 results.

For the binary mixtures simulations, an exhaustive search cannot be performed, as it would result in almost 100k simulations varying the 45 oils with 99 different proportions. So, only the genetic algorithm was applied, together with its sensitivity analysis.

Regarding population size, the higher the total population, the best results that can be obtained. However, this also increases computational cost, as process will take longer to run, so this modification but be done with caution. In this work, population size was increased only until 100 individuals, as it already increased running time to 12 hours in almost all simulations.

For the Replacement Elitism Rate, which derives from a fixed qualitative parameter, the optimal value found was 20%; lower and higher values resulted in worse mixtures. Crossover rate showed the best result in a value of 90%. Mutation rate seems to not affect so much the results, as 3 out of the 4 simulations converged to the same best mixture.

With the optimized parameters in the binary mixture optimization, it was already possible to see an increase in the objective function when compared to single VO tests. With population size of 100, an average of 900 mixtures were evaluated in each simulation in approximately 12 hours, which shows the biggest advantage of this method: analysis of a large of number of oils and fats, in a short period of time, with no experiments needed.

In a real product development scenario, experiments are required, and this tool would be useful by choosing previously the best mixtures before going into the laboratory, which would fasten the launch of new products or raw material substitution. Also, innovative, and non-previously tested materials can be identified.

7. Future Works

As future works, the following points are highlighted:

- Run Genetic Algorithm for ternary mixtures.

In this work, only single and binary oil mixtures were tested. Another possibility would be to test ternary mixtures to achieve an even higher objective function.

- Enhance Genetic Algorithm:

1. Modify objective function.

The current objective function considers only values of solid fat content at three different temperatures, with the final performance being an average of them. With this, it was not possible to obtain a curve with a similar shape to cocoas. So, other types of aggregation could be done over the SFC values, such as weighted average, and other properties could be added to the genetic algorithm optimization.

2. Perform a Sensitivity Analysis on GA Qualitative Parameters

In this project, only quantitative parameters were varied and tested. To achieve even better results, a sensitivity analysis on qualitative parameters could be performed to find better results for the optimization process.

3. Change the percentage of each oil/fat in binary mixtures.

In the current simulation, the percentage of each oil could be any integer number from 1 to 99. However, very similar results were obtained because of this. So, in a future simulation, percentage could be limited to go from 1% to 95%, by 5 p.p. steps.

8. Scientific Events Participation

The current work was presented in the V International Meeting on Fats and Oils, which happened in Campinas on May 10, 11 and 12th of 2022.

BEARZI, L. F.; TELES DOS SANTOS, M. **Computer-Aided Product Design for Cocoa Butter Alternatives**. International Meeting on Fats and Oils - Challenges in a Changing World. **Anais**.Campinas: Sociedade Brasileira de Óleos e Gorduras, 2022.

9. References

ABIGOR, R. D. et al. Production of Cocoa Butter-like Fats by the Lipase-Catalyzed Interesterification of Palm Oil and Hydrogenated Soybean Oil. **JAOCS**, v. 80, n. 12, 2003.

ACHENIE, L. E. K.; GANI, R.; VENKATASUBRAMANIAN, V. **Computer Aided Molecular Design: Theory and Practice**. [s.l: s.n.]. v. 12

AFOAKWA, E. O. World cocoa production, processing and chocolate consumption pattern. Em: **Chocolate Science and Technology**. Chichester, UK: John Wiley & Sons, Ltd, 2016. p. 17–48.

AHN, J. H. et al. Optimization of microencapsulation of seed oil by response surface methodology. **Food Chemistry**, v. 107, n. 1, p. 98–105, 1 mar. 2008.

ALHIJAWI, B.; AWAJAN, A. Genetic algorithms: theory, genetic operators, solutions, and applications. **Evolutionary Intelligence**, 2023.

ARRIETA-ESCOBAR, J. A. et al. Integration of Consumer Preferences and Heuristic Knowledge in the Design of Formulated Products: Application to a Cosmetic Emulsion. **Computer Aided Chemical Engineering**, v. 46, p. 433–438, 1 jan. 2019.

AUGUSTO, P. E. D. et al. Modelling the effect of temperature on the lipid solid fat content (SFC). **Food Research International**, v. 45, n. 1, p. 132–135, jan. 2012.

AUSTIN, N. D.; SAHINIDIS, N. V.; TRAHAN, D. W. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. **Chemical Engineering Research and Design**, v. 116, p. 2–26, 1 dez. 2016.

AZIAN, M. N. et al. Viscosity Estimation of Triacylglycerols and of Some Vegetable Oils, Based on Their Triacylglycerol Composition. **JAOCS**, v. 78, p. 1001–1005, 2001.

BAKER, J. E. **An analysis of the effects of selection in genetic algorithms**. [s.l: s.n.].

BELSLEY, D. A. **Handbook of Computational Econometrics**. [s.l: s.n.].

BLOCK, J. M.; FIGUEIREDO, M. F.; GOMIDE, F. A. C. Blending Process Optimization into Special Fat Formulation by Neural Networks. **JAOCS**, v. 74, n. 12, p. 1537–1541, 1997.

BUDIANTO; KUSMARDINI, D. THE COMPARISON OF THE EFFECTIVENESS OF COCOA BUTTER (CB) AND COCOA BUTTER ALTERNATIVES (CBA) IN CHOCOLATE MANUFACTURE. **Bulletin of the Transilvania University of Brasov, Series II: Forestry, Wood Industry, Agricultural Food Engineering**, v. 14–63, n. 2, p. 85–98, 2021.

BURKE, E. K.; KENDALL, G. (EDS.). **Search Methodologies**. Boston, MA: Springer US, 2005.

CERIANI, R. et al. Group contribution model for predicting viscosity of fatty compounds. **Journal of Chemical and Engineering Data**, v. 52, n. 3, p. 965–972, maio 2007.

CERIANI, R.; GONÇALVES, C. B.; COUTINHO, J. A. P. Prediction of viscosities of fatty compounds and biodiesel by group contribution. **Energy and Fuels**, v. 25, n. 8, p. 3712–3717, 18 ago. 2011.

CHUMPITAZ, L. D. A.; COUTINHO, L. F.; MEIRELLES, A. J. A. Surface Tension of Fatty Acids and Triglycerides. **JAOCS**, v. 76, n. 3, 1999.

CHURI, N.; ACHENIE, L. E. K. Novel Mathematical Programming Model for Computer Aided Molecular Design. p. 3788–3794, 1996.

CIGNITTI, S.; ZHANG, L.; GANI, R. Computer-aided Framework for Design of Pure, Mixed and Blended Products. **12th International Symposium on Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering**, 2015.

CONTE, E. et al. Design of Formulated Products: Experimental Component. **AIChE Journal**, v. 58, p. 173–189, 2012.

CONTE, E.; GANI, R.; NG, K. M. Design of formulated products: A systematic methodology. **AIChE Journal**, v. 57, n. 9, p. 2431–2449, set. 2011.

COSTA, R.; MOGGRIDGE, G. D.; SARAIVA, P. M. Chemical product engineering: An emerging paradigm within chemical engineering. **AIChE Journal**, v. 52, n. 6, p. 1976–1986, jun. 2006.

DALENOGARE, L. S. et al. The expected contribution of Industry 4.0 technologies for industrial performance. **International Journal of Production Economics**, v. 204, p. 383–394, 1 out. 2018.

DAVIS, S. M. From “Future Perfect”: Mass Customizing. p. 16–21, 1989.

DEAN, A.; VOSS, D. **Design and Analysis of Experiments**. [s.l: s.n.].

DEPOORTERE, L. **The use and applicability of cocoa butter equivalents (CBEs) in chocolate products**. [s.l.] Universiteit Gent, 2010.

DÍAZ-TOVAR, C. A.; GANI, R.; SARUP, B. Lipid technology: Property prediction and process design/analysis in the edible oil and biodiesel industries. **Fluid Phase Equilibria**, v. 302, n. 1–2, p. 284–293, 15 mar. 2011.

EIBEN, A. E.; SMIT, S. K. **Parameter tuning for configuring and analyzing evolutionary algorithms**. **Swarm and Evolutionary Computation**Elsevier B.V., , 2011.

ESTEBAN, B. et al. Temperature dependence of density and viscosity of vegetable oils. **Biomass and Bioenergy**, v. 42, p. 164–171, jul. 2012.

FARMANI, J. Modeling of solid fat content of chemically interesterified fully hydrogenated soybean oil and canola oil blends as a function of temperature and saturated fatty acids. **Journal of Food Measurement and Characterization**, v. 9, n. 3, p. 281–289, 11 set. 2015.

FASINA, O. O. et al. Predicting melting characteristics of vegetable oils from fatty acid composition. **LWT - Food Science and Technology**, v. 41, n. 8, p. 1501–1505, nov. 2008.

FERRER, G. et al. Empirical equations for viscosity and specific heat capacity determination of fatty acids. **Journal of Energy Storage**, v. 10, p. 20–27, 1 abr. 2017.

FOGLIATTO, F. S.; DA SILVEIRA, G. J. C.; BORENSTEIN, D. The mass customization decade: An updated review of the literature. **International Journal of Production Economics**, v. 138, n. 1, p. 14–25, jul. 2012.

FORTUNE BUSINESS INSIGHTS. Cocoa and Chocolate Market Size, Share and COVID-19 Impact Analysis, By Type, By Application, and Regional Forecast, 2022-2029. Available in: https://www.fortunebusinessinsights.com/industry-reports/cocoa-and-chocolate-market-100075. Accessed in 08 sept. 2022

GANI, R. Chemical product design: Challenges and opportunities. **Computers and Chemical Engineering**, v. 28, n. 12, p. 2441–2457, 15 nov. 2004.

GANI, R. et al. Computer-Aided Molecular Design and Property Prediction. Em: **Computer Aided Chemical Engineering**. [s.l.] Elsevier B.V., 2016. v. 39p. 153–196.

GANI, R. Group contribution-based property estimation methods: advances and perspectives. **Current Opinion in Chemical Engineering**, v. 23, p. 184–196, 1 mar. 2019.

GANI, R.; O'CONNELL, J. P. Properties and CAPE: from present uses to future challenges. **Computers and Chemical Engineering**, v. 25, p. 3–14, 2001.

HARPER, P. M. et al. Computer-aided molecular design with combined molecular modeling and group contribution. **Fluid Phase Equilibria**, v. 158, p. 337–347, 1999.

HARPER, P. M.; GANI, R. A multi-step and multi-level approach for computer aided molecular design. **Computers and Chemical Engineering**, v. 24, 2000.

HEINTZ, J. et al. Computer aided product design tool for sustainable product development. **Computers and Chemical Engineering**, v. 71, p. 362–376, 2014.

HILL, M. Chemical Product Engineering-The third paradigm. **Computers and Chemical Engineering**, v. 33, n. 5, p. 947–953, 21 maio 2009.

HJORTH, J. L. et al. Thermodynamic Modeling of Multi-phase Solid–Liquid Equilibria in Industrial-Grade Oils and Fats. **Journal of the American Oil Chemists' Society**, v. 92, n. 1, p. 17–28, jan. 2015.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems**. [s.l.] The MIT Press, 1992.

HUANG, C. Y.; LAI, C. Y.; CHENG, K. T. Fundamentals of Algorithms. Em: **Electronic Design Automation**. [s.l.] Elsevier Inc., 2009. p. 173–234.

INTERNATIONAL COCOA ORGANIZATION (ICCO). Cocoa Monthly Average Prices. Available in: https://www.icco.org/statistics/. Accessed in 08 sept. 2022.

JAHURUL, M. H. A. et al. Cocoa butter fats and possibilities of substitution in food products concerning cocoa varieties, alternative sources, extraction methods, composition, and characteristics. **Journal of Food Engineering**, v. 117, n. 4, p. 467–476, 2013.

JEBARI, K.; MADIAFI, M. Selection Methods for Genetic Algorithms. **Int. J. Emerg. Sci**, v. 3, n. 4, p. 333–344, 2013.

JORGE A. W. GUT. **Programação Matemática Para Otimização de Processos**. 1. ed. São Paulo: Edusp - Editora da Universidade de São Paulo, 2021.

KALAKUL, S. et al. Computer aided chemical product design – ProCAPD and tailor-made blended products. **Computers and Chemical Engineering**, v. 116, p. 37–55, 4 ago. 2018.

KARUNANITHI, A. T.; ACHENIE, L. E. K.; GANI, R. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. **Industrial and Engineering Chemistry Research**, v. 44, n. 13, p. 4785–4797, 22 jun. 2005.

KATOCH, S.; CHAUHAN, S. S.; KUMAR, V. A review on genetic algorithm: past, present, and future. **Multimedia Tools and Applications**, v. 80, n. 5, p. 8091–8126, 1 fev. 2021.

KONTOGEORGIS, G. M.; GANI, R. **Computer Aided Property Estimation for Process and Product Design**. [s.l: s.n.]. v. 19

KOZENY, V. Genetic algorithms for credit scoring: Alternative fitness function performance comparison. **Expert Systems with Applications**, v. 42, n. 6, p. 2998–3004, 15 abr. 2015.

LI, L.; LIU, G. Corn oil-based oleogels with different gelation mechanisms as novel cocoa butter alternatives in dark chocolate. **Journal of Food Engineering**, v. 263, p. 114–122, 1 dez. 2019.

LIN, B. et al. Computer-aided molecular design using Tabu search. **Computers and Chemical Engineering**, v. 29, n. 2, p. 337–347, 15 jan. 2005.

LIPP, M.; ANKLAM, E. Review of cocoa butter and alternative fats for use in chocolate - Part A. Compositional data. **Analytical. Nutritional and Clinical Method Section**, v. 62, n. 1, p. 73–97, 1998.

LOULLIS, A.; PINAKOULAKI, E. Carob as cocoa substitute: a review on composition, health benefits and food applications. **European Food Research and Technology**, v. 244, n. 6, p. 959–977, 1 jun. 2018.

MASHWANI, W. K.; HAIDER, R.; BRAHIM BELHAOUARI, S. A Multiswarm Intelligence Algorithm for Expensive Bound Constrained Optimization Problems. **Complexity**, v. 2021, p. 1–18, 27 fev. 2021.

MATTEI, M.; KONTOGEORGIS, G. M.; GANI, R. A comprehensive framework for surfactant selection and design for emulsion based chemical product design. **Fluid Phase Equilibria**, v. 362, p. 288–299, 25 jan. 2014.

MELO-ESPINOSA, E. A. et al. Surface tension prediction of vegetable oils using artificial neural networks and multiple linear regression. **Energy Procedia**, v. 57, p. 886–895, 2014.

MOGGRIDGE, G. D.; CUSSLER, E. L. An introduction to chemical product design. **Chemical Engineering Research and Design**, v. 78, n. 1, p. 5–11, 2000.

MOGGRIDGE, G. D.; CUSSLER, E. L. **Chemical Product Design**. [s.l.] Cambridge University press, 2001.

MONTGOMERY, D. C. **Design and analysis of experiments**. [s.l: s.n.].

NATHÁLIA CASSIA FERREIRA DIAS. **O papel do hormônio auxina na virulência do patógeno de cacau Moniliophthora perniciosa**. [s.l: s.n.].

NG, K. M.; GANI, R.; DAM-JOHANSEN, K. **Chemical Product Design: Toward a Perspective Through Case Studies**. [s.l: s.n.]. v. 23

NILSON, J. Measuring solid fat content. **The Manufacturing Confectioner**, p. 88–91, 1986.

NOUREDDINI, H.; TEOH, B. C.; CLEMENTS, D. Viscosities of vegetable oils and fatty acids. **JAOCS**, v. 69, p. 1189–1191, 1992.

O'BRIEN, R. D. **Fats and Oils: Formulating and Processing for Applications**. [s.l.] CRC Press, 1988.

ÓLAFSSON, S. Chapter 21 Metaheuristics. Em: **Handbooks in Operations Research and Management Science**. [s.l: s.n.]. v. 13p. 633–654.

PEREIRA, E. et al. Prediction of the melting behavior of edible fats using UNIFAC and UNIQUAC models. **Fluid Phase Equilibria**, v. 493, p. 58–66, 1 ago. 2019.

PEREIRA, E.; MEIRELLES, A. J. A.; MAXIMO, G. J. Predictive models for physical properties of fats, oils, and biodiesel fuels. **Fluid Phase Equilibria**, v. 508, 15 mar. 2020.

PINE II, B. J.; VICTOR, B.; BOYNTON, A. C. Making Mass Customization Work. **Harvard Business Review**, 1993.

RABELO, J. et al. Viscosity Prediction for Fatty Systems. **JAOCS**, v. 77, p. 1255–1262, 2000.

RODENBUSH, C. M.; HSIEH, F. H.; VISWANATH, D. S. Density and Viscosity of Vegetable Oils. **JAOCS**, p. 1415–1419, 1999.

SANTOS, O. O. et al. A simple, fast and efficient method for transesterification of fatty acids in foods assisted by ultrasound energy. **Journal of the Brazilian Chemical Society**, v. 25, n. 9, p. 1712–1719, 1 set. 2014.

SEDAGHAT BOROUJENI, L. et al. Optimization of sunflower oil bleaching parameters: Using response surface methodology (RSM). **Food Science and Technology (Brazil)**, v. 40, p. 322–330, 1 jun. 2020.

SEIDER, W. D. et al. Perspectives on chemical product and process design. **Computers and Chemical Engineering**, v. 33, n. 5, p. 930–935, 21 maio 2009.

SILVA, T. J.; BARRERA-ARELLANO, D.; RIBEIRO, A. P. B. The impact of fatty acid profile on the physicochemical properties of commercial margarines in Brazil. **JAOCS, Journal of the American Oil Chemists' Society**, v. 99, n. 6, p. 469–483, 1 jun. 2022.

SIVAKANTHAN, S.; JAYASOORIYA, A. P.; MADHUJITH, T. Optimization of the production of structured lipid by enzymatic interesterification from coconut (Cocos nucifera) oil and sesame (Sesamum indicum) oil using Response Surface Methodology. **LWT - Food Science and Technology**, v. 101, p. 723–730, 1 mar. 2019.

SMITH, K. W. Structured and modified lipids - Cocoa butter and cocoa butter equivalents. Em: [s.l: s.n.]. p. 401–422.

SMITH, B. V.; IERAPEPRITOU, M. G. Integrative chemical product design strategies: Reflecting industry trends and challenges. **Computers and Chemical Engineering**, v. 34, n. 6, p. 857–865, 10 jun. 2010.

SOARES, F. A. S. D. M. et al. Effects of chemical interesterification on physicochemical properties of blends of palm stearin and palm olein. **Food Research International**, v. 42, n. 9, p. 1287–1294, nov. 2009.

TEIXEIRA, P. J. P. L.; THOMAZELLA, D. P. DE T.; PEREIRA, G. A. G. **Time for Chocolate: Current Understanding and New Perspectives on Cacao Witches' Broom Disease Research**. **PLOS Pathogens**Public Library of Science, , 2015.

TELES DOS SANTOS, M. et al. Thermal properties of palm stearin, canola oil and fully hydrogenated soybean oil blends: Coupling experiments and modeling. **Journal of Food Engineering**, v. 185, p. 17–25, 1 set. 2016.

TELES DOS SANTOS, M.; GERBAUD, V.; LE ROUX, G. A. C. Comparison of predicted and experimental DSC curves for vegetable oils. **Thermochimica Acta**, v. 545, p. 96–102, 10 out. 2012.

TELES DOS SANTOS, M.; GERBAUD, V.; LE ROUX, G. A. C. Solid Fat Content of vegetable oils and simulation of interesterification reaction: Predictions from thermodynamic approach. **Journal of Food Engineering**, v. 126, p. 198–205, abr. 2014.

TELES DOS SANTOS, M.; GERBAUD, V.; ROUX, G. A. C. L. Modeling and simulation of melting curves and chemical interesterification of binary blends of vegetable oils. **Chemical Engineering Science**, v. 87, p. 14–22, 14 jan. 2013.

TELES DOS SANTOS, M.; MORGAVI, P.; LE ROUX, G. A. C. Exploring amazonian fats and oils blends by computational predictions of solid fat content. **OCL - Oilseeds and fats, Crops and Lipids**, v. 25, n. 1, 1 jan. 2018.

TERZI, S. et al. Product lifecycle management - From its history to its new role. **International Journal of Product Lifecycle Management**, v. 4, n. 4, p. 360–389, 2010.

TOSCANO, G. et al. Vegetable oil and fat viscosity forecast models based on iodine number and saponification number. **Biomass and Bioenergy**, v. 46, p. 511–516, nov. 2012.

VALERI, D.; MEIRELLES, A. J. A. Viscosities of Fatty Acids, Triglycerides, and Their Binary Mixtures. **JAOCS**, v. 74, n. 10, p. 1221–1226, 1997.

VENKATASUBRAMANIAN, V.; CHAN, K.; CARUTHERS, J. M. COMPUTER-AIDED MOLECULAR DESIGN USING GENETIC ALGORITHMS. **Computers chem. Engng**, v. 18, n. 9, p. 833–844, 1994.

VITOLO, M.; NEVES RODRIGUES RACT, J.; ANUNCIAÇÃO BRAGA GUEBARA, S. Feasibility on the Use of Acid Value for Evidencing the Esterification of Glycerol with Caprylic Acid Catalyzed by Lipase. **ATHENS JOURNAL OF SCIENCES**, v. 4, n. 1, p. 21–36, 28 fev. 2017.

WESDORP, L. H. **Liquid-Multiple Solid Phase Equilibria in Fats - theory and experiments**. [s.l: s.n.].

WON, K. W. Thermodynamic Model of Liquid-Solid Equilibria for Natural Fats and Oils. **Fluid Phase Equilibria**, v. 82, p. 261–273, 1993.

YANG, Y. S. et al. Quality Function Deployment-based optimization and exploration for ambiguity. **Journal of Engineering Design**, v. 14, n. 1, p. 83–113, mar. 2003.

YUNUS, N. A. et al. A systematic methodology for design of tailor-made blended products. **Computers and Chemical Engineering**, v. 66, p. 201–213, 4 jul. 2014.

ZANAKIS, S. H.; EVANS, J. R. Heuristic “Optimization”: Why, When, and How to Use It. **Interfaces**, v. 11, n. 5, p. 84–91, out. 1981.

ZEBERG-MIKKELSEN, C. K.; STENBY, E. H. Predicting the melting points and the enthalpies of fusion of saturated triglycerides by a group contribution method. **Fluid Phase Equilibria**, v. 162, p. 7–17, 1999.

ZHANG, L. et al. A Computer-Aided Methodology for Mixture-Blend Design. Applications to Tailor-Made Design of Surrogate Fuels. **Industrial and Engineering Chemistry Research**, v. 57, n. 20, p. 7008–7020, 23 maio 2018.

ZHANG, L.; CIGNITTI, S.; GANI, R. Generic mathematical programming formulation and solution for computer-aided molecular design. **Computers and Chemical Engineering**, v. 78, p. 79–84, 2 jul. 2015.

ZONG, L.; RAMANATHAN, S.; CHEN, C. C. Fragment-based approach for estimating thermophysical properties of fats and vegetable oils for modeling biodiesel production processes. **Industrial and Engineering Chemistry Research**, v. 49, n. 2, p. 876–886, 20 jan. 2010.

APPENDIX A – Table with vegetable oils and fats' fatty acid composition.

VO ID	VO Name	% saturated FA	Hexanoic	Octanoic	Decanoic	Dodecanoic	Tetradecanoic	Hexadecanoic	Hexadecenoic	Octadecanoic	Octadecenoic 9	Octadecenoic 11	Octadecadienoic	Octadecatrienoic	Icosano	Icosatetraeno	Docosano	Tetracosano
1	Açaí Oil	25%	0%	0%	0%	0%	0%	22%	2%	2%	60%	0%	12%	0%	1%	0%	0%	0%
2	Argan Oil	19%	0%	0%	0%	0%	0%	13%	2%	6%	46%	0%	33%	0%	0%	0%	0%	0%
3	Avocado Oil	13%	0%	0%	0%	0%	0%	11%	8%	1%	61%	0%	12%	1%	0%	0%	0%	0%
4	Babassu Oil	84%	0%	6%	4%	45%	17%	9%	0%	3%	13%	0%	3%	0%	0%	0%	0%	0%
5	Brazilian Nut Oil	18%	0%	0%	0%	0%	0%	13%	0%	5%	48%	0%	34%	0%	0%	0%	0%	0%
6	Buriti Pulp Oil	19%	0%	0%	0%	0%	0%	19%	0%	0%	69%	0%	11%	1%	0%	0%	0%	0%
7	Canola Oil	7%	0%	0%	0%	0%	0%	5%	0%	3%	68%	0%	18%	7%	0%	0%	0%	0%
8	Coconut Oil	93%	0%	6%	5%	46%	19%	9%	0%	3%	6%	0%	1%	0%	0%	0%	0%	0%
9	Corn Oil	14%	0%	0%	0%	0%	0%	11%	0%	2%	25%	0%	60%	1%	0%	0%	0%	0%
10	Cotton Seed Oil	28%	0%	0%	0%	0%	1%	24%	1%	3%	17%	0%	53%	0%	0%	0%	0%	0%
11	Cupuassu Oil	51%	0%	0%	0%	0%	0%	8%	0%	34%	43%	0%	6%	0%	9%	0%	0%	0%
12	Andiroba Oil	35%	0%	0%	0%	0%	0%	24%	0%	10%	57%	0%	6%	0%	0%	0%	0%	0%
13	Vaccenic-enriched Oil	4%	0%	0%	0%	0%	0%	2%	0%	2%	6%	84%	6%	0%	0%	0%	0%	0%
14	Copalba Oil	18%	0%	0%	0%	0%	1%	12%	0%	5%	43%	0%	33%	2%	0%	0%	0%	0%
15	Fully Hydrogenated Soy Fat	100%	0%	0%	0%	0%	0%	12%	0%	88%	0%	0%	0%	0%	0%	0%	0%	0%
16	Grapeseed Oil	13%	0%	0%	0%	0%	0%	8%	0%	4%	20%	0%	67%	0%	0%	0%	0%	0%
17	Illipe Butter	64%	0%	0%	0%	0%	0%	20%	0%	43%	36%	0%	0%	0%	0%	0%	0%	0%
18	Jatropa Butter	21%	0%	0%	0%	0%	0%	14%	1%	7%	45%	0%	33%	0%	0%	0%	0%	0%
19	Kokum Kernel Fat	60%	0%	0%	0%	0%	0%	0%	0%	60%	40%	0%	0%	0%	0%	0%	0%	0%
20	Licuri Oil	88%	0%	11%	7%	45%	15%	7%	0%	4%	8%	0%	4%	0%	0%	0%	0%	0%

VO ID	VO Name	% saturated FA	Hexanoic	Octanoic	Decanoic	Dodecanoic	Tetradecanoic	Hexadecanoic	Hexadecenoic	Octadecanoic	Octadecenoic 9	Octadecenoic 11	Octadecadienoic	Octadecatrienoic	Icosano	Icosatetraeno	Docosano	Tetracosano
21	Linseed Oil	13%	0%	0%	0%	0%	0%	7%	0%	6%	25%	0%	14%	47%	0%	0%	0%	0%
22	Macadamia Nut Oil	22%	0%	0%	0%	0%	0%	11%	28%	5%	45%	0%	2%	0%	4%	0%	1%	0%
23	Macauba Kernel Oil	70%	1%	6%	4%	41%	8%	7%	0%	2%	27%	0%	3%	0%	0%	0%	0%	0%
24	Mango Oil	53%	0%	0%	0%	0%	11%	0%	0%	41%	45%	0%	0%	0%	0%	0%	0%	0%
25	Mango Seed Butter	53%	0%	0%	0%	0%	0%	9%	0%	39%	41%	0%	6%	0%	2%	0%	1%	0%
26	Microalgae Oil	45%	0%	0%	0%	0%	3%	39%	44%	3%	3%	0%	1%	0%	0%	6%	0%	0%
27	Murumuru Butter	90%	0%	1%	1%	49%	30%	7%	0%	3%	7%	0%	3%	0%	0%	0%	0%	0%
28	Olive Oil	12%	0%	0%	0%	0%	0%	9%	1%	3%	80%	0%	6%	1%	0%	0%	0%	0%
29	Palm Oil	49%	0%	0%	0%	0%	1%	43%	0%	4%	41%	0%	10%	0%	0%	0%	0%	0%
30	Palm Stearin	62%	0%	0%	0%	0%	1%	56%	0%	5%	32%	0%	6%	0%	0%	0%	0%	0%
31	Passion Fruit Oil	14%	0%	0%	0%	0%	0%	11%	0%	3%	17%	0%	66%	3%	0%	0%	0%	0%
32	Passion Fruit Seed Oil	10%	0%	0%	0%	0%	0%	10%	0%	0%	17%	0%	70%	0%	0%	0%	0%	0%
33	Peanut Oil	20%	0%	0%	0%	0%	0%	11%	0%	2%	48%	0%	33%	0%	1%	0%	3%	2%
34	Pequi Oil	40%	0%	0%	0%	0%	0%	36%	1%	2%	47%	0%	11%	0%	0%	0%	0%	0%
35	Pomegranate Oil	15%	0%	0%	0%	0%	0%	11%	0%	4%	27%	0%	53%	5%	0%	0%	0%	0%
36	Pracaxi Oil	40%	0%	0%	0%	1%	0%	1%	0%	3%	47%	0%	12%	1%	12%	0%	23%	0%
37	Rice Brain Oil	24%	0%	0%	0%	0%	0%	22%	0%	2%	43%	0%	31%	2%	0%	0%	0%	0%
38	Rubber Seed Oil	14%	0%	0%	0%	0%	0%	9%	0%	5%	28%	0%	34%	24%	0%	0%	0%	0%
39	Sal Fat	56%	0%	0%	0%	0%	0%	6%	0%	39%	41%	0%	2%	0%	9%	0%	0%	0%
40	Shea Butter	54%	0%	0%	0%	0%	0%	6%	0%	48%	42%	0%	5%	0%	1%	0%	0%	0%

VO ID	VO Name	% saturated FA	Hexanoic	Octanoic	Decanoic	Dodecanoic	Tetradecanoic	Hexadecanoic	Hexadecenoic	Octadecanoic	Octadecenoic 9	Octadecenoic 11	Octadecadienoic	Octadecatrienoic	Icosano	Icosatetraeno	Docosano	Tetracosano
41	Soybean Oil	20%	0%	0%	0%	0%	0%	12%	0%	7%	29%	0%	44%	3%	0%	0%	0%	0%
42	Sunflower Oil	10%	0%	0%	0%	0%	0%	6%	0%	4%	26%	0%	64%	1%	0%	0%	0%	0%
43	Tucumã Pulp	29%	0%	0%	0%	0%	0%	26%	0%	3%	65%	0%	2%	4%	0%	0%	0%	0%
44	Tucumã Kernel	82%	0%	0%	0%	49%	25%	6%	0%	2%	14%	0%	4%	0%	0%	0%	0%	0%
45	Ucuúba Butter	95%	0%	0%	0%	16%	73%	4%	0%	1%	4%	0%	1%	0%	0%	0%	0%	0%

APPENDIX B – Pseudocode of the Genetic Algorithm used in the Binary Mixture Optimization

```
n_elem #Number of available oils in the database
n_pop #Total population of each generation
n_gen #Total number of generations (termination criteria)
elit #Replacement Elitism Rate
selection_elit = n_pop - elit
mutation_rate
crossover_rate
dataset_with_oil_list
j = 0

while j < n_gen:
        dataset_evaluated = calculate_objective_function(dataset_with_oil_list)
        parents_list = elitism_selection(dataset_evaluated)

############################## Crossover operation ################################

   i = 0
   while i < (selection_elit):
      if crossover_prob <= crossover_rate:
         child_1, child_2 = crossover(parents_list[i], parents_list[i+1])

      else:
         child_1, child_2 = parents_list[i], parents_list[i+1]

      i = i + 2

############################## Mutation operation ################################
   k = 0
   while k < (selection_elit):
      if mutation_prob <= mutation_rate:
         child_mut = mutation(children[k], n_elem)

      else:
         child_mut = children[k]

      k = k + 1

best_mixtures_kept_in_next_pop = elitism_replacement(dataset_with_oil_list)
final_dataset = best_mixtures_kept_in_next_pop ++ child_mut
final_dataset_with_sfc_values = run_programs(final_dataset)
```