

Article

An Optimal Distillation Process for Turpentine Separation Using a Firefly Algorithm

Gustavo Mendes Platt ^{1,2,*} , Otávio Knevez de Azevedo ^{1,†} and Francisco Bruno Souza Oliveira ^{2,†} 

¹ Graduate Program in Agroindustrial Systems and Processes, School of Chemistry and Food, Federal University of Rio Grande (PPGSPA/FURG), Santo Antônio da Patrulha 95500-000, Brazil; otavioazevedo8@gmail.com

² Graduate Program in Computational Modelling in Science and Technology, Department of Engineering and Computing, State University of Santa Cruz (PPGMC/UESC), Ilhéus 45662-900, Brazil; fbsoliveira@uesc.br

* Correspondence: gmplatt@furg.br

† These authors contributed equally to this work.

Abstract: The optimal design of distillation separation processes has become a fundamental tool in industries in order to minimize operating costs and investments. In many cases, the optimization stage has been carried out using metaheuristics, with the process simulation stage carried out externally to the optimization. This paper presents an optimal design methodology for separating the components of turpentine, a raw material of natural origin, based on coupling a distillation process simulator with the Firefly metaheuristic as an optimizer. Results were obtained for a distillation process to obtain α -pinene and β -pinene (two of the main components of turpentine), meeting purity criteria in the top products of the equipment while minimizing a measure of the total annualized cost. The results show that the tool developed—together with the Firefly algorithm—is capable of obtaining optimized results (although there is no guarantee of a global optimum) from a small set of initial design configurations.

Keywords: optimization; metaheuristic; Firefly algorithm; distillation



Academic Editor: Wei Gao

Received: 13 March 2025

Revised: 16 April 2025

Accepted: 19 April 2025

Published: 22 April 2025

Citation: Platt, G.M.; Azevedo, O.K.d.; Oliveira, F.B.S. An Optimal Distillation Process for Turpentine Separation Using a Firefly Algorithm. *Modelling* **2025**, *6*, 34. <https://doi.org/10.3390/modelling6020034>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Process modeling and simulation have become extremely important in today's industry and for the development of new industrial technologies [1,2]. In separation processes, mathematical modeling plays a key role [3–5]. In fact, the implementation of the so-called Industry 4.0 depends on modeling, simulation and optimization tools [6].

There are currently many studies devoted to the separation of substances from natural raw materials in environmentally sustainable processes [7]. In this context, turpentine is an oil obtained from pine trees in the pine chemical industry, and is a natural source of monoterpenes, such as α -pinene and β -pinene [8]. In addition to α -pinene and β -pinene, turpentine contains δ -3-carene, limonene and p-cymene, among others. Its composition varies depending on the type of tree analyzed [9,10].

Turpentine is usually used as a raw material for the pharmaceutical industry (production of perfumes and deodorants, for example). On the other hand, it may be economically interesting to separate the components present in turpentine. According to Ariono et al. [9], the selling prices of α -pinene and β -pinene exceed the selling price of turpentine by 45% and 74%, respectively, which justifies the study of separation processes of turpentine into its components. The world's largest exporters of turpentine are the United States, China and Brazil [11]. In Latin American countries, such as Brazil, resin-harvesting activities are

proving to be important from the point of view of family farming [12] as part of a relevant integration between farmers and the agro-industry.

In recent studies of distillation processes design, the most modern approaches contemplate optimal design structures, where the purity specifications of the products of interest are achieved while simultaneously minimizing the operating costs and an annualization of the investment in the equipment (commonly referred to as the total annualized cost (TAC)). In many cases, the minimization procedure is carried out with the help of a metaheuristic employed externally to the routine for simulating the distillation process. As pointed by Tian et al. [13], the use of sequential modular process simulators (the most common type of process simulator) is suitable for algorithms that are not based on derivatives (such as metaheuristics). In this situation, as discussed by Murrieta-Duenas et al. [14], the simulation environment (as well as the calculation of the objective function) appears as a “black-box” model. Javaloyes-Antón et al. [15] coupled a commercial simulator with the Particle Swarm Optimization algorithm in order to obtain an optimal design configuration for distillation processes, minimizing the TAC. Li et al. [16] applied a Simulated Annealing algorithm to the optimal design of extractive dividing wall columns in the ternary benzene–isopropanol–water system. Hu et al. [17] used a hybrid Differential Evolution–NSGA-II algorithm in the optimal design of separation sequences while considering two conflicting objectives: minimizing the total annual cost and the carbon dioxide emissions. Chia et al. [18] employed a Genetic Algorithm and the fast and elitist nondominated sorting Genetic Algorithm (NSGA-II) to solve mono- and multi-objective optimization problems for a hybrid distillation–pervaporation process. In the mono-objective problem, the total annualized cost (TAC) was minimized. The two components of the TAC—Capital expenditure (CAPEX) and Operational expenditure (OPEX)—were considered as conflicting objectives in the multi-objective problem. Chen et al. [19] employed an NSGA-III algorithm in a multi-objective problem for methanol distillation, where the conflicting objectives were the total annual cost and the carbon dioxide emissions. Tian et al. [13] proposed an Enhanced Logic-Based Bender’s Decomposition Algorithm with the Proximity Principle to be used in conjunction with process simulators. Results were obtained for a methanol distillation column and an extraction/distillation process for the separation of cyclohexane/cyclohexene. Murrieta-Duenas et al. [14] presented a comparison between a Differential Evolution approach and Boltzmann-based distribution algorithms in the optimization of distillation processes. The methodologies were compared using a single distillation column for a binary mixture, a separation train (three sequential columns) and a quaternary mixture separation in a single column.

In the context of turpentine separation, few studies have been published, and without direct applications of optimization methodologies. For instance, Becerra et al. [20] analyzed the separation of d-limonene and α -pinene from citrus and turpentine oils by continuous and vacuum distillation with an economical analysis. Valencia et al. [8] presented a study with an experimental and simulation approach to obtain α -pinene and β -pinene from turpentine using vacuum batch distillation. Chalier et al. [21] compared the extraction of turpentine essential oil considering conventional- and microwave-assisted hydro-distillation and vacuum distillation.

Proposed Current Study and Research Gap

The present study analyzed the feasibility of the separation process of turpentine compounds through a distillation sequence (two columns) in order to obtain α -pinene (top of the first column) and β -pinene (top of second column) from turpentine oil. Based on this scenario, computational simulation and optimization codes were used to assess the technical and economic feasibilities of the process, with the aim to minimize the project

costs. All codes were developed in the Scilab language (©2025 Dassault Systèmes—all rights reserved), an open-source platform. The developed framework was composed by a simulation module and an optimization structure, based on the Firefly algorithm [22] (a population-based metaheuristic). An important feature of the methodology implemented is the use of a small number of individuals in the population due to the complexity of the algorithm.

The main differences between the present research and the previous works were as follows:

- The use of the Firefly metaheuristic with a small population size in the optimal design of a distillation process: as far as we know, this is the first application of Firefly to this type of problem, although there are publications on the use of Firefly in tuning PID controllers in distillation columns [23]. The Firefly algorithm has some peculiar characteristics [24] that justify its study in real engineering problems, although the comparisons of metaheuristics in the present problem were beyond the scope of this work.
- The vast majority of studies using metaheuristics to optimize distillation processes used proprietary tools (commercial simulators) to simulate the process. These tools are closed source and, in many cases, are not accessible to most industries. In this work, open-source and free tools were used, and all the modeling necessary to implement the methodology is presented.
- The application of the optimal design methodology for the separation of turpentine components is also not a subject that is usually addressed in the scientific literature. The work of Valencia et al. [8] addressed only simulation aspects. Becerra et al. [20] presented a sensitivity analysis in order to evaluate the economic feasibility of the process, but not a project-optimization approach. In both cases [8,20], commercial simulation codes were used in the simulation step.

2. Mathematical Modeling

In this section, we present the mathematical modeling of the entire process, including the thermodynamic models, the distillation process equations and the economic modeling.

2.1. Thermodynamic Modeling

This work considered a mixture made up of α -pinene, β -pinene and p-cymene. Other substances could be included in the model, but we opted for these three substances for computational cost reasons.

In this work, we considered the modified Raoult's law for phase equilibrium (a nonideal liquid phase and an ideal vapor phase). In this situation, the vapor–liquid equilibrium ratio k_i was represented by [25]

$$k_i = \frac{\gamma_i P_i^{sat}}{P} = \frac{y_i}{x_i} \quad (1)$$

where the index i refers to a substance, and x_i and y_i refer, respectively, to the molar fractions of the liquid and vapor phases. The choice of an ideal vapor phase model was justified by the low operating pressure (P) in the process.

The vapor pressures of pure fluids (P^{sat}) were calculated by an Antoine-type equation:

$$\ln P_i^{sat} = A_i - \frac{B_i}{T - C_i}, \quad (2)$$

where A_i , B_i and C_i are the Antoine constants for substance i ; P_i^{sat} is in kPa; and T is the temperature in Kelvin. For α -pinene and β -pinene, these values were provided by Bernardo-Gil and Barreiros [26]. The Antoine parameters for p-cymene were tabulated by Bernardo-Gil and Ribeiro [27].

The activity coefficients in a mixture with C components were calculated by the Wilson model [25]:

$$\ln \gamma_i = 1 - \ln \left(\sum_{j=1}^C x_j \Lambda_{ij} \right) - \sum_{k=1}^C \frac{x_k \Lambda_{ki}}{\sum_{j=1}^C x_j \Lambda_{kj}}, \quad (3)$$

where

$$\Lambda_{ij} = \frac{V_j}{V_i} \exp \left(\frac{-a_{ij}}{RT} \right), \quad (4)$$

where a_{ij} are nonsymmetric binary interaction parameters and R is the universal gas constant. Also, in Equation (4), V_i refers to the molar volume of component i . For the binary pair α -pinene + β -pinene, the binary interaction parameters were obtained from [26]. The binary parameters for α -pinene and p-cymene were tabulated by Bernardo-Gil and Ribeiro [28]. Finally, the data for β -pinene + p-cymene came from Bernardo-Gil and Ribeiro [27]. The molar volume of α -pinene was 158.5 mL/mol [29]; for β -pinene, the molar volume was 156.2 mL/mol [30]. Finally, for p-cymene, the molar volume was 155.5 mL/mol [31]. The Wilson model was chosen by taking into account the availability of binary interaction parameters between the components of the mixture, although there were no strong deviations from ideality.

Since the reduced temperature $T_r \ll 1$ (where $T_r = T/T_c$, and T_c is the critical temperature of a pure component), it was reasonable to consider that the latent heats of vaporization were not functions of temperature. In this sense, we considered $\Delta H_{vap}^{\alpha\text{-pinene}} = 44.57$ kJ/mol and $\Delta H_{vap}^{\beta\text{-pinene}} = 45.80$ kJ/mol [32]. Furthermore, we used a constant value for the heat of vaporization of the mixtures (in order to obtain the constant molar overflow assumptions). Considering that the mixture that was formed was mainly composed of α -pinene and β -pinene, it was reasonable to assume that $\Delta H_{vap} \approx \frac{\Delta H_{vap}^{\alpha\text{-pinene}} + \Delta H_{vap}^{\beta\text{-pinene}}}{2}$.

2.2. Column Modeling

In the following subsections, we detail how the flow rates, thermal duties, compositions and temperatures in the columns were obtained. In addition, the hydrodynamic aspects of the column are discussed.

A typical stage j of the distillation column is presented in Figure 1. In this stage, we considered a feed with flowrate F_j , vaporized fraction β_j and composition z_j . In order to represent the condenser and the reboiler of the column, a heat duty Q_j was also permitted. The stage received liquid from an upper stage ($j - 1$), with a flowrate L_{j-1} and liquid composition x_{j-1} . Similarly, a vapor flow V_{j+1} from stage $j + 1$ with vapor composition y_{j+1} entered in stage j . The liquid and vapor flows left stage j with compositions x_j and y_j . Finally, we considered sidedraw removals of liquid U_j and vapor W_j . Each stage was an equilibrium stage with temperature T_j and pressure P_j .

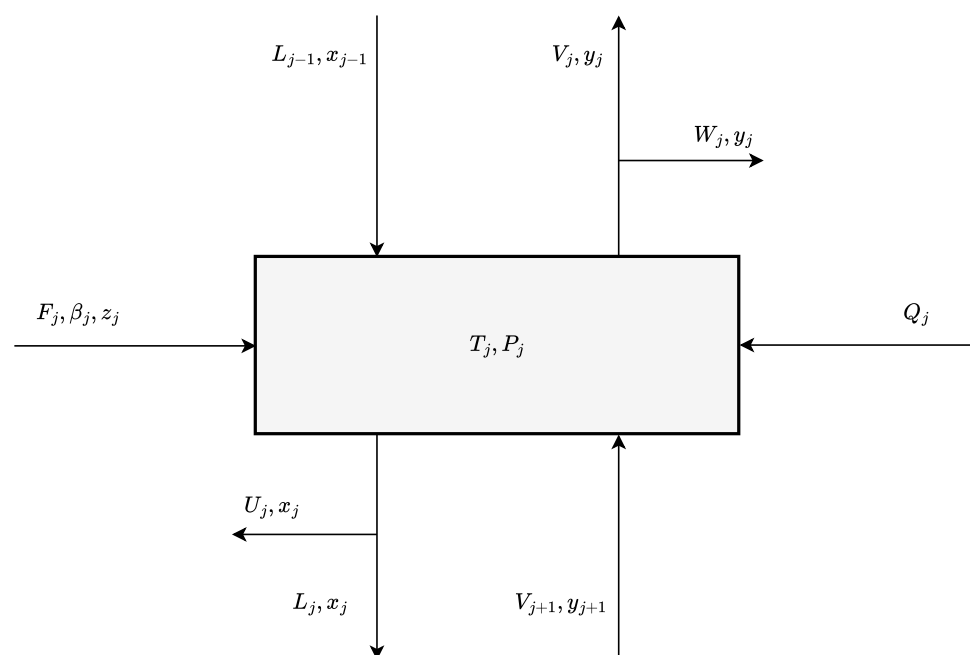


Figure 1. Column stage j .

2.2.1. The Flowrate and Heat Duty Problem— (L, V, Q) Problem

In this work, we considered the constant molar overflow assumptions [33,34] as a consequence of a linear dependence of the saturation enthalpies on the compositions. In this situation, the changes in the molar flowrates in a section j of the column were represented by

$$V_{j+1} = V_j + W_j - F_j\beta_j - \frac{Q_j}{\Delta H_{vap}}, \quad (5)$$

and

$$L_j = L_{j-1} - U_j + F_j(1 - \beta_j) - \frac{Q_j}{\Delta H_{vap}}. \quad (6)$$

One of the specifications usually made in distillation column simulations is the external top reflux ratio R_r :

$$R_r = \frac{L_1}{D}, \quad (7)$$

where L_1 represents the liquid returning to the column and D is the distillate flowrate.

Thus, considering that the distillate flowrate D was known (obtained from the material balance of the entire column), as well as the value of R_r , we could calculate the value of L_1 , and then, the value of L_2 , etc., for all the column sections. In this situation, the (L, V, Q) problem appeared as a linear algebraic system that could be solved before solving the compositional and temperature problems, which reduced the size of the non-linear problem.

2.2.2. Composition Problem

A component material balance in stage j represented in Figure 1 gave

$$L_{j-1}x_{j-1} + V_{j+1}y_{j+1} - (L_j + U_j)x_j - (V_j + W_j)y_j = -F_jz_j. \quad (8)$$

Furthermore, in each stage, we had the phase equilibrium condition in a vectorial version of Equation (1):

$$y_j = \text{diag}(k_j)x_j, \quad (9)$$

where the operator *diag* refers to a diagonal matrix. It must be clear that in the last expression, k_j is a vector with length C . Similarly, y_j and x_j refer to the vectors of molar fractions for the vapor and liquid phases, respectively. Since k_j is a nonlinear function of x_j and T , and also considering the relationship between the composition of the liquid and the vapor in equilibrium, it is clear that the composition problem is a nonlinear algebraic system of equations.

The constraints regarding the molar fraction vector needed to also be considered for each equilibrium stage j :

$$\sum_{i=1}^C x_{i,j} = 1, \sum_{i=1}^C y_{i,j} = 1. \quad (10)$$

Under constant molar overflow assumptions, the energy balances and a constraint regarding molar fractions were substituted by the two equations representing the changes in the molar flowrates—Equations (5) and (6).

At the end of solving the compositional problem, the liquid and vapor composition values were obtained at each stage of the column, as well as the temperature values.

2.2.3. Column Hydrodynamics

With the compositions, flowrates and temperatures calculated for each stage of the equipment, the next step was to conduct the hydrodynamic calculations in the column. These calculations were important to evaluate the column diameter, an essential parameter in the economic evaluation of the column shell and trays. In this work, we followed the procedure described by Smith [25].

The first step was to calculate the flow parameter F_{LV} , as follows:

$$F_{LV} = \frac{M_L L}{M_V V} \left(\frac{\rho_V}{\rho_L} \right)^{0.5}, \quad (11)$$

where M_L and M_V are the liquid and the vapor molar mass, respectively. Since the main components of the mixture (α -pinene and β -pinene) were isomers, we considered $M_L = M_V$. The density of the liquid was considered as $\rho_L = 858 \text{ kg/m}^3$. The vapor density was calculated using the ideal gas law with a temperature of 400 K (considering the typical temperature values in this process). The quantities L and V are the molar flow rates of liquid and vapor in the column, and were obtained by solving the (L, V, Q) problem (a linear problem).

The terminal velocity parameter K_T was then calculated using [25]

$$K_T = \left(\frac{\sigma}{20} \right) \exp \left[-2.979 - 0.717 \ln F_{LV} - 0.0865 (\ln F_{LV})^2 + 0.997 \ln H_T - 0.07973 \ln F_{LV} \ln H_T + 0.256 (\ln H_T)^2 \right], \quad (12)$$

where σ is the dynamic viscosity and H_T is the tray spacing (m). Here, we used $\sigma = 25.87 \text{ mN/m}$.

The vapor flooding velocity (at the bottom of the column) was [25]

$$v_T = 0.9 K_T \left(\frac{\rho_L - \rho_V}{\rho_V} \right). \quad (13)$$

Then, one could calculate the column diameter:

$$D_c = \left(\frac{4 M_V V}{0.9 \times 0.8 \times \pi \rho_V v_T} \right)^{0.5}, \quad (14)$$

where the factors 0.9 and 0.8 stand for the downcomer area and the fraction of the flooding velocity, respectively [25].

The column height was represented by

$$H_c = 0.7010N_e, \quad (15)$$

where the factor 0.7010 represents a slack of 15% in relation to the number of stages in the column. The tray spacing considered was 2 ft (0.6096 m).

2.3. Economic Modeling

The objective function to be minimized was a measure of the total annual cost (TAC), represented by [17]

$$TAC = \frac{\text{capital investment}}{\text{payback}} + \text{operation cost}, \quad (16)$$

where the term “payback” refers to the payback time (in years). In this work, we considered a payback time of three years [17].

The capital investment was the sum of the column costs, the column internals (trays, in this work) and the heat exchangers (condensers and reboilers). We used some correlations to estimate these investments.

The column cost (shell) was represented by [17]

$$C_{shell} = \left(\frac{M\&S}{280} \right) 937.6 D_c^{1.066} H_c^{0.802} (2.18 + F_c), \quad (17)$$

where $M\&S$ is the Marshall and Swift index ($M\&S = 1448.3$ [17]; the Marshall and Swift index is a dimensionless parameter used to take into account inflationary effects over time), D_c is the column diameter (m) and H_c is the column height (m). The value of F_c was 3.67 (stainless steel, low pressure).

The tray costs were calculated using [17]

$$C_{tray} = \left(\frac{M\&S}{280} \right) 97.2 D_c^{1.55} H_c F_c, \quad (18)$$

where $F_c = (1 + 1.7)$ (sieve tray, stainless steel).

For the heat exchangers, we used [17]

$$C_{he} = \left(\frac{M\&S}{280} \right) 474.67 A^{0.65} (2.29 + F_c), \quad (19)$$

where A is the heat exchanger area in m^2 and $F_c = 3.75$. The heat exchanger area was calculated using the thermal duties of the condenser and the reboiler:

$$A = \frac{Q}{U\Delta T}. \quad (20)$$

In Equation (20), Q is the condenser/reboiler thermal duty, U is the global heat transfer coefficient and ΔT is the logarithm mean temperature difference. For the condenser, we used $U = 568.1 \text{ W/m}^2\text{K}$ [35], and cooling water entered at 32°C and left the equipment at 47°C . For the reboiler, we used $U\Delta T = 35,487 \text{ W/m}^2$ [35].

The utility costs were the medium pressure steam cost and the cooling water cost. We considered only medium-pressure steam, with the cost calculated by [15]

$$C_{steam} = Q_{reb} \times 8.22 \times 8000, \quad (21)$$

where Q_{reb} is the reboiler thermal duty in GJ/s. The factor 8.22 represents the cost of steam (USD/GJ). The plant operation factor was 8000 h/year. In a similar way, the cost of the cooling water was [17]

$$C_{cw} = Q_{cond} \times 0.354 \times 8000, \quad (22)$$

where the cost of the cooling water was 0.354 USD/GJ.

3. Computational Modeling

This section is devoted to the description of the computational tools employed in the solution of the mathematical models of the process. Essentially, two main blocks were developed: a simulation module and an optimization module (based on a metaheuristic called the Firefly method [22]).

3.1. Column Simulation Strategy

The first step of the column simulation was the solution of the (L, V, Q) problem. Since this problem is linear, we employed a simple matrix inversion of the Scilab based on the Lapack DGETRF and DGETRI routines [36].

The compositional problem was solved by a damped Newton method [37], with the Jacobian matrix evaluated by finite differences (with the perturbation size $h = 1 \times 10^{-6}$). We used a constant damping factor, with an initial value of 0.7, in order to avoid divergence in the initial iterations. When the absolute error between two consecutive iterations was less than 1 (also an arbitrary value), the damping factor was increased to 1 in order to speed up the convergence of Newton's method in the vicinity of the solution. If the inverse of the Jacobian matrix did not exist (divergence condition), the objective function value was set to 1×10^7 . Considering the size of the problem, where Jacobian matrices with dimensions greater than 200 by 200 were easily obtained, the compositional solution stage was one of the "slow" steps in the process.

3.2. Optimization Strategy—Metaheuristic Method

After developing the simulation tool, the next step involved implementing the optimizer using a stochastic optimization approach (metaheuristics).

In this work, we used the Firefly metaheuristic [22] as the optimization algorithm for the process optimal design. The Firefly algorithm mimics the behavior of fireflies in nature, where a "brighter" element attracts other, less "bright" elements. A brighter individual, in a minimization context, refers to a point with a lower objective function value. A very important feature in the Firefly algorithm is its innate ability to divide the population into subpopulations in a multiswarm scheme, as pointed out by Yang and He [24]. Platt et al. [38] demonstrated this pattern in the calculation of a double azeotrope (a problem with two solutions). Details of the Firefly algorithm can be found in Yang [22].

In the situation where firefly i is attracted by firefly j , its movement is represented by the following equation:

$$\theta_i^{t+1} = \theta_i^t + \beta_0 \exp(-\gamma r_{ij}^2) \odot (\theta_j^t - \theta_i^t) + \alpha^t \odot \epsilon_i^t, \quad (23)$$

where β_0 is the attractiveness for two collapsing fireflies ($r_{ij} = 0$), γ is the light absorption coefficient, α_t is the randomization parameter vector and ϵ_i presents a uniform distribution in the interval $[-0.5; 0.5]$. In this equation, t represents the number of generations (or movements) for the fireflies and the symbol $\odot$ is a Hadamard product.

The algorithmic structure of Firefly is presented in Algorithm 1. As pointed out by Zhang et al. [39], the time complexity of the Firefly algorithm is $O(m^2 t_{max})$ (where m is the number of fireflies and t_{max} is the number of generations), as a consequence of the double

loop in the internal structure. Thus, considering the computational cost of the simulation step, the number of fireflies has to be small. Some authors have recently addressed the issue of time complexity for the Firefly algorithm. For instance, Wang et al. [40] proposed a Neighborhood attraction Firefly algorithm (NaFa), avoiding the so-called “full attraction” model of the original proposal. Bei et al. [41] presented an Improved Hybrid Firefly algorithm with a Probability Attraction Model (IHFAPA), also devoted to the reduction in the time consumed in the attraction-between-fireflies stage. Cheng et al. [42] proposed a Hybrid Firefly algorithm with group attraction (HFA-GA), avoiding oscillations caused by a large number of attractions. Therefore, the issue of the high number of attractions in the algorithm can be addressed, but the central point of this work was to verify whether Firefly is capable of solving—in its canonical form—the optimal design problem of the separation system under study. The ranking of the fireflies (considering the values of the objective functions) is performed by a typical bubble sort algorithm [43].

Algorithm 1: The Firefly algorithm

Require: $f(\theta), m, \beta_0, t_{max}, \gamma, \alpha^0, t$

- 1: **for** $i = 1$ to m **do**
- 2: $\theta_i \leftarrow$ generate initial solution
- 3: **end for**
- 4: **while** $t < t_{max}$ **do**
- 5: **for** $i = 1$ to m **do**
- 6: **for** $j = 1$ to m **do**
- 7: $r_{ij} \leftarrow$ calculate the Euclidean distance between θ_i and θ_j
- 8: $\epsilon_i^t \leftarrow$ generate random vector in the interval $[-0.5; 0.5]$
- 9: **if** $f(\theta_j) < f(\theta_i)$ **then**
- 10: $\theta_i^{t+1} \leftarrow \theta_i^t + \beta_0 \exp(-\gamma r_{ij}^2) \odot (\theta_j^t - \theta_i^t) + \alpha^t \odot \epsilon_i^t$
- 11: **end if**
- 12: **end for**
- 13: **end for**
- 14: $\alpha^{t+1} \leftarrow 0.997\alpha^t$
- 15: $t \leftarrow t + 1$
- 16: Rank the fireflies and find the best element (bubble sort).
- 17: **end while**

One point to note is that this work was not intended to compare metaheuristics in the optimal design problem, but to demonstrate how this class of algorithms can be used to obtain practical results in engineering problems without specific adaptations for the problem.

3.3. Proposed Framework

Considering the implementation of simulation and optimization tools, a structured approach was proposed to solve the problem.

Each vector θ was then formed by the optimization variables $N_1, F_{s,1}, R_{r,1}, N_2, F_{s,2}$ and $R_{r,2}$. Since the number of stages and the feed stages were integer variables, we simply considered rounding to the nearest integer in these cases. Thus, Firefly worked with real variables, and the simulation module performed the rounding procedure.

The proposed framework was then composed of a simulation module—responsible for the simulation of the entire process (the two distillation columns)—an economic module and an optimization module (the Firefly algorithm), according to Figure 2. At first, a set of initial estimates of stage numbers (N_1 and N_2), feed stages ($F_{s,1}$ and $F_{s,2}$) and reflux ratios

($R_{r,1}$ and $R_{r,2}$) was generated by the Firefly algorithm (the optimization module). With this information and the process design specifications (turpentine feed flow rate, turpentine feed composition and bottom flow rates of the two columns), it was then possible to carry out the process simulation. As already mentioned, the simulation was performed in two stages: initially, the problem of calculating flow rates and thermal duties was solved in the form of a linear algebraic system. Then, the compositional problem was solved using a damped Newton method. At the end of the simulation stage, all the variables relevant to the process had been calculated, which then allowed the total annualized cost to be calculated. If the molar fraction of α -pinene in the first column was less than 0.9 or the molar fraction of β -pinene in the second column was less than 0.9, the project was considered unfeasible, and a value of 1×10^8 was assigned to the objective function. The Firefly algorithm then ranked the elements of the population using a bubble sort algorithm. If the maximum number of generations t_{max} was reached, the best firefly was the optimal solution to the problem. Otherwise, the algorithm moved the fireflies to the next iteration.

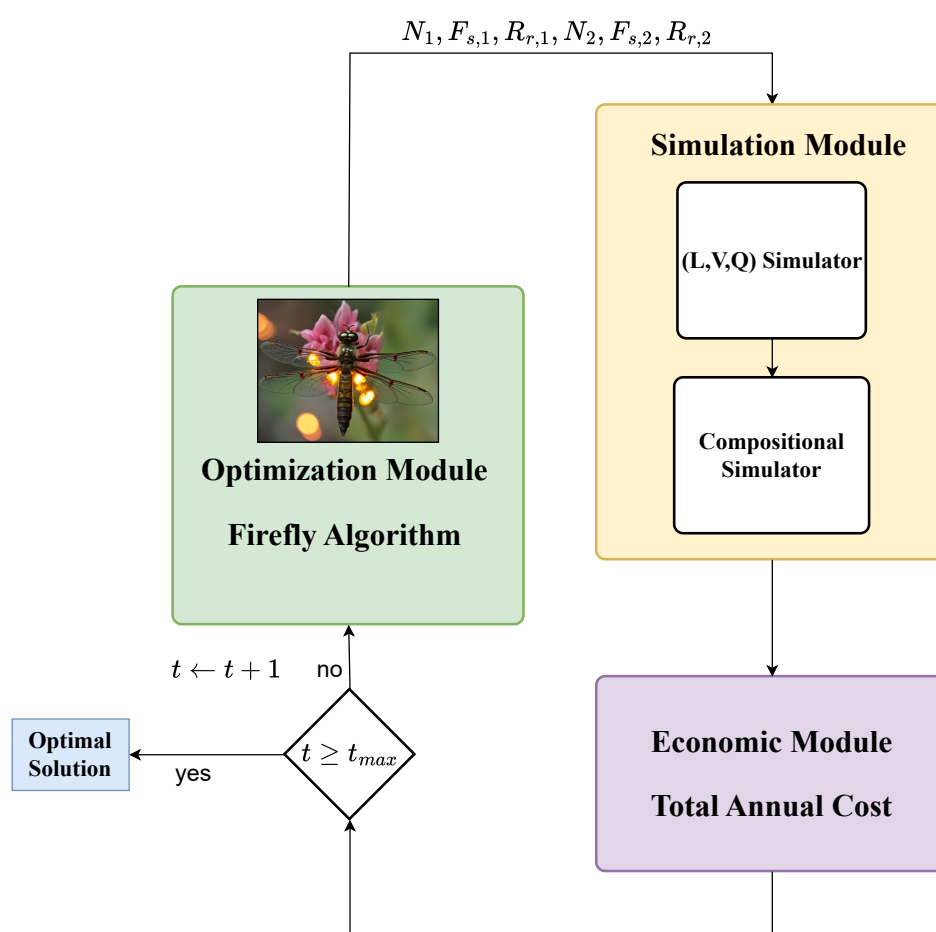


Figure 2. Block diagram of the proposed framework.

4. Results

In this section, we present the computational results for a process devoted to obtaining α - and β -pinene from turpentine. As pointed out previously, we considered only three components in turpentine (α -pinene, β -pinene and p-cymene). Considering a greater number of components (many of them in very low concentrations) implies a significant increase in the computational cost without adding important differences in the process design. The objective was to obtain α - and β -pinene with minimum purities of 90% (in molar quantities) in the top of the two distillation columns with a minimum TAC.

We considered the following parameters for the process:

- Turpentine flow rate (saturated liquid): 100 kmol/h;
- Turpentine molar composition (z): $z = [0.45 \ 0.45 \ 0.10]^t$;
- Bottom flow rate (first column): 52 kmol/h;
- Bottom flow rate (second column): 40 kmol/h;
- Operation pressure: 60 kPa;
- Minimum purity— α -pinene: 90% (molar);
- Minimum purity— β -pinene: 90% (molar).

It should be made clear that these parameters were simulation specifications (in this study) and were not subject to optimization. Any changes to these specifications will produce different design configurations.

4.1. Firefly Configuration/Calibration

The first step to obtain results was to determine the Firefly control parameters. As pointed out by Joy et al. [44], adjusting the parameters of the Firefly algorithm can be performed in different ways, including Monte Carlo simulations, the Latin hypercube method, manual adjustment and empirical adjustment. Considering that the aim was to avoid configurations that did not meet the design specifications (top α - and β -pinene purities in the columns), the algorithm's control parameters were adjusted in such a way as to produce "neighbors" of the initial estimates, avoiding large "jumps". Therefore, an empirical adjustment of control parameters was used. The variables involved in the problem had similar dimensions (number of stages, feed stage and reflux ratios) so that the values of the elements of β_0 and α were similar. The number of elements in the population (m , the number of fireflies) and the number of generations were essentially limited by the computing time, mainly considering that each element of the population represented an entire simulation of the process. As pointed out previously, the number of fireflies is a critical parameter considering the time complexity of the algorithm. In this sense, we opted for a small quantity of fireflies—five elements in the population—also considering that the simulation step involved solving high-dimensional non-linear algebraic systems (the compositional problem).

The number of generations was another important control parameter in the algorithm. Figure 3 illustrates a typical convergence pattern for a run with 50 generations. It was clear that after 30 generations, the reductions in the objective function were less significant (nearly marginal). The number of 30 generations was therefore adopted.

The Firefly algorithm, like all other metaheuristics, should not be run only once. In this case, the method was run 20 times in order to obtain a minimum statistical relevance. The parameters used were as follows:

- Number of runs: 20;
- Number of fireflies (m): 5;
- Number of generations (t_{max}): 30;
- $\gamma = 0.02$ (light absorption coefficient);
- $\beta_0 = [2 \ 2 \ 1 \ 2 \ 2 \ 1]^t$ (attractiveness for two collapsing fireflies);
- $\alpha^0 = [1 \ 1 \ 1 \ 1 \ 1 \ 1]^t$ (randomization parameter vector).

We also considered random initial estimates in the vicinity of a "feasible" (but non-optimized) design in order to avoid a situation where all fireflies represented a non-feasible condition with respect to the minimum purities. The ranges used in the initial estimates are presented in Table 1.

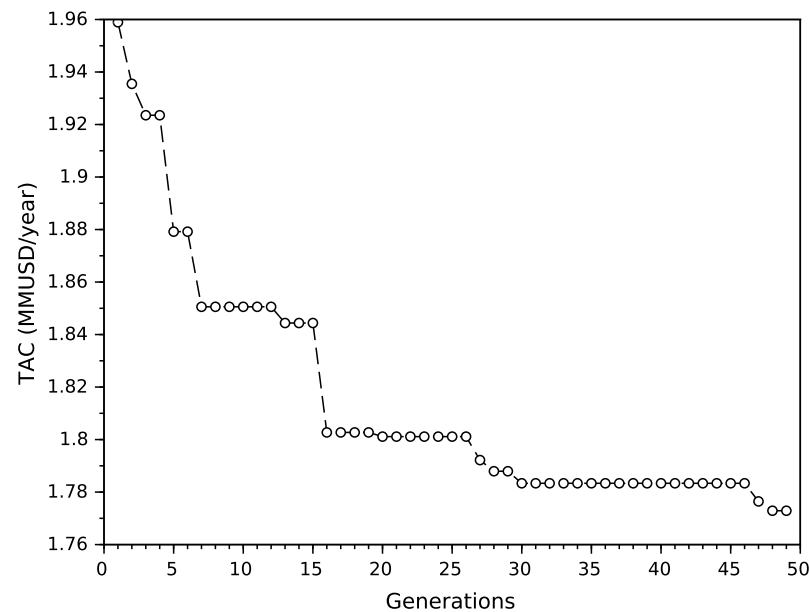


Figure 3. Typical convergence pattern from 50 generations.

Table 1. Initial estimates (uniform distribution).

	Ne_1	$F_{s,1}$	$R_{r,1}$	Ne_2	$F_{s,2}$	$R_{r,1}$
Interval	[45, 55]	[20, 30]	[5, 7]	[55, 65]	[30, 40]	[4, 6]

4.2. Computational Results

Table 2 contains the computational results for 20 independent runs.

Table 2. Computational results (20 independent runs).

	Mean	Standard Deviation	Best Solution	Worst Solution
Objective function (TAC) (USD/year)	1,782,095.3	8930.4	1,765,566.6	1,804,858.8

We also conducted a Shapiro–Wilk test [45] to verify the normality of the results obtained in the 20 runs of the algorithm. The Shapiro–Wilk test was appropriate given the small number of elements in the sample. We obtained $W = 0.899$ (the Shapiro–Wilk statistic) and the p -value = 0.039, which indicates that there was a significant deviation from the normal distribution (since p -value ≤ 0.05).

Figure 4 presents a typical convergence pattern considering the swarm of fireflies. For reasons of graphic visualization, the maximum value of the objective function was set to 2.2×10^6 USD. We can observe that some elements of the population (typically firefly 4 and firefly 5) reached the maximum value of the objective function several times, which corresponded to projects that did not meet the minimum purity criteria for α - and β -pinene at the tops of the columns and/or when there were convergence problems in the algorithm. This situation was desirable during the convergence process, as it denoted the diversification characteristic of the Firefly method.

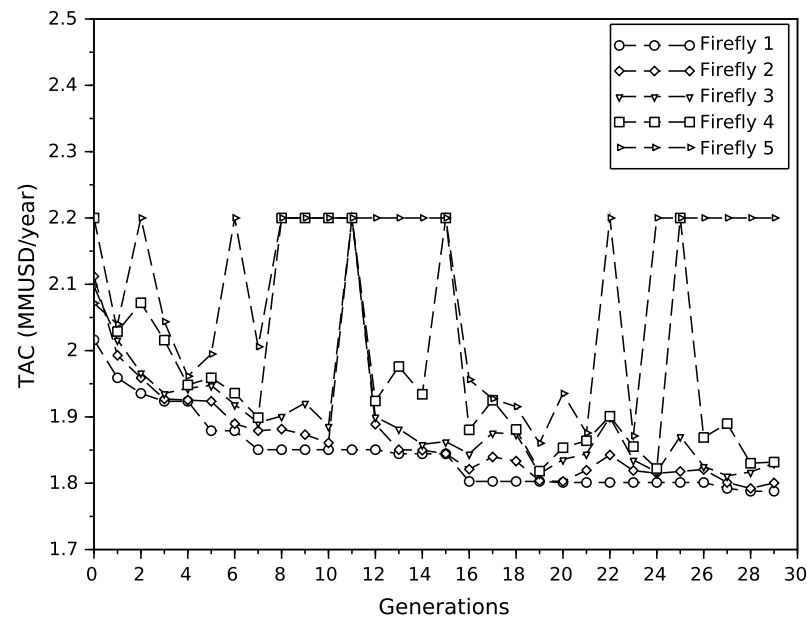


Figure 4. Convergence patterns (objective functions) for a typical swarm of fireflies.

A three-dimensional projection of the same convergence trajectory is shown in Figure 5. The dots in green represent higher values of the objective function, whereas those in dark blue indicate lower values. Once again, it can be seen that fireflies 4 and 5 were far from those with the lowest values of the objective function, which once again pointed to a diversification characteristic. The axes in Figure 5 refer to the parameters of the first column (N_1 is the number of stages, $F_{s,1}$ is the feed stage and $R_{r,1}$ represents the reflux ratio).

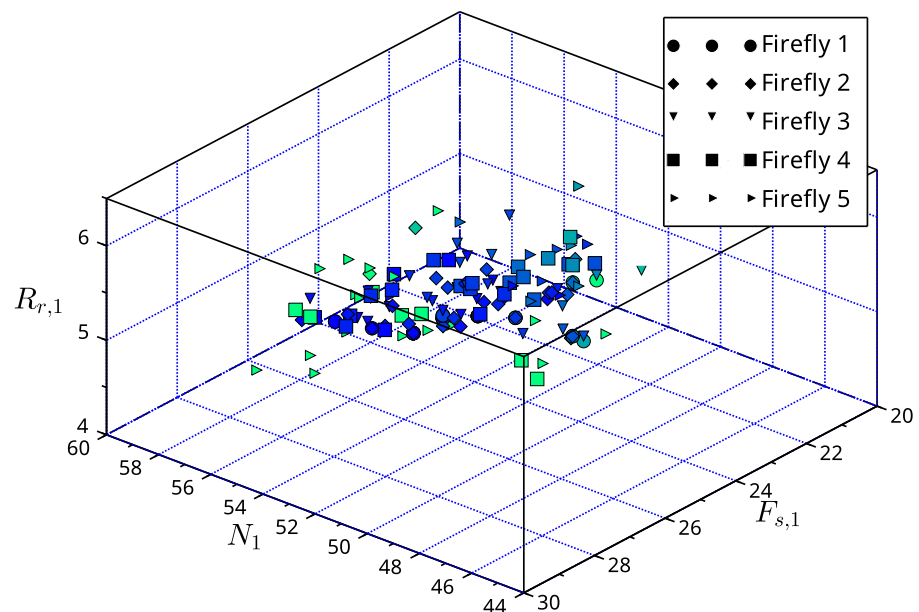


Figure 5. Projection in the space $N_1 \times F_{s,1} \times R_{r,1}$ of the convergence patterns for a typical swarm of fireflies.

Figure 6 shows a comparison in percentage terms of the costs involved in capital investment and operating costs for the best and worst solutions found. It can be seen that the percentages of operating costs and investment capital were close when comparing the best and worst solutions found.

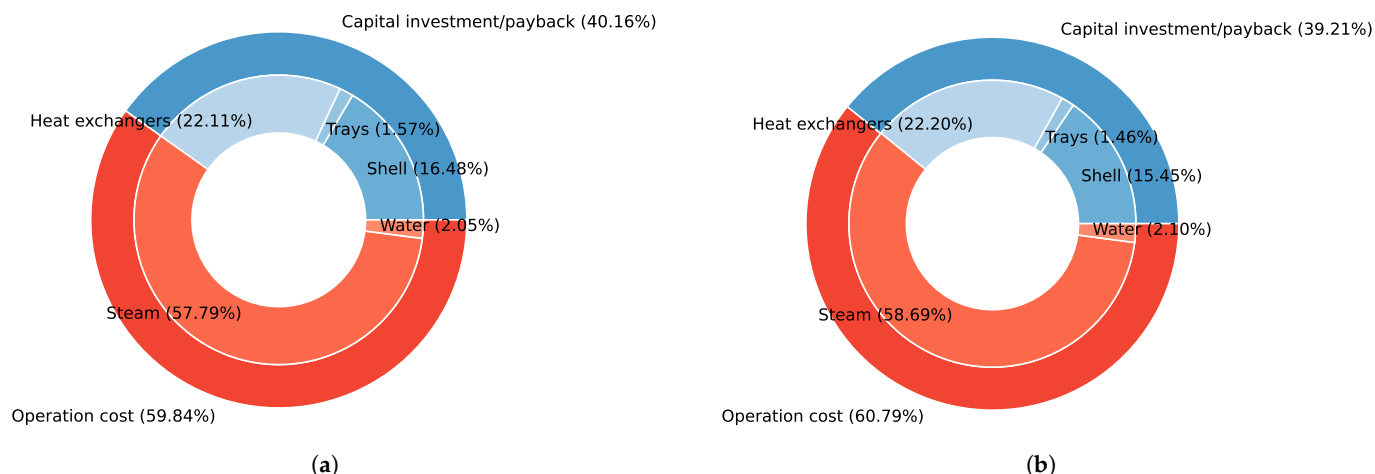


Figure 6. Donut charts for capital investment and operation cost elements: (a) best solution; (b) worst solution.

The convergence patterns for the best and worst solutions are depicted in Figure 7. It can be seen that both the best and worst solutions started from high objective function values when compared with the final solutions.

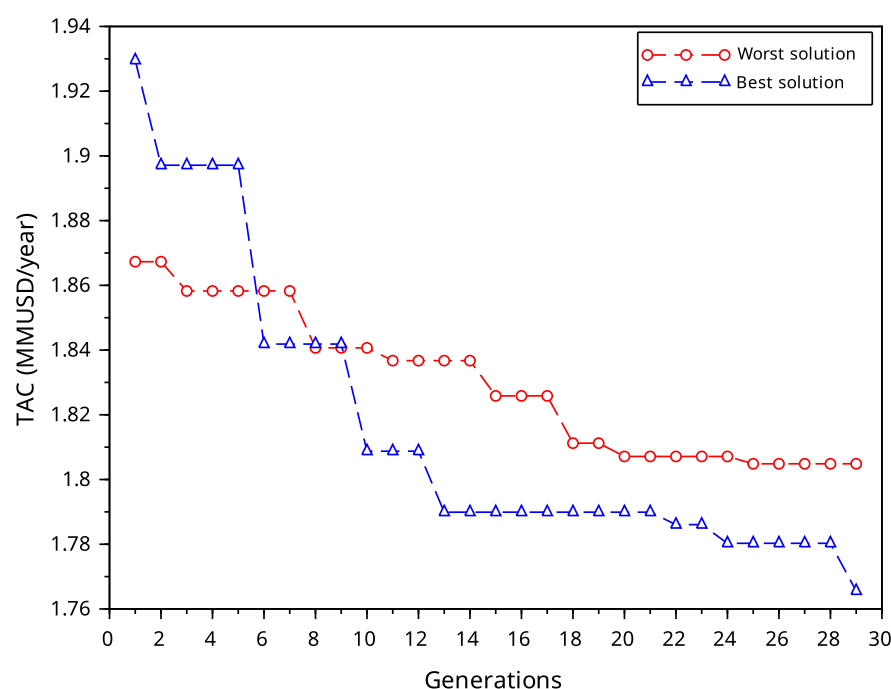


Figure 7. Convergence patterns for the best and worst solutions.

The results for each column, corresponding to the best solution found, are detailed in Figure 8. We can observe that the constraints regarding the purities of α - and β -pinene at the top of the two distillation columns were satisfied. In the first column (Figure 8a), practically all the α -pinene was removed at the top of the equipment. The temperature profile in the first column is presented in Figure 8b. There was a considerable difference in temperature between the top and bottom of the column, which was consistent with the large difference in composition between the ends of the equipment. Figure 8c illustrates the composition profile for the second column. In this case, the differences in composition between the top and bottom were less pronounced compared with the first column, mainly

because there were reduced amounts of α -pinene. Consistently, the temperature differences were narrow (less than 2 °C), as indicated by Figure 8d.

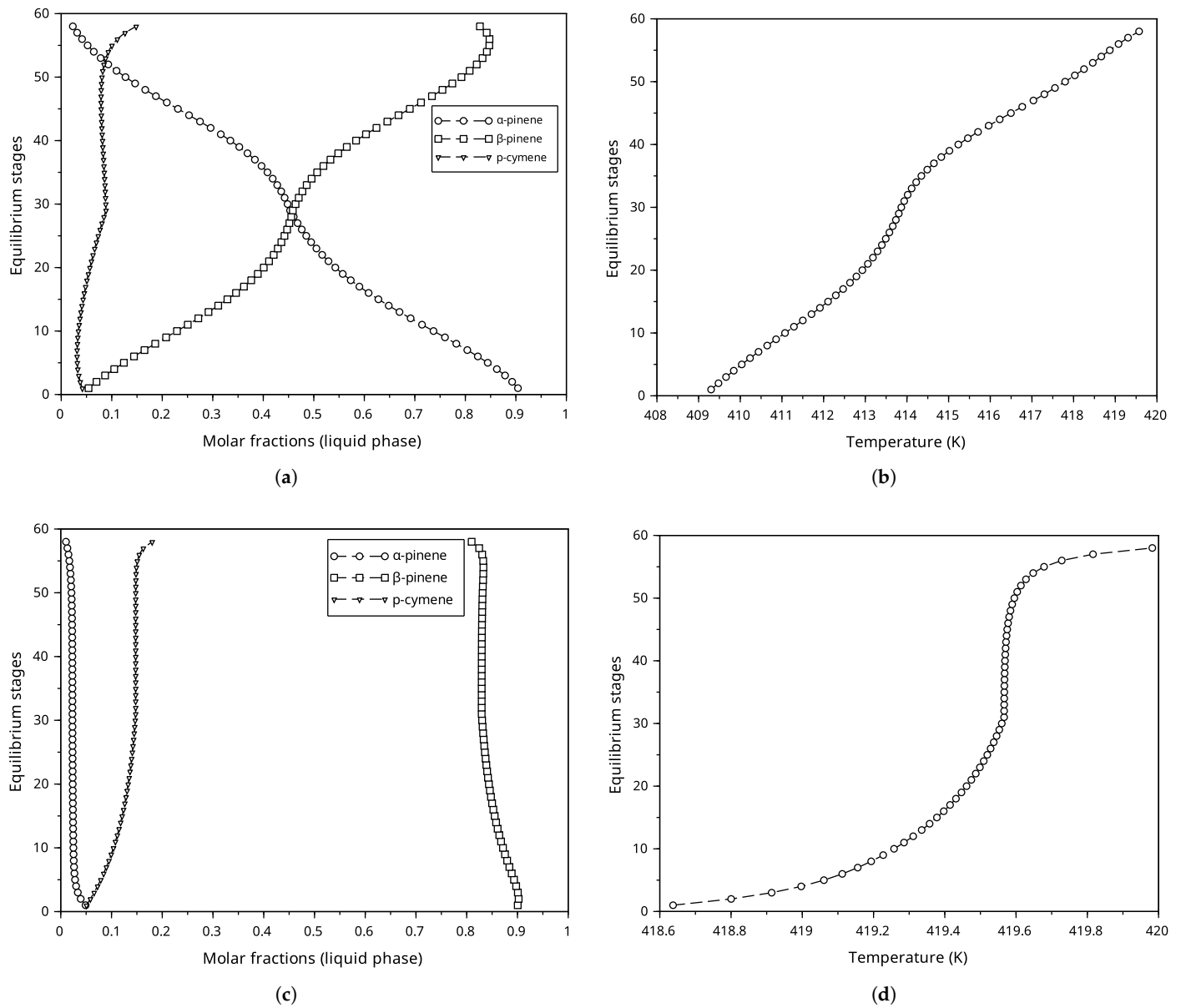


Figure 8. Composition and temperature profiles for the optimal design: (a) composition profile—column 1; (b) temperature profile—column 1; (c) composition profile—column 2; (d) temperature profile—column 2.

The final flowsheet, with details of each of equipment (number of stages and feed stages) and energy consumption, is shown in Figure 9.

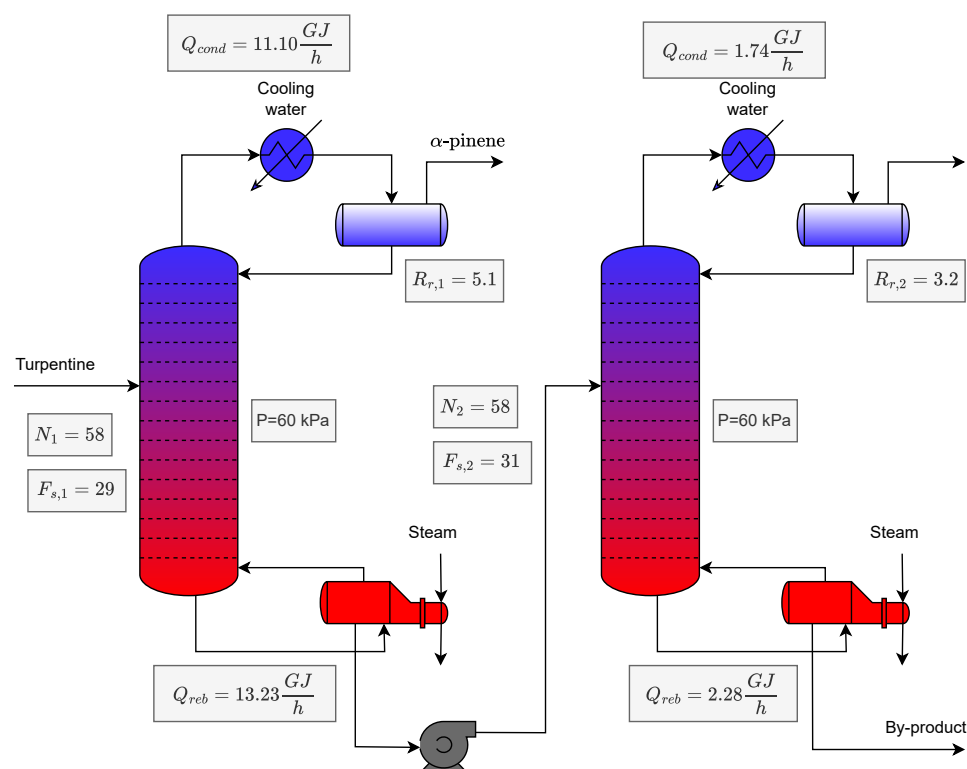


Figure 9. Final flowsheet—optimal design.

5. Discussion

The results obtained show that even with a small number of elements in the population (i.e., a small initial set of design configurations), the optimization strategy promoted a significant reduction in the total annualized cost in comparison with the initial estimates. Since there was a compromise between the size of the equipment (number of equilibrium stages) and the operating costs (in fact, several authors have treated this problem as a multi-objective optimization [18]), a diversity of solutions was to be expected, even with the evolution of the algorithm's generations. Another point to be highlighted is that Firefly's characteristic is a certain diversification of the population (even at the end of the process) [24], unlike what occurs, for example, with the Differential Evolution algorithm (where all the elements of the population tend to collapse, reducing the diversity [46]). As already mentioned, one strategy for reducing population diversity would be to increase the number of generations—since $\alpha^{t+1} \leftarrow \zeta \alpha^t$, with $\zeta = 0.997$, the algorithm lost its diversification characteristic and increased its intensification at the end of the process—but such a measure significantly increases the computational cost with marginal reductions in the objective function. Another possibility would be to reduce the parameter ζ in order to reduce the portion corresponding to the random movement of the Fireflies more quickly. On the other hand, such a strategy can cause the population to become trapped in regions of local minima. Thus, as with metaheuristics, there is a difficult trade-off between diversification and intensification [47]. Because the simulation stage is computationally costly, this balance becomes even more complicated to achieve.

In addition, the difference between the objective function value for the best and worst solution is around 2% of the TAC value, indicating that the algorithm has found a promising region in terms of minimization. Naturally, the best solution found can be included as one of the candidates in subsequent optimization strategies in order to refine the results found. It should also be noted that the optimization problem in question could be described in a multi-objective format, with a conflict between minimizing the capital investment while

minimizing the operating costs. In such a situation, a Pareto curve would represent the set of nondominated solutions to the problem and, ultimately, the project engineer would be able to determine the number of stages of the equipment and determine the reflux ratios from them.

With regard to the performance of the proposed framework, a comparison between open-source and proprietary tools in this context would be complicated for a number of reasons (although it is a pertinent question). Scilab is an interpreted programming language (typical of prototyping), while commercial process simulators are implemented in other programming languages. Furthermore, the connection between an optimization tool (typically a metaheuristic) and a commercial simulation code is also a non-trivial task. Therefore, we considered that performance comparisons between this proposal and other approaches were beyond the scope of this work.

6. Conclusions

In this paper, we present a simulation and optimization methodology for the optimal design of distillation separation systems based on the Firefly algorithm. The developed framework was applied in a process to obtain α -pinene and β -pinene from turpentine, a raw material of natural origin.

The results indicate that the application of the methodology was effective at obtaining a process flowsheet that was able to obtain α - and β -pinene with the purities established while minimizing the total annualized cost.

Author Contributions: Conceptualization, G.M.P. and O.K.d.A.; methodology, G.M.P., O.K.d.A. and F.B.S.O.; software, G.M.P., O.K.d.A. and F.B.S.O.; validation, G.M.P., O.K.d.A. and F.B.S.O.; formal analysis, G.M.P., O.K.d.A. and F.B.S.O.; investigation, G.M.P., O.K.d.A. and F.B.S.O.; resources, G.M.P. and F.B.S.O.; data curation, G.M.P., O.K.d.A. and F.B.S.O.; writing—original draft preparation, G.M.P., O.K.d.A. and F.B.S.O.; writing—review and editing, G.M.P., O.K.d.A. and F.B.S.O.; visualization, G.M.P., O.K.d.A. and F.B.S.O.; supervision, G.M.P. and F.B.S.O.; project administration, G.M.P. and F.B.S.O.; funding acquisition, G.M.P. and F.B.S.O. All authors have read and agreed to the published version of this manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. O'Brien, A.; Zhang, H.; Allwood, D.M.; Rawsthorne, A. From Data to Draught: Modelling and Predicting Mixed-Culture Beer Fermentation Dynamics Using Autoregressive Recurrent Neural Networks. *Modelling* **2024**, *5*, 201–222. [\[CrossRef\]](#)
2. Grznár, P.; Gregor, M.; Krajčovič, M.; Mozol, S.; Schickerle, M.; Vavřík, V.; Ďurica, L.; Marschall, M.; Bielik, T. Modeling and Simulation of Processes in a Factory of the Future. *Appl. Sci.* **2020**, *10*, 4503. [\[CrossRef\]](#)
3. Hubach, T.; Pillath, M.; Knaup, C.; Schlüter, S.; Held, C. Li⁺ Separation from Multi-Ionic Mixtures by Nanofiltration Membranes: Experiments and Modeling. *Modelling* **2023**, *4*, 408–425. [\[CrossRef\]](#)
4. Lazarev, S.; Protasov, D.; Konovalov, D.; Khorokhorina, I.; Abonosimov, O. Modeling of the Nanofiltration Process Based on Convective Diffusion Theory. *Modelling* **2024**, *5*, 1729–1744. [\[CrossRef\]](#)
5. Platt, G.M.; Yang, X.S.; Silva Neto, A.J. *Computational Intelligence, Optimization and Inverse Problems with Applications in Engineering*, 1st ed.; Springer Publishing Company, Incorporated: Berlin/Heidelberg, Germany, 2019. [\[CrossRef\]](#)
6. Cimino, A.; Gnani, M.G.; Longo, F.; Barone, G.; Fedele, M.; Piane, D.L. Modeling & Simulation as Industry 4.0 enabling technology to support manufacturing process design: A real industrial application. *Procedia Comput. Sci.* **2023**, *217*, 1877–1886. [\[CrossRef\]](#)
7. Salha, G.B.; Abderrabba, M.; Labidi, J. A status review of terpenes and their separation methods. *Rev. Chem. Eng.* **2021**, *37*, 433–447. [\[CrossRef\]](#)
8. Valencia, L.F.; Arias-Velandia, J.; Barrera, R.; Villa, A.L. Tailoring turpentine for fine chemicals: Experimentation and simulation of fractional vacuum distillation for selective α - and β -pinene enrichment. *Braz. J. Chem. Eng.* **2024**. [\[CrossRef\]](#)

9. Ariono, D.; Aldiyana, G.; Adrian, R.; Indarto, A. Fractionation of Turpentine. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *742*, 012029. [CrossRef]
10. Rubini, M.; Clopeau, A.; Sandak, J.; Dumarcay, S.; Sandak, A.; Gerardin, P.; Charrier, B. Characterization and classification of Pinus oleoresin samples according to Pinus species, tapping method, and geographical origin based on chemical composition and chemometrics. *Biocatal. Agric. Biotechnol.* **2022**, *42*, 102340. [CrossRef]
11. Available online: https://assets.noviams.com/novi-file-uploads/pcai/PCA-_Global_Impact_of_the_Modern_Pine_Chemical_Industry_-Study_Final-adffb0c6.pdf (accessed on 17 April 2025).
12. Pastl, C. Income Extracted from the Forest. 2021. Available online: <https://www.correiodopovo.com.br/especial/renda-extra%C3%ADda-da-floresta-1.673931> (accessed on 17 April 2025). (In Portuguese)
13. Tian, C.; Zhang, X.; Lan, Y.; Sun, J. An Enhanced Logic-Based Bender's Decomposition Algorithm with Proximity Principle for Simulator-Based Distillation Process Optimization. *Processes* **2025**, *13*, 977. [CrossRef]
14. Murrieta-Dueñas, R.; Cortez-González, J.; Segovia-Hernández, J.; Hernández-Aguirre, A.; Gutiérrez-Guerra, R.; Hernández, S. A comparative analysis of differential evolution and Boltzmann-based distribution algorithms with constraint handling techniques for distillation process optimization. *Chem. Eng. Res. Des.* **2025**, *214*, 39–53. [CrossRef]
15. Javaloyes-Antón, J.; Ruiz-Femenia, R.; Caballero, J.A. Rigorous Design of Complex Distillation Columns Using Process Simulators and the Particle Swarm Optimization Algorithm. *Ind. Eng. Chem. Res.* **2013**, *52*, 15621–15634. [CrossRef]
16. Li, M.; Cui, Y.; Shi, X.; Zhang, Z.; Zhao, X.; Zhu, X.; Gao, J. Simulated annealing-based optimal design of energy efficient ternary extractive dividing wall distillation process for separating benzene-isopropanol-water mixtures. *Chin. J. Chem. Eng.* **2021**, *33*, 203–210. [CrossRef]
17. Hu, Z.; Li, P.; Liu, Y. Enhancing the Performance of Evolutionary Algorithm by Differential Evolution for Optimizing Distillation Sequence. *Molecules* **2022**, *27*, 3802. [CrossRef]
18. Chia, D.N.; Duanmu, F.; Sorensen, E. Single- and multi-objective optimisation of hybrid distillation-pervaporation and dividing wall column structures. *Chem. Eng. Res. Des.* **2023**, *194*, 280–305. [CrossRef]
19. Chen, W.; Hu, Z.; Gao, X.; Liu, Y. Simulation and Multi-Objective Optimization of Three-Column Double-Effect Methanol Distillation by NSGA-III Algorithm. *Processes* **2023**, *11*, 1515. [CrossRef]
20. Becerra, J.A.; Villa, A.L. Techno-Economic Evaluation of d-Limonene and α -Pinene Separation from Citrus and Turpentine Oils. *Chem. Eng. Technol.* **2020**, *43*, 2295–2306. [CrossRef]
21. Chaliar, P.; Martinez-Lopez, B.; Lacour, M.A.; Rigou, P. Extraction of turpentine essential oil from *Pinus Pinaster* Ait: Comp. Yield Compos. Conv.- Microw. Assist.-Hydro Vac. Distill. *Sustain. Chem. Pharm.* **2024**, *41*, 101702. [CrossRef]
22. Yang, X.S. Firefly Algorithms for Multimodal Optimization. In *Proceedings of the Stochastic Algorithms: Foundations and Applications*; Watanabe, O., Zeugmann, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 169–178.
23. dos Santos Coelho, L.; Mariani, V.C. Firefly algorithm approach based on chaotic Tinkerbell map applied to multivariable PID controller tuning. *Comput. Math. Appl.* **2012**, *64*, 2371–2382. [CrossRef]
24. Yang, X.S.; He, X.S., Why the Firefly Algorithm Works? In *Nature-Inspired Algorithms and Applied Optimization*; Yang, X.S., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 245–259. [CrossRef]
25. Smith, R. *Chemical Process Design and Integration*; John Wiley & Sons Ltd: London, UK, 2005.
26. Bernardo-Gil, M.; Barreiros, A. Vapour - liquid equilibria of α -pinene + β -pinene at 53.3 and 80kPa. *Fluid Phase Equilibria* **1994**, *100*, 283–291. [CrossRef]
27. Bernardo-Gil, M.; Ribeiro, M.A. Vapor-liquid equilibria of binary mixtures of β -pinene with limonene and P-cymene at atmospheric pressure. *Fluid Phase Equilibria* **1993**, *85*, 153–160. [CrossRef]
28. Bernardo-Gil, M.; Ribeiro, M.A. Vapor-liquid equilibria of binary system based on pine resin. *Fluid Phase Equilibria* **1989**, *53*, 15–22. [CrossRef]
29. Available online: <http://www.stenutz.eu/chem/solv6.php?name=%28%2B%29-a-pinene> (accessed on 17 April 2025).
30. Available online: [http://www.stenutz.eu/chem/solv6%20\(2\).php?name=%28%2B%29-b-pinene](http://www.stenutz.eu/chem/solv6%20(2).php?name=%28%2B%29-b-pinene) (accessed on 17 April 2025).
31. Available online: https://www.stenutz.eu/chem/solv6.php?name=p-cymene#google_vignette (accessed on 17 April 2025).
32. Xuwu, A.; Riheng, H.; Hu, W.; Mali, W.; Yongkuang, Z. Enthalpies of Vaporization of α - and β -pinene. *Acta Phys.-Chim. Sin.* **1987**, *3*, 668–671. [CrossRef]
33. Mathew, T.J.; Tawarmalani, M.; Agrawal, R. Relaxing the constant molar overflow assumption in distillation optimization. *AIChE J.* **2023**, *69*, e18125. [CrossRef]
34. Platt, G.M.; Medeiros, J.L. Phase rule calculations and the thermodynamics of reactive systems under chemical equilibrium. *Braz. J. Chem. Eng.* **1999**, *16*, 247–265. [CrossRef]
35. Douglas, J. *Conceptual Design of Chemical Processes*; Chemical Engineering Series; McGraw-Hill: New York, NY, USA, 1988.
36. Available online: <https://www.netlib.org/lapack/> (accessed on 16 April 2025).
37. Platt, G.M.; Lobato, F.S.; Libotte, G.B.; Moura Neto, F.D. Basins of attraction and critical curves for Newton-type methods in a phase equilibrium problem. *Int. J. Comput. Sci. Eng.* **2020**, *23*, 91. [CrossRef]

38. Platt, G.M.; Domingos, R.P.; Andrade, M.O. Application of the Firefly and Luus–Jaakola algorithms in the calculation of a double reactive azeotrope. *Comput. Sci. Discov.* **2014**, *7*, 015002. [\[CrossRef\]](#)
39. Zhang, L.; Liu, L.; Yang, X.S.; Dai, Y. A Novel Hybrid Firefly Algorithm for Global Optimization. *PLoS ONE* **2016**, *11*, e0163230. [\[CrossRef\]](#)
40. Wang, H.; Wang, W.; Zhou, X.; Sun, H.; Zhao, J.; Yu, X.; Cui, Z. Firefly algorithm with neighborhood attraction. *Inf. Sci.* **2017**, *382–383*, 374–387. [\[CrossRef\]](#)
41. Bei, J.L.; Zhang, M.X.; Wang, J.Q.; Song, H.H.; Zhang, H.Y. Improved Hybrid Firefly Algorithm with Probability Attraction Model. *Mathematics* **2023**, *11*, 389. [\[CrossRef\]](#)
42. Cheng, Z.; Song, H.; Wang, J.; Zhang, H.; Chang, T.; Zhang, M. Hybrid firefly algorithm with grouping attraction for constrained optimization problem. *Knowl.-Based Syst.* **2021**, *220*, 106937. [\[CrossRef\]](#)
43. Astrachan, O. Bubble sort: An archaeological algorithmic analysis. *ACM SIGCSE Bull.* **2003**, *35*, 1–5. [\[CrossRef\]](#)
44. Joy, G.; Huyck, C.; Yang, X.S. Parameter tuning of the firefly algorithm by three tuning methods: Standard Monte Carlo, quasi-Monte Carlo and latin hypercube sampling methods. *J. Comput. Sci.* **2025**, *87*, 102588. [\[CrossRef\]](#)
45. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591. [\[CrossRef\]](#)
46. Opara, K.R.; Arabas, J. Differential Evolution: A survey of theoretical analyses. *Swarm Evol. Comput.* **2019**, *44*, 546–558. [\[CrossRef\]](#)
47. Yang, X.S.; Deb, S.; Fong, S. Metaheuristic Algorithms: Optimal Balance of Intensification and Diversification. *Appl. Math. Inf. Sci.* **2014**, *8*, 977–983. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.