International Journal of Sciences
(ISSN 2305-3925)
http://www.ijSciences.com

Research Article
Volume 2, Issue Mar 2013

Mathematical Model and Software for Multiple Location of Co-Generation Energy Plants: A Case Study in Southern Regions of Chile

Juan Alfredo Gómez[1], Juan Pablo Concha[2], Bruno Neumann[3]

[1]Center of Modeling and Scientific Computation, Mathematical Engineering Department, Universidad de La Frontera, Francisco Salazar 1145, Temuco, Chile
[2]jconcha@ufro.cl
[3]Systems Engineering Department, Universidad de La Frontera, Francisco Salazar 1145, Temuco, Chile

Abstract: This paper addresses the issue of locating and determining the area of supply of various cogeneration energy plants, based on forest biomass. Two models are proposed, binary and mixed programming, depending on whether or not to allow the intersection between supply areas. Each model computes the ideal places to install biomass plants and their respective supply areas of raw material, since the latter are handled implicitly by the decision variables. We also propose two solution strategies depending on the size of the problem: Branch and cut algorithm for problems of medium size and heuristics associated with a genetic algorithm for large problems. We also develop software that automates the construction of the appropriate model, based on information provided by the user, delivering the optimal locations together with supply areas for the number and type of plants desired.

Keywords: Cogeneration energy, Biomass forestry, Multiple Location, Mixed and Binary Programming, Genetic Algorithm

Introduction

Alternative management options for energy production, which is also sustainable, are the combined heat and power (CHP) plants based on forest biomass. Plants based on biomass forestry transform this material into heat energy and then electricity. It is imperative optimality in the location of each plant and in the identification of resource areas, defined as part of the region where the plant extracts the needed biomass. The optimality in the location of plants is strongly linked to the costs of sale, processing, and transportation of biomass within the region.

We develop an automatic methodology to find a combination of locations, feasible in the region, for installation of biomass plants, whether or not having common supply areas, and which are profitable for the investors. The model we propose assumes as data only all possible locations and select at the same time the optimal location and the corresponding supply areas. The idea was inspired by the classical models "Uncapacitated Facility Location" (UFL) and "Set Covering", (Wolsey, 1998).

The structure of this paper is as follows: Section 2 deals with background research, addressing the problem description, the detailed of what we call strong and weak requirements, and goals. Section 3 explains the binary and mixed programming models. Section 4 discusses the implementation of the model, which takes place in two ways: with a mixed programming solver and a heuristic approach. In Section 5, we describe the heuristic method for the problem and implementation of binary mixed model GUROBI Section 6 discusses some of the computational results obtained in experiments with real data. Finally, Section 7 presents the conclusions of this study.

State Of Art

Since 1988, the model "facility location" was introduced, which focuses on the location of plants using an abstract mathematical modeling concept. (Guerrero and Carrazo, 1998) show an application of the model to the installation and expansion of biomass plants based on olive, which imposes as data the supply area of a plant. In the present study we also used the "facility location" type model, but it is addressed by assuming a non determined supply area and allowing the intersection of the maximal areas of supply for the plants, that is, competition between them. The objective function of (Guerrero and Carrazo, 1998) is a facility location type without constraints. Our model provides an alternative objective function, where an extraction area exceeding 25 km is considered and where it is proportional to the transportation costs of each cell of biomass. This is determined by the Euclidean distance from the cell to the plant. A different treatment of plant location, when there is competition for resources, can

Juan Alfredo Gómez (Correspondence)
jagomez@ufro.cl OR **jagomezfdez@gmail.com**

also be seen in the work of (Panichelli and Gnansounou, 2008).

The application of satellite technology (GIS) is an innovation in the estimation of biomass availability as reflected in the work of (Xun Shi and Elmore, 2008) Bernetti et al. (2004), (Krukanont and Prasertsan, 2004), and (Pontt, 2005). We find a large number of publications that use the facility location model to determine the location of a single plant, for example, Esteban et al. (2000),Graham et al. (2000) and Voivontas et al. (2001), but was hardly worked on the location of various plants. The location of biomass plants based on wood waste resembles the problem of locating forest plants (Troncoso and Garrido, 2001) and our approach is also useful for this problem.

Research Background

Problem Description

To determine the optimal plant locations we will assume that there is enough information about the region, by means of geographical and/or statistical studies, geo-referenced using satellite technology and other means. The information is stored in data files and each of these files we shall call "information layer". Therefore, in what follows we shall assume that the total region is divided into elements of approximately equal area (eg. 10 km^2 rectangles) and refer to each element of this division as a "cell". All layers or data files have the same size and each contain a particular information in all cells. The different data files, which is assumed available, are detailed below:

- **Biomass**: A file containing an estimate of the usable amount of biomass in each cell in the region.
- **Biomass type:** Indicating the tree species found in each cell, classifying it by heat capacity.
- **Roads:** Indicating whether each cell contains a (part of) road of total region.
- **Distance to nearest road:** Indicating how far away each cell is from the nearest road.
- **Distance to nearest electric connection:** Indicates how far each cell is from the nearest connection point to the electric network.
- **Elevation:** Shows the elevation average on each of the cells.
- **Constraints:** Lists the prohibition of locating a plant or using biomass in the cells of the region. These sectors may be associated with rivers or lakes, indigenous communities, protected forests and national parks, among others.

The candidate sites for locating biomass plants are assumed explicitly designated by the planner. They are scattered around the region and are associated with each cell, thus constituting a new layer:

- **Feasible locations**: Cells shall be encoded, identifying those potential sites of localization.

Obviously these points are subject to restrictions imposed by the layer of constraints. The candidate sites correspond to particular requirements of the planner. This provides several scenarios that can be formulated. For example, choosing only locations:

- Contained in a certain part of the region,
- Close to roads and highways,
- Near industries or towns,
- Near electrical connections
- With slope less than X degrees

For the formulation of mixed programming model, we need a new concept. So far we have referred to the "Supply Area" of a plant (denoted SA) and now we specify its definition as a set of cells in the biomass layer which serve the necessary biomass to operate the plant during the study period. The "maximum supply area" (MSA) shall be defined as the set of cells in the layer of biomass, where the transportation to the plant of the amount of biomass available is economically viable. One way to calculate the MSA is to establish a maximum radius r, i.e. the maximum length allowed in the transportation of biomass from cells to the plant. Note that the feasible location MSA are calculated a priory and they are considered data for the model. The SA of each plant will almost always be a subset of the MSA, but they are unknown and will be determined by solving the model.

The problem of establishing the SA of each plant is dictated by two approaches: the first allows the intersection of the SA of each plant and the second strictly prohibits the intersection of these. This is why we develop two models.

User Requirements

The following describes the requirements imposed to the problem of choosing optimal sites and supply area for each plant. The requirements are categorized into two groups: strong and soft. Strong requirements must be met mandatory and soft requirements are managed externally by the programmer and represent desirable conditions, but not mandatory for the choice of installation sites.

Strong Requirements

1. The amount of biomass available in the SA should be enough to keep operating the plant (of any kind) during the entire period.
2. The SA of two or more plants may (or may not) intersect, but the biomass removed as a whole must not exceed the amount available in each cell. This particular restriction gives rise to two models, one that tolerates other intersections and another one not allowing it.
3. In a cell of the feasible locations layer can be installed only one particular type of plant size and technology.

4. It is necessary to install at least a minimum number and at most, a maximum number of plants of each type.

The choice of places to install the plants would be done:

1. trying to minimize transportation costs,
2. giving preference to locations near concentrations of biomass,
3. priorizing the locations near consumer communities,
4. looking for sites closer to the electrical interconnection system,
5. trying to be located on the sides of roads,
6. causing the least harm to the environment,
7. looking for an investment with lower taxes.

Note that weak requirements go side by side with the objectives of the investors, and they cannot be met simultaneously. So in practice they must choose, add and / or remove soft requirements. The soft requirements are involved in the definition of the objective function, which is constructed by calculating the Net Present Value (NPV) for a given set of locations. This economic indicator is detailed below and the optimal solution (maximum) obtained using it, produces (theoretically) the most profitable investment to carry out the project with the resulting locations. The next section details the models proposed to address the described problem.

Model Description

Objective Function

To establish a model that reflects reality as best as possible, we introduce as objective function the economic indicator NPV (Net Present Value). This is a measurement tool to estimate how profitable is to install a plant in any given location. To calculate the NPV, information on its main components is required, which are described in detail below:

Initial Investment

- *Investment in plant technology*.
- *Storage Investment.*
- *Investment in electricity network connection*.
- *Investment thermal network connection*.

Revenues

Revenues can be classified into:

- *Sales of electricity*.
- *Sales of thermal energy*.

Expenses

Expenses can be classified into:

- Costs relating to biomass, such as:
  Purchase cost of biomass.
  Cost of collection.
  Shipping cost.
  Cost of storage at the plant.
  Cost of processing.
  Cost of waste disposal.
- Expenses relating to the plant:
  Cost of maintenance of the plant,
  Cost of maintenance of the electric network,
  Cost of network maintenance of thermal grid.
  Cost of taxes on buildings.

Annual Cash Flow

With all the above elements for each cell of feasible locations layer, assuming that it has already been designated the site and associated supply area (SA) of a plant, we can calculate an annual cash flow (ACF) based on the usual equation:

$$ACF = R - E - D.$$

Here R is the sum of revenues, E the sum of the expenses and D is depreciation of all equipment (including transportation). The cash flow is calculated for all the predetermined time period (eg. 10 years) and then, subtracting the initial investment, the NPV is obtained for a plant. Note that we are not considering the problem of how to perform the annual exploitation of biomass and the criteria for optimal operation of the plant during the entire period of time. This requires another optimization model.

Definition of the Data

The model requires definitions of different types of data sets, which are assumed known and described in detail below, with their respective notations:

- Cells throughout the region are a finite number (the same for each layer), they will be numbered (with positive integers) consecutively and shall be represented by the index set Θ. We denote by N the number of elements of Θ.
- The vector B has dimension N and is given by the cells in the data file **Biomass**. Each element of vector B indicates the type and amount (in tons.) of available biomass existing in the cell $j \in \Theta$ of the region. The elements of the matrix B will be called "points of biomass" and denoted by $b_j,\ j = 1, 2, ..., N.$
- The set $I \subset \Theta$, consists of the indices of cells being possible locations of the plants in the region. Denote by N_I the number of elements of I. The indices $i \in I$ are therefore associated with the candidate sites to locate a plant.
- The set $J_i \subset \Theta$, consists of the indices of cells belonging to the maximum supply area (MSA_i), corresponding to the feasible plant location at position $i \in I$. Each of these sets depends on the manner in which the MSA is calculated for plants, for example, using the maximum radius r. Denote by N_{J_i} the number of elements of J_i.
- The vector G_i will be called vector of benefits

(for the position $i \in I$) and has dimension N_{J_i}. This vector is determined by the criteria that the developer has chosen to soft requirements, reflected in the objective function. It is obviously directly related to the coefficients of income and costs at which the NPV is calculated. The element $g_j^i \in G_i$ represents the profit (revenue - expense) obtained for each ton. of biomass extracted from the cell $j \in J_i$ to be processed in a plant located at position $i \in I$.

- The set $\Gamma_i \subset \Theta$, consists of the indices of the cells MSA_i of position $i \in I$, which also belong to MSA_k of other positions $k \neq i$. In other words, they are the indices of the cells belonging to J_i which also belongs to J_k with $k \in I,\ k \neq i$. We denote by N_{Γ_i} the number of elements Γ_i. Note that Γ_i can be empty and this means that each cell $j \in J_i$ belongs only to MSA_i. In symbols:
$\Gamma_i = \bigcup_{\substack{s \in I \\ s \neq i}} (J_i \cap J_s).$
- The index set T represents the types of plants, which are defined by existing technologies and the possible sizes of plants. Denote by N_T the number of elements of T.
- The vector E has size N_T and each component $E_k,\ k \in T$, denotes the amount of biomass required by a k-type plant to operate for the preassigned time.
- Each type of plant has a fixed cost for installation. This information is stored in an array C of size $N_I \times N_T$, whose elements are c_{ik} where $i \in I$ and $k \in T$.
- The minimum and maximum quantities of plants of type k, to be installed in the region will be represented by the symbols $\eta_{\min}^k$ and $\eta_{\max}^k$.

Definition of Variables

- y_{ik} is a binary variable indicating whether the plant with technology type $k \in T$ must be installed ($y_{ik} = 1$) or not ($y_{ik} = 0$) at position $i \in I$,
- x_{ij} is a real variable, which indicates what fraction of the available biomass in cell $j \in J_i$ will be used by the plant located at position $i \in I$.

The variable x_{ij} is bounded above by 1 and below by 0. Assuming that variables x_{ij} are real, they allow the intersection between the supply areas (SA_i) of different plants. If we assume that $x_{ij} \in \{0, 1\}$, i.e. if they are binary variables, this means that all the available biomass in cell j will be used ($x_{ij} = 1$) or not ($x_{ij} = 0$) for plant i, so that implicitly prohibits the intersection of SA_i.

Binary Model

Constraints (Strong Requirements)

Strong requirements must be included in the model as constraints. First establish the constraints for the binary programming model that solves the problem with banning intersections.

No installation of more than one plant in a candidate site for location:

$$\sum_{k \in T} y_{ik} \leq 1, \quad i \in I. \tag{1}$$

Satisfaction of the amount of biomass required by the plants at positions $i \in I$:

$$\sum_{j \in J_i} b_j x_{ij} \geq \sum_{k \in T} E_k y_{ik}, \quad i \in I. \tag{2}$$

If $y_{ik} = 0,\ \forall k \in T$, this means that no plant is located at position i, but the inequality in (2) does not prevent some positive quantity of biomass is extracted for the plant in position i. Therefore, we add a third constraint:

$$x_{ij} \leq \sum_{k \in T} y_{ik}, \quad j \in J_i,\ \forall i \in I. \tag{3}$$

Used biomass can not exceed the amount of biomass b_j available at point j.

$$\sum_{i \in I} x_{ij} \leq 1, \quad j \in \Gamma_s, \ s \in I. \tag{4}$$

The number and type of plants to be installed should also be taken into account:

$$\eta^k_{\min} \leq \sum_{i \in I} y_{ik} \leq \eta^k_{\max}, \quad k \in T. \tag{5}$$

Finally, the restriction of binary variables:

$$x_{ij}, y_{ik} \in \{0, 1\}, \ j \in J_i, \ i \in I, \ k \in T. \tag{6}$$

Objective Function (Soft Requirements)

The objective function maximizes the total profits (NPV):

$$\max \ F(\mathbf{x}, \mathbf{y}) = \sum_{i \in I} \sum_{j \in J_i} g^i_j b_j x_{ij} - \sum_{i \in I} \sum_{k \in T} c_{ik} y_{ik}. \tag{7}$$

To calculate the vector $G_i = (g^i_j)$ of benefits we use the expression:

$$g^i_j = R_{ij} - K_{ij}, \quad j \in J_i, \quad i \in I,$$

where:

- R_{ij} is the revenue per ton. of biomass produced by the cell j processed in the plant i:
- K_{ij} is the cost per ton. invested in the purchase, collection, transportation, storage, processing and waste disposal of the biomass available in cell j processed in plant i. The coefficient c_{ik} of F includes the cost of installing a plant of type k on site i, plus the costs of plant maintenance, maintenance of electrical and thermal networks and payment of taxes on commercial buildings. Depreciation should be taken into account. Finally, the model without intersections is as follows:

$$\left(\begin{array}{ll} \max & F(\mathbf{x}, \mathbf{y}) = \sum_{i \in I} \sum_{j \in J_i} g_{ij} b_j x_{ij} - \sum_{i \in I} \sum_{k \in T} c_{ik} y_{ik}, \\ s.a. & \sum_{k \in T} y_{ik} \leq 1, \quad i \in I, \\ & \sum_{j \in J_i} b_j x_{ij} \geq \sum_{k \in T} E_k y_{ik}, \quad i \in I, \\ & x_{ij} \leq \sum_{k \in T} y_{ik}, \quad \forall j \in J_i, \ \forall i \in I, \\ & \sum_{i \in I} x_{ij} \leq 1, \quad \forall j \in \Gamma_s, \ s \in I, \\ & \eta^k_{\min} \leq \sum_{i \in I} y_{ik} \leq \eta^k_{\max}, \quad k \in T, \\ & x_{ij}, y_{ik} \in \{0, 1\}, \ j \in J_i, \ i \in I, \ k \in T. \end{array} \right) \tag{8}$$

Mixed Model

The mixed model is nearly the same as binary, with the difference that allows intersections between SA_i. What we do is simply relax the variables x_{ij} of the binary model and define them as real numbers between 0 and 1. Then x_{ij} is a continuous variable that indicates what percentage of the biomass, available at the biomass point $j \in J_i$ will be used by the plant located at position $i \in I$. Thus the restriction (2) of biomass needed, takes the following form:

$$\sum_{j \in J_i} b_j x_{ij} = \sum_{k \in T} E_k y_{ik}, \quad i \in I.$$

It is easy to see that the constraints of type (3) are not necessary. Mixed programming model is considerably reduced in the amount of constraints.

$$\left(\begin{array}{ll} \max & F(\mathbf{x},\mathbf{y}) = \sum_{i\in I}\sum_{j\in J_i} g_{ij} b_j x_{ij} - \sum_{i\in I}\sum_{k\in T} c_{ik} y_{ik}, \\ s.a. & \sum_{k\in T} y_{ik} \le 1, \quad i \in I, \\ & \sum_{j\in J_i} b_j x_{ij} = \sum_{k\in T} E_k y_{ik}, \quad i \in I, \\ & \sum_{i\in I} x_{ij} \le 1, \quad \forall j \in \Gamma_s, \ s \in I, \ s \ne i, \\ & \eta^k_{\min} \le \sum_{i\in I} y_{ik} \le \eta^k_{\max}, \quad k \in T, \\ & x_{ij} \in [0,1], \ y_{ik} \in \{0,1\}, \ j \in J_i, \ i \in I, \ k \in T. \end{array}\right) \tag{9}$$

where:

In what follows we discuss the different approaches considered for the solution of both models.

Implementation of the Solution

Heuristics Method for Binary Model

The binary model has N_V variables and N_R constraints,

$$N_V = \sum_{i\in I} (N_I \times N_{J_i}) + (N_I \times N_T),$$

$$N_R = 2N_I + \sum_{i\in I} (N_I \times N_{J_i}) + \sum_{s\in I,\ s\ne i} N_{\Gamma_s} + 1.$$

If the problem is not very large, you can use an exact type of optimization algorithm as Branch and Bound or Branch & Cut since (8) is a linear problem in binary variables. When N_V and/or N_R are too large, it is impossible to find the optimal solution with computational reasonable time, and a practical solution is to use heuristic methods. From this point of view the problem is combinatorial in nature and we can deal with it using a Genetic Algorithm (GA), which essentially randomly generates different "populations" of solutions and keeps the best. For a detailed description of this type of algorithm see in (Goldberg, 1989).

The main difficult for the implementation of GA is the necessary generation of many populations with a sufficient number of feasible solutions. The random generation of (x_{ij}, y_{ik}) does not guarantee feasibility and design of a random algorithm to obtain feasible solutions of (8) is needed.

The idea we propose is to generate feasible solutions of (8) by a heuristic procedure, which allows to define feasible values of the variables x_{ij} from given feasible values of y_{ik}. In other words, given a feasible location of the plants (defined values of y_{ik}), which respects the constraints (1) and (5), we apply a heuristic procedure to find feasible values of x_{ij} and therefore, we find a feasible solution (x_{ij}, y_{ik}) of (8). Under these conditions, Genetic Algorithm (GA) only have to generate populations of feasible values for the variables y_{ik} (which is much easier) and the heuristic procedure (HP) is responsible for complete a feasible values for the rest x_{ij} of the variables. The calculation of the objective function value (NPV) is then possible for each solution (x_{ij}, y_{ik}) generated and the Genetic Algorithm works.

Heuristic Procedure

The study region, is supposed to be divided into rectangular cells with approximately the same area. Each of these cells have an associated order index $i \in \Theta$ and in turn, to each cell is associated also a pair of coordinates (x_i, y_i) = (row, column) indicating the position of the cell in a crosstab, containing information of the geographical site that defines this cell. This is the usual way to work with geographic data (Xun Shi and Elmore, 2008).

- It is given a certain location of plants, i.e. a set of values $y_{ik} \in \{0, 1\}$ that satisfy the constraints (1) and (5). Denote by:

$$\eta_{\min} = \sum_{k\in T} \eta^k_{\min}, \quad \eta_{\max} = \sum_{k\in T} \eta^k_{\max}.$$

Let p, $\eta_{\min} \le p \le \eta_{\max}$, be the number of plants located by the $y_{ik} = 1$, let be $i_1, i_2, ..., i_p \in I$

indexes of the corresponding cells and $(x_{i_1}, y_{i_1}), (x_{i_2}, y_{i_2}), ..., (x_{i_p}, y_{i_p})$ its corresponding coordinates in the crosstab.

- For each of the chosen cell i_r, with coordinates (x_{i_r}, y_{i_r}) in the crosstab and with type of plant k_r, we define a "provisional" supply area PSA_{i_r}, building a series of boxing rectangles (R_n), centered in the cell of coordinates (x_{i_r}, y_{i_r}) and where R_{n+1} is obtained adding cells from the edge of R_n with the cells surrounding it:

$$\boxed{R_{n+1}} = \begin{bmatrix} (x_{n-1}, y_{n+1}) & (x_n, y_{n+1}) & (x_{n+1}, y_{n+1}) \\ (x_{n-1}, y_n) & \boxed{R_n} & (x_{n+1}, y_n) \\ (x_{n-1}, y_{n-1}) & (x_n, y_{n-1}) & (x_{n+1}, y_{n-1}) \end{bmatrix}$$

When this process reaches one end of the crosstab, the addition continues in all directions where is possible (on the sides where there are adjacent cells to R_n). The process ends up in step n_{i_r}, when the amount of biomass available in all cells in $R_{n_{i_r}}$ is sufficient for the operation of the plant of type k_r located at i_r, i.e. it is greater or equal than E_{k_r}. The rectangle $R_{n_{i_r}}$ defines a provisional supply area (PSA_{i_r}) to the plant in i_r. This part of the heuristic we call *construction of provisional areas*.

During construction of the rectangles R_{n_s}, $s = i_1, i_2, ..., i_p$, it is not taken into account the positions of the other plants, and therefore, the (PSA_{i_r}) may or may not have intersections with each other. Since the intersections are prohibited by the constraints (3) and (6), the cells of the intersection must be assigned to one (and only one) of the plants. This is done in the next step:

- The NPV is calculated for each of the locations $i_r, r = 1, 2, ..., p$, taking into account the provisional areas (PSA_{i_r}) and values $V_{i_1}, V_{i_2}..., V_{i_p}$ are obtained. If the plant getting the highest NPV is V_{i_s} then, the cells given by R_{i_s} are definitely assigned to i_s and therefore (PSA_{i_s}) is defined as the supply area SA_{i_s} (no longer provisional) of i_s. This part of the heuristic we call *assignment of supply area*.

- If SA_{i_s} has not intersected other PSA_{i_r}, $r \neq s$, the *assignment of supply area* step is repeated for the rest of the plant positions i_r, $r = 1, ..., p$, $r \neq s$.

- If SA_{i_s} has intersection with any of the other PSA_{i_r}, $r \neq s$, we comeback to the previous step, and the *construction of provisional area* and *assignment of supply area* steps are repeated for the remaining plants i_r, $r = 1, ..., p$, $r \neq s$, but putting to zero the amount of biomass available in the cells of R_{i_s}. This avoid that biomass in R_{i_s} is assigned to other plants.

- In each cycle a definitely supply area (a number of cells) is assigned to one or more of the plants i_r, $r = 1, ..., p$, leaving less and less "unassigned" plants at each step. The process ends when all plants are assigned a supply area without intersection with the others. This also means that we are implicitly assigning values 0 or 1 to the variables x_{ij}, $j \in J_i$, $i \in I$.

Remarks

- In the end, the final supply areas are not necessarily rectangles, but parts of them. During the process it is necessary to keep the information on which cells are assigned to which plants. Moreover, the heuristic ensures that satisfy the constraints (2), (3), (4) and (6).

- It is clear that the procedure is not optimal and highly dependent on the order in which the SA are assigned. Here we choose the descending order of the NPV, but for a given set of locations another different order may provide a higher total NPV. Nevertheless, the numerical experiences show satisfactory results.

- Note that we have also built a function that, given a set of feasible values $\mathbf{y} = (y_{ik})$, determines the supply areas SA_{i_r} corresponding to the components $y_{i_r k_r} = 1$, $r = 1, 2, ..., p$ and therefore, the values of variables $\mathbf{x} = (x_{ij})$, and calculates the total NPV = sum of all NPV_{i_r}. This value is exactly the objective function value (7) of the binary model for $(\mathbf{x}, \mathbf{y})$. We denote this function γ, which can be written:

$$\gamma(\mathbf{y}) = \sum_{i \in I} \sum_{k \in T} c_{ik} y_{ik} + \sum_{i \in I} \sum_{j \in J_i} g_j^i b_j x_{ij}(\mathbf{y}),$$

In general, it is a nonlinear function of $\mathbf{y} = (y_{ik})$. This is because the SA of a position $i \in I$ is not fixed, by the way the SA is constructed in the heuristic, and it depends on the relative positions of the other plants, given by the vector $\mathbf{y}$.

Reduced Binary Model

The problem is reduced to maximize the nonlinear function γ as a function of variables $\mathbf{y} = (y_{ik})$, subject to feasibility constraints:

$$
\left(\begin{array}{ll}
\max & \gamma(\mathbf{y}) = \sum_{i\in I}\sum_{k\in T} c_{ik}y_{ik} + \sum_{i\in I}\sum_{j\in J_i} g_{ij}b_j x_{ij}(\mathbf{y}), \\
s.a. & \sum_{k\in T} y_{ik} \le 1, \quad i \in I, \\
 & \eta^k_{\min} \le \sum_{i\in I} y_{ik} \le \eta^k_{\max}, \quad k \in T, \\
 & \mathbf{x} = (x_{ij}) := \text{Heuristics}(\mathbf{y}), \\
 & \mathbf{y} = (y_{ik}) \in \{0,1\}^{N_I \times N_T}
\end{array}\right)
$$

(10)

Solving this nonlinear programming problem is carried out using a genetic algorithm (Mitchell, 1996) and the results can be seen later in the Numerical Experiences section.

Optimal Solution of Mixed Model

The implementation was done in $C++$ using the function library of solver GUROBI (GUROBI, 2012), developing a collection of classes and functions that were used for data processing, automated construction and solution of the mixed model. The information is assumed provided by several double entry text files (layers), which must contain the amount of biomass, elevation, distance to roads and distance to the nearest electrical connection at each cell of the file. The most important functions developed for the implementation of the main program, comprising the following parts:

1. *Load Data*.
2. *Variable Definition.*
3. *Constraint Definition.*
4. *Solution.*
5. *Solution Display.*

In the next section we give a description of the results that we consider relevant to an assessment of mixed programming model and its solution. Feasible sites for plants locations were generated randomly for the experiments and the basic information belongs to a southern part of Chile.

Numerical Experiences And Discussion

Numerical Experiences

In this section we discuss the results obtained in solving a set of plant location problems with feasible locations layer randomly generated. We use the developed software for the Binary (8) and Mixed (9) models and a Genetic Algorithm for the Reduced Binary model (10).

The data from the region, used for the experiments, were provided by the Department of Forestry at the University La Frontera, within an ongoing jointly project developed in the southern part of Chile. Figure 1 is a map of the region, indicating the areas of available biomass (in green), areas where there are restrictions (red and blue) and the feasible places where to locate a plant (yellow dots).

Financial data were estimated taking into account the reports by (Xun Shi and Elmore, 2008) and Rentizelas et al. (2009).

We designed a set of experiments, in which is fixed the amount of places to locate plants, but varying the locations of the installation sites, which were chosen randomly and are uniformly distributed in the region. An example (mauve points) is shown in Figure 1. The number of plants to be installed in each of the experiments varied from 1 to 10. We considered two different technologies, which define two types of plants characterized by the biomass required for its operation for 10 years.

Figure 1: indicating the areas of available biomass (in green), areas where there are restrictions (red and blue), feasible places where to locate a plant (yellow dots) and locations of the installation sites, which were chosen randomly and are uniformly distributed in the region (mauve points).

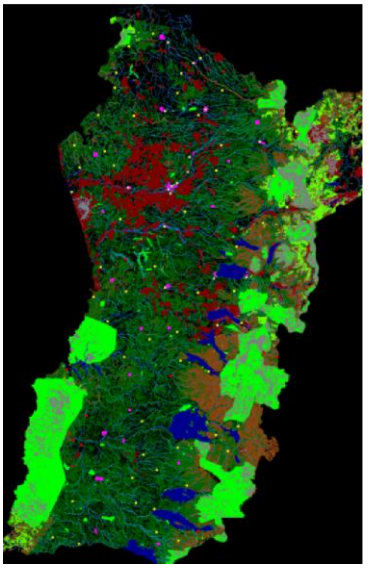


Figure 2 illustrates the relationship between the value of the statistical means of the execution time of each algorithm with respect to the number of plants to be installed. Furthermore, another graph shows the relative error of the NPV, for each number of plants

Figure 2: Relationship between the value of the statistical means of the execution time of each algorithm with respect to the number of plants to be installed.

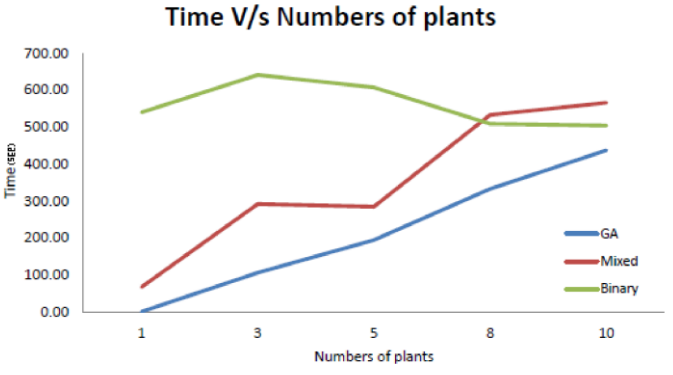


Figure 3: Relative error of the NPV (ErV), for each number of plants.

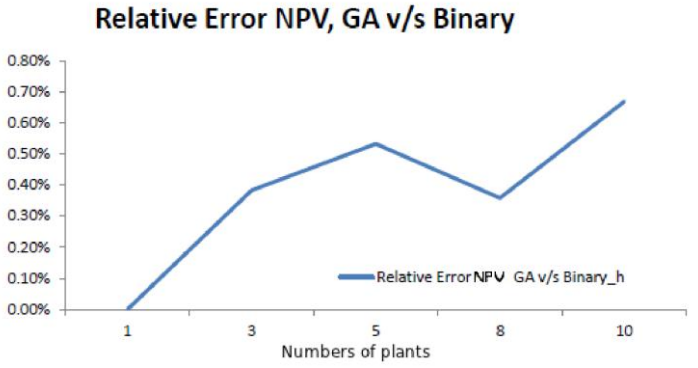
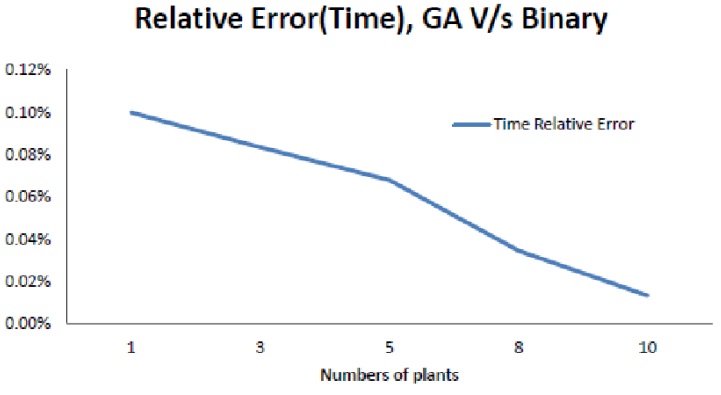


Figure 4: Relative error of the $time$, for each number of plants.



We conducted two additional experiments, the first is to modify the feasible locations, to investigate what happens when they are concentrated in a small area. We choose 20 possible locations concentrated in the center of the region. The results can be seen in Tables 1.

	GA	Mixed	Binary	Relat Error
1	0.13	222.87	2991.8	0.38%
3	82.7	2726.86	7048.61	0.33%
5	192.19	-	4488.8	0.7%
8	402.98	-	12101.54	0.82%

Table 1: Time GA, Mixed, Binary and Averange Relative Error NPV(GA, Binary) of nearby plants

In the second experiment, we exaggerated the number of possible locations, in order to study what happens when there are a large number of possible locations. The results are shown in Table 2.

	GA	Binary	Relat Error
1	0.3615	3063.33	0.28%
3	71.40	3544.46	0.45%
5	128.85	3485.48	0.48%
8	229.5	3704.95	0.53%
10	329.12	3495.25	0.56%

Table 2: Time GA, Binary, and Averange Relative Error NPV(GA, Binary) of 60 feasible plants.

To improve the Genetic Algorithm (GA), we performed a series of experiments aimed to determine the best range of parameters: mutation, crossover, number of generations and population size. The results are shown in Tables 3 and 4.

	Gen 60	Gen 120	Gen 240
1	0%	0%	0%
3	1.12%	0.2%	0.11%
5	1.50%	0.8%	0.36%
10	3.43%	1.9%	0.73%

Table 3: Crossing mutation parameters, for population 20

	Pop 20	Pop 40	Pop 60
1	0%	0%	0%
3	1.12%	0.87%	0.37%
5	1.50%	1.33%	1.35%
10	3.43%	2.32%	2.83%

Table 4: Crossing mutation parameters , for 60 generation

If the solution obtained in the GA is given as initial solution to GUROBI algorithm in the Binary model, no improvement in the convergence was obtained. A combined method with GA and GUROBI was more effective. The mode of operation is as follows: Run GA to find a feasible solution with a high repetition rate, then assign values to the variables y_{ik} of the Binary model, according to the solution of GA, then use GUROBI to solve the Binary model only for binary variables x_{ij}. We call this the "hybrid method". The results of this hybrid method are given in Table 5.

	GA	Hybrid
1	0.13	4.43
3	82.7	93.3
5	192.19	206.5

Table 5: Time GA and Hybrid

Discussion

From the results in experiments with homogeneous distribution of plants we can say that:

1. The average of GA for (RBM) is very close to the optimal NPV achieved by GUROBI in (BM) ($\approx$ 0.5% difference), which is reflected in Figure $2(b)$.
2. From Figure $2(a)$ we see that both, the GA in (RBM) and GUROBI in (MM), increase the running time with increasing the number of plants to install, while (BM) is relatively constant. Also the increasing of GA in (RBM) becomes approximately linear (see Figure $2(a)$).
3. With the increasing of the number of plants to be installed, the running times of GA in (RBM) approach those of GUROBI in (BM). See Figure $2(c)$.
4. When the number of plants to be installed is greater than 5, the difference of NPV between GA in (RBM) and GUROBI in (BM) increases, exceeding 0.6%. See Figure $2(b)$.

From the results in experiments with concentration of plants in a relatively small area we can say that:

1. Running times of GUROBI for (BM) are very sensitive to the distribution of feasible locations of the plants. When they are concentrated running times increase considerably (over one hour computation time), while the runtime for GA in (RBM) never exceeds 7 minutes. See Table 1.
2. We recommend to use the genetic algorithm in (RBM) when the number of plants to be installed is not greater than 5 and when the feasible locations of the plants are concentrated in a small region. This will give us an approximate solution with 99% of closeness to the optimal value of the NPV and a much smaller runtime. See Table 1.
3. If there is a lot of feasible installation sites, the genetic algorithm performs remarkably well in time (see Table 2) and is approximately $0,5$% of the optimal NPV value (see Table 2). Therefore it is preferable to use the genetic algorithm in (RBM) when there is a large number of candidate locations to install a set of plants.
4. We consider that after 240 generations with 60 population size, the solution found with GA in (RBM) becomes significant. Increasing the size of the population does not imply a better behavior of the GA, but an increasing number of generations results in a significant advance to the optimal

solution. The best results were obtained with parameters of crossover and mutation of the order 0.6 and 0.25 respectively. See Table 3 and 4.

With respect to a combined method GA+GUROBI we can say:

1. If we give the solution of GA in (RBM) to GUROBI in (BM) as initial solution, there is not improvement, the time differences are not significant.
2. Fixing values y_{ik} of the GA solution in (BM) when you want to install less than 5 plants, the hybrid method always achieves optimal or very satisfactory solutions. But if you want to find 10 or more plant locations, the GA does not ensure convergence, allowing a nonreliable initial solution to (BM) and hence, the hybrid method becomes inaccurate (see Table 5). In the case of concentrated feasible locations of plants, it is extremely advisable to use the hybrid method, since GA quickly find a feasible solution. When plants to install are less than 5, the GA solution is confident and thus, finding the solution of (BM) without the varibles y_{ij} is a simpler task that ensures the exact optimal solution. (See Table 5).

Conclusions

This paper presents a binary and mixed programming model for locating biomass forestry plants. Both models determine the location and simultaneously define the supply area of each plant. In addition we propose two approaches to their resolution: the first is the coding of the Binary (BM) and mixed (MM) models in a commercial software (GUROBI) for medium-scale problems and the second is the application of a genetic algorithm to a Reduced Binary model (RBM) for larger problems.

Acknowledgement

This applied research has been funded through FONDEF Research Project Number D08I1056 The authors wish to acknowledge the financial support and continuous motivation.

References

[1] Bernetti, I., C. Fagarazzi, R. Fratini., 2004. A methodology to analyze the potential development of biomass energy sector: an application in Tuscany, Forest Policy and Economics. 415-432.

[2] Esteban, L.S., P. Pérez Ortiz, P. Ciria, J. Carrasco., 2000. Evaluación de los recursos de biomasa forestal en la provincia de Soria. Análisis de alternativas para su aprovechamiento energético. Colección Documentos CIEMAT, Centro de Investigaciones Energéticas Medioambientales y Teconológicas. 124.

[3] Goldberg, D., 1989. Genetic algorithm in search, optimization and machine learning. Addison Wesley.

[4] Graham, R.L., B.C. English, C.E. Noon.,2000. A Geographic Information System-based modelling system for evaluating the cost of delivered energy crop feedstock. Biomass and Bioenergy, 309-329.

[5] Guerrero F., A. Carrazo., 2005. Localización de centrales de energìa eléctrica a partir de biomasa procedente de olivar. Revista de Estudios Regionales. 153-175.

[6] GUROBI, Mathemathical Programming Software, http://www.gurobi.com.

[7] Krukanont, P., S. Prasertsan., 2004. Geographical distribution of biomass and potential sites of rubber wood fired power plants in Southern Thailand. Biomass and Bioenergy. 47-59.

[8] Mitchell, M., 1996. An introduction to genetic algorithms. MIT Edit.

[9] Panichelli, L., E. Gnansounou., 2008. GIS-based approach for defining bioenergy facilities location: A case study in Northern Spain based on marginal delivery costs and resources competition between facilities. Biomass and Bioenergy. 289-300

[10] Pontt, C. Estudio de contribución de las ERNC al SIC al 2025 (Potencial de Biomasa en Chile), Technical Report, Universidad Técnica Federico Santa Marìa, 2008

[11] Rentizelas A.A., I.P. Tatsiopoulos, A. Tolis., 2009. An optimization model for multi-biomass tri-generation energy supply. Biomass and Bioenergy. 223-233.

[12] Troncoso J., R. Garrido., 2002. Modelos de localización de instalaciones: una aplicación para la producción y logìstica forestal. Scielo. 57-67.

[13] Voivontas, D., D. Assimacopoulos, E.G. Koukios., 2001. Assessment of biomass potential for power production: a GIS based method. Biomass and Bioenergy. 101-112.

[14] Wolsey, L. A.,1998 , Integer Programming, Wiley-Interscience publication.

[15] Xun Shi, A. Elmore., 2008. Using spatial information technologies to select sites for biomass power plants: A case study in Guangdong Province, China. Biomass and Bioenergy. 35-43.