

Population-Based Methods in the Optimization of Stand Management

Timo Pukkala

Pukkala, T. 2009. Population-based methods in the optimization of stand management. *Silva Fennica* 43(2): 261–274.

In Finland, the growth and yield models for tree stands are simulation programs that consist of several sub-models. These models are often non-smooth and non-differentiable. Direct search methods such as the Hooke-Jeeves algorithm (HJ) are suitable tools for optimizing stand management with this kind of complicated models. This study tested a new class of direct search methods, namely population-based methods, in the optimization of stand management. The tested methods were differential evolution, particle swarm optimization, evolution strategy, and the Nelder-Mead method. All these methods operate with a population of solution vectors, which are recombined and mutated to obtain new candidate solutions. The management schedule of 719 stands was optimized with all population-based methods and with the HJ method. The population-based methods were competitive with the HJ method, producing 0.57% to 1.74% higher mean objective function values than HJ. On the average, differential evolution was the best method, followed by particle swarm optimization, evolution strategy, and Nelder-Mead method. However, differences between the methods were small, and each method was the best in several stands. HJ was alone the best method in 7% of stands, and a population based method in 3% (Nelder-Mead) to 29% (differential evolution) of stands. All five methods found the same solution in 18% of stands.

Keywords differential evolution, evolution strategy, particle swarm optimization, Nelder-Mead method, amoeba search, polytope search, evolutionary computing

Addresses University of Joensuu, Faculty of Forest Sciences, P.O. Box 111, FI-80101 Joensuu, Finland **E-mail** timo.pukkala@joensuu.fi

Received 16 September 2008 **Revised** 22 December 2009 **Accepted** 3 April 2009

Available at <http://www.metla.fi/silvafennica/full/sf43/sf432261.pdf>

1 Introduction

The most important decisions in forest stand management are the timing, intensity and type of cuttings, and regeneration. If the vocation of forestry is to produce economic benefits, cutting decisions should be made so as to maximize the present value of all future incomes and costs (NPV). The whole cutting schedule of a stand is defined by a set of decision variables such as the time points and intensities of cuttings. In the best cutting schedule the values of decision variables are set so that the NPV is maximized. Optimization is an automated way of searching for the best set of decision variables. Since there are often several decision variables, methods for multidimensional optimization must be used.

Calculation of the net present value of a cutting schedule uses growth and yield models of tree stands. These are complicated models which are difficult to maximize analytically. Most growth and yield models are simulators that consist of several sub-models for instance for diameter increment, height increment, tree survival, and stem taper. In most Finnish simulators the tree stand is represented by a set of individual trees, each tree representing a diameter class of a tree species. The use of a rather small number of individual trees adds non-smoothness to the growth and yield model because the value of a tree, and therefore the whole stand, may increase suddenly when the tree reaches minimum dimensions required, for instance, for saw log.

Direct search methods are suitable tools for maximizing or minimizing functions that are non-smooth and non-differentiable. In their comprehensive book about non-linear programming, Bazaraa et al. (1993) mention three such methods (multidimensional search methods without using derivatives): the cyclic coordinate method, the Hooke-Jeeves method, and the Rosenbrock method. The cyclic coordinate method alters the value of one decision variable at a time, i.e. coordinate axes are the search directions. The Hooke-Jeeves method uses two search modes: exploratory search in the direction of coordinate axes (one decision variable altered at a time) and pattern search in other directions (more than one decision variable altered simultaneously). The method of Rosenbrock tests m (m is the number of optimized decision variables)

orthogonal search directions which are equal to coordinate axes only at the first iteration.

Another type of method, namely dynamic programming, has been used extensively in stand management optimization (see Hoganson et al. 2008 for a review). It differs from direct search methods so that, instead of seeking optimal values for continuous decision variables, it finds the optimal sequence of stand states defined by classified values of several growing stock characteristics (for instance, basal area, mean diameter, and stand age). Direct search methods can be used in both even-aged and uneven-aged forestry (Trasobares and Pukkala 2005). Uneven-aged management can also be optimized using linear programming and a transition matrix model (e.g. Kaya and Buongiorno 1987).

Of the direct search methods mentioned in Bazaraa et al. (1993) the Hooke-Jeeves algorithm has been used much in forestry, at least in Finland (e.g. Valsta 1992a, 1992b, Møykkynen et al. 2000, Miina and Pukkala 2000, Pukkala and Miina 2005). In the other fields a new type of methods, called as population-based methods or evolutionary computation methods, have also been used to solve complicated optimization problems. Optimization uses a population of solutions instead of a single vector of decision variables. Three such methods are commonly mentioned in literature: differential evolution (Storn and Price 1997), particle swarm optimization (Kennedy and Eberhart 1995), and evolution strategy (Bayer and Schwefel 2002). The method of Nelder and Mead (also called polytope search and amoeba search) may be added to the list although it is rarely called as population-based method (Nelder and Mead 1965). However, because it operates with several solution vectors which are combined to form new vectors, it does belong to the category of population-based methods.

The above population-based methods are direct search methods which do not require any guesses of derivatives of the objective function and have no assumptions concerning it. Therefore, they seem lucrative tools for the optimization of stand management schedules when the function to be maximized is a complicated simulation model. However, population based methods have been used very little, maybe not at all, in forestry. The purpose of this study was to test the population-based methods in the optimization of stand man-

agement. The method of Hooke and Jeeves (1961) was used as a reference method to which the population-based methods were compared.

2 Materials and Methods

In the following, symbol \mathbf{x}_i is used for a solution, i.e. a vector of m decision variables. A population consists of n solution vectors. The objective function value of a solution is denoted as $f(\mathbf{x}_i)$. Each member of the population is characterized by \mathbf{x}_i and $f(\mathbf{x}_i)$. Depending on the method, the members may also have other characteristics such as strategy parameter vector or velocity vector. All methods except Hooke-Jeeves start with an initial population of n vectors of m decision variables. In all population-based methods the elements of the initial vectors were randomly drawn from uniform distribution:

$$\mathbf{x}_i = \{x_{ij} \mid 1 \leq j \leq m, x_{ij} = U[a_j, b_j]\}$$

where x_{ij} is the j th decision variable of vector \mathbf{x}_i , a_j and b_j are, respectively, the lower and upper limit of the j th variable, and U stands for uniform distribution. The stopping criteria were also the same in all population-based methods. The search was stopped when a pre-defined maximum number of iterations were completed. The search was stopped earlier if the difference in the objective function value between the best and the worst solution vector was less than one percent of the best objective function value: Stop if $[f(\mathbf{x}_{\text{best}}) - f(\mathbf{x}_{\text{worst}})] < 0.01 \times f(\mathbf{x}_{\text{best}})$.

2.1 Hooke and Jeeves Method (HJ)

The HJ method needs an initial solution vector \mathbf{x}_1 . The search begins with exploratory search in the directions of coordinate axes (one decision variable altered at a time). After completing one round of exploratory searches the algorithm goes to pattern search if the exploratory search is successful. The pattern search alters the values of more than one decision variable simultaneously. If exploratory search cannot improve the solution the step size used in exploratory search is halved

and the search is repeated. The search is stopped once the step size becomes smaller than a pre-defined stopping criterion.

Let $\mathbf{d}_1, \dots, \mathbf{d}_m$ be the coordinate directions and let $\mathbf{y}_1 = \mathbf{x}_1$ and $k=j=1$. The algorithm works as follows (Bazaraa et al. 1993):

Exploratory search

- 1) If $f(\mathbf{y}_j + \Delta \mathbf{d}_j) > f(\mathbf{y}_j)$ (success), let $\mathbf{y}_{j+1} = \mathbf{y}_j + \Delta \mathbf{d}_j$ and go to Step 2. Otherwise if $f(\mathbf{y}_j - \Delta \mathbf{d}_j) > f(\mathbf{y}_j)$ (success), let $\mathbf{y}_{j+1} = \mathbf{y}_j - \Delta \mathbf{d}_j$ and go to Step 2. Otherwise let $\mathbf{y}_{j+1} = \mathbf{y}_j$ and go to Step 2 (no improvement found in direction j).
- 2) If $j < m$, replace j by $j+1$ and repeat Step 1 (go to next decision variable). Otherwise go to *Pattern search* if $f(\mathbf{y}_{m+1}) > f(\mathbf{x}_k)$ (at least one successful change detected in the directions of coordinate axes). If $f(\mathbf{y}_{m+1}) \leq f(\mathbf{x}_k)$ go to *Step size reduction*.

Pattern search

- 3) Let $\mathbf{x}_{k+1} = \mathbf{y}_{m+1}$ and let $\mathbf{y}_1 = \mathbf{x}_{k+1} + \alpha(\mathbf{x}_{k+1} - \mathbf{x}_k)$. Replace k by $k+1$, let $j=1$, and go to Step 1.

Step size reduction

- 4) If $\Delta \leq \epsilon$, stop. Otherwise replace Δ by $\Delta/2$. Let $\mathbf{y}_1 = \mathbf{x}_k$, $\mathbf{x}_{k+1} = \mathbf{x}_k$, $j=1$, replace k by $k+1$, and repeat Step 1.

The method has three parameters: initial steps size (Δ), stopping criterion (ϵ), and α . The initial step size was different for different decision variables; it was 0.1 times the range $[a_j, b_j]$ of the variable. The stopping criterion was equal to 0.01 times the initial steps size. The search was stopped once the step size was smaller than the stopping criterion for every decision variable. Parameter α was taken as 1.0.

2.2 Differential Evolution (DE)

In DE, an iteration consists of n tournaments, one for every solution. In a tournament, solution \mathbf{x}_i is compared to a trial solution \mathbf{x}_t , which is obtained by combining \mathbf{x}_i and a noise vector \mathbf{y}_i . The combination is produced by taking elements randomly from vectors \mathbf{x}_i and \mathbf{y}_i so that the probability is α for \mathbf{x}_i and $1 - \alpha$ for \mathbf{y}_i . The noise vector \mathbf{y}_i is a combination of tree vectors (\mathbf{x}_a , \mathbf{x}_b and \mathbf{x}_c) randomly drawn from the current population:

$$\mathbf{y}_i = \mathbf{x}_c + \beta(\mathbf{x}_a - \mathbf{x}_b)$$

If $f(\mathbf{x}_i) > f(\mathbf{x}_i)$ the trial vector \mathbf{x}_t replaces \mathbf{x}_i . The method has four parameters: α , β , n , and maximum number of iterations. Parameter α was taken as 0.9, parameter β as 0.7, n as $5 \times m$ (m is the number of optimized decision variables), and the maximum number of iterations as 100. The optimal solution is the best member at the last iteration.

2.3 Particle Swarm Optimization (PS)

In PS the population consists of n particles. Each particle is characterized by $\mathbf{x}_i, f(\mathbf{x}_i), \mathbf{x}_i^b, f(\mathbf{x}_i^b)$ and \mathbf{v}_i , where \mathbf{x}_i is the current solution, $f(\mathbf{x}_i)$ the current objective function value, \mathbf{x}_i^b the best solution found by the particle so far (“particle best”), $f(\mathbf{x}_i^b)$ the objective function value of particle best, and \mathbf{v}_i a vector of m velocities. In addition, the whole population is characterized by the best solution found so far by the swarm (\mathbf{x}^g , called as “global best”) and the objective function value $f(\mathbf{x}^g)$ of global best. Vectors \mathbf{v}_i were initialized with zeroes and vectors \mathbf{x}_i with uniform random numbers $U[a_j, b_j]$ as explained above. In PS, vectors \mathbf{x}_i are called as particle positions.

An iteration of PS begins with the production of two random numbers r_1 and r_2 , which are uniformly distributed between 0 and 1. Then, particle positions are updated by adding velocities to the current values (element by element):

$$\mathbf{x}_i^{\text{updated}} = \mathbf{x}_i + \mathbf{v}_i$$

Velocities are up-dated as follows:

$$\mathbf{v}_i^{\text{updated}} = w\mathbf{v}_i + c_1r_1(\mathbf{x}_i^b - \mathbf{x}_i) + c_2r_2(\mathbf{x}^g - \mathbf{x}_i)$$

where w , c_1 and c_2 are parameters. Vector \mathbf{v}_i determines the direction of movement for particle i . The above equation shows that the direction depends on both the particle best and the global best.

Solutions $\mathbf{x}_i^{\text{updated}}$ replace \mathbf{x}_i . If $f(\mathbf{x}_i) > f(\mathbf{x}_i^b)$, \mathbf{x}_i replaces \mathbf{x}_i^b (new particle best) and $f(\mathbf{x}_i)$ replaces $f(\mathbf{x}_i^b)$. If $f(\mathbf{x}_i) > f(\mathbf{x}^g)$, \mathbf{x}_i replaces \mathbf{x}^g (new global best) and $f(\mathbf{x}_i)$ replaces $f(\mathbf{x}^g)$. Velocity vectors $\mathbf{v}_i^{\text{updated}}$ replace \mathbf{v}_i . The optimal solution of the PS method is vector \mathbf{x}^g at the end of the last iteration.

Particle swarm optimization has five parameters; w , c_1 , c_2 , n , and maximum number of iterations. Parameter w , which is called as inertial constant, was taken as 0.95. Parameters c_1 and c_2 were equal to 2. They determine how much the particle is directed towards the particle best (cognitive component) and global best (social component). The number of particles (n) was $50 + 10m$ where m is the number of decision variables. The maximum number of iterations was 50.

2.4 Evolution Strategy (ES)

In ES, every solution of the population is characterized by $\mathbf{x}_i, \mathbf{s}_i$, and $f(\mathbf{x}_i)$, where \mathbf{x}_i is the current solution (vector of current values of decision variables), \mathbf{s}_i a vector of strategy parameters, and $f(\mathbf{x}_i)$ the objective function value calculated for the solution. The strategy parameters, one for each decision variable, determine how much the values of decision variables of a recombinant are mutated. The solution vectors were initialized using $U[a_j, b_j]$ as explained above. The initial values of strategy parameters were obtained from

$$\mathbf{s}_i = \alpha \mathbf{x}_i$$

where α is a parameter set as 0.2. During every generation, one offspring was produced as a mutated recombination of two parents. If the offspring was better (had a better objective function value) than the worst solution of the current population, the offspring replaced the worst solution. This kind of ES is referred to as a steady state ES, which is an ES without a generation gap.

The best solution of the current population was always used as a parent (parent a). The other parent (b) was selected randomly from among the remaining solutions. The selected parents produced a recombinant as follows:

$$\mathbf{x}_r = 0.5(\mathbf{x}_a + \mathbf{x}_b)$$

$$\mathbf{s}_r = 0.5(\mathbf{s}_a + \mathbf{s}_b)$$

The recombinant was mutated to obtain an offspring:

$$\mathbf{s}_o = \mathbf{s}_r \times e^{\tau \times N(0,1)}$$

$$\mathbf{x}_o = \mathbf{x}_r + \mathbf{s}_o \times N(0,1)$$

where τ is called a learning parameter which was calculated as $\tau = 1/\sqrt{n}$, and $N(0,1)$ is a normally distributed random number with mean equal to 0 and standard deviation equal to 1. The above formulas show that the mutated strategy parameters control the amount by which the elements of the recombinant solution are mutated. The mutated solution $(\mathbf{x}_o, \mathbf{s}_o, f(\mathbf{x}_o))$ replaces the worst solution of the parent population. Since the learning parameter τ is a function of population size, the ES has only three parameters: α , n , and maximum number of generations. Parameter α was set as 0.2 and population size (n) as $50 + 10m$ where m is the number of decision variables. The maximum number of generations was 2000. The best solution after the last generation was taken as the optimal solution found by the ES method.

2.5 Nelder-Mead Method (NM)

In the NM method the worst, second-worst and best solutions, denoted as \mathbf{x}_w , \mathbf{x}_s , and \mathbf{x}_b , respectively, are found in the beginning of a new iteration. The worst solution is replaced by a new candidate solution which is a transformation of all solutions. However, if the candidate solution is worse than the current worst (\mathbf{x}_w), there is no replacement. Instead, a shrinking operation is done, in which all solutions (points) except the best are moved closer to the best solution (\mathbf{x}_b).

The production of the candidate solution begins with the calculation of the centroid solution (denoted as \mathbf{x}_m), which is the mean of all solutions except the worst:

$$\mathbf{x}_m = 1/(n-1) \times \sum_{i \neq w} \mathbf{x}_i$$

Then a reflection point is computed as a function of the centroid and the worst point:

$$\mathbf{x}_r = \mathbf{x}_m + \alpha(\mathbf{x}_m - \mathbf{x}_w)$$

If $f(\mathbf{x}_b) > f(\mathbf{x}_r) \geq f(\mathbf{x}_s)$, \mathbf{x}_r replaces \mathbf{x}_w and the iteration is completed. If $f(\mathbf{x}_r) > f(\mathbf{x}_b)$, i.e. the reflection point is better than the current best, an expansion point is calculated as

$$\mathbf{x}_e = \mathbf{x}_m + \gamma(\mathbf{x}_r - \mathbf{x}_m)$$

If $f(\mathbf{x}_e) > f(\mathbf{x}_r)$, \mathbf{x}_e replaces \mathbf{x}_w and the iteration is terminated. Otherwise (if $f(\mathbf{x}_r) \geq f(\mathbf{x}_s)$) \mathbf{x}_r replaces \mathbf{x}_w and the iteration is terminated. If $f(\mathbf{x}_r) < f(\mathbf{x}_s)$, i.e. the reflection point is poorer than the second-worst, a contraction point is calculated. The way of calculation depends on whether $f(\mathbf{x}_r)$ is between $f(\mathbf{x}_w)$ and $f(\mathbf{x}_s)$ (contract outside) or if $f(\mathbf{x}_r) \leq f(\mathbf{x}_w)$ (contract inside). In outside contraction the trial point is calculated from the reflection and centroid points:

$$\mathbf{x}_c = \mathbf{x}_m + \beta(\mathbf{x}_r - \mathbf{x}_m)$$

If $f(\mathbf{x}_c) \geq f(\mathbf{x}_r)$, the point is accepted (\mathbf{x}_c replaces \mathbf{x}_w) and the iteration is completed. Inside contraction point is calculated from the best and centroid points:

$$\mathbf{x}_c = \mathbf{x}_m + \beta(\mathbf{x}_b - \mathbf{x}_m)$$

If $f(\mathbf{x}_c) \geq f(\mathbf{x}_r)$, \mathbf{x}_c replaces \mathbf{x}_w and the iteration is completed. If the reflection, expansion and contracting operations cannot find a point better than \mathbf{x}_w , a shrinking operation is done for all points except the best:

$$\mathbf{x}_i^{\text{updated}} = \mathbf{x}_i + \delta(\mathbf{x}_i - \mathbf{x}_b) \text{ for } i = 1, \dots, n, \text{ with } i \neq b.$$

The new solution vectors $\mathbf{x}_i^{\text{updated}}$ replace \mathbf{x}_i and a new iteration begins.

The NM algorithm has four parameters in addition to population size and maximum number of iterations: α for reflection, γ for expansion, β for contraction, and δ for shrinkage. The usual parameter values were used: $\alpha = 1$, $\gamma = 2$, $\beta = 1/2$, and $\delta = 1/2$. The number of points (population size) was not $m + 1$ as usual but $5 \times m$, i.e. five times the number of optimized decision variables. This change reduced the tendency of the method to converge to a local optimum. The maximum number of iterations was 1000.

2.6 Optimization Problems

This study optimized the cutting schedules of stands. A cutting schedule was defined by the number, timing, intensity and type of thinnings, and the timing of clear fellings. The number of thinnings was not optimized since it is not

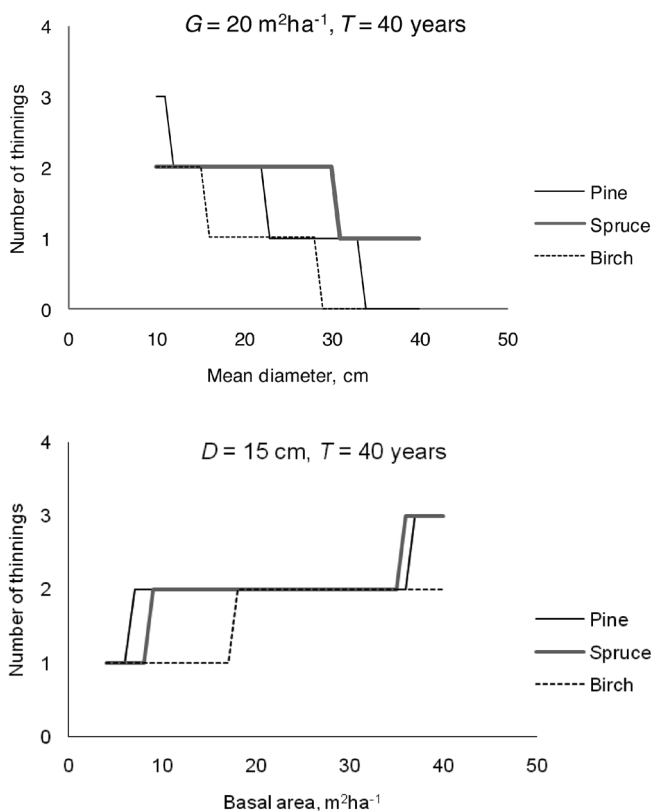


Fig. 1. Dependence of the number of remaining thinnings on the mean diameter (D) and basal area (G) of the stand when stand age (T) is 40 years.

a continuous variable. Instead, the optimal number of remaining thinnings was calculated with models, which are based on the optimization of the management schedule of many young stands with zero, one, two and three thinnings. The best number of thinnings (with the highest net present value) was modeled as a function of stand characteristics:

Scots pine:

$$n_{\text{thin}} = 3.629 - 0.091D + 0.574\ln(G) - 0.495\ln(T)$$

Norway spruce:

$$n_{\text{thin}} = 3.011 - 0.839\ln(D) + 0.684\ln(G) - 0.017T$$

Birch:

$$n_{\text{thin}} = 1.629 - 0.078D + 0.363\ln(G)$$

where D is basal-area-weighted mean diameter (cm), T is basal-area-weighted mean tree age

(years) and G is stand basal area (m^2ha^{-1}). The models show that the optimal number of remaining thinnings decreases with increasing mean diameter and tree age, but increases with increasing stand basal area (Fig. 1). The models are based on 3% discounting rate and the same dimensions and prices of timber assortments as used in this study (Table 1). The predicted number of thinnings was rounded to the closest integer. If the stand was interpreted as a two-storey stand (the height of the tallest stratum was at least twice the height of the shortest stratum), one thinning was added to the result because it was assumed that an additional thinning may be required to remove the over-storey and the predicted number of thinnings are required to grow the remaining stand to the end of rotation. It is noteworthy that the net present value of the optimized schedule does not

Table 1. Stumpage prices and minimum top diameters of timber assortments.

	Pine	Spruce	Birch
Log price, € m ⁻³	46	43	44
Pulpwood price, € m ⁻³	14	22	14
Minimum top diameter of log, cm	15	16	18
Minimum top diameter of pulpwood, cm	8	8	9

change much if the number of thinnings deviates by one from the optimal number.

The decision variables which were optimized were:

For each thinning:

- number of years since the start of simulation (first thinning) or previous thinning (other thinnings)
- harvest percentage, specified separately for all tree strata that were inventoried separately

For final felling:

- number of years since the last thinning

The number of decision variables was therefore $n_{DV} = n_{thin} \times (1 + n_{strata}) + 1$. For instance, if there were two thinnings and three strata (e.g. pine, birch and spruce), the number of decision variables was $2 \times (1 + 3) + 1 = 9$. The type of thinning was also optimized, at least to some extent, since the thinning intensity was optimized separately for every stratum. In a two-storey stand, different strata may represent different canopy layers of the same species, which means that optimization had freedom to use high thinning or low thinning. Within a stratum the thinning intensity was the same in all diameter classes.

The starting values of cutting intervals were obtained as follows:

$$c = (R - T) / (n_{thin} + 1)$$

where c is the number of years since the beginning of simulation (1st cutting) or since previous cutting (other cuttings), R is a guess of the rotation length ($R = 100$), T is the age of the initial stand, and n_{thin} is the number of thinnings. The initial value of harvest percentage was taken as 30 for

all strata in all thinnings.

These initial values were used in the HJ method as the starting solution (\mathbf{x}_1). The other methods required a range for each decision variable $[a_j, b_j]$. The range was $[5, 2c]$ for cutting intervals and $[0, 100]$ for thinning percentages. These ranges were also used to calculate the initial step size of the HJ method: $\Delta_j = 0.1 \times (b_j - a_j)$.

2.7 Objective Function

The objective variable was net present value of all future incomes and costs (NPV) calculated with 3% discounting rate. The stumpage prices of timber assortments are given in Table 1. The NPV was calculated as the sum of the net present values of the remaining cutting treatments plus the discounted soil expectation value of bare land:

$$NPV = \sum_{y=0}^Y \frac{N_y}{(1+i)^y} + \frac{SEV}{(1+i)^Y}$$

where y is the number of years since the beginning of simulation, Y is number of years to the end of rotation, N_y is net income in year y , and i is discounting rate ($i = 0.03$). The SEV of bare land, which represents the value of all future rotations, was calculated with the models of Pukkala (2005). These models are based on the optimized management schedules for full rotation of a great number of stands on different sites and with different timber prices and discounting rates (3500 optimizations altogether with the HJ method). It is noteworthy that the discounted value of SEV, calculated with 3% rate, is rather small compared to the net present value of the remaining rotation.

2.8 Yield Model

A simulation model was used to simulate stand development to the end of rotation with different values of decision variables. The simulator used the models of Hynynen et al. (2002) for site index, dominant height development, diameter increment, crown height, height increment and tree survival. The taper models of Laasasenaho (1982) were used to calculate assortment volumes of removed trees. The saw-log volume obtained

from the taper model was corrected using the log reduction models of Mehtätalo (2002). The purpose was to transfer a part of log volume to pulpwood volume due to defects which are present but not taken into account in the taper models. The predictions of Mehtätalo's models were multiplied with correction factors which were 0.7 for pine, 0.4 for spruce, and 1.2 for birch (Malinen et al. 2007). This is because it has been noticed (Malinen et al. 2007) that Mehtätalo's models overestimate the reduction in pine and spruce, and underestimate it in birch.

The input data for the simulation model were a set of model trees that represented the stand. However, the inventory data consisted of stand level characteristics (number of trees per hectare, stand basal area, mean height, mean age, mean diameter) assessed separately for different tree strata. Therefore, the model trees were generated by first predicting the diameter distribution of each stratum. Then, the distributions were divided into five diameter classes of equal width and the mid-point tree of each class was taken as a model tree. The frequencies of the mid-point trees were calibrated using goal programming as explained in Pukkala and Miina (2005). For example, if the stand had four strata, $4 \times 5 = 20$ model trees represented the stand in simulation. Each tree was characterized by species, dbh, age, height, and frequency. The stratum number of each model tree was maintained, to be able to use different thinning intensities in different strata.

2.9 Stands

The management schedule was optimized for one stand with all methods to illustrate the development of the population of solutions in different methods. This stand (referred to as test stand) had four strata in two canopy layers (Table 2). Norway spruce was present in both canopy layers. This kind of stand structure is common in Finnish forests. The temperature sum of the region was assumed to be 1200 d.d, elevation 150 m above sea level, and the proportion of lakes in the region 0.2. The test stand represented medium site fertility (*Myrtillus* site).

The second set of data consisted of compartment inventory data from 719 stands in North Karelia, Finland. These stands represented all common

Table 2. Tree strata of the test stand. *G* is stand basal area, *N* is number of trees per hectare, *H* is mean height, *T* is mean age, and *D* is mean diameter. A dash (-) means that the variable (*G* or *N*) was not assessed in the field.

Species	<i>G</i> , m ² /ha	<i>N</i> , trees/ha	<i>H</i> , m	<i>T</i> , years	<i>D</i> , cm
Pine	8	-	15	40	17
Spruce	8	-	11	40	14
Birch	8	-	16	40	18
Spruce	-	800	4	20	7

stand structures and included one-species one-layer stands, one-layer mixtures of two to four species, and two-layer stands where both canopy layers could be mixtures of two to four species. The temperature sum of the region was assumed to be 1200 d.d. for all stands, elevation 150 m above sea level, and the proportion of lakes 0.2.

3 Results

3.1 Results for the Test Stand

The HJ, DE, and NM methods found practically the same optimal management schedule for the test stand (Table 3): The first thinning was conducted immediately, the second after 18–19 years, the third after 30–31 years, and final felling after about 45 years. PS and ES converged to somewhat different solutions. The NPV of the solution was the highest for NM, followed by DE, HJ, PS, and ES.

All methods removed all birches in the first thinning which was always conducted immediately (Table 3, Fig. 2). ES removed also some pine. The second thinning removed only pine in all methods. PS left some pines to continue growing whereas the other methods removed all of them. The third thinning removed trees mainly from the larger spruce layer in all methods except PS, which removed the remaining pines and no spruces at all. The clear cutting, which was 12–15 years later, removed the remaining spruces. From mean the tree size in different strata (Table 2) and the order of removing the strata it can be concluded that all thinnings were high thinnings.

Table 3. Cutting years (no. of years since the beginning of simulation) and net present value (NPV) in the optimal solution found for the test stand by Hooke-Jeeves method (HJ), differential evolution (DE), particle swarm optimization (PS), evolution strategy (ES) and Nelder-Mead method (NM).

Cutting	Optimization method				
	HJ	DE	PS	ES	NM
1. thinning	0	0	0	0	0
2. thinning	19.0	19.1	12.3	14.3	18.3
3. thinning	31.1	31.5	26.2	40.7	30.0
Final felling	45.7	44.9	41.2	52.2	45.4
NPV, €/ha	7118	7139	7070	7057	7147

The thinnings decreased stand age since youngest trees were maintained longest. The upper layer was completely removed at tree age of about 70 years (40+30=70) in the three best methods (DE, NM, HJ). The average tree age at final felling was 65 years in these methods (20+45=65).

Fig. 3 shows that the range of objective function values among the population of solutions gets narrower during the search process in all methods. However, there are certain differences between the methods: in ES and NM the worst and best solutions approach to each other quite fast whereas in DE and PS this happens mainly in the beginning of the search. This difference is because of the fact that, at every iteration, ES and NM concentrate on replacing the worst solution by a better one whereas PS updates all solutions at every iteration. DE compares every solution to a new trial solution which is only slightly different, and replaces the solution if the trial is better.

3.2 Results for the Forest

The mean NPV of the optimal management schedule for the 719 stands was the highest when DE was used as the optimization method (Table 4). The second best was PS, followed by ES, NM and HJ. However, differences in the mean NPV

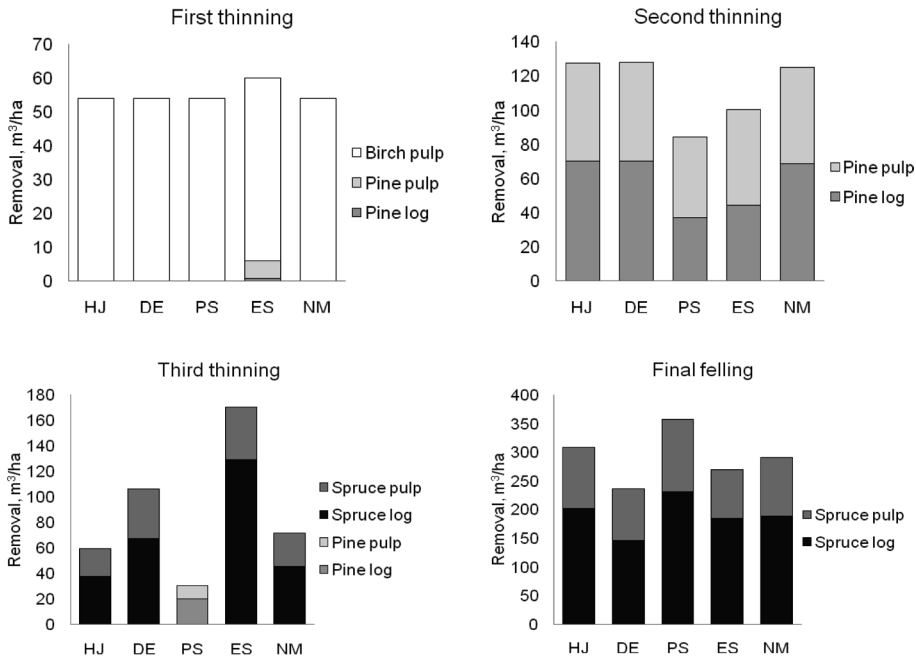


Fig. 2. Assortment removals in the optimal cutting schedules for the test stand found by the Hooke-Jeeves (HJ), differential evolution (DE), particle swarm optimization (PS), evolution strategy optimization (ES) and Nelder-Mead (NM) methods.

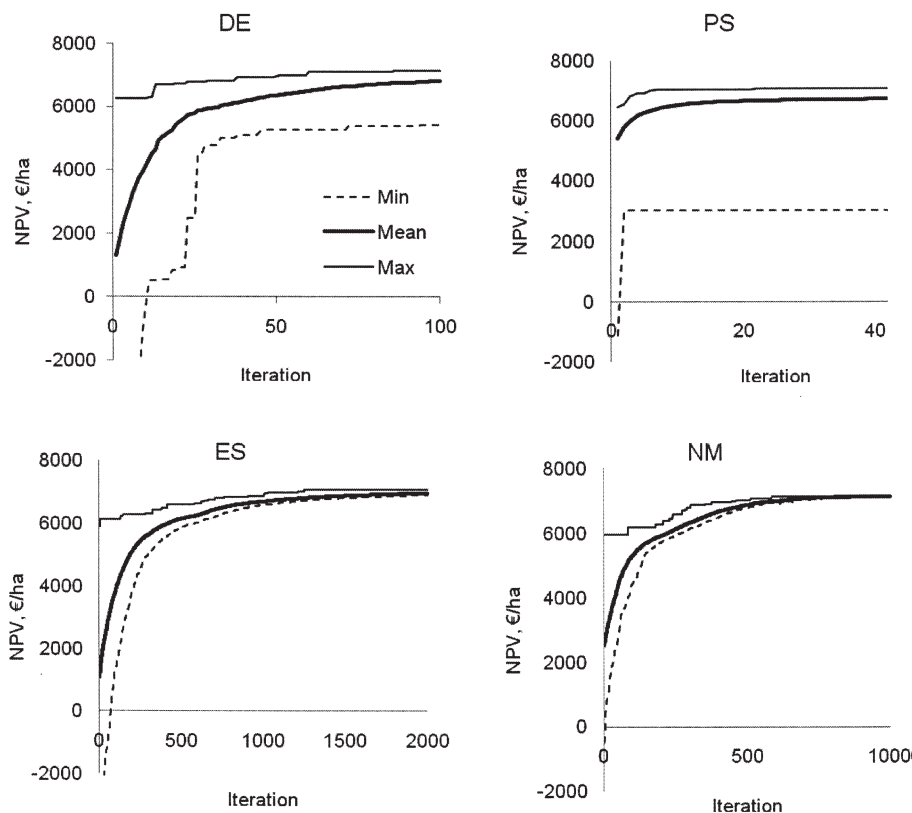


Fig. 3. Development of the objective function value (NPV) of the worst and best solutions and the mean NPV of all solutions in the test stand during the search process of differential evolution (DE), particle swarm optimization (PS), evolution strategy optimization (ES) and Nelder-Mead (NM) method.

were small, maximally less than two percent. DE was better than HJ in 64% of stands and worse than HJ in 11% of stands. A NPV difference of 1 €/ha was used as the limit to rank a method better than another. NM, which had the poorest average performance among the population-based methods, was better than HJ in 43% of stands and worse than HJ in 34% of stands. It is noteworthy that every method was many times the best: DE was the best in 54% of stands and NM in 24% of stands. Sometimes several methods shared the first position. If these cases are not counted, DE was (alone) the best in 29%, PS in 13%, HJ in 7%, ES in 6% and NM in 3% of stands. All five methods gave the same objective function value (with an accuracy of 1 €/ha) for 18% of stands. HJ was the fastest method, followed by NM and

ES. DE and PS were 5 to 10 times slower than the other methods.

The NPVs of the optimal solutions found by the population-based methods correlate closely with the NPV of the HJ-solution (Fig. 4). There were several stands for which DE found a clearly better solution than HJ but only one stand for which HJ was clearly superior to DE. PS was also clearly better than HJ in several stands but only in a few stands clearly worse than HJ. ES was clearly superior to HJ in some stands but in no stand clearly worse than HJ. NM was clearly superior to HJ in four stands and never clearly poorer than HJ. From the clear superiority of population-based methods in several stands it may be concluded that sometimes HJ gets trapped to a local optimum.

Table 4. Summary of the optimization results for 719 stands with the Hooke-Jeeves method (HJ), differential evolution (DE), particle swarm optimization (PS), evolution strategy (ES) and Nelder-Mead method (NM).

Variable	Optimization method				
	HJ	DE	PS	ES	NM
Mean NPV, €/ha	4815	4899	4881	4874	4847
NPV, % of HJ	100.00	101.74	101.37	101.22	100.57
Better than HJ, %	-	64	56	58	43
Worse than HJ, %	-	11	20	18	34
Best method, %	27	54	41	33	24
Best alone, %	7	29	13	6	3

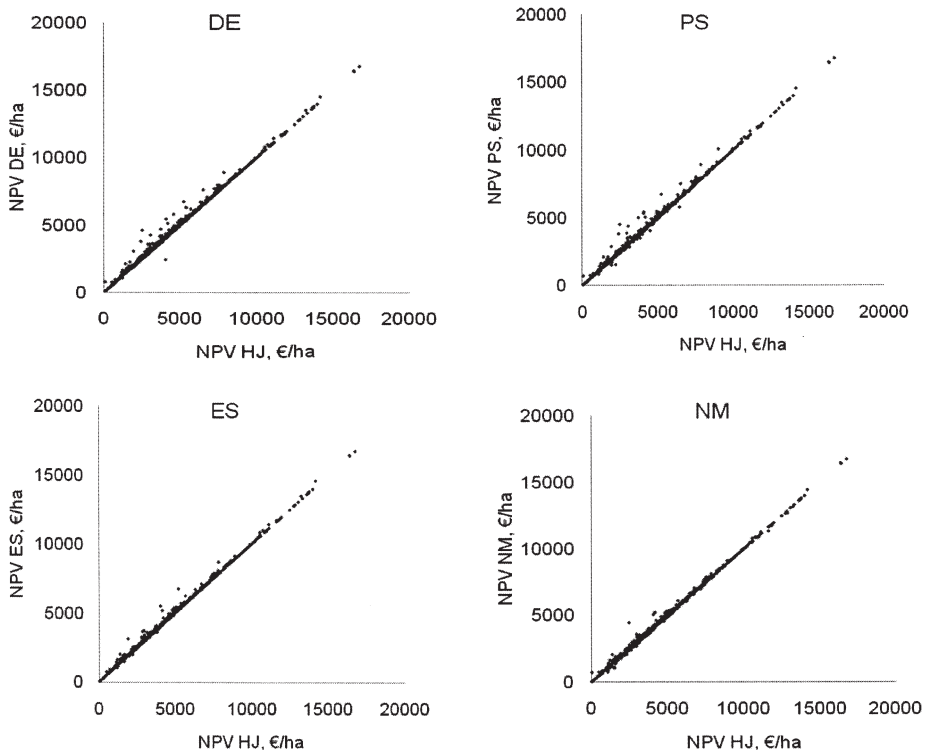


Fig. 4. Correlation of the objective function value found by differential evolution (DE), particle swarm optimization (PS), evolution strategy (ES) and Nelder-Mead method (NM) with the objective function value found by Hooke-Jeeves method (HJ) in 719 stands.

4 Discussion

The main conclusion that can be drawn from this study is that all of the tested population-based methods work well in the optimization of stand management. All are competitive with the Hooke-Jeeves method which has been used much in forestry. The advantage of the population-based methods is that they are easy to use, program, and implement. DE worked best according to several criteria but differences between the methods were minimal. Therefore, any of the five methods tested in this study could be used. It is noteworthy that the method that ranked worst among the population-based methods, namely the NM method, was the best method in 24% of stands. The HJ method, which produced a slightly lower mean NPV for the 719 stands than the other methods, was the best in 27% of stands. An interesting topic for further research would be to analyze how the ranking of methods depends on stand characteristics.

The performance of each method depends on its parameters. This study used parameter values recommended in literature. An exception was that a larger than recommended population was used in the NM method: the population size was $5m$ instead of the usual $m+1$ (m is the number of decision variables). This modification slightly improved the method: the mean NPV of the 719 stands was 4847 €/ha when the method operated with $5m$ solution vectors and 4767 €/ha (1.7% less) when $m+1$ solution vectors were used. The maximum number of iterations was not taken from literature but such a value was used for each method which was found sufficient for convergence. The maximum number of iteration varied from 50 to 2000 among the four population-based methods. However, iteration has a different meaning in different methods: in DE and PS, all solutions are tested or updated during every iteration whereas ES and NM produce only one new trial per iteration (except when shrinking operation is done in NM).

Of the tested methods, HJ was the fastest and DE and PS by far the slowest, approximately 10 times slower than HJ. The NM method with a population size of $m+1$ would have been slightly faster than HJ. The long computing times of DE and PS are mainly due to the high maxi-

imum number of iterations and because the other stopping criterion (small difference between the best and worst solution) was rarely met in these methods. Therefore, in many problems, most of the computing time was probably wasted for the generation of inferior trial solutions. As can be seen from Fig. 2, the objective function value of the worst solution vector approaches to the best value very slowly in DE and PS, implying that DE and PS maintain population diversity much longer than ES and NM, which always try to replace the worst solution by a better one. Therefore, if computing time is an issue, other stopping criteria should be developed for DE and PS such as the difference between the mean and the best solution or the number of iterations that go without any improvement in the best solution. In this study, computing time was not regarded important, and the maximum number of iterations was given a high enough value to guarantee that the search process was not terminated prematurely.

Evolution strategy resembles genetic algorithms (Reeves 1993) in the sense that recombination and mutation are used to produce offspring. However, genetic algorithms are used mainly for combinatorial problems (to optimize integer variables) and ES is designed for continuous decision variables. Another difference is that, opposite to GA, there is no stochasticity in ES when selecting members for the next generation. Also, the population members of GA do not have their own evolution strategy parameters. The other population-based methods also have similarities with GA: solution vectors are combined to obtain a new vector. Random variation may be added to the elements of the recombinant, this operation corresponding to mutation. The only deterministic method (after generating the initial solutions) is the Nelder-Mead method. However, also this method uses recombination: all solutions except the worst are combined to obtain a centroid solution, which is then recombined with the worst solution to obtain a reflected, expanded, or contracted solution.

Particle swarm optimization (PS) resembles ant colony optimization (Dorigo 1997, Zeng et al. 2007) in the sense that both methods use memory and social component in the search process. Ant colony optimization implements this by pheromone trails which are stronger for good solutions. When new trial solutions are formed

(ants look for new paths) the process is affected by the amount of pheromone in different places, the effect of pheromone trails corresponding to the cognitive and social components of the PS method. Opposite to PS, ant colony optimization is designed for combinatorial problems.

One way to further improve the methods tested in this study would be to optimize the parameters of each method (Pukkala and Heinonen 2006). Most probably the improvements would be not much in most cases since the used parameter values were based on earlier experience. Of the four population-based methods, evolution strategy is the most versatile since it can be implemented in many different ways. Evolution strategies are commonly denoted as $(n/\rho + \lambda)$ -ES, where n is the size of the parent population, ρ the number of parents that are recombined to make an offspring, and λ the number of offspring produced in one generation (Bayer and Schwefel 2002). The ES used in this study may therefore be denoted as $(n/2 + 1)$ -ES. In addition to population size, changeable parameters include the number of offspring produced in one generation, number of parents used to make a recombinant, the manner of selecting the parents, the method of recombining the parents, and mutation parameters. Therefore, evolution strategy may be the method that offers most room for further experimentation in forestry applications.

The standard metaheuristics that are used in combinatorial optimization, namely genetic algorithm, simulated annealing and tabu search (Reeves 1993, Borges et al. 2002), can be modified so that they work with continuous decision variables (e.g. Corana et al. 1987, Wikström and Erikson 2000, Chelouah and Siarry 2000). Therefore, one subject of further research would be a systematic analysis of the performance of these metaheuristics in the optimization of stand management.

References

- Bazaraa, M.S., Sherali, H.D. & Shetty, C.M. 1993. Nonlinear programming. Theory and algorithms. Second edition. John Wiley & Sons, Inc., Hoboken. 639 p. ISBN 0-471-55793-5.
- Bayer, H.-G. & Schwefel, H.-P. 2002. Evolution strategies. A comprehensive introduction. *Natural Computing* 1: 3–52.
- Borges, J.G., Hoganson, H.M. & Falcão, A.O. 2002. Heuristics in multi-objective forest planning. In: Pukkala, T. (ed.). *Multi-objective forest planning*. Kluwer Academic Publishers, Dordrecht. p. 119–151.
- Chelouah, R. & Siarry, P. 2000. A continuous genetic algorithm designed for the global optimization of multimodal functions. *Journal of Heuristics* 6(2): 191–213.
- Corana, A., Marchesi, M., Martini, C. & Ridella, S. 1987. Minimising multimodal functions of continuous variables with the "simulated annealing" algorithm. *ACM Transactions on Mathematical Software (TOMS)* 13(3): 262–280.
- Dorigo, M. & Gambardella, L.M. 1997. Ant colonies for the traveling salesman problem. *BioSystems* 43: 73–81.
- Hoganson, H.M., Borges, J.G. & Wei, Y. 2008. Coordinating management decisions of neighboring stands with dynamic programming. In: von Gadow, K. & Pukkala, T. (eds.). *Designing green landscapes. Managing Forest Ecosystems* 15: 187–221.
- Hooke, R. & Jeeves, T.A. 1961. "Direct search" solution of numerical and statistical problems. *Journal of Association for Computing Machinery* 8: 212–229.
- Hynynen, J., Ojansuu, R., Hökkä, H., Siipilehto, J., Salminen, H. & Haapala, P. 2002. Models for predicting stand development in MELA system. *Finnish Forest Research Institute Research Papers* 835. 116 p.
- Kaya, I. & Buongiorno, J. 1987. Economic harvesting of uneven-aged northern hardwood stands under risk. *Forest Science* 33: 889–907.
- Kennedy, J. & Eberhart, R.C. 1995. Particle swarm optimization. *Proceedings of the 1995 IEEE International Conference on Neural Networks (Perth, Australia)*, IEEE Service Center, Piscataway, NJ, IV. p. 1942–1948.
- Laasasenaho, J. 1982. Taper curve and volume equations for pine, spruce and birch. *Communications of the Finnish Forest Research Institute* 108. 74 p.
- Malinen, J., Kilpeläinen, H., Piira, T., Redsvén, V., Wall, T. & Nuutinen, T. 2007. Comparing model-based approaches with bucking simulation-based approach in the prediction of timber assortment recovery. *Forestry* 80(3): 309–321.

- Mehtätalo, L. 2002. Valtakunnalliset puukohtaiset tukkivähennysmallit männyille, kuuselle, koivulle ja haavalle. Metsätieteen aikakauskirja 4/2002: 575–591.
- Miina, J. & Pukkala, T. 2000. Using numerical optimization for specifying individual-tree competition models. *Forest Science* 46(2): 277–283.
- Möykkynen, T., Miina, J. & Pukkala, T. 2000. Optimizing the management of a *Picea abies* stand under risk of butt rot. *Forest Pathology* 30(2): 65–76.
- Nelder, J.A. & Mead, R. 1965. A simplex method for function minimization. *The Computer Journal* 7: 308–313.
- Pukkala, T. 2005. Metsikön tuottoarvon ennustemallit kivennäismaan männiköille, kuusikoille ja rauduskoivikoille. *Metsätieteen aikakauskirja* 3/2005: 311–322.
- & Heinonen, T. 2006. Optimizing heuristic search in forest planning. *Nonlinear Analysis: Real World Applications* 7(2006): 1284–1297.
- & Miina, J. 2005. Optimising the management of a heterogeneous stand. *Silva Fennica* 39(4): 525–538.
- Reeves, C.R. 1993. *Modern heuristic techniques for combinatorial problems*. John Wiley & Sons, Inc. 320 p.
- Storn, R. & Price, K. 1997. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization* 11: 341–359.
- Trasobares, A. & Pukkala, T. 2005. Optimising the management of uneven-aged *Pinus sylvestris* L. and *Pinus nigra* Arn. mixed stands in Catalonia, north-east Spain. *Annals of Forest Science* 61: 747–758.
- Valsta, L. 1992a. A scenario approach to stochastic anticipatory optimization in stand management. *Forest Science* 38: 430–447.
- 1992b. An optimization model for Norway spruce management based on individual-tree growth models. *Acta Forestalia Fennica* 232. 20 p.
- Wikström, P. & Erikson, L.-O. 2000. Solving the stand management problem under biodiversity-related considerations. *Forest Ecology and Management* 126(3): 361–376.
- Zeng, H., Pukkala, T., Peltola, H. & Kellomäki, S. 2007. Application of ant colony optimization for the risk management of wind damage in forest planning. *Silva Fennica* 41(2): 315–332.

Total of 27 references