Canadian Journal of Forest Research
Revue canadienne de recherche forestière

Meta optimization of stand management with population based methods

Journal:	<i>Canadian Journal of Forest Research</i>
Manuscript ID	cjfr-2017-0404.R1
Manuscript Type:	Article
Date Submitted by the Author:	24-Jan-2018
Complete List of Authors:	Jin, Xingji; Northeast Forestry University School of Forestry, Department of Forest Management Pukkala, Timo; University of Eastern Finland Li, Fengri; Northeast Forestry University School of Forestry, Department of Forest Management
Keyword:	population-based method, Hooke-Jeeves method, Korean pine, multifunctional forestry, multi-objective
Is the invited manuscript for consideration in a Special Issue? :	N/A

SCHOLARONE™
Manuscripts

1 **Meta optimization of stand management with population based methods**

2 Xingji Jin

3 Department of Forest Management, School of Forestry, Northeast Forestry University,
4 Harbin, Heilongjiang, P.R. China 150040; email: xingji_jin@163.com

5

6 Timo Pukkala

7 University of Eastern Finland, P.O. Box 111, 80101 Joensuu, Finland;
8 email: timo.pukkala@uef.fi

9

10 Fengri Li

11 Corresponding author

12 Department of Forest Management, School of Forestry, Northeast Forestry University,
13 Harbin, Heilongjiang, P.R. China 150040; email: fengrili@126.com

14 **Abstract**

15 The amount of different products and services obtained from forests depends on several
16 management decisions such as thinning years, thinning intensity, thinning type and rotation
17 length. The relationships between management actions and the various outputs obtained
18 from forests are complicated. This makes stand management optimization challenging,
19 especially if the number of simultaneously maximized outputs and the number of optimized
20 variables are high. The direct search method of Hooke and Jeeves (HJ) has been used much
21 in stand management optimization. In recent years, population-based methods have been
22 proposed as an alternative to the HJ method. The performance of a population-based
23 method depends on its parameters such as number iterations and population size (number
24 of solution vectors used in the population-based method). This study used two-level meta
25 optimization to simultaneously optimize the parameters of a population-based method and
26 the management schedule of a stand. Four population-based methods were analysed:
27 differential evolution (DE), particle swarm optimization (PS), evolution strategy optimization
28 (ES) and the method of Nelder and Mead (NM). With optimal parameter values, DE and PS
29 found the best stand management schedules, followed by ES and NM. DE and PS performed
30 better than HJ. Therefore, DE and PS should be used more in forest management and their
31 search algorithms should be further developed.

32 **Keywords:** population-based method, Hooke-Jeeves method, Korean pine, multifunctional
33 forestry, multi-objective optimization

34

35 Introduction

36 Optimization of the management of an even-aged stand typically consists of finding the best
37 thinning years, thinning intensities, thinning types and rotation length for one full rotation.

38 It is then assumed that similar rotations are repeated to infinity. The objective variable in
39 most optimization cases is the net present value of future incomes and costs. However, as
40 forestry is seen more and more as multifunctional activity, multi-objective utility functions
41 have been increasingly used as the objective function (e.g., Jin et al. 2017). Another
42 approach to multi-objective optimization is to convert all benefits into monetary units and
43 then maximize the total net present value of all products and services (Hartmann 1976;
44 Koskela et al. 2007).

45 Typical of stand management optimization is that the relationships between management
46 parameters (e.g., cutting years) and objective function are not smooth. There may be
47 several local optima and even instant changes in objective function value, for instance when
48 trees reach the minimum dimensions for a valuable timber assortment. Therefore, the
49 optimization problem is rather complicated. The production model is often a simulation
50 model, consisting of sub-models for diameter increment, survival, stem taper, biomass, etc.
51 The simulation model may include step-wise, non-smooth relationships.

52 The method of Hooke and Jeeves (1961) has been used much during the past decades to
53 optimize stand management (Haight and Monserud 1990; Valsta 1992; Pasalodos-Tato et al.
54 2009; Jin et al. 2017). It belongs to the category of non-linear optimization methods, which
55 does not require that the objective function is continuous and smooth. The method does
56 not calculate derivatives (Bazaraa et al. 1993). However, the method of Hooke and Jeeves
57 (HJ) does not find the optimal solution with certainty as it may be trapped to local optima.

58 Therefore, it is common to combine the HJ direct search with random search and repeat the
59 optimization several times to increase the likelihood that the global optimum is found.

60 Because of the above limitations of the HJ method, it is worthwhile to inspect alternative
61 methodologies to optimize stand management. One category of alternative methods are
62 population-based methods, which operate with several solution vectors (population of
63 solutions), instead of only one vector as in HJ. Examples of population-based methods are
64 the Nelder and Mead method (Nelder and Mead 1965), evolution strategy optimization
65 (Bayer and Schwefel 2002), particle swarm optimization (Kennedy and Eberhart 1995) and
66 differential evolution (Storn and Price 1997). Population-based methods have been used in
67 forestry only a few times (Pukkala 2009; Pukkala et al. 2010; Arias-Rodil et al. 2015). There
68 are only two studies in the field of forest management, which compare the performance of
69 population-based methods to HJ and to each other (Pukkala 2009; Arias-Rodil et al. 2015).

70 Each population-based method has a few parameters, such as population size and number
71 of iterations. Population size is the number of simultaneous solution vectors used in the
72 search algorithm of the population-based method. The performance of the method depends
73 on its parameter values. In previous studies (Pukkala 2009; Pukkala et al. 2010; Arias-Rodil
74 et al. 2015), the parameters were taken from literature or they were based on trial-and-
75 error. It was not guaranteed that the parameter values were optimal. Therefore, the earlier
76 studies may not give the right picture on the performance of different methods when they
77 are used in the best possible way.

78 This study used a two-stage optimization approach to stand management optimization. The
79 upper-level optimization problem consisted of finding the best set of parameters for a
80 population-based method. The lower-level problem consisted of solving the stand

management problem with every parameter combination tested during an upper-level optimization run (Fig. 1). In earlier literature, this procedure has been called as super-optimization, hyper-optimization, or meta-optimization (Mercer and Sampson 1978; Grefenstette 1986). It produces both the optimal parameter values for the population based method, and the optimal solution for the stand management problem. The comparison of alternative population-based methods is “fair”, since each method is used with optimal parameter values. Another advantage of “meta optimization” is that it finds better stand management schedules, compared to cases in which population-based methods or other optimization algorithms are used with less appropriate parameter values.

90

91 **Methods**

92 *Stand management problem*

The stand management optimization problem of this study consisted of finding the optimal even-aged management schedule for a Korean pine (*Pinus koraiensis* Siebold & Zucc.) plantation in multifunctional forestry. The same young initial stand and the same objective function were used as in Jin et al. (2017). Stand development was simulated using individual-tree models for diameter increment, tree height, survival, stem taper and biomass. Carbon stock was calculated from biomass. As Korean pine produces edible seeds, the annual seed yield of the stand was also calculated for every stand state, using an empirical model (see Jin et al. 2017 for details).

101 A multi-objective utility function was maximized, in which the utility consisted of net
 102 present value, wood production, seed yield, and average carbon stock of the stand during
 103 the rotation. The objective function was

$$104 \quad U = 0.25 \times (NPV/NPV_{\max}) + 0.25 \times (WP/WP_{\max}) + 0.25 \times (S/S_{\max}) + 0.25 \times (C/C_{\max}) \quad (1)$$

105 where U is utility, NPV is net present value calculated with a 2% discount rate (RMB/ha), WP
 106 is mean annual wood harvest (m^3/ha), S is mean annual seed yield (kg/ha), and C is mean
 107 annual carbon stock of live biomass, including both below- and above-ground parts of trees
 108 (tons/ha). $NPV_{\max}$, $WP_{\max}$, $S_{\max}$ and $C_{\max}$ are the highest possible values of objective
 109 variables, found by single-objective optimization.

110 The management of one rotation was optimized with three thinning treatments. For each
 111 thinning, the optimized variables were: number of years from beginning or previous
 112 thinning, and two parameters of a logistic curve that expressed the thinning intensity as a
 113 function of diameter (see Jin et al. 2017 for details):

$$114 \quad p(d) = \frac{1}{1 + \exp(a_1(a_2 - d))} \quad (2)$$

115 where $p(d)$ is the proportion of harvested trees when dbh is d cm, and a_1 and a_2 are
 116 parameters to be optimized. Parameter a_2 gives the diameter at which thinning intensity is
 117 0.5, and a_1 defines the type of thinning. The last optimized variable was the number of
 118 years since last thinning to final felling. The number of simultaneously optimized variables
 119 was 10.

120

121

122 *Overview of the meta optimization method*

123 The meta optimization comprised two simultaneous optimization processes (Fig. 1). The
124 upper level optimization problem consisted of finding optimal parameter values for those
125 population-based methods, which were used to optimize the management of the Korean
126 pine stand:

127

128 $\text{Max } U = f(\mathbf{y})$
129 subject to (3)
130 $\mathbf{y} \in \mathbb{R}^k$

131

132 where U is utility calculated with Equation 1 for an optimized stand management schedule,
133 and $\mathbf{y}$ is a vector of k real numbers (in this study, a vector of parameter values of the
134 population-based method used to optimize stand management). The method of Hooke and
135 Jeeves (1961) was used in the upper level optimization. During a HJ run, HJ passed different
136 sets of parameter values to the population-based method, which solved the stand
137 management problem with each combination of parameter values (Fig. 1). The objective
138 function value of the optimal stand management schedule found by the population-based
139 method was passed back to HJ. The same procedure was repeated many times, until the
140 stopping criterion of HJ was met.

141 Figure 1

142 *Optimized population-based methods*

143 The tested population-based methods were the Nelder and Mead (1965) method (NM),
144 evolution strategy optimization (ES, Bayer and Schwefel 2002), particle swarm optimization
145 (PS, Kennedy and Eberhart 1995) and differential evolution (DE, Storn and Price 1997). All

four methods were used in the same way as described in Pukkala (2009). The parameters optimized for each method were the same as used in Pukkala (2009).

The population size of the Nelder and Mead method (NM) is usually one more than the number of optimized decision variables, which would be $10+1=11$ in our stand management optimization problem. However, the population size can also be different from this rule (Pukkala 2009). We tested both variants of NM. In the latter case, population size was optimized together with the other parameters of NM. It turned out that NM with a fixed population size (11 in our case) was clearly inferior to NM with optimized population size. Therefore, only NM with optimized population size was included in the analyses of this study.

In all four methods, the initial population was generated by drawing uniformly distributed random numbers from a range [DVmin, DVmax] defined for each decision variable (DV) and shown in Table 1. Then, the stand management problem was solved with each DV combination (for each member, or solution vector, of the population) and the obtained utility value of the stand management schedule was used as the “fitness” value of the member. The solution found by the population based method was the best solution vector of the population at the end of last iteration.

Table 1

Nelder and Mead method (NM)

166 After generating the initial population of solution vectors, NM finds the best ($\mathbf{x}_b$), the worst
 167 ($\mathbf{x}_w$) and the second-worst ($\mathbf{x}_s$) solution vector. Then, a new candidate solution ($\mathbf{x}_r$) is
 168 generated using reflection:

$$169 \quad \mathbf{x}_r = \mathbf{x}_m + \alpha(\mathbf{x}_m - \mathbf{x}_w)$$

170 where $\mathbf{x}_m$ is the centroid of all solutions except $\mathbf{x}_w$:

$$171 \quad \mathbf{x}_m = \sum \mathbf{x}_i (i \neq w) / (n-1)$$

172 If $\mathbf{x}_r$ is at least equally good as $\mathbf{x}_s$ but worse than $\mathbf{x}_b$, $\mathbf{x}_r$ replaces $\mathbf{x}_w$ and the iteration is
 173 terminated. If $\mathbf{x}_r$ is better than $\mathbf{x}_b$, an expansion point $\mathbf{x}_e$ is calculated from

$$174 \quad \mathbf{x}_e = \mathbf{x}_m + \gamma(\mathbf{x}_r - \mathbf{x}_m)$$

175 If $\mathbf{x}_e$ is better than $\mathbf{x}_r$, $\mathbf{x}_e$ replaces $\mathbf{x}_w$ and the iteration is terminated. Otherwise, if $\mathbf{x}_r$ is better
 176 than $\mathbf{x}_s$, $\mathbf{x}_r$ replaces $\mathbf{x}_w$ and the iteration is completed. If $\mathbf{x}_r$ is worse than $\mathbf{x}_s$, a contraction
 177 point is calculated using one of the following two equations

$$178 \quad \mathbf{x}_c = \mathbf{x}_m + \beta(\mathbf{x}_r - \mathbf{x}_m), \text{ if the fitness of } \mathbf{x}_r \text{ is between those of } \mathbf{x}_w \text{ and } \mathbf{x}_s$$

$$179 \quad \mathbf{x}_c = \mathbf{x}_m + \beta(\mathbf{x}_b - \mathbf{x}_m), \text{ if } \mathbf{x}_r \text{ is worse than } \mathbf{x}_w$$

180 The contraction point $\mathbf{x}_c$ replaces $\mathbf{x}_w$ and the iteration is completed. If none of the
 181 operations (reflection, expansion, contraction) finds a point better than $\mathbf{x}_w$, shrinking takes
 182 place, in which all $\mathbf{x}_i$ except $\mathbf{x}_b$ are updated as follows

$$183 \quad \mathbf{x}_i^{\text{updated}} = \mathbf{x}_i + \delta(\mathbf{x}_i - \mathbf{x}_m)$$

184 A new iteration starts after the shrinking operation. The solution vector of NM is $\mathbf{x}_b$ at the
 185 end of the last iteration. The NM algorithm has six parameters

- 186 • Population size (m)

- 187 • Number of iterations (n)
- 188 • Reflection parameter (α)
- 189 • Expansion parameter (γ)
- 190 • Contraction parameter (β)
- 191 • Shrink parameter (δ)

192

193 *Evolution strategy optimization (ES)*

194 In ES, each solution vector $\mathbf{x}_i$ of the population has an associated strategy vector $\mathbf{s}_i$. The
 195 initial values of the elements of strategy vector are obtained from

$$196 \quad \mathbf{s}_i = \sigma \mathbf{x}_i$$

197 In the ES algorithm used in this study, iteration consisted of replacing the worst member of
 198 the population ($\mathbf{x}_w, \mathbf{s}_w$) by a combination of the best member ($\mathbf{x}_b, \mathbf{s}_b$) and a randomly
 199 selected member ($\mathbf{x}_a, \mathbf{s}_a$). These parents produced a recombinant ($\mathbf{x}_f, \mathbf{s}_f$) as follows:

$$200 \quad \mathbf{x}_f = 0.5(\mathbf{x}_b + \mathbf{x}_a)$$

$$201 \quad \mathbf{s}_f = 0.5(\mathbf{s}_b + \mathbf{s}_a)$$

202 This recombinant was mutated to obtain an offspring ($\mathbf{x}_o, \mathbf{s}_o$), which replaced the worst
 203 member of the population:

$$204 \quad \mathbf{s}_o = \mathbf{s}_f \times e^{\tau \times N(0,1)}$$

$$205 \quad \mathbf{x}_o = \mathbf{x}_f + \mathbf{s}_o \times N(0,1)$$

206 where $\tau = 1/\sqrt{n}$, and $N(0,1)$ is a normally distributed random number with mean equal to 0
 207 and standard deviation equal to 1. ES has only three parameters

- Population size (m)
- Number of iterations (n)
- Initialization parameter of the strategy vectors (σ)

Particle swarm optimization (PS)

In PS, all solution vectors $\mathbf{x}_i$ (which in PS are called particles) are updated during every iteration. Each particle has an associated vector of velocities ($\mathbf{v}_i$), which tells how much different elements of the solution vector (decision variables of the stand management optimization problem) are changed when $\mathbf{x}_i$ is updated. In addition, the best solution vector found so far by particle i ($\mathbf{x}_i^b$, “particle best”) is kept in the memory since it affects the movement of the particle. The movement also depends on the best solution found so far by all particles. This solution is called as “global best” and denoted as $\mathbf{x}^g$.

All initial velocities were zeros. The velocities and locations of particles were updated as follows

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

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

where w is co-called inertial constant, c_1 and c_2 are parameters which determine how much the particle is directed towards its own best-so-far solution ($\mathbf{x}_i^b$) and the global best solution ($\mathbf{x}^g$), and r_1 and r_2 are random numbers uniformly distributed between 0 and 1.

The parameters of PS are:

- Population size (m)

- 229 • Number of iterations (n)
- 230 • Inertial constant (w)
- 231 • Effect of particle best on particle's movement (c_1)
- 232 • Effect of global best on particle's movement (c_2)

233

234 *Differential evolution* (DE)

235 In DE, every solution vector $\mathbf{x}_i$ of the population is compared to a trial vector $\mathbf{x}_t$ at every
 236 iteration. If $\mathbf{x}_t$ is better than $\mathbf{x}_i$, it replaces $\mathbf{x}_i$. Therefore, opposite to NM and ES, several
 237 solution vectors may change during one iteration. The test vector is produced by selecting
 238 elements randomly from $\mathbf{x}_t$ and a noise vector ($\mathbf{y}_i$). The probability of selecting an element
 239 from $\mathbf{x}_i$ is p and the probability of selecting from $\mathbf{y}_i$ is $1-p$. The noise vector is a combination
 240 of three randomly selected vectors, $\mathbf{x}_A$, $\mathbf{x}_B$ and $\mathbf{x}_C$:

$$241 \quad \mathbf{y}_i = \mathbf{x}_C + \lambda(\mathbf{x}_A - \mathbf{x}_B)$$

242 DE has four parameters:

- 243 • Population size (m)
- 244 • Number of iterations (n)
- 245 • Probability to select from $\mathbf{x}_i$ (p)
- 246 • Parameter of the formula for noise vector (λ)

247

248 Figures 2 and 3 contain visualizations of the four population-based methods in a single
 249 optimization run. In each method, the population members become gradually more similar

to each other when the optimization proceeds (Fig. 2). There are, however, some differences between the methods. In ES and NM, the variation within the population often decreases quite soon so that the final population consists of almost similar members. Typical of PS are stray particles, some of which may be quite different from the main group of particles.

Figures 2 and 3

In NM and ES, only one population member (the worst) is changed during an iteration. Since most members do not change, the number of times a certain member changes is low (Fig. 3). In DE, all members which are worse than a test vector are replaced by the test vector during an iteration, which leads to a higher number of changes than in NM and ES. The number of changes is the highest in PS, in which every member (particle) moves at every iteration. Contrary to all other methods, the best location found by a certain particle ("particle best") is not necessarily the ending location of the particle.

Parameter optimization method

The direct search method of Hooke and Jeeves (1961; Bazaraa et al. 1993) (henceforth referred to as HJ) was used to find the optimal parameters for the population-based methods. Each parameter had a starting value (Start in Table 2) and a range, which was used to calculate the step size of the parameter in HJ search ([Min,Max] in Table 2). The initial step size was $0.1 \times (\text{Max}-\text{Min})$. In addition, every parameter had a lowest and highest allowed value (Low and High in Table 2). If the HJ direct search attempted to test values beyond [Low, High], they were replaced by Low or High.

272 Table 2

273 The HJ direct search consists of alternating exploratory search and pattern search modes
274 (Oczyszka 1984). One optimized variable at a time is changed during exploratory search
275 whereas several variables may change simultaneously in pattern search, depending on how
276 they were changed during the previous exploratory search. The step size is halved after the
277 pattern search, after which a new exploratory search is started. The search is stopped when
278 the step size is smaller than a predefined stopping criterion, which in our study was equal to
279 0.01 times the initial step size.

280 Since HJ may be trapped to a local optimum, the direct search was repeated another time.
281 The second direct search was started from the best of 100 sets of randomly selected
282 combinations of optimized variables (parameters of the population-based methods). The
283 random values were uniformly distributed between the Min and Max values shown in Table
284 2.

285 The population-based method optimized the stand management optimization problem with
286 each combination of parameters tested during a HJ run. Since the population-based
287 methods use random numbers, repeated optimizations do not necessarily find the same
288 solution, which makes the HJ direct search more unreliable. Therefore, the stand
289 management optimization problem was solved 10 times with every parameter combination,
290 and the mean utility value of these searches was returned to the HJ algorithm. If the
291 population-based method used more than 100 seconds to solve the stand management
292 problem, the parameter combination was penalized, to prevent HJ from suggesting
293 parameters, which would lead to very slow search. Only very few parameter combinations

were penalized, which means that the 100-second limitation did not have any major effect on the optimal parameter values of the population-based methods.

The two repeated HJ direct searches produced two sets of “optimal” parameters for each population-based method. Since the two sets differed, it was concluded that the process of determining the optimal values involves uncertainty. To find out which parameters are the most critical for a good search process and what are their suitable ranges, all direct search steps of the two HJ runs and all the 100 random searches were used as a data set to fit a regression model, which indicated how the utility value of the stand management schedule depended on the parameters of the population-based methods. Different transformations of parameters were used as potential predictors in linear regression analysis, and statistically significant parameters (untransformed or transformed) were selected as predictors.

Comparison of population-based methods

To further compare the performance of the four population-based methods to each other, the management schedule of the Korean pine plantation was solved with each method, using 1, 2, 3, 4 or 5 thinning treatments during a rotation. The utility value and the time consumption were recorded for each solution. The HJ method was also included in this comparison as a reference. Note that we used HJ for two purposes: first, to optimize the parameters of the population-based methods, and second, to optimize the management of Korean pine stand.

Results

Optimal parameter values

The optimal parameter values of NM were: population size 30, number of iterations 599, and the parameters for reflection, expansion, contraction and shrink 1.00, 2.71, 0.51 and 0.51, respectively (column 1st DS, Table 3). They were found in the first HJ direct search. The values found in the second direct search (2nd DS in Table 3) differed considerably for some parameters, particularly population size and expansion, which may indicate either that the performance of NM is not sensitive to these parameters, or the two HJ direct searches found different local optima.

Table 3

The two optimizations for ES parameters produced rather similar results (Table 3). The largest difference was in the number of iterations, which was either 1287 or 1888. The optimal population size was about 100, and the strategy parameter was about 0.1.

The two direct searches for optimizing PS produced different results, suggesting that they found different local optima (Table 3). The second direct was clearly better, and it resulted in a population size of 92 with 72 iterations. The “global best” had a stronger effect on particles’ movements than the “particle best” since parameter c_2 had a larger value (2.61) than parameter c_1 (2.14).

Also in DE the two HJ direct searches resulted in clearly different optima, the second being clearly better. The optimal parameter set consisted of 99 solution vectors with 110 iterations. When a test vector was produced, the probability to select its element from the

current solution vector of the member was 0.18 and the probability to select from noise vector was $1-0.18=0.82$. The parameter for producing a noise vector was 0.60.

Models for utility

To find out how sensitive the results (in our case, the utility index calculated for the stand management schedule) of a population-based method are to its parameter values, all parameter value combinations tested during the two direct searches and the 100 random searches (Fig. 4) were used to model the relationships between parameters and utility.

Figure 4

The following models were fitted:

$$\text{NM: } U = 0.567 + 0.0761\ln(m) - 0.000936m - 0.335/n + 0.0157\ln(\alpha) - 0.0167\alpha + 0.00708\ln(\beta) - 0.0274\beta + 0.00631\ln(\delta) - 0.00332\delta^2$$

$$\text{ES: } U = 0.555 + 0.0736\ln(m) - 0.000892m - 1.47/n + 0.0102\ln(\sigma) - 0.0929\sigma$$

$$\text{PS: } U = 0.717 + 0.0156\ln(m) - 0.000136m - 0.219/n + 0.00285\ln(w) + 0.00058c_1 + 0.00856c_2 - 0.00136c_2^2$$

$$\text{DE: } U = 0.785 + 0.0619\ln(m) - 0.0157Vm - 0.133/n + 0.0534\ln(\rho) - 0.207V\rho + 0.213V\lambda - 0.180\lambda$$

where m is population size, n is number of iterations; α , β , δ are the reflection, contraction and shrink parameters of NM; σ is the strategy parameter of ES, w , c_1 and c_2 are PS parameters describing the inertial effect (w) and the effect of particle best (c_1) and global best (c_2); and ρ and λ are the test vector and noise vector parameters of DE.

According to these models, NM is most sensitive to population size and contraction parameter (Fig. 5). The number of iterations should be 300–400, after which the improvement obtained with higher number of iterations is very small. The effect of expansion parameter was not significant, suggesting that the performance of NM was not sensitive to this parameter.

Figure 5

The model for ES utility suggests that population size should be around 75, strategy parameter should be 0.1, and the number of iterations should be 700 or more (Fig. 6). The model is well in line with the results of the two HJ direct searches (Table 3).

Figure 6

The model developed for PS did not reveal clear optima for the parameter values as increasing value of all parameters except c_2 increased the predicted utility (Fig. 7). For c_2 (effect of global best), the optimal parameter value was 3.0. However, the performance of PS was rather insensitive to all its parameters if population size was 80 or more and the number of iterations was at least 50.

Figure 7

The model for DE suggests that population size should be around 60 and the number of iterations should be at least 100. The optimal value of the test vector parameter (p) is 0.3 and the noise vector parameter (λ) should be about 0.35. The model reveals that the performance of DE is sensitive to large deviations from these values (Fig. 8).

Figure 8

Table 3 also contains the parameter values derived from modelling and the values used in Pukkala (2009). The mean utility of 10 repeated optimizations is also shown for all parameter combinations. Parameter values derived from modelling were the best for ES and PS whereas the second HJ direct search produced the best parameter values for DE. The DE parameters used in Pukkala (2009) were clearly inferior to those found in the optimizations of this study. This result is in line with the diagrams of Figure 8, which shows that too large values for the test and noise vector parameters greatly decrease the performance of DE. For NM, none of the parameter combinations found in this study outperformed the parameters used in Pukkala (2009).

Utility and time consumption of different methods

The four population-based methods were compared to each other by optimizing stand management with 1, 2, 3, 4 and 5 thinning treatments using the best direct search values of the parameters. Each problem was solved 10 times with each of the methods, to have information about the similarity of repeated optimization runs. The stand management problems were also solved with the HJ method.

The results show clearly that NM was not competitive with the other methods (Fig. 9, Table 4): the utilities produced by repeated runs varied much, the mean utilities of 10 repeated runs were low, and the highest utility of the 10 runs was lower than in the other methods. Typical of HJ was high variation between repeated direct searches, especially in cases where the number of optimized variables was high (4 or 5 thinnings, with 13 or 16 simultaneously optimized variables). The best solutions found by ES were worse than the best solutions

found by HJ (Fig. 9). On the other hand, ES solutions varied less than HJ solutions (Fig. 9, Table 4).

Figure 9, Table 4

DE produced the highest utility values with the smallest variation between repeated runs. PS also performed well, and its competitiveness seemed to improve when the optimization problem became more complicated; when the number of thinnings was five, PS found better solutions than DE (Fig. 9). The computing time of PS increased more slowly than with DE with increasing difficulty of the stand management problem (Table 4).

Optimal management schedule obtained with different methods

When the stand was managed according to the optimal 3-thinning schedule found by different methods (the best of the 10 repeated optimization runs), the growing stock volume was predicted to develop as shown in Figure 10. All five “optimal” management schedules were different. The suggested optimal rotation length was approximately 80 years with all five methods, and the thinning treatments suggested by PS and ES almost coincided. The thinning prescriptions found by DE were also close to those found by PS and ES.

Figure 10

Although Figure 10 suggests that ES found almost the same solution as DE and PS (which produced the best solutions), the utility value of the ES solution was lower than in the solutions found by PS and DE. This may be related to thinning type, which may have been optimized better in PS and DE.

423

424 **Conclusions**

425 The optimal parameter values of the population-based methods are of the same magnitude
426 as used earlier by Pukkala (2009) and Arias-Rodil et al. (2015). A clear difference are the test
427 vector and noise vector parameters of DE, which should be clearly smaller than suggested
428 by Pukkala (2009). Arias-Rodil et al. (2015) used parameter values, which are close to the
429 optimized DE parameter values found in this study.

430 In Arias-Rodil et al. (2015), DE outperformed all the other methods. Pukkala (2009) also
431 concluded that DE was the best of the four population-based methods, and better than HJ.
432 However, the margin of DE to the other methods was very small, most probably because the
433 parameter setting of DE was poorer in Pukkala (2009), as compared to Arias-Rodil et al.
434 (2015) and the current study.

435 Opposite to the current study, Arias-Rodil et al. (2015) found that PS was clearly inferior to
436 DE and HJ. The most significant difference in its parameters was that the inertial parameter
437 (w) was lower in Arias-Rodil et al. (2015), as compared to the current study and that of
438 Pukkala (2009). Otherwise, the conclusions about the ranking of the five different methods
439 (HJ and four population-based methods) is fairly similar in all three studies: DE is the best
440 with small variation between repeated runs, followed by HJ, PS, ES and NM. According to
441 the current study, PS may be ranked as the second best, and its performance seems to
442 improve when the optimization problem becomes more complicated.

443 In the management problems of Pukkala et al. (2010), representing uneven-aged forestry,
444 the overall ranking of the four methods was as follows: PS (best) > ES > DE > NM (worst).

The ranking depended on problem formulation, but a clear result was that NM was not competitive with the other methods, and PS was good in all problem formulations.

Based on the current study, two of the population-based methods, namely DE and PS, can be ranked better than the much-used HJ, which means that forest managers should adopt these methods and use them more. The superiority of the population-based methods is most evident in complicated optimization problems involving more than 10 decision variables; HJ becomes unreliable in these problems.

This study used optimization to identify the best possible parameter values of the population-based methods. Modelling was used as an additional tool to analyse the sensitivity of the quality of the solution to different parameters. The models also show what is a suitable population size in different methods and what is the minimum number of iterations required for a good result. In addition, the models show that some parameters must not differ much from their optimal values. For example, the test and noise vector parameters of DE (p and λ), should be 0.15–0.6, and the contraction parameter of NM should be 0.2–0.3.

Previous studies (Pukkala 2009, Arias-Rodil et al. 2015) used earlier literature and simple experimentation to set the parameters of the population-based methods. Compared to Pukkala (2009), our optimizations were able to improve the parameter set of three out of four methods. The improvement was very clear for DE, which did not perform well in our problem with the parameters proposed by Pukkala (2009). On the other hand, no improvement was found for NM parameters. This outcome is most probably related to the high variability between repeated NM runs (high degree of stochasticity in NM), which made it difficult for HJ to find good parameters for NM.

The optimal parameters of population-based methods most probably depend on the type and complexity of the problem. However, the effect of problem may not be drastic in stand management optimization since the performance and ranking of the methods was nearly the same when the number of optimized variables ranged from 4 (1 thinning) to 16 (5 thinnings). However, further studies on the relationships between parameter values and type and complexity of the optimization problem may bring new light to this issue.

It is also possible to develop the population-based methods further. For example, Pukkala et al. (2010) suggested that the solutions will improve if the problem is solved repeatedly so that the best member of the previous run becomes a member of the initial population of a new run. In addition, fine-tuning the methods so that the parameter values change during the optimization run, may also offer possibilities to improve the methods. It is also possible to develop various “hybrids” of the four methods analysed in this study, for instance by producing a part of the initial solution vectors of certain method by using another optimization method.

There are also other optimization methods inspired by phenomena occurring in nature, such as bee colony optimization (Pham et al. 2015) and intelligent water drops (Shah-Hosseini 2009). The development of these kind of “novel” optimization methods has been criticized since their theoretical background is often weak (Iztok et al. 2013; Sörensen 2013). The four methods tested in this study are all used widely and at least two of them seem to outperform the method of Hooke and Jeeves (1961). Therefore, analysing the performance of these methods and developing them further continue to be relevant research topics also in the future.

Acknowledgments

This research was financially supported by the National Natural Science Foundation of China (31600511), and the Fundamental Research Funds for the Central Universities of the People's Republic of China (2572017CA04).

References

- Arias-Rodil, M., Pukkala, T., Gonzalez-Gonzalez, J.R., Barrio-Anta, M. and Dieguez-Aranda, U. 2015. Use of depth-first search and direct search methods to optimize even-aged stand management: a case study involving maritime pine in Asturias (northwest Spain). *Can. J. For. Res.* 45(10): 1269-1279.
- Bayer, H.-G. and Schwefel, H.-P. 2002. Evolution strategies. A comprehensive introduction. *Natural Computing* 1: 3-52.
- Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. 1993. Nonlinear programming. Theory and algorithms. Second edition. John Wiley & Sons, Inc., Hoboken. pp. 1-639. ISBN 0-471-55793-5.
- Grefenstette, J.J. 1986. Optimization of control parameters for genetic algorithms. *IEEE Transactions Systems, Man, and Cybernetics* 16 (1): 122-128.
doi:10.1109/TSMC.1986.289288.
- Haight, R.G. and Monserud, R.A. 1990. Optimizing any-aged management of mixed-species stands. I. Performance of a coordinate search process. *Can. J. For. Res.* 20(1): 15-25.
- Hartman, R. 1976. The harvesting decision when a standing forest has value. *Econ. Inquiry* 14(1): 52-58.

- 513 Hooke, R. and Jeeves, T. 1961. "Direct search" solution of numerical and statistical
514 problems. *J. Assoc. Comput. Mach.* 8(2): 212-229.
- 515 Fister, I. Jr, Yang, X-S, Fister, T., Brest, J. and Fister, D. 2013. A Brief Review of Nature-
516 Inspired Algorithms for Optimization. *Elektroniški Vestnik* 80(3): 1-7.
- 517 Jin, X., Pukkala, T. and Li, F. 2016. Fine-tuning heuristic methods for combinatorial
518 optimization in forest planning. *Eur. J. For. Res.* 135(4): 765-779.
- 519 Jin, X., Pukkala, T. and Li, F. 2017. Optimal management of Korean pine plantations in
520 multifunctional forestry. *J. For. Res.* 28(5): 1027-1037.
- 521 Kennedy, J. and Eberhart, R.C. 1995. Particle swarm optimization. *Proceedings of the 1995*
522 *IEEE International Conference on Neural Networks (Perth, Australia)*, IEEE Service Center,
523 Piscataway, NJ, IV. pp. 1942-1948.
- 524 Koskela, E., Ollikainen, M. and Pukkala, T. 2007. Biodiversity Conservation in Commercial
525 Boreal Forestry: The Optimal Rotation Age and Retention Tree Volume. *For. Sci.* 53(3): 443-
526 452.
- 527 Mercer, R.E. and Sampson, J.R. 1978. "Adaptive search using a reproductive meta-plan".
528 *Kybernetes*. 7(3): 215-228. doi:10.1108/eb005486.
- 529 Nelder, J.A. and Mead, R. 1965. A simplex method for function minimization. *The Computer*
530 *Journal* 7(4): 308-313.
- 531 Osyczka, A. 1984. Multicriterion optimization in engineering with FORTRAN programs. Ellis
532 Horwood, Chichester. pp. 1-178.

- 533 Pasalodos-Tato, M., Pukkala, T. and Castedo-Dorado, F. 2009 Models for the optimal
534 management of *Pinus radiata* D. Don in Galicia (north-western Spain) under risk of fire. Allg.
535 Forst Jagd. Z. 180(11/12): 238-249.
- 536 Pham, D.T. and Castellani, M. 2015. A comparative study of the bees algorithm as a tool for
537 function optimisation. Cogent Engineering 2(1): 1091540.
- 538 Pukkala, T. 2009. Population-based methods in the optimization of stand management.
539 Silva Fenn. 43(2): 261-274.
- 540 Pukkala, T., Lähde, E. and Laiho, O. 2010. Optimizing the structure and management of
541 uneven-sized stands in Finland. Forestry 83(2): 129-142.
- 542 Shah-Hosseini, H. 2009. The intelligent water drops algorithm: a nature-inspired swarm-
543 based optimization algorithm. Int. J. Bio-Inspired Comp. 1 (1/2): 71–79.
544 doi:10.1504/ijbic.2009.022775.
- 545 Sörensen, K. 2013. Metaheuristics—the metaphor exposed. Int. Trans. Oper. Res. 22(2015):
546 3-18. doi:10.1111/itor.12001.
- 547 Storn, R. and Price, K. 1997. Differential evolution – a simple and efficient heuristic for global
548 optimization over continuous spaces. Journal of Global Optimization 11(4): 341-359.
- 549 Valsta, L. 1992. An optimization model for Norway spruce management based on individual-
550 tree growth models. Acta For. Fenn. 232:1-20.

1 Table 1. Ranges of the decision variables of the stand management optimization problem. These
2 ranges were used to generate the initial solution vectors of the population-based search methods.

Decision variable	DVmin	DVmax
First thinning		
• Years to 1 st thinning	0	40
• Parameter a_1 of the thinning intensity curve	-2	2
• Parameter a_2 of the thinning intensity curve	10	50
Second thinning		
• Years to 2 nd thinning	5	40
• Parameter a_1 of the thinning intensity curve	-2	2
• Parameter a_2 of the thinning intensity curve	10	50
Third thinning		
• Years to 3 rd thinning	5	40
• Parameter a_1 of the thinning intensity curve	-2	2
• Parameter a_2 of the thinning intensity curve	10	50
Final felling		
• Years to final felling	5	40

3 See Equation 2 for the explanation of parameters a_1 and a_2 of the thinning intensity curve.

4

5

Table 2. Ranges and starting values for the optimization of population-based search method. “Start” is the starting value of the first Hooke and Jeeves direct search; [Min, Max] is the range from which the initial step size was calculated (Initial step = $0.1 \times (\text{Max} - \text{Min})$); “Low” is the lowest and “High” is the highest accepted value of the parameter.

Parameter	Low	Min	Start	Max	High
Nelder-Mead method					
Population size (m)	2	3	20	100	200
Iterations (n)	3	10	500	1000	2000
Reflection (α)	0.01	0.1	1.0	5.0	10.0
Expansion (γ)	0.1	0.5	2.0	10.0	20.0
Contraction (β)	0.001	0.01	0.5	5.0	10.0
Shrink (δ)	0.001	0.01	0.5	5.0	10.0
Evolution strategy optimization					
Population size (m)	2	3	20	100	200
Iterations (n)	3	10	100	1500	10000
Strategy parameter (σ)	0.001	0.01	0.2	2.0	5.0
Particle swarm optimization					
Population size (m)	2	3	20	100	200
Iterations (n)	1	1	50	200	500
Inertial parameter (w)	0.01	0.5	1.0	1.5	3.0
Effect of particle best (c_1)	0.01	0.2	2.0	4.0	10.0
Effect of global best (c_2)	0.01	0.2	2.0	4.0	10.0
Differential evolution					
Population size (m)	2	3	20	100	200
Iterations (n)	1	1	50	200	500
Test vector parameter (ρ)	0.0001	0.001	0.1	0.8	0.99
Noise vector parameter (λ)	0.01	0.1	0.7	0.99	0.9999

12 Table 3. Optimal values of the parameters of population-based search methods in the 1st and 2nd
 13 direct search (DS) and according to model expressing the objective function value as a function of
 14 the parameters. The best parameter combination is shown in boldface. The utility value is the mean
 15 utility of 10 repeated optimizations.

Parameter	1 st DS	2 nd DS	Model	Pukkala
Nelder-Mead method				
Population size (m)	30	86	75	50
Iterations (n)	599	566	400	1000
Reflection (α)	1.00	4.78	0.90	1.00
Expansion (γ)	2.71	3.76	(2.00)	2.00
Contraction (β)	0.51	0.62	0.25	0.5
Shrink (δ)	0.51	0.41	0.90	0.5
Utility	0.7646	0.7511	0.7589	0.7725
Evolution strategy optimization				
Population size (m)	107	96	75	150
Iterations (n)	1888	1287	700	2000
Strategy parameter (σ)	0.10	0.11	0.12	0.20
Utility	0.7718	0.7660	0.7734	0.7712
Particle swarm optimization				
Population size (m)	30	92	100	150
Iterations (n)	70	72	100	50
Inertial parameter (w)	0.997	0.555	1.00	0.95
Effect of particle best (c_1)	2.38	2.14	2.50	2.00
Effect of global best (c_2)	2.00	2.61	2.50	2.00
Utility	0.7758	0.7860	0.7890	0.7833
Differential evolution				
Population size (m)	29	99	60	50
Iterations (n)	50	110	100	100
Test vector parameter (ρ)	0.18	0.18	0.30	0.9
Noise vector parameter (λ)	0.79	0.60	0.35	0.7
Utility	0.7721	0.7906	0.7844	0.7438

Table 4. Mean and standard deviation (Sdev) of the utility index (U) and average time consumption in 10 repeated optimizations for different stand management problems solved with different methods (HJ = Hooke and Jeeves, NM = Nelder and Mead, ES = evolution strategy optimization, PS = particle swarm optimization, DE = differential evolution). The highest mean utility, smallest standard deviation and shortest time consumption for each problem are shown in boldface.

	HJ	NM	ES	PS	DE
1 thinning (4 optimized variables)					
Mean U	0.7499	0.7540	0.7541	0.7574	0.7615
Sdev of U	0.0062	0.0024	0.0040	0.0045	0.0003
Time, s	2.5	4.0	17.5	56.9	25.9
2 thinnings (7 optimized variables)					
Mean U	0.7771	0.7694	0.7654	0.7848	0.7845
Sdev of U	0.0102	0.0122	0.0077	0.0040	0.0009
Time, s	4.7	5.3	22.7	47.0	44.1
3 thinnings (10 optimized variables)					
Mean U	0.7792	0.7664	0.7721	0.7887	0.7913
Sdev of U	0.0087	0.0105	0.0061	0.0022	0.0024
Time, s	7.2	6.5	21.2	53.7	92.1
4 thinnings (13 optimized variables)					
Mean U	0.7033	0.7446	0.7754	0.7803	0.7913
Sdev of U	0.1997	0.0301	0.0041	0.0067	0.0014
Time, s	12.9	8.1	24.3	73.0	124.5
5 thinnings (16 optimized variables)					
Mean U	0.7303	0.7084	0.7724	0.7824	0.7828
Sdev of U	0.0849	0.0515	0.0115	0.0101	0.0040
Time, s	16.0	9.1	28.4	80.3	138.3

21

1 **Figure captions**

2 Figure 1. Flow chart of the meta optimization method. The upper level optimization employs Hooke
3 and Jeeves algorithm to find optimal parameters for a population-based (PB) algorithm. The stand
4 management problem is solved with every tested combination of parameter values (PV) of the PB
5 algorithm. The purpose is to find such parameters for the PB method, which yield the best possible
6 management schedule for the stand. The stand management schedule is defined by a set of
7 management variables (MV), such as thinning years, thinning intensity and rotation length.

8 Figure 2. Development of the population in different search methods during one optimization run.
9 The large open circles show the initial population, the smaller open circles indicate the population
10 after one third of the iterations, and the black dots show the final population. Two decision variables
11 out of 10 are shown (years to 1st thinning and years from 1st to 2nd thinning).

12 Figure 3. Development of five members of the population in different search methods during one
13 optimization run. The ending values are shown with bullets (except particle swarm optimization).
14 Two decision variables out of 10 are shown (years to 1st thinning and years from 1st to 2nd thinning).

15 Figure 4. Examples of data points used to develop models for the relationship between the
16 parameters of the population-based methods and objective function value. The dataset includes all
17 search steps of two Hooke and Jeeves direct searches and all points of one random search consisting
18 on 100 random combinations of parameters. The examples show the effect of number of iterations
19 and test vector parameter (p) on the performance of differential evolution.

20 Figure 5. Dependence of objective function value (Utility) on the parameters of Nelder-Mead
21 method.

22 Figure 6. Dependence of objective function value (Utility) on the parameters of evolution strategy
23 optimization.

24 Figure 7. Dependence of objective function value (Utility) on the parameters of particle swarm
25 optimization.

26 Figure 8. Dependence of objective function value (Utility) on the parameters of differential
27 evolution.

28 Figure 9. Range of utility values obtained with different methods when stand management was
29 optimized with 1 to 5 thinning treatments and each optimization was repeated 10 times. The darker
30 part of the bar shows the range after excluding 2 worst and 2 best solutions.

31 Figure 10. Development of stand volume in the optimal three-thinning schedule found with different
32 optimizations methods.

33

Draft

34
35

36
37
38
39
40
41
42
43

Figure 1. Flow chart of the meta optimization method. The upper level optimization employs Hooke and Jeeves algorithm to find optimal parameters for a population-based (PB) algorithm. The stand management problem is solved with every tested combination of parameter values (PV) of the PB algorithm. The purpose is to find such parameters for the PB method, which yield the best possible management schedule for the stand. The stand management schedule is defined by a set of management variables (MV), such as thinning years, thinning intensity and rotation length.

44 Figure 2. Development of the population in different search methods during one optimization run.
 45 The large open circles show the initial population, the smaller open circles indicate the population
 46 after one third of the iterations, and the black dots show the final population. Two decision variables
 47 out of 10 are shown (years to 1st thinning and years from 1st to 2nd thinning).

48

49 Figure 3. Development of five members of the population in different search methods during one
50 optimization run. The ending values are shown with bullets (except particle swarm optimization).
51 Two decision variables out of 10 are shown (years to 1st thinning and years from 1st to 2nd thinning).

52

53

54 Figure 4. Examples of data points used to develop models for the relationship between the
 55 parameters of the population-based methods and objective function value. The dataset includes all
 56 search steps of two Hooke and Jeeves direct searches and all points of one random search consisting
 57 on 100 random combinations of parameters. The examples show the effect of number of iterations
 58 and test vector parameter (p) on the performance of differential evolution.

59

60
61 Figure 5. Dependence of objective function value (Utility) on the parameters of Nelder-Mead
62 method.
63

64

65

66 Figure 6. Dependence of objective function value (Utility) on the parameters of evolution strategy
67 optimization.

68

69

70

71 Figure 7. Dependence of objective function value (Utility) on the parameters of particle swarm
72 optimization.

73

74

75

76 Figure 8. Dependence of objective function value (Utility) on the parameters of differential
77 evolution.

78

79

Figure 9. Range of utility values obtained with different methods when stand management was optimized with 1 to 5 thinning treatments and each optimization was repeated 10 times. The darker part of the bar shows the range after excluding 2 worst and 2 best solutions.

Figure 10. Development of stand volume in the optimal three-thinning schedule found with different optimizations methods. The utility value of the management schedule is shown in parentheses.