

Expansion of the minimum-inbreeding seed orchard design to operational scale

Milan Lstibůrek · Jan Stejskal · Alfonsas Misevicius ·
Jiří Korecký · Yousry A. El-Kassaby

Received: 14 November 2014 / Revised: 12 January 2015 / Accepted: 20 January 2015 / Published online: 3 February 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The minimum-inbreeding (MI) seed orchard layout, formulated originally as a global quadratic assignment problem, was expanded into realistic problem sizes that are often encountered in operational forestry, where two modifications were tested: (1) the merging algorithm of independent MI's solutions (i.e., smaller blocks) and (2) the extended global (genetic-tabu) algorithm. Extending the global heuristic algorithm of the quadratic assignment problem seems to be the most efficient strategy. The reported minimum-inbreeding distance of the extended MI scheme was the lowest in comparison to the completely randomized and the randomized, replicated, staggered clonal-row (R^2 SCR) seed orchard design schemes. These conclusions also hold for more complex scenarios when added relatedness among orchard's parents or unequal deployment was considered. This improved MI scheme is suitable to large and complex advanced-generation seed orchards, where many practical constraints have to be jointly considered.

Keywords Forest tree improvement · Advanced-generation seed orchard · Genetic gain and diversity · Quadratic assignment problem

Introduction

The minimum-inbreeding (MI) seed orchard design was introduced by Lstibůrek and El-Kassaby (2010) to accommodate the multiple constraints often encountered during designing advanced-generation orchards, specifically those related to common ancestry (relatedness) among clones, the desire for preferential mating (assortative mating), and the utilization of variable clonal size. These constraints are often present jointly with complex distributional properties (e.g., non-linear deployment) and their high-order interactions (e.g., Hodge and White 1993; El-Kassaby 2003). MI seed orchard layout is formulated as a quadratic assignment problem (QAP), where combination of decisions on clonal assignment is treated jointly during their allocation to the orchard's grid and the impact of one decision cannot be assessed before the other decisions are resolved. In the original QAP formulation, the flow matrix depicts flow volumes between facilities and the distance matrix, the distances between their respective assigned locations (Rardin 1998). The advantages of this approach include (1) single-objective optimization, (2) inclusion of a single flow matrix with individual weighting applied to all genetic entries on the grid, thus meeting all constraints in a single optimization run, (3) global assignment (the resulting solution is theoretically the best combinatorial match between flow and distance matrices), (4) there is no need to consider separation zones or additional constraints, and (5) there are no distributional assumptions on the formulation of flow and distance matrices; thus, the dispersion of planting positions is not

Communicated by J. Beaulieu

Electronic supplementary material The online version of this article (doi:10.1007/s11295-015-0842-5) contains supplementary material, which is available to authorized users.

M. Lstibůrek (✉) · J. Stejskal · J. Korecký
Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamýcká 129, 165 21 Praha 6, Czech Republic
e-mail: lstiburek@fld.czu.cz

A. Misevicius
Department of Multimedia Engineering, Kaunas University of Technology, Studentu st. 50-400a/416a, LT-51368 Kaunas, Lithuania

Y. A. El-Kassaby
Department of Forest Sciences, Faculty of Forestry,
University of British Columbia, 2424 Main Mall,
Vancouver, BC V6T 1Z4, Canada

geometrically constrained and any configurations (including missing spots) are treated optimally. Correspondingly, one can impose any hierarchical structure on individual genetic entries using the flow matrix formulation with any level of complexity. Practically speaking, given that the objective function is meeting the design expectation and proper weighting is applied in the flow matrix, then, the resultant seed orchard grid is optimum, thus giving rise to the minimum expected average inbreeding across the life span of the orchard, where inbreeding is defined as the sum of the inverse distances between all ramets of the same clone over all clones (Eq. 5, “Comparison to alternative seed orchard designs” section).

Quadratic assignment problem is inherently complex (non-deterministic polynomial-time hard), and exact mathematical solutions (e.g., Fischetti et al. 2012; Nyberg and Westerlund 2012) are limited to relatively small problem sizes ($n < 30$), even though new heuristic algorithms (e.g., Davendra and Zelinka 2009; Tosun et al. 2013) are being developed for handling much larger instances (ant algorithms, simulated annealing, genetic algorithms, iterative improvement, or tabu search). The original implementation of the quadratic assignment problem by Lstibůrek and El-Kassaby (2010) was associated with a limitation on problem size. In their comparison, they demonstrated the concept using 10×10 seed orchard block (100 planting positions) in a hypothetical seed orchard. They postulated that problem size is the single most important variable influencing the problem complexity of the MI design and correspondingly proposed two measures: (1) dividing the seed orchard to separate blocks, each solved independently, or (2) using novel meta-heuristic methods to address very large instances of quadratic assignment problems. Both proposals were considered in the present study.

The main objectives of the current study are to (1) expand the MI design concept to operationally feasible large-scale applications, (2) evaluate the impact of relatedness and proportional deployment on the efficiency of the MI scheme, and (3) to compare the resultant MI solutions to existing seed orchard layouts.

Algorithm development

Minimum-inbreeding seed orchard layout is based on a quadratic assignment problem with the following objective function:

$$\min \rightarrow \sum_{i,j=1}^n \sum_{k,l=1}^n c_{i,j} d_{k,l} x_{i,k} x_{j,l} \quad (1)$$

where x is a binary variable, i,j is a pair of clones ($i \neq j$), k,l is a pair of positions ($k \neq l$), and $c_{i,j}$ and $d_{k,l}$ are elements of the C (flow) and D (distance) input matrices, respectively, as

introduced by Lstibůrek and El-Kassaby (2010). The problem is optimized assuming the following constraints:

$$\sum_{i=1}^n x_{i,j} = 1 \quad \forall \quad 1 \leq i \leq n \quad (2)$$

$$\sum_{j=1}^n x_{i,j} = 1 \quad \forall \quad 1 \leq j \leq n \quad (3)$$

$$x_{i,j} \in \{0, 1\} \quad \forall \quad 1 \leq i \wedge j \leq n \quad (4)$$

In the original minimum-inbreeding formulation, elements of the material-flow matrix C denote linear transformations of pairwise coancestries between trees i and j (lower genetic relationship is preferred over the lower distance), and correspondingly, elements of the distance matrix D denote physical distance between the i th and the j th planting position on the orchard's grid (with no assumption on the distribution of planting positions). The objective is therefore to minimize the planting of closely related trees at a close distance (with elevated probability of pollen exchange).

Preparation of input matrices was automated (software tool available from ML) following the identical approach in Lstibůrek and El-Kassaby (2010) with some minor adjustments to handle larger problem sizes. In his tabu search algorithm, Misevicius (2005) merged several existing algorithms into “intensification and diversification” procedures, where the intensification and diversification are applied iteratively to solutions. This original algorithm was deterministically tailored and found suitable to scenarios similar to the 10×10 orchard grid problems as illustrated in Lstibůrek and El-Kassaby (2010). In the present study, two alternatives were tested to handle larger instances: (1) merging algorithm (dividing the seed orchard to separate blocks, each solved independently), and (2) extended global (genetic-tabu) algorithm (using novel meta-heuristic methods to address very large instances of quadratic assignment problems).

Minimum-inbreeding scheme: merging algorithm

We developed a computerized procedure (R code, Supplementary material 1), where the orchard's grid is subdivided into independent blocks; each is solved independently using the original MI algorithm (Lstibůrek and El-Kassaby 2010), and later, blocks are rotated and merged, using a random operator. The algorithm is divided into following steps:

1. Seed orchard layout is divided into independent blocks of equal size N . Clones are randomly assigned into blocks. This often results in unequal numbers of ramets, where the variation in clonal size is minimized to meeting the best integer solution and the total number of genetic entries.

2. Powerful genetic-tabu search algorithm is applied in each block to optimize spatial layout. Note, for the MI design, only one optimum design (heuristic solution) is produced.

Resulting optimized grids are merged into the final seed orchard layout. The merging algorithm (see Fig. 1, for illustration) is divided into two steps: (a) random and independent rotation of each grid and (b) random allocation of genetic entries residing at the same position in all grids into a corresponding $N \times N$ submatrix in the resulting layout. Large number of iterations is run for each scenario, and the resulting layout is selected based on the minimum distance ($d_{\min}$; Lstibůrek and El-Kassaby 2010) with the lowest minimum-inbreeding value attained from the different iterations conducted. The minimization of $d_{\min}$ was performed assuming 10,000 iterations of the merging algorithm.

Minimum-inbreeding scheme: extended global (genetic-tabu) algorithm

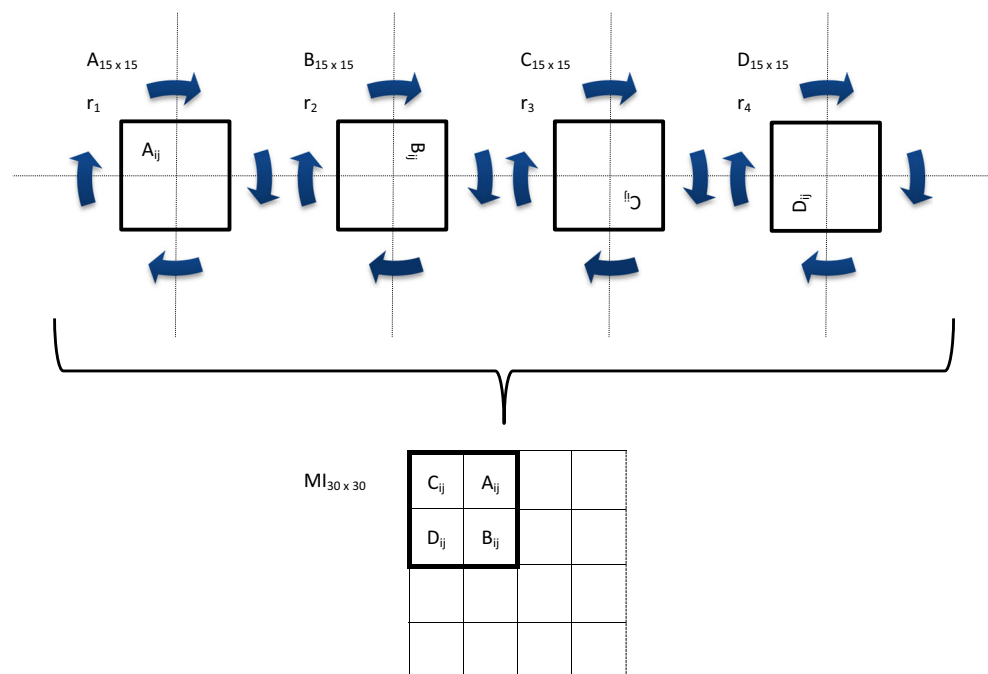
The QAP problem size was extended to accommodate the entire declared seed orchard grid. Solutions are improved by starting from carefully selected mutated local optima, rather than assuming randomly scattered points. Several controlling parameters (Table 1) were tested following the recommendations by Misevicius (2004; p. 70). The recommended values of controlling parameters were chosen on the basis of the experience of previous experimentation with the instances from the library of the QAP instances—QAPLIB (<http://anjos.mgi.polymtl.ca/qaplib/>). For more details on the

Table 1 Input seed orchard's design parameters of the powerful genetic-tabu search algorithm

Parameter	Value	Description
P_SIZE	20	Population size
GEN_N	3	Number of generations
OFFSPR_N	1	Number of offspring per generation
ITER_N	7	Number of improvement iterations
τ	n	Number of tabu search iterations
RS_N	6	Number of restarts
S_FACT	1	Selection factor
LTLS_FACT	0.030	Lower tabu list size factor
HTLS_FACT	0.070	Higher tabu list size factor
M_RATE_FACT_1	0.070	Mutation rate factor (1)
M_RATE_FACT_2	0.120	Mutation rate factor (2)
ET	0.001	Population entropy threshold

calibration of the controlling parameters, the reader is referred to Misevicius et al. (2009). Resultant deterministic parameters were obtained to meeting very large MI scenarios. Correspondingly, the QAP computer code was modified by A.M. to allow for even larger MI scenarios instances and the software tool is available from the same author. In the present study, we used a standard PC with Windows 7 OS, Intel Core i7 processor 2.93 GHz and 4 GB of RAM and each MI orchard global solution required approximately 8 h of computing time (meeting the declared parameters).

Fig. 1 Merger of four independent MI blocks (schemes). Blocks are depicted as A , B , C , and D , each with 15×15 (rows $\times$ columns). These matrices were independently rotated (original, 90° , 180° , or 270°) using a random operator r . Following the rotation, corresponding j th elements in all matrices were randomly merged into 2×2 submatrices in the resultant $MI_{30 \times 30}$ matrix



Comparison to alternative seed orchard designs

Comparison to alternative seed orchard designs was conducted utilizing both MI algorithms (merging algorithm (MA) and extended global (genetic-tabu) algorithm (EGA)), and it was based on a square 30×30 orchard grid (900 planting positions) (note: the developed algorithm is not limited to this particular size). Throughout the study, the task was to allocate the 900 genetic entries (trees) to their respective planting positions.

As the genetic-tabu search algorithm is a heuristic procedure (the true mathematical solution remains unknown), it is important (particularly under large scenarios) to (1) set a reference point to any comparison and (2) develop an evaluation criterion to compare the MI solution to the reference point. Accordingly, as a reference, we used the completely randomized scheme (RD, evaluated over a large number of independent iterations), where all entries were randomly allocated on the orchard's grid.

Resultant seed orchard designs from all algorithms (900 planting positions) were evaluated by the minimum distance ($d_{\min}$) which is estimated as follows (Lstibůrek and El-Kassaby 2010):

$$d_{\min} = \sum_{l=1}^{N_C} \sum_{i=1}^{N_R} \sum_{j=i+1}^{N_R} \frac{1}{d_{il,jl}^2} \quad (5)$$

where d is the distance between the i th and the j th ramets of the l th clone, N_C is the number of clones, and N_R is the number of ramets within a clone. This statistics is used as a proxy to potential inbreeding (selfing: mating among ramets of the same clone). The sum is made across the entire orchard's grid; therefore, it accounts for all pairwise relationships among all genetic entries. More explanatory remarks are provided by Lstibůrek and El-Kassaby (2010).

When half-siblings ($h-s$) were present, $d_{\min}$ calculation was modified as follows:

$$d_{\min} = d_{\min}(\text{selfing}) + \frac{1}{2} d_{\min}(h-s) \quad (6)$$

which corresponds to weighting constants used in the objective function of the MI scheme. While the $d_{\min}$ (selfing) is identical to that used for clonal material (Eq. 5); the $d_{\min}$ ($h-s$) accounts for all pairwise distances among all ramets of related clones, summed across all pairwise relationships.

The following scenarios were considered throughout the study (orchard size=900 positions):

- Scenario 1 All clones were unrelated and non-inbred with two scenarios using $N_C=30$ ($N_R=30$) and $N_C=60$ ($N_R=15$).
- Scenario 2 Mixture of related (four half-sibs) and unrelated (remaining clones were treated as unrelated and

non-inbred) with $N_R=30$ and $N_R=15$ for $N_C=30$ and $N_C=60$, respectively.

- Scenario 3 Unrelated with unequal clonal contribution with three groups of 10 and 20 clones for the 30- and 60-clone scenarios, respectively. The former produced variable clonal sizes of 25, 30, and 35 ramets/clone (sum=250+300+350=900 ramets) while the latter with clonal sizes of 12, 15, and 18 ramets/clone (sum=240+300+360=900 ramets).

Additionally, the MI seed orchards with the MA and EGA algorithms were compared to the randomized, replicated, staggered clonal-row layout (R^2 SCR) (El-Kassaby et al. 2014, software tool available from Y.A.E.) and the randomized (RD) seed orchard layout. Comparisons to the R^2 SCR followed the program instructions where a single-tree plot layout and a separation zone of three ramets were applied across all scenarios. Due to the large stochastic variation, we report averages across 100 independent iterations. In the RD layout, all genetic entries were randomly allocated to respective planting positions using a pseudo-random number generator with no constraint imposed in this layout (i.e., no separation zones). R code to generate RD layout and to calculate $d_{\min}$ for the resultant SO schemes is included (Supplementary Material 2).

Results

As anticipated from the optimization theory (Rardin 1998), the merging algorithm of the MI scheme proved only advantageous under simple scenario #1, where all clones were assumed unrelated with equal representation. Comparing to the MI scheme, the randomized allocation produced 2.6 and 4.2 % higher $d_{\min}$ values for the 30 and 60 clones, respectively. Under more complicated scenarios (2 and 3), the advantage of the MI over the randomized scheme was considerably diminished as indicated by the large $d_{\min}$. We attributed this to the reduced capacity of the merging algorithm (MA) to producing efficient orchards' layouts. Therefore, any further comparisons will be exclusively based on the extended global (genetic-tabu) algorithm (EGA).

Following modifications of the quadratic assignment problem (QAP), the EGA appeared to be an efficient algorithm to handle very large instances of the MI scheme. Results of the 30×30 MI orchard scheme are included in Figs. 2, 3, and 4 (30 clones, scenarios 1–3, respectively). Simple geometrical inspection of the MI scheme under scenario 1 (Fig. 2) reveals symmetrical pattern of separating ramets of the same clone. There are no visible groupings or gaps across the entire seed orchard grid. As a heuristic solution, it should be possible to improve this solution by further extension of the computing time; however, we have seen only negligible gains due to

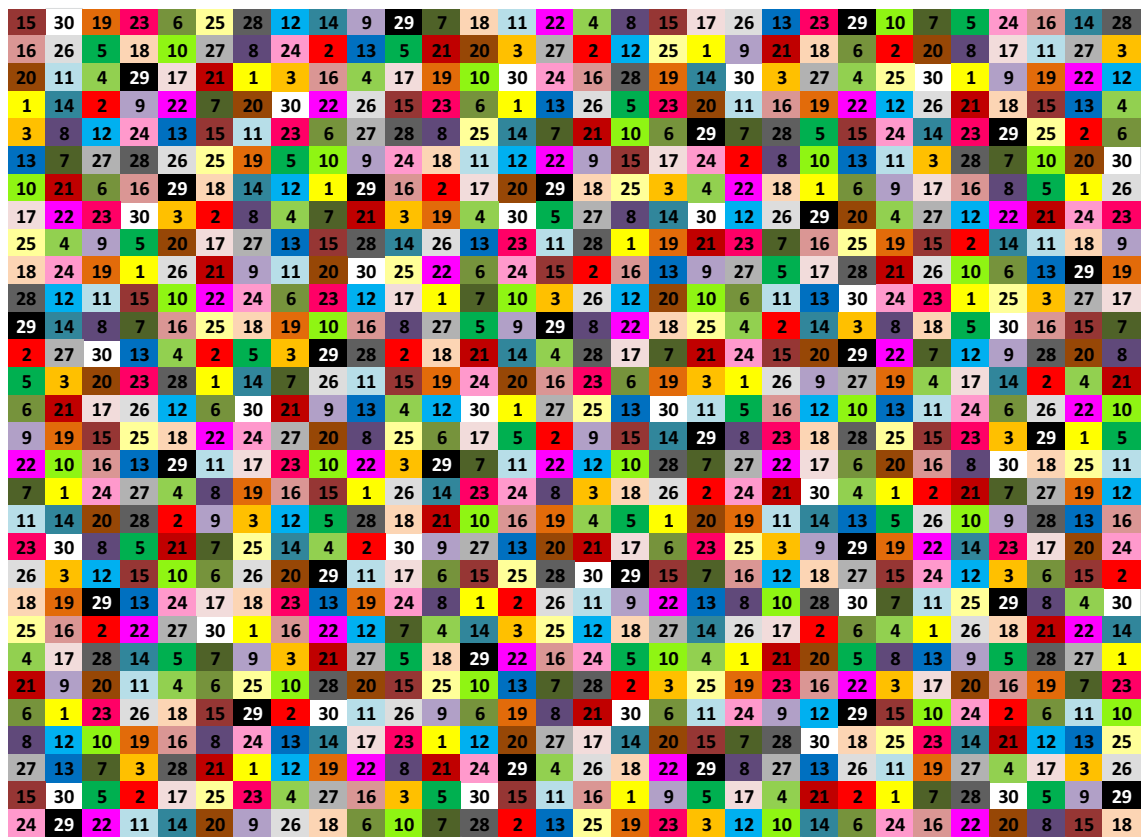


Fig. 2 Optimum global MI-EGA scheme assuming scenario 1 and 30 clones

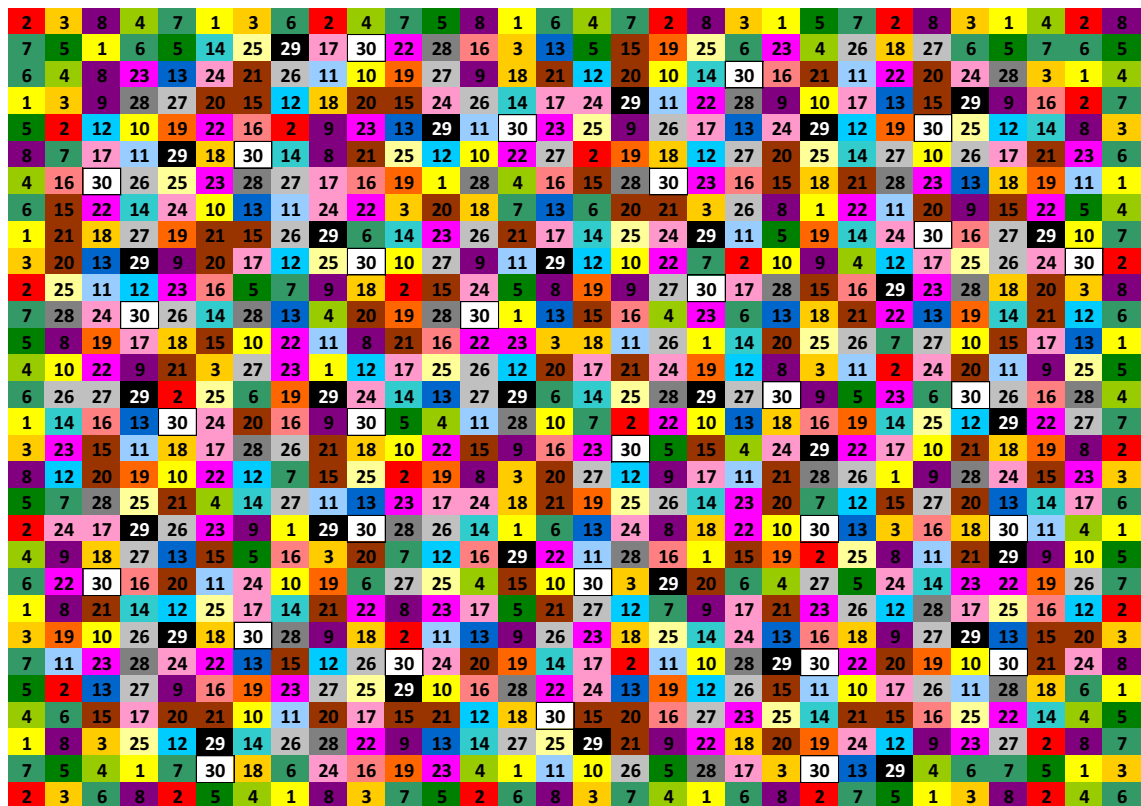


Fig. 3 Optimum global MI-EGA scheme assuming scenario 2 and 30 clones

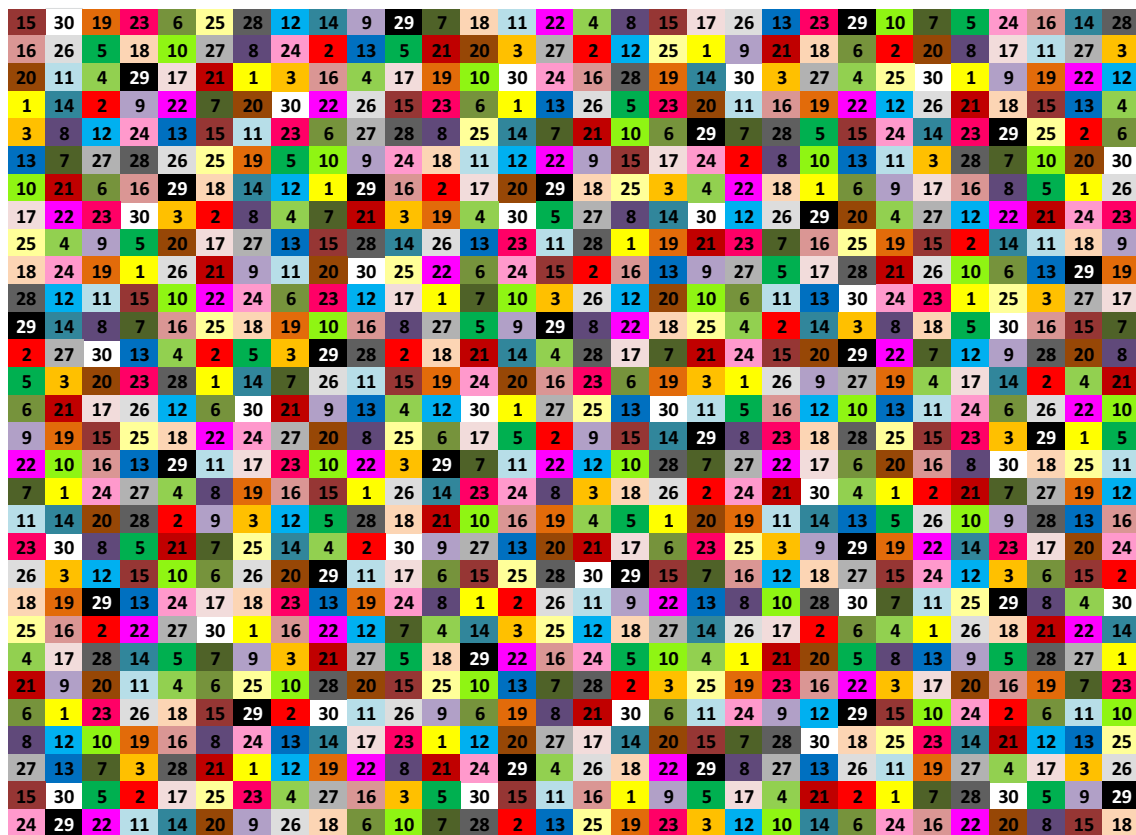


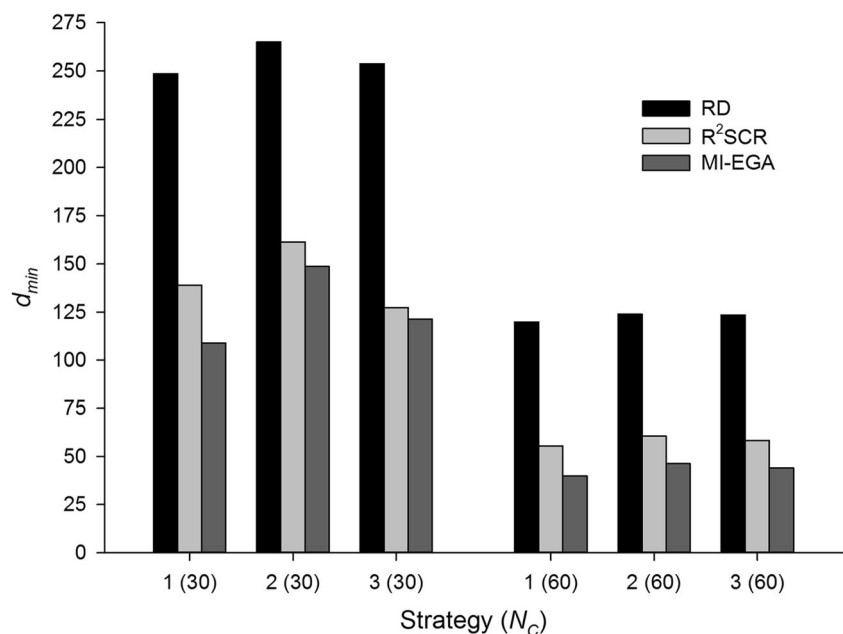
Fig. 4 Optimum global MI-EGA scheme assuming scenario 3 and 30 clones

added computing time (small fractions of the $d_{\min}$) under such instances. Under scenario 2 (Fig. 3), when related clones are present, majority of ramets of the related clones are alternated along the edge. Isolation among the ramets of related clones is visible from the inspection of the same figure. Under scenario 3 (variable clonal sizes), clones with larger sizes tend to prevail

along the orchard's edge, while clones with smaller sizes are more frequent at the orchard's center, following the minimization of the expected inbreeding in seed orchard's crop (Fig. 4).

Resulting $d_{\min}$ values were calculated for the 30×30 seed orchard layouts (R²SCR, MI) and compared to the RD mean (Fig. 5).

Fig. 5 Minimum-inbreeding ($d_{\min}$) values plotted for the RD, R²SCR, and MI-EGA seed orchard layouts for three comparisons scenarios (1 all clones are unrelated and non-inbred, 2 mixture of related and unrelated, 3 unrelated with unequal clone size) for two clonal sizes (30 30×30, 60 60×60)



In the first strategy (no relatedness, equal contributions), the MI design resulted in 78.4 % ($N_C=30$) and 72.2 % ($N_C=60$) of the R^2SCR 's d_{min} . When relatedness among four clones was assumed, the same relative difference was 92.3 % ($N_C=30$) and 76.6 % ($N_C=60$). In the third strategy (unequal clonal contributions), the difference was 95.2 % ($N_C=30$) and 75.4 % ($N_C=60$).

Discussion

This study was initiated to expand the MI scheme of Lstibůrek and El-Kassaby (2010) to accommodate realistic seed orchards' sizes with the added advantage of minimizing the design underlying criteria; namely, minimum-inbreeding layouts (d_{min}). This is particularly important as advanced-generation orchards often harbor common ancestry and are usually established with variable clone sized. Additional advantage of the MI design is capitalizing on the knowledge gained during breeding and testing, thus providing opportunity to increase the probabilities for assortative matings. We developed two MI algorithms, namely the merging algorithm (MA) and the extended global (genetic-tabu) algorithm (EGA) for orchards' layout expansion, and realized that an advantage of the former diminishes under complex scenarios (relatedness and/or unequal contributions); however, the latter was proven to be universally effective under various scenarios. Quantitative comparisons supported the extended global algorithm superiority to the randomized orchard scheme under various orchard sizes and clonal make-up as it produced the lowest d_{min} values. With the current model, additional design improvement is likely possible through extending the computing time to multiple days or by using supercomputing facilities. Additional design improvement (i.e., lower d_{min} values) is also possible as future heuristic algorithms of the quadratic assignment problem become available.

The reported differences in R^2SCR and MI schemes are relatively lower, when assessed on the RD basis, and it seems that both R^2SCR and MI schemes are efficient in separating related genetic entries on the seed orchard grid. The MI scheme outperforms the R^2SCR scheme in this single measure as the d_{min} corresponds to the objective function of the MI scheme. However, from a practical point of view, both seed orchard layouts will significantly reduce mating among related entries, even under more complicated scenarios (i.e., advanced-generation seed orchard layouts).

Probably, the main benefit of the current study is that it provides a theoretical insight into seed orchard layout optimization. There are two theoretical limits underlying the spatial distribution, one provided by the completely randomized scheme (RD) where all restrictions are relaxed and all genetic entries are treated equally. Such a scenario, when

averaged across high number of independent iterations, mimics what could be called a random mating (union of gametes) or panmixia. Clear disadvantage is the lack of separation among the same (ramets of the same clone) or similar (i.e., half-siblings) genetic entries. Second and equally important is the minimum-inbreeding (MI) seed orchard layout, which is a global optimization protocol, leading to the minimum expected inbreeding. There are no separation zones either, but the layout is optimized using global objective function minimization, where the resulting separation is optimized to a given set of input parameters. Thus, theoretically, it should always lead to the best configuration of genetic entries given the objective function is meeting the actual realistic breeder's objective (separation of similar or same genetic entries by distance).

This study also highlights few important points to seed orchards' designs. Specifically, in advanced generations, smaller number of clones are desired (higher selection differential), and in many practical scenarios, a subset of the selected clones would inevitably share common ancestry. As the MI design implies, attention is often directed to intentional selfing avoidance (large weight on separation of ramets of the same clone); however, as the number of clones goes down, this is accompanied by increase in the number of ramets per clone which adds additional levels of complexity. This problem could be further intensified by relatedness among clones or unequal deployment. One logical solution is to reduce the weight on selfing in the MI's objective function, resulting in larger optimization space to accommodate additional parameters. Another possible solution is to merge ramets of the same clone together in clonal-rows without imposing constraints on selfing (El-Kassaby 2003; El-Kassaby et al. 2007, 2014). This makes biological sense as evidence of low selfing rates in conifers supports this notion (El-Kassaby 1995; O'Connell 2003). These low selfing rates are mitigated by the unique reproductive biology of conifers which characterizes by high genetic load, severe inbreeding depression, and polyembryony (Buchholz 1962; Sorensen 1982; Griffin and Lindgren 1985; Williams and Savolainen 1996).

Acknowledgments Support from the National Agency for Agriculture Research (NAZV; grant QJ1320013; M. Lstibůrek), the Ministry of Education, Youth and Sports (Kontakt II; grant LH13021; M. Lstibůrek), the Internal Grant Agency of the Czech University of Life Sciences (CIGA; grant 20144301; J. Korecký), and the Natural Sciences and Engineering Research Council of Canada (Discovery and IRC Grants) and the Johnson's Family Forest Biotechnology Endowment to Y. A. El-Kassaby are highly appreciated.

Data archiving statement In our study, we were developing mathematical model to optimize spatial allocation of trees in seed orchards. We did not use any real-world data of any species, nor any sequencing data. As far as the actual mathematical model, it is thoroughly described in the text. All output data have been prepared and published in figures that are included in the manuscript (the resulting spatial allocation).

References

- Buchholz JT (1962) Origin of cleavage polyembryony in conifers. *Bot Gaz* 81:55–71
- Davendra D, Zelinka I (2009) Optimization of quadratic assignment problem using self organising migrating algorithm. *Comput Inform* 28:1001–1012
- El-Kassaby YA (1995) Evaluation of the tree improvement delivery system: factors affecting the genetic potential. *Tree Physiol* 15:545–550
- El-Kassaby YA (2003) Clonal-row vs. random seed orchard designs: mating pattern and seed yield of western hemlock (*Tsuga heterophylla* (Raf.) Sarg.). *For Genet* 10:121–127
- El-Kassaby YA, Stoehr MU, Reid D, Walsh CG, Lee TE (2007) Clonal-row vs. random seed orchard designs: interior spruce mating system evaluation. *Can J For Res* 37:690–696
- El-Kassaby YA, Fayed M, Klápště J, Lstibůrek M (2014) Randomized, replicated, staggered clonal-row (R^2 SCR) seed orchard design. *Tree Genet Genomes* 10:555–563
- Fischetti M, Monaci M, Salvagnin D (2012) Three ideas for the quadratic assignment problem. *Oper Res* 60:954–964
- Griffin AR, Lindgren D (1985) Effect of inbreeding on production of filled seed in *Pinus radiata* experimental results and a model of gene action. *Theor Appl Genet* 71:334–343
- Hodge GR, White TL (1993) Advanced-generation wind-pollinated seed orchard design. *New For* 7:213–236
- Lstibůrek M, El-Kassaby YA (2010) Minimum inbreeding seed orchard design. *For Sci* 56:603–608
- Misevicius A (2004) An improved hybrid genetic algorithm: new results for the quadratic assignment problem. *Knowl-Based Syst* 17: 65–73
- Misevicius A (2005) A tabu search algorithm for the quadratic assignment problem. *Comput Optim Appl* 30:95–111
- Misevicius A, Rubliauskas D, Barkauskas V (2009) Some further experiments with the genetic algorithm for the quadratic assignment problem. *Inform Tech Control* 38:325–332
- Nyberg A, Westerlund T (2012) A new exact discrete linear reformulation of the quadratic assignment problem. *Eur J Oper Res* 220: 314–319
- O’Connell LM (2003) The evolution of inbreeding in western redcedar (*Thuja plicata*: Cupressaceae). Ph.D. Thesis, University of British Columbia
- Rardin RL (1998) Optimization in operations research. Prentice Hall Inc., 919p
- Sorensen FC (1982) The roles of polyembryony and embryo viability in the genetic system of conifers. *Evolution* 36:725–733
- Tosun U, Dokeroglu T, Cosar A (2013) A robust island parallel genetic algorithm for the quadratic assignment problem. *Int J Prod Res* 51: 4117–4133
- Williams CG, Savolainen O (1996) Inbreeding depression in conifers: implications for breeding strategy. *For Sci* 42:102–117