Contents lists available at ScienceDirect

Forest Ecology and Management

Democratising forest management: Applying multiwinner approval voting to tree selection

Arne Pommerening[a,*], Markus Brill[b], Ulrike Schmidt-Kraepelin[b], Jens Haufe[c]

[a] *Swedish University of Agricultural Sciences SLU, Faculty of Forest Sciences, Department of Forest Ecology and Management, Skogsmarksgränd 17, SE-901 83 Umeå, Sweden*
[b] *Technische Universität Berlin, Efficient Algorithms Research Group, 10587 Berlin, Germany*
[c] *Forest Research, Technical Development, DG1 1QB Ae Village, United Kingdom*

ARTICLE INFO

ABSTRACT

Climate change, biodiversity losses, global health threats, changing recreation patterns are but a few of the many challenges that currently, and for some time to come, the world has to cope with. To address these challenges and to mitigate some of them, ecosystem and particularly conservation management increasingly have to adopt strategies never considered before. One such new possibility is crowdsourcing, a variant of public consultation, where a number of experts are invited and, for example, asked to mark trees that – in their opinion – should be removed in order to improve or restore a forest ecosystem. This type of crowdsourcing has recently been carried out in many European countries and overseas as part of what commonly is referred to as marteloscope. In this paper, we addressed the question of how the rating or voting of such a crowd of experts is best aggregated to obtain one final, consolidated list of trees to be evicted. Standard approval voting often leads to a domination by the majority of voters and important contributions by minority experts are largely ignored. To avoid this and to better represent the pluralism of expertise and opinions in matters, where currently no best-practice guidelines exist, we analysed the effects of three proportional multiwinner rules used in political science by applying them to 50 marteloscope experiments in Great Britain. Our results indicated that proportional rules – particularly in situations where the invited expert markers disagree – achieved a better representation of different opinions than standard approval voting. Proportional rules also act as a safety mechanism reducing risks when the majority decisions prove inappropriate and as a consequence forest development could completely go astray.

1. Introduction

Many global changes such as climate change, global diversity loss and pandemics require forest ecosystem management and particularly forest conservation to pursue completely new, unprecedented avenues (Wagner et al., 2014; Jandl et al., 2019; Ontl et al., 2020). In addition new social trends, partly related to the current COVID-19 pandemic, lead to additional shifts in the demand for ecosystem goods and services. For example, the provision of community forests near urban centres mainly used for recreation (Lacaze, 2000; Petucco et al., 2018; Riccioli et al., 2019) may now be more important than ever to mitigate adverse psychological and social effects of pandemics and to channel forest visitors away from conservation hotspots.

Since many of these new directions in ecosystem management are unprecedented, best-practice guidelines for how to design forest landscapes that are capable of providing the properties needed to achieve

mitigation and conservation are not available. In such situations it can be a useful strategy to invite several experts as part of a public consultation exercise (Petucco et al., 2018) and to ask them to provide their professional opinion through a hands-on selection of trees in the forest (Vítková et al., 2016; Bravo-Oviedo et al., 2020).

An important part of modern, sustainable forest management including forest conservation is concerned with the selective marking and subsequent removal of some of the trees of a forest stand to promote residual trees and other vegetation by providing them with resources and opportunities. In forestry, such operations are traditionally referred to as thinnings (Helms, 1998) and they are meant to steer interactions between trees, and trees and other vegetation to meet clearly defined objectives. When evicting trees in such a way, usually a major decision is taken that typically affects the dynamics of a forest stand for many years if not decades to come (Pommerening et al., 2018). Until recently, forest managers and researchers have often assumed that tree marking

---

for eventual eviction leads to almost unanimous results with hardly any variation given that the staff in question had the same education and thinning instructions. Research starting in the 1990s and particularly recent studies have clearly shown that this assumption is not true (Zucchini and Gadow, 1995; Füldner et al., 1996; Daume et al., 1997; Vítková et al., 2016; Pommerening et al., 2018).

Studying agreement between professionals judging an object is common practice in medicine and part of assessing reliability and reproducibility of decision making as well as quality assurance (Cao et al., 2016; Pommerening et al., 2018). Agreement between individuals selecting trees has not been much considered in forest science and practice until recently. In a forest management and forest science context studying agreement raises the valid question, why experienced forest managers and operators should not be allowed to deviate from one another in terms of marking trees as long as overall objectives and targets are met. Popular opinion in the forestry profession suggests that the general trend rather than the individual tree is important and allowing people's decisions to differ may even reduce the risk of fatal decisions (Pommerening and Grabarnik, 2019).

In some countries such as Switzerland it is not uncommon that several forest managers select trees in one and the same forest stand to discuss their choices in order to eventually arrive at balanced and consolidated final decisions (Junod, pers. communication). This approach of employing the "wisdom of the crowd" (Surowiecki, 2004), i.e. deliberately allowing professionals to "vote" independently for trees they wish to evict, may prove very useful in applications where new, unprecedented management directions are taken. In such novel cases, knowledge about best-practice management is not readily available and crowdsourcing (Ghezzi et al., 2018) by synthesising the opinions of several experts may in fact be the best way forward. In such an approach, the question arises how best to ensure that all tree markers contribute to the final, aggregated list of approved trees for eviction so that this list is representative of all forest managers taking part in the marking and no crucial information is lost.

The objective of this interdisciplinary study was to analyse how approaches from political sciences, specifically multiwinner approval voting (Brams et al., 2019; Brill et al., 2017), applied to tree selection can contribute to an aggregation of individual suggestions for solutions to environmental management which is proportional to the expert forest managers' preferences. We also explored how these proportional rules are likely to influence residual forest structure compared to multiwinner approval voting. In a voting context, these methods were developed to ensure that the outcome of a political election proportionally reflects the opinion of the voters (i.e. the electorate) and not just the opinion of a narrow majority and a very similar intention exists in public consultations on best-practice environmental management.

2. Materials and methods

2.1. Multiwinner approval voting

Approval voting is an electoral system where each voter may select or vote for (= approve of) any number of candidates by including the assessments "yes" (approved) or "no" (not approved), which is commonly translated to "1" and "0" for computer processing. Originally approval voting was devised for single-winner elections (Brams and Fishburn, 1978) and can also easily be extended to multiwinner elections, e.g. to elect the members of a council or a committee. Given the voters' approval information, a *multiwinner approval voting rule* selects a subset of all candidates to form a committee or other body. The most straightforward and commonly applied voting rule simply ranks all candidates according to the number of voters approving of them, also known as a candidate's *approval score*. A committee or other elected body is then formed by iteratively adding candidates, starting with the candidate with the highest approval score until the desired committee size has been reached. We referred to this rule as *standard approval*

voting (AV). Although standard AV is well motivated in the single-winner case (Brams and Fishburn, 1978), the method can lead to undesirable outcomes in the multiwinner case, when potentially candidates put forward by a small majority can be disproportionately elected and the candidates of minority voters are then largely unrepresented. This effect, which is often referred to as *tyranny of the majority*, has been challenged as being incompatible with democratic principles such as proportional representation.

In response to this challenge, several multiwinner approval voting rules have been proposed with the objective of electing more representative committees or other bodies. We referred to these rules as proportional multiwinner rules, as they can be interpreted as generalisations of apportionment methods which are used to assign seats in parliament to political parties following an election (Balinski and Young, 1982; Brill et al., 2018). In this paper, we mainly focussed on proportional rules originally proposed by Thiele (1895) and Phragmén (1895). These rules and their variants have recently gained much attention and have been extensively studied in the voting literature for their potential to produce committees with proportional representation (Aziz et al., 2017; Brill et al., 2017).

2.1.1. Thiele's proportional rules

The Danish polymath Thorvald N. Thiele (1895) proposed a remarkably general class of proportional multiwinner rules. In the context of our paper, we focussed on sequential methods: Starting with an empty committee, these methods iteratively add candidates until the committee is complete. Thiele's sequential methods progressively reduce the weight of a voter's approval, as more and more of his or her approved candidates are elected. To define and specify these rules, we employed the so-called depreciation weights of the two most prominent apportionment methods, one of which was proposed independently by Thomas Jefferson and Victor D'Hondt, the other independently by Daniel Webster and André Saint-Laguë. The former method is also referred to as *sequential proportional approval voting* (Aziz et al., 2017). Following Brams et al. (2019) we referred to these two Thiele methods as "Jefferson" and "Webster" in the remainder of the text.

For describing Jefferson and Webster more formally, let R_j denote the set of voters that voted for candidate j and in any iteration let n_i be the number of candidates voted for by voter i that have already been elected in one of the previous iterations. In each iteration, the *weight* or *deservingness score* w_j of a yet unelected candidate j is given by

$$w_j = \sum_{i \in R_j} \frac{1}{n_i + h} \tag{1}$$

where h is a parameter that is set to $h = 1.0$ when applying the Jefferson method and to $h = 0.5$ when applying Webster. Following a suggestion by Brams et al. (2019) we also tested $h = \frac{1}{10}$, $h = \frac{1}{4}$ and $h = \frac{3}{4}$ in this study.

When computing the deservingness score of a candidate, the contribution of each voter is initially 1 and is then subsequently reduced by an amount that reflects the number of candidates voted for by i that already have been elected (i.e. added to the committee). In the first iteration, no candidate has yet been elected, hence $n_i = 0$ for every voter i. As a consequence for Jefferson $w_j = r_j$ and for Webster $w_j = 2r_j$, where r_j denotes the number of voters that voted for candidate j. Therefore the first candidate elected according to both methods is the candidate with the largest approval score. The contributions of voters to weights w_j are devalued more and more as candidates of whom they approve are elected. This devaluation is stronger for small h, i.e. for small h the trend is greater to favour candidates whose voters have not yet had an approved candidate elected.

2.1.2. Phragmén's proportional rule

Phragmén (1895) instead suggested an approach based on the idea that a candidate requires a certain amount of support from the

electorate to be elected. For election results that are proportional to voter preferences, the voter support should be distributed as evenly as possible among the voters (Brill et al., 2017). In other words, the maximum support provided by a single voter should be as small as possible. To achieve this, in each iteration, that candidate is added to the committee/body to be elected that minimises the maximum voter support. At first all voters i have a vote support value of 0, i.e. $s_i = 0$. In the first iteration, the candidate with the highest approval score is chosen, as this minimises the maximum support value. In the next iteration, another candidate is elected such that the resulting maximum voter support value is as small as possible, however, now some voters already carry a support value $s_i > 0$ from earlier iterations. In each iteration, the maximum support value of a candidate j potentially to be elected is calculated as

$$s_j^{max} = \frac{1 + \sum_{i \in R_j} s_i}{r_j} \qquad (2)$$

Here R_j again denotes the set of voters that voted for candidate j and r_j is the number of voters that voted for candidate j, i.e. the size of R_j. As a result, the required support value of 1 (associated with the potential election of j) is distributed among all voters in R_j in such a way that all r_j voters carry the same total voter support. Then a candidate j with smallest s_j^{max} is elected and the voter support values are updated for all voters i that voted for the newly elected candidate j (Brill et al., 2017):

$$s_i = s_j^{max} \quad \text{for all} \quad i \in R_j \qquad (3)$$

Just like Thiele's rules introduced in Section 2.1.1, Phragmén's rule is sequential in the sense that candidates are iteratively added to a committee/body one at a time. As an added advantage this rule provides theoretical guarantees for proportional representation: Phragmén's rule satisfies a property known as "proportional justified representation" (Brill et al., 2017), whereas Thiele's sequential rules do not satisfy this property (Aziz et al., 2017). Informally speaking, proportional justified representation guarantees that every group of voters with similar preferences is adequately represented in the committee in the sense that the committee includes sufficiently many candidates that are voted for by this group of voters.

2.1.3. Application to tree selection

In our application, the test persons or forest managers are considered as voters and the trees are considered as candidates. Similar applications, for example, include items selected by customers for purchase in a shop or picking juvenile amateur players by a jury of parents to form a local football team. Contrary to the classic situation in political elections, the number of voters, r, in our tree research always is comparatively small whilst the number of candidates to be elected, k, is comparatively large. However, this difference has no implication for the multiwinner methods, as they were based on generic, theoretical considerations to produce proportional outcomes and do not depend on particular ratios of voters to candidates.

Tree selection by multiple forest managers, experts or test persons has, as previously mentioned, so far only rarely been considered in forest science, however, basing decisions on multiple raters is quite common in medicine, psychology, sociology and political science (Fleiss et al., 2003; Cicchetti, 1994; Hallgren, 2012; Brams et al., 2019).

2.2. Study sites

For this study, data from 50 marteloscope experiments from all over GB were analysed (Fig. 1). The Technical Development Department of Forest Research at Ae (Scotland, UK) regularly holds forest management training seminars and as part of these events marteloscope experiments are carried out. Marteloscopes are forest research and training sites where all trees are measured and numbered. During the experiment a number of test persons (also referred to as raters or voters)

Fig. 1. Location of the UK marteloscope sites managed by the Ae Training Centre and Forest Research.

independently walk through these sites and note trees to be evicted from the forest on a sheet of paper or in a software application on a field computer (Pommerening and Grabarnik, 2019). The Technical Development Department of Forest Research in the UK and the Swedish University of Agricultural Sciences have teamed up in the spirit of citizen science for the purpose of quantifying indicators of quality assurance of forestry training with a view to diagnose behavioural trends and to eventually improve the training provision.

Most of the sites include forest stands of Sitka spruce (*Picea sitchensis* (Bong.) Carr.), hybrid larch (*Larix × marschlinsii* Coaz), Japanese larch (*Larix kaempferi* (Lamb.) Carr.) and Scots pine (*Pinus sylvestris* L.). In some of these stands, other species have later colonised the site, however, the aforementioned species represent the main species in terms of density. Peckett Stone at the Welsh-English border is a beech (*Fagus sylvatica* L.) forest and Dean (in the Forest of Dean) is a Norway spruce (*Picea abies* L.) forest, i.e. they are exceptions from the aforementioned species composition (Pommerening et al., 2018).

All marteloscopes were located in even-aged forests that were originally planted as monocultures with only one species. Other species occasionally occur, but they are minorities and were not included in the thinning instructions. With the notable exception of Ae, each marteloscope had a size of 0.1 ha. The size of the Ae marteloscope was 0.133 ha. For each tree the following variables were measured: diameter at breast height (d) (measured in centimetres at 1.3 m height), total tree height [m] and Cartesian coordinates in metres. We calculated basic summary characteristics and presented them in Table 1.

All sites represent early forest development stages, e.g. the stem exclusion phase according to Oliver and Larson (1996) or the early

Table 1
Description of the forest sites and marteloscopes included in this research. N – density, calculated as number of trees per hectare, G – basal area, calculated as the sum of cross-sectional tree stem areas at 1.3 m above soil level), d_g – quadratic stem diameter at 1.3 m above soil level, h_{100} – stand top height, calculated as the mean height of the largest 100 trees per hectare, v_d – coefficient of variation of stem diameters 1.3 m above soil level, k_d – skewness of the empirical stem diameter distribution, r – number of forest managers marking trees separately for the low and crown thinning experiments and n – number of trees eligible for selection. Several numbers of r indicate that several experiments have taken place in the same and/or in different years as specified.

Site	Main species	Year(s)	N [trees ha^{-1}]	G [m²·ha^{-1}]	d_g [cm]	h_{100} [m]	v_d	k_d	r Low	r Crown	n
Ae	*Picea sitchensis*	2011	1321	41.9	20.1	21.2	0.35	0.17	10	11	176
Ardross	*Larix × marschlinsii*	2012, 2013	2180	32.3	13.7	13.5	0.37	0.49	7, 8	7, 8	218
Bin	*Picea sitchensis*	2010	1540	59.3	22.1	22.4	0.30	0.12	8	8	154
Black Isle	*Pinus sylvestris*	2013	2010	26.0	12.8	11.0	0.24	0.18	11	–	201
Cannock Chase	*Larix × marschlinsii*	2012, 2013	2040	35.8	14.9	14.8	0.29	0.07	6, 20	6, 20	204
Cannock Chase	*Larix × marschlinsii*	2014	2040	36.7	15.1	17.0	0.31	0.15	16, 11, 9	16, 11, 9	204
Craigvinean	*Picea sitchensis*	2013	3000	53.0	15.0	15.0	0.22	−0.07	15	–	300
Craigvinean	*Picea sitchensis*	2015	3000	56.7	15.5	16.6	0.24	0.07	8, 7	8, 7	300
Crychan	*Larix × marschlinsii*	2010	1930	41.2	16.5	16.2	0.28	−0.04	6	6	193
Crychan	*Larix × marschlinsii*	2013	1610	41.5	18.1	17.8	0.26	−0.17	8	10	161
Dalby	*Larix kaempferi*	2011	1900	46.2	17.6	18.8	0.28	0.31	9	10	190
Dean	*Picea abies*	2016, 2017	3050	36.2	12.3	13.2	0.34	0.37	18, 11, 9, 15	18, 11, 9, 15	305
Dean	*Picea abies*	2018	2830	41.8	13.7	16.7	0.35	0.36	11	11	283
Glentress	*Picea sitchensis*	2013	1760	58.1	20.5	23.5	0.30	0.06	13	13	176
Haldon	*Picea sitchensis*	2014	1780	43.9	17.7	18.8	0.35	0.39	16	16	178
Loch Ard	*Picea sitchensis*	2015	2450	43.3	15.0	18.2	0.35	0.36	14	14	245
Peckett Stone	*Fagus sylvatica*	2011	830	34.7	23.1	24.8	0.29	0.33	11	10	83
Tummel	*Picea sitchensis*	2019	3230	42.4	12.9	13.3	0.28	−0.18	8	7	323

biostatic phase according to Emborg et al., (2000) with associated high tree densities both in terms of trees per hectare and basal are per hectare. Only Ae and Peckett Stone are middle-aged stands. Stem size diversity as described by the coefficient of variation and skewness is comparatively low, which is typical of plantations at the brink of being transformed to continuous cover or near-natural forest management (Pommerening, 2004).

The study included 50 groups of test persons (voters) rating the trees as part of training sessions. Each group was comprised of a number of test persons varying from a minimum of 6 (Cannock Chase, Crychan) to a maximum of 20 (Cannock Chase, see Table 1). About 95% of the test persons were employed by the state forestry service (Forestry Commission, Natural Resources Wales) in different capacities ranging from machine operators to work supervisors and also included woodland officers and forest managers. The remaining 5% of the test persons mainly worked as forestry contractors (Pommerening et al., 2018). These test persons rated between 83 (Peckett Stone) and 323 (Tummel) trees.

The experiments conducted on each site included two different thinning types, i.e. strategies for evicting trees from the site. The first experiment involved low thinnings, otherwise known as thinnings from below, where trees are removed mainly from the lower canopy and from among the smaller diameter trees (Helms, 1998). The main objective of this type of thinning is to promote the growth of larger trees by removing smaller ones. The second type of experiment involved crown thinnings, also referred to as thinnings from above, where trees are removed that are part of the main stand canopy in order to favour the best among the most dominant trees by removing their direct competitors (Helms, 1998). The test persons were provided with broad thinning instructions, which slightly varied from site to site depending on local conditions. In most cases, experiments involving low and crown thinnings were conducted with the same test persons on the same marteloscope sites and some of the experiments (Cannock Chase, Craigvinean, Crychan, Dean, see Table 1) were repeated in the same and/or in subsequent years (involving the same trees but different test persons), which contributed data from a total of 50 experiments to this study. For the purpose of aggregating final lists of trees to evict we determined k, the number of trees to finally select, as that number corresponding with a removal of 30% of initial basal area (as indicated in Table 1), i.e. the trees were selected iteratively according to the sequential multiwinner rules described in Section 2.1 until the residual basal area fell below 70% of the initial basal area. This percentage was consistent with the instructions given to the test persons and with the forest stand conditions.

2.3. Statistical measures of voting behaviour

In order to quantify the effects of standard AV and proportional multiwinner rules, we included a number of measures in our study.

One of these measures is Fleiss' kappa (Fleiss, 1971; Fleiss et al, 2003), which is frequently used in applied statistics. The concept of kappa is based on pairwise comparisons and has its roots in the one-way analysis of variance. Fleiss' kappa can be expressed as

$$\kappa = \frac{p_0 - p_e}{1 - p_e}, \tag{4}$$

where p_0 is the observed proportion of ratings in agreement and p_e is the expected proportion of ratings in agreement (see Pommerening et al., 2018 for details). The values of κ usually lie between 0 and 1 and agreement increases with increasing κ. Agreement here is defined as similarity in votes.

We defined representativeness as the number m_i of the candidates approved by each test person i that were finally elected by standard AV or by proportional multiwinner rules, i.e. m_i is the number of candidates that test person i successfully put forward. Considering a larger degree of intersections between the sets R_j we assumed that with increasing representativeness numbers m_i should ideally approach a uniform distribution. The deviation of the empirical distribution of numbers m_i from the uniform distribution on the set $\{1, 2, ..., r\}$ can be quantified by applying the test statistic of the χ^2 goodness-of-fit test. The test statistic is calculated according to

$$\hat{\chi}^2 = \sum_i^r \frac{\left(m_i - \frac{\sum_{i=1}^r m_i}{r}\right)^2}{\frac{\sum_{i=1}^r m_i}{r}}. \tag{5}$$

However, our aim was not to carry out a test, but simply to characterise the extent of deviation of the numbers m_i from the uniform distribution to identify significant differences between standard AV and proportional multiwinner rules. The value of the $\hat{\chi}^2$ can be compared to

the critical value $\chi_{r-1,0.05}$ (5% quantile of the χ^2 distribution with $r - 1$ degrees of freedom) by calculating

$$\chi = \frac{\widehat{\chi}^2}{\chi_{r-1,0.05}}.$$

(6)

Small values of χ indicate small deviations from the uniform distribution, i.e. a high degree of representativeness (Pommerening et al., 2018).

An alternative to χ is the coefficient of variation r_m of the proportions m_i/k of trees approved by test person i and elected as part of the final, definite list of trees to be evicted. Small values of r_m indicate a high degree of representativeness.

Finally we included the ratio of the proportion of number of trees (N) marked with "1" and the proportion of basal area (G, derived from stem diameter using the area equation of the circle) of these trees (Kassier, 1993) in the analysis, see Pommerening et al. (2018) and Vítková et al. (2016).

$$B = \frac{\text{Proportion of the number of trees selected}}{\text{Proportion of the basal area of selected trees}} = \frac{P_N}{P_G}$$

(7)

In our case, this measure quantifies the human tree selection strategy of the aggregated tree list by comparing numbers of trees selected with their cumulative size. If $B < 1$, a smaller proportion of trees has been selected compared to their proportion of cumulative basal area. In a management context, this typically indicates a crown thinning and the trees selected show a tendency of being in the upper part of the empirical diameter distribution. A larger proportion of trees is selected compared to their proportion of basal area, if $B > 1$. In a management context, this is consistent with a thinning from below and trees were preferably selected in the lower part of the empirical diameter distribution (Pommerening et al., 2018).

For each experiment, χ, B and r_m were calculated separately for the lists of trees elected by standard AV and by proportional rules whilst κ was calculated based on the approval votes of each individual test person.

2.4. Characteristics of forest structure

To broadly characterise differences in forest structure caused by removing the finally elected trees we also quantified the coefficient of variation of stem diameters v_d and the skewness k_d of the empirical diameter distribution of the residual forest stands, i.e. after removing the elected trees. All calculations were performed using R (version 3.5.1, R Development Core Team 2019) based on our own code.

3. Results

3.1. Differences in the number of finally selected trees

Particularly striking was the significant difference in the number of finally selected trees between experiments involving low thinnings and those that were associated with crown thinnings (Fig. 2) despite the same relative reduction in basal area in both types of experiments. This outcome was expected, since it is part of the difference in the definitions of the two management strategies. Whilst for the former the median was around 47% of the initial tree number, the median corresponded with only 25% of the initial tree number in the case of crown thinnings. Variance of the number of finally selected trees was larger in low than in crown thinnings. Overall proportional rules led to lower variances in the finally selected trees compared to standard AV (Fig. 2), particularly in crown thinnings. Independent of thinning type the choice of h (Eq. (1)) did not make any difference in the number of finally selected trees (including $h = \frac{1}{10}$, $h = \frac{1}{4}$ and $h = \frac{3}{4}$ for which results are not shown here).

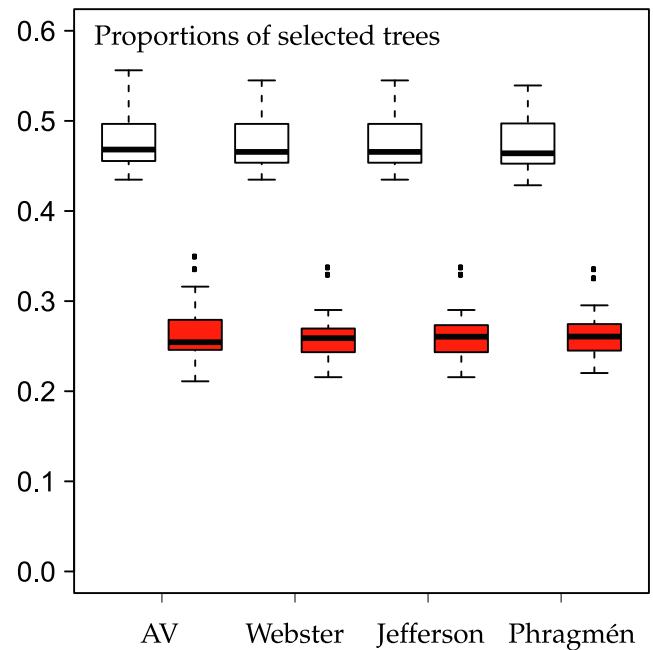Next we studied the number of trees that differed between the set of

Fig. 2. Box plots of the proportions of trees selected in low-thinning experiments (white) and crown-thinning experiments (red) according to standard AV and the proportional rules Webster ($h = 0.5$), Jefferson ($h = 1.0$) and Phragmén. h is the parameter in Eq. (1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
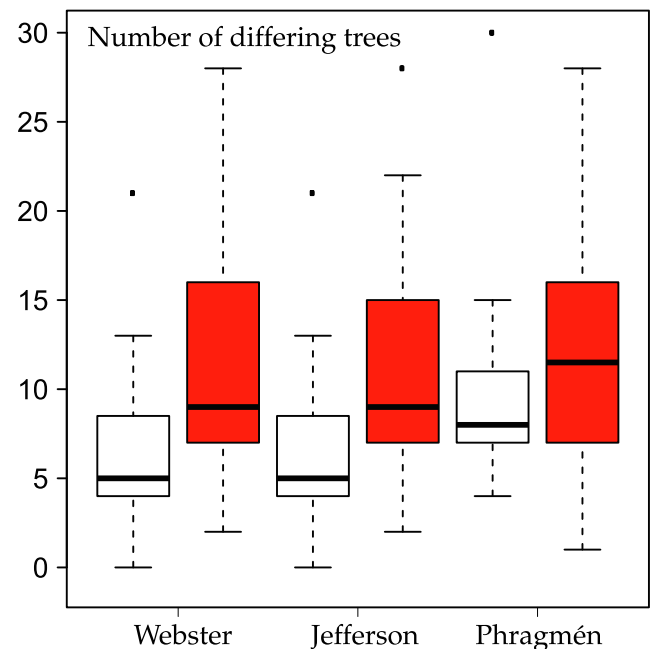

Fig. 3. Box plots of the absolute number of trees different between standard AV and the Webster ($h = 0.5$), Jefferson ($h = 1.0$) and Phragmén methods applied to trees selected in low-thinning experiments (white) and crown-thinning experiments (red). h is the parameter in Eq. (1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

trees elected by standard AV and the set resulting from proportional rules (Fig. 3). Again, the results showed that these numbers of differing trees did not much depend on h, only for $h = 1.0$ the variance was slightly smaller in the crown-thinning experiments than it was for other
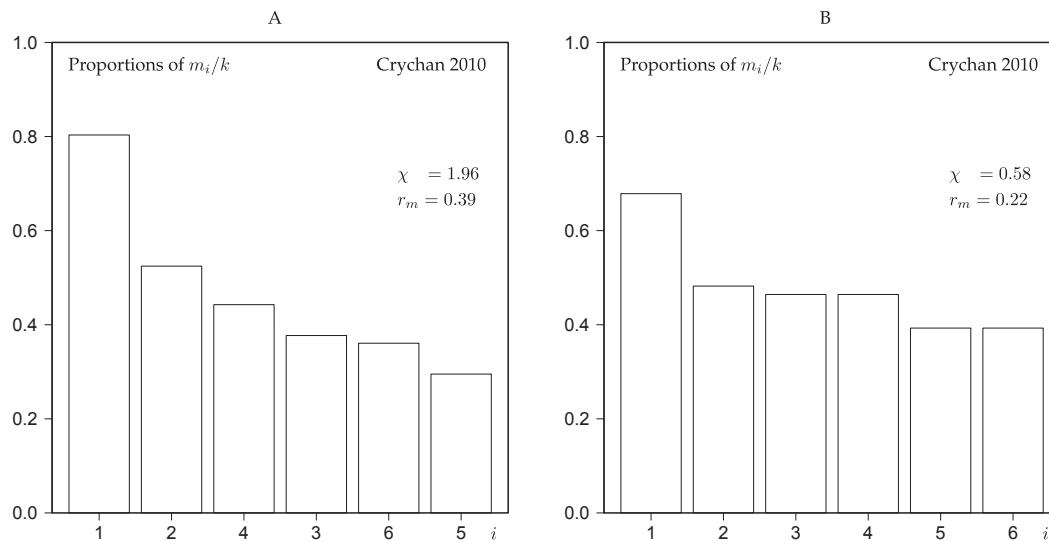
Fig. 4. Bar charts of the proportions of the numbers m_i/k of the candidates approved by each test person i that were finally elected by (A) standard AV and by (B) Jefferson (using $h = 1.0$). The numbers on the abscissa denote the test persons 1, …, 6 of the Crychan-2010 (crown-thinning) experiment. The numbers were originally assigned according to the number of votes given by each test person, i.e. test person #1 provided the largest number of votes whilst test person #6 the smallest. Here the test persons and the corresponding bars were re-ordered according to the proportions of the numbers m_i. i denotes test persons, χ a measure related to the χ^2 goodness-of-fit tests and explained in Eqs. (5) and (6) and r_m is the coefficient of variation of the proportions m_i/k of trees approved by test person i and elected as part of the final, definite list of trees to be evicted.

values of h and the same thinning type and for the Phragmén method. There were clear differences between standard AV and proportional rules: For Jefferson and Webster the medians were much lower in the case of low-thinning experiments (5 trees) compared to those of crown thinnings (9 trees). In the case of the Phragmén method the median numbers of differing trees were larger, i.e. 8 trees in low-thinning experiments and 11.5 trees in crown-thinning experiments. The differences between low and crown thinnings were less for Phragmén than they were for Jefferson and Webster. Also the variance was generally lower for low thinnings than it was for crown thinnings.

3.2. Representativeness

We were also interested to learn whether proportional rules helped to increase the number m_i of the candidates approved by each test person i that were finally elected by standard AV and by proportional rules. To study this we calculated χ (Eq. (6)) and r_m separately for standard AV and for proportional rules. To explain the methodology and to show the details for one of the experiments we briefly discuss the results of the experiment Crychan 2010 in detail (Fig. 4). The bars in the bar chart show the proportions of candidates put forward by test persons 1–6. In panel A, the bars sharply decline from left to right whilst in panel B, the decline is much more moderate with smaller proportions on the left-hand side compared to those in the standard AV-chart in panel A. Thus the proportional rules have helped to redistribute some of the votes in favour of test persons whose candidates were less represented in the AV list of elected candidates. This trend is also reflected by the change in values of measures χ and r_m. Both values were markedly reduced for proportional rules, i.e. they helped to approach the uniform distribution of m_i a bit more and have reduced the differences between the bars as shown by the coefficient of variation.

Following on from this expected outcome we computed the differences $\Delta\chi$ and Δr_m between standard AV and proportional rules for all 50 experiments and were interested in learning how they related to other characteristics introduced in Sections 2.3 and 2.4. First, we plotted $\Delta\chi$ and Δr_m over κ (Eq. (4); Fig. 5). There clearly was a significant relationship between $\Delta\chi$ and Δr_m and κ. In both cases the relationship was strongest for crown-thinning experiments, where there often is confusion among British forest managers (Pommerening et al., 2018), and for

Jefferson/Webster. The overall trend of the relationship tells us that the effectiveness of proportional rules increases with decreasing agreement among the test persons, which is consistent with the fact that decreasing agreement means an increase in the variability of opinions and votes. As noted in previous research, in low thinnings, κ is in Britain traditionally much higher and varies less than in crown thinnings and the relationships with both $\Delta\chi$ and Δr_m and κ in low thinnings were not so clear in the data we analysed. For Phragmén's method, $r' = -0.46^{***}$ and $r' = -0.60^{***}$ for $\Delta\chi$ and Δr_m, respectively. These markedly lower values may partly be related to the finding in Fig. 4, i.e. that with the Phragmén method the numbers of finally selected trees do not differ between low- and crown-thinning experiments as much as they do with Jefferson/Webster.

In addition we checked whether the three proportional rules satisfied proportional justified representation. As explained in Section 2.1.2, Phragmén's method always produces tree selections satisfying this property. Interestingly, neither Jefferson nor Webster or any of the other values of parameter h led to a violation of this property for any of the 50 experiments. Given such a comparatively large number of experiments, this was a somewhat unexpected result and we take this as an indication that in the context of our and similar experiments the advantage of Phragmén's rule over Thiele's rules is rather of a theoretical nature.

3.3. Influence of proportional rules on tree and forest characteristics

As ratio B measures the tree selection strategy of people, it is interesting to see how the application of proportional rules influences tree selection strategies (Fig. 6). Here we have only shown the results for AV and Jefferson, as the latter ones are much the same for Webster and Phragmén. The results clearly emphasised the typical divide between crown and low-thinning experiments, where low thinnings ideally should score above the horizontal line running through 1 and crown thinnings rather below that line. Whilst the medians for low thinnings were in the expected range for both AV and proportional rules (AP), those for crown thinnings were too close to 1 or even above the demarcation line.

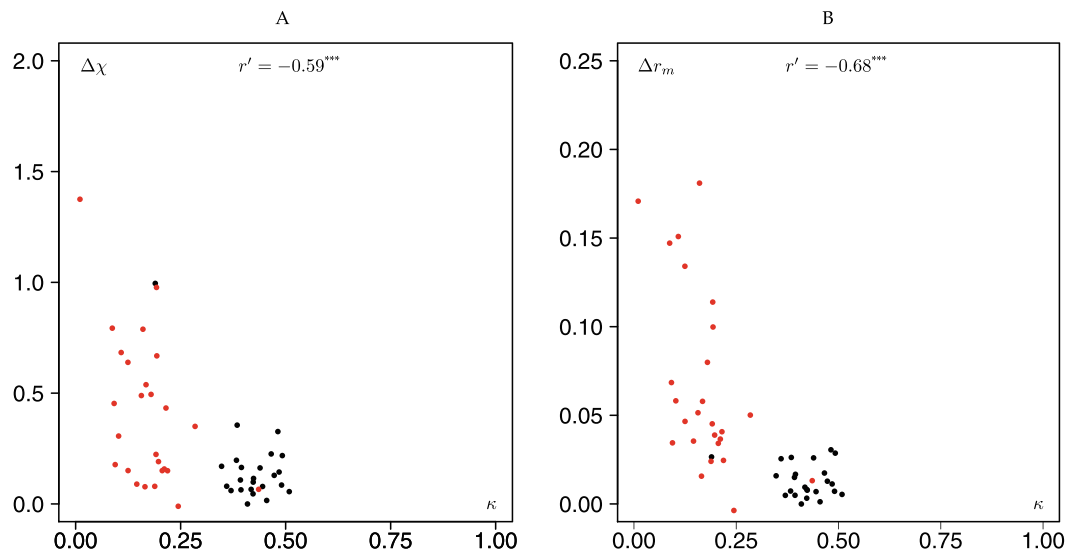This result is not surprising, since the crown-thinning method is not

Fig. 5. Relationship between Fleiss' κ (Eq. (4)) and $\Delta\chi$ (A) and Δr_m (B) using Jefferson apportionment with $h = 1.0$. The results relating to crown-thinning experiments are highlighted in red, those relating to low-thinning experiments are in black. r' is the Pearson correlation coefficient and "***" indicates p values < 0.001. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
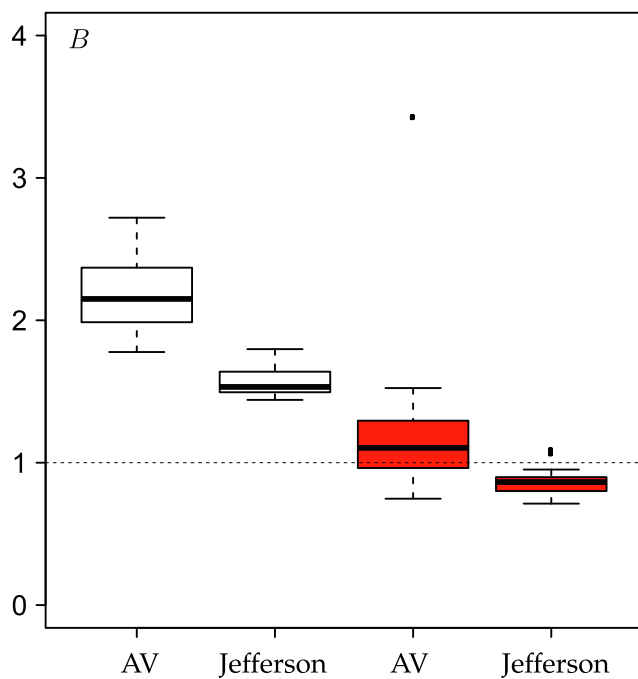

Fig. 6. Effect of proportional rules on ratio B (Eq. (7)) measuring the tree selection strategy. The results relating to crown-thinning experiments are highlighted in red, those relating to low-thinning experiments are in white. AV – standard multiwinner approval voting, Jefferson – proportional Thiele rule with Jefferson depreciation weights ($h = 1.0$). The horizontal line through 1 marks the boundary between low- and crown-thinning tree selections. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

well established in Britain and many test persons probably unintentionally fell back into "old habits" when participating in the crown thinning experiments rather than that they implemented what they had just learned in the preceding training, see Pommerening et al. (2018). The effect of proportional rules clearly was that the value of B was significantly reduced for both tree-selection strategies. Interestingly the reduction was bigger for low thinnings than it was for crown thinnings.

At the same time proportional rules also markedly reduced the variance of B. The results imply that the finally elected tree lists produced by proportional rules lead to more reasonable and consistent outcomes as far as measure B is concerned, i.e. they better reproduce the two tree selection experiments.

Finally we analysed how the removal of evicted trees would affect the structure of the forest stands under consideration (Fig. 7). Here again the results did not differ much between the three proportional rules Jefferson, Webster and Phragmén. Therefore we only presented the results for standard AV and Jefferson in Fig. 7. Interestingly the stem-diameter coefficient of variation significantly dropped in low thinnings when using proportional rules compared to the situation of standard AV (Fig. 7A). This implies that the size structure of stands would simplify, i.e. become less diverse, than for AV. In crown thinnings, proportional rules would slightly but insignificantly increase the stem-diameter coefficient of variation. For both thinning strategies proportional rules apparently lead to more typical results in terms of what is expected from theory and literature. On the other hand, skewness significantly increased in low thinnings as a result of proportional rules but only insignificantly in crown thinnings (Fig. 7B). As a consequence stem diameter distributions would increasingly become right-skewed, which is a size-diversity gain.

4. Discussion

On first glance it may seem odd to apply methods from theoretical politics to problems in managing forest ecosystems that are based on natural sciences. What has voting to do with the behaviour of humans selecting trees and how can applying any of these methods help to solve environmental and conservation problems?

Traditionally in forestry it was mainly one person at a time who would decide about which trees to evict from a given forest site, e.g. one forest manager or one machine operator. Still it is largely true what Gadow wrote back in 1996, i.e. that the process of selective tree marking up to now has not been taken seriously and given much attention neither in practice nor in research (Gadow, 1996) despite being crucial to modern concepts of ecological forest management and particular of forest conservation (Pommerening, 2004). Often tree marking is even delegated to insufficiently qualified or experienced staff, which may prove fatal in times of rapid climate change and rapid diversity loss. Interestingly, recent research has shown that agreement even
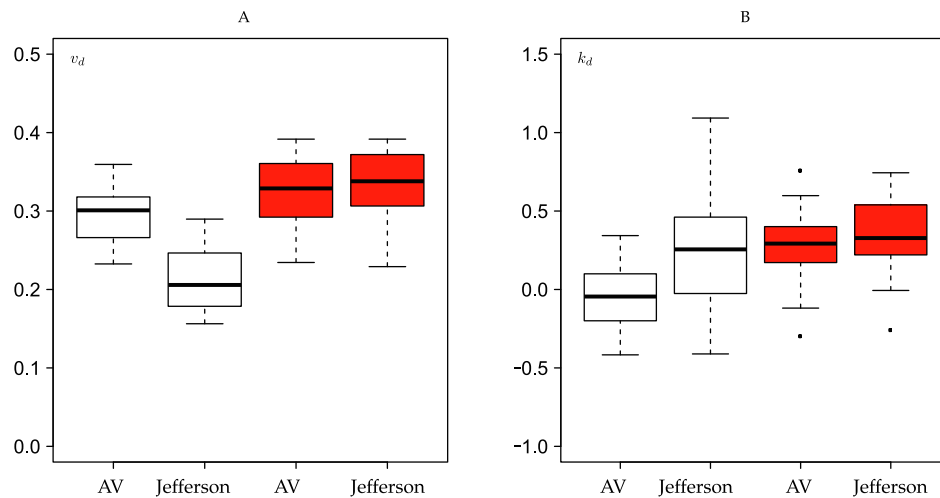
Fig. 7. Effect of proportional rules on the stem-diameter coefficient of variation v_d and on skewness k_d of the residual forest stands, i.e. after removing the evicted trees. The results relating to crown-thinning experiments are highlighted in red, those relating to low-thinning experiments are in white. AV – standard multiwinner approval voting, Jefferson –proportional Thiele rule with Jefferson depreciation weights ($h = 1.0$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

among trained forestry staff is not high, although forest practice and forest science usually assume the opposite (Vítková et al., 2016; Pommerening et al., 2018; Pommerening and Grabarnik, 2019). This can be viewed as a problem, but in times of multiple changes it most importantly is an opportunity: The apparent lack of agreement among trained forest managers is an expression of a diversity of opinions which is quite natural in what, after all, is complex decision making in complex environments. It is natural that quite a few people produce similar marking results, whilst those of others may differ markedly from this group or another. Often enough it is hard to say *ad hoc* what is the right decision and strategy, particularly where new directions are taken and strategies are still under development such as in conservation and in strategies of mitigating climate change. In such situations, proportional multiwinner rules used in political science can help to ensure that important minority expertise and opinion do not go under or are unduly suppressed by majorities. If, for example, the majority of forest managers mark trees leading to an inappropriate strategy of supporting a forest ecosystem, standard approval voting would reflect the majority opinion only and differing opinions that may offer better alternatives would be ignored.

Our results showed that proportional multiwinner rules can be applied to tree marking, even though the structure of the data, i.e. the ratio of voters to elected individuals, markedly differs from situations in political elections. The application leads to meaningful results where between 5 and 12 trees on average differ compared to standard approval voting. Overall the number of trees that were selected by one method but not by another were comparatively small. This can be explained by the fact that there are no clear factions among the test persons as they often occur in politics. Nevertheless the few differing trees can sometimes be quite significant for their size and other differing properties.

There are significant differences in the numbers of trees selected in crown-thinning and low-thinning experiments. In general terms the differences suggest that proportional rules are particularly effective in situations where the voters or tree markers show differing behaviour and disagree (see Fig. 5). This is intuitively plausible and an important confirmation of the effectiveness of proportional rules, since an increasing lack of agreement is typically caused by greater diversity in the marking results and proportional rules attempt to represent this diversity as best as possible, hence in such situations we can expect the greatest differences between the results of standard AV and those of proportional rules.

The application of proportional rules clearly impacts on the residual

forest and therefore also on future forest development. Ratio *B* is a key characteristic in marteloscope research for reconstructing the marking strategy that a test person has followed, i.e. selecting many small as opposed to fewer but bigger trees. The medians of this ratio were significantly reduced by proportional rules (Fig. 6). As a result the medians resulting from proportional rules were more realistic than those from standard AV and the consistency of the results also increased due to reduced variances.

The results with regard to the stem-diameter structure of the residual forest showed that proportional rules clearly had an impact on residual forest structure thus confirming that the varied opinions of the test persons can affect forest development (Fig. 7). The application of proportional rules significantly reduced the coefficient of variance of stem diameters in low thinning experiments whilst this measure more or less stayed the same in crown thinning experiments. This implies that applying proportional rules increases the effect expected from the two tree selection strategies. On the other hand skewness increased consistently with proportional rules.

Overall our results showed that proportional rules are useful concepts in situations where individual choices concerning trees proposed to be felled or to be retained need to be aggregated to arrive at one definite, consolidated list or set. In that situation, proportional rules lead to better results than standard AV, but it is less important which particular proportional method is used. The Jefferson, Webster and Phragmén methods employed in this study led to nearly the same results and even the property of proportional justified representation was satisfied for all three methods. This most likely is due to the typical data structure in marteloscope research where, as alluded to before, the number of voters, *r*, always is comparatively small whilst the number of candidates to be elected, *k*, is comparatively large. Hence in contrast to political elections there are potentially always quite a few identical tree choices that are put forward by many test persons and this increases the average representation of voter groups.

Multiwinner approval voting can be easily applied in practice. All that is needed is to temporarily number all trees in a given forest stand so that the numbers can be seen on the stem surface from afar. The numbered trees can occur in a continuous area or in small sample plots, i.e. it is also possible to only include a sample of all trees, as this is mostly done in forest inventory. Then specialists (e.g. in forest conservation) are individually asked to independently mark trees for removal according to clear instructions. The results are then a group decision which is computed as explained in this article and subsequently the actual tree fellings are determined by these results.

5. Conclusions

Proportional multiwinner rules from theoretical politics are an intriguing tool for aggregating the marking results of multiple forest managers that independently select trees in the same forest stand for improving ecosystem goods and services including resilience to climate change and maintaining tree diversity. This is a common situation in marteloscope experiments and exercises that are carried out in many countries (Pommerening et al., 2018). The aggregation can then serve as a reference for assessing individual selections similar to the idea of the Vorobyov mean in Stoyan et al. (2018), but more importantly the aggregation can also be directly implemented in the forest as the result of a group decision based on a vote among specialists. This makes particular sense when an optimum solution is unknown or difficult to derive *ad hoc*. Such a public consultation would ultimately increase professional creditability and decision transparency. Proportional rules then make sure that differing but important minority opinions are represented along with majority opinions. This is particularly useful in situations, where conventional wisdom-of-the-crowd approaches such as standard approval voting would allow forest development to go completely astray, if the majority decisions later prove to be inappropriate. As such, proportional rules act as safety mechanisms to reduce risks much in the same way as these methods were intended to safeguard democracy in political processes. Aggregation of individual tree-selection scores by multiwinner voting rules can also be interpreted as crowdsourcing (Ghezzi et al., 2018), where a group of specialists is employed to offer solutions to a problem in ecosystem management and an efficient algorithm synthesises the individual responses. A good representation of differing opinions in forest management is particularly important when new management directions are proposed, e.g. near-natural or continuous cover forestry, conservation management, managing woodlands for recreation and for mitigating climate change.

CRediT authorship contribution statement

A.P. conceived the research idea, designed the data analysis. J.H. ran the marteloscope experiments and contributed the corresponding data. A.P. and U.S.-K. programmed the required R code and carried out the data analysis. M.B. and U.S.-K. advised on the proportional multiwinner rules. All authors interpreted and discussed the results and contributed to the text.

Declaration of Competing Interest

There was no conflict of interest.

Acknowledgements

The authors thank the Technical Development Department of Forest Research at Ae (Scotland, UK) for our citizen-science partnership that enabled the collection of the unique and valuable data set that was used in this study. Inspiring initial email discussions with Steven Brams (New York University, USA) have helped to shape the idea of this research. M. B. and U. S.-K. were supported by the Deutsche Forschungsgemeinschaft (grant #BR 4744/2-1).

References

Aziz, H., Brill, M., Conitzer, V., Elkind, E., Freeman, R., Walsh, T., 2017. Justified representation in approval-based committee voting. Soc. Choice Welf. 48 (2), 461–485.
Balinski, M., Young, H.P., 1982. Fair Representation: Meeting the Ideal of One Man, One Vote. Yale University Press, Yale.
Brams, S.J., Fishburn, P.C., 1978. Approval voting. Am. Polit. Sci. Rev. 72 (3), 831–847.
Brams, S.J., Kilgour, D.M., Potthoff, R.F., 2019. Multiwinner approval voting: an apportionment approach. Public Choice 178 (1–2), 67–93.
Bravo-Oviedo, A., Marchi, M., Travaglini, D., Pelleri, F., Manetti, M.C., Corona, P., Cruz, F., Bravo, F., Nocentini, S., 2020. Adoption of new silvicultural methods in

Mediterranean forests: the influence of educational background and socio-demographic factors on marker decisions. Ann. Forest Sci. 77 (2). https://doi.org/10.1007/s13595-020-00947-z.
Brill, M., Freeman, R., Janson, S., Lackner, M., 2017. Phragmén's voting methods and justified representation. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4-9 February 2017, pp. 406-413. Palo Alto, AAAI Press.
Brill, M., Laslier, J.-F., Skowron, P., 2018. Multiwinner approval rules as apportionment methods. J. Theoretical Polit. 30 (3), 358–382.
Cao, H., Sen, P.K., Peery, A.F., Dellon, E.S., 2016. Assessing agreement with multiple raters on correlated kappa statistics: Dependent kappa with multiple raters. Biom. J. 58 (4), 935–943.
Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6 (4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284.
Daume, S., Füldner, K., Gadow, K.v., 1997. Zur Modellierung personenspezifischer Durchforstungen in ungleichaltigen Mischbeständen. [Modelling person-specific thinnings in uneven-aged mixed stands.]. Allgemeine Forst- und Jagdzeitung 169, 21–26.
Emborg, J., Christensen, M., Heilmann-Clausen, J., 2000. The structural dynamics of Suserup Skov, a near-natural temperate deciduous forest in Denmark. For. Ecol. Manage. 126 (2), 173–189.
Füldner, K., Sattler, S., Zucchini, W., Gadow, K.v., 1996. Modellierung personenabhängiger Auswahlwahrscheinlichkeiten bei der Durchforstung. [Modelling person-specific tree selection probabilities in a thinning.]. Allgemeine Forst- und Jagdzeitung 167, 159–162.
Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76 (5), 378–382.
Fleiss, J.L., Levin, B., Paik, M.C., 2003. Statistical Methods for Rates and Proportions. John Wiley & Sons, New York.
Gadow, K.v., 1996. Modelling growth in managed forests — realism and limits of lumping. Sci. Total Environ. 183 (1–2), 167–177.
Ghezzi, A., Gabelloni, D., Martini, A., Natalicchio, A., 2018. Crowdsourcing: a review and suggestions for future research: crowdsourcing. International Journal of Management Reviews 20 (2), 343–363.
Hallgren, K.A., 2012. Computing inter-rater reliability for observational data: an overview and tutorial. TQMP 8 (1), 23–34. https://doi.org/10.20982/tqmp.08.1.p023.
Helms, J.A., 1998. The Dictionary of Forestry. Society of American Foresters, Bethesda, MD.
Jandl, R., Spathelf, P., Bolte, A., Prescott, C.E., 2019. Forest adaptation to climate change—is non-management an option? Ann. Forest Sci. 76 (2). https://doi.org/10.1007/s13595-019-0827-x.
Kassier, H., 1993. Dynamics of diameter and height distributions in commercial timber plantations. PhD thesis. University of Stellenbosch.
Lacaze, J.-F., 2000. Forest management for recreation and conservation. Forestry 73, 137–141.
Oliver, C.D., Larson, B.C., 1996. Forest stand dynamics, Update edition. John Wiley & Sons, New York.
Ontl, T.A., Janowiak, M.K., Swanston, C.W., Daley, J., Handler, S., Cornett, M., Hagenbuch, S., Handrick, C., McCarthy, L., Patch, N., 2020. Forest management for carbon sequestration and climate adaptation. J. Forestry 118, 86–101.
Petucco, C., Søndergaard Jensen, F., Meilby, H., Skovsgaard, J.P., 2018. Visitor preferences of thinning practice in young even-aged stands of pedunculate oak (*Quercus robur* L.): comparing the opinion of forestry professionals in six European countries. Scand. J. For. Res. 33 (1), 81–90.
Phragmén, E., 1895. Proportionella val. En valteknisk studie. [Proportional elections. A theoretical study.] Svenska spörsmål 25. Lars Hökersbergs förlag, Stockholm.
Pommerening, A., Pallarés Ramos, C., Kędziora, W., Haufe, J., Stoyan, D., 2018. Rating experiments in forestry: How much agreement is there in tree marking? PLOS ONE 13, e0194747.
Pommerening, A., 2004. A review of the history, definitions and methods of continuous cover forestry with special attention to afforestation and restocking. Forestry 77 (1), 27–44.
Pommerening, A., Grabarnik, P., 2019. Individual-based methods of forest ecology and management. Springer, Cham.
R Development Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria http://www.r-project.org.
Stoyan, D., Pommerening, A., Wünsche, A., 2018. Rater classification by means of set-theoretic methods applied to forestry data. J. Environ. Statist. 8, 2.
Surowiecki, J., 2004. The wisdom of crowds. Random House, New York.
Riccioli, F., Marone, E., Boncinelli, F., Tattoni, C., Rocchini, D., Fratini, R., 2019. The recreational value of forests under different management systems. New Forest. 50 (2), 345–360.
Thiele, T.N., 1895. Om flerfoldsvalg. Oversigt over det Kongelige Danske Videnskabernes Selskabs Forhandlinger. [About multiple choice. Overview of the Royal Danish Science Society's Negotiations], Copenhagen, pp. 415-441.
Vítková, L., Ní Dhubháin, Á., Pommerening, A., 2016. Agreement in tree marking: what is the uncertainty of human tree selection in selective forest management? Forest Sci. 62 (3), 288–296.
Wagner, S., Nocentini, S., Huth, F., Hoogstra-Klein, M., 2014. Forest management approaches for coping with the uncertainty of climate change: trade-offs in service provisioning and adaptability. E&S 19 (1). https://doi.org/10.5751/ES-06213-190132.
Zucchini, W., Gadow, K.v., 1995. Two indices of agreement among foresters selecting trees for thinning. Forest Landscape Res. 1, 199–206.