Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance

Moilanen, Atte

2022-05

Moilanen , A , Lehtinen , P , Kohonen , I , Jalkanen , J , Virtanen , E A & Kujala , H 2022 , ' Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance ' , Methods in Ecology and Evolution , vol. 13 , no. 5 , pp. 1062-1072 . https://doi.org/10.1111/2041-210X.13819

http://hdl.handle.net/10138/343584
https://doi.org/10.1111/2041-210X.13819

Received: 13 July 2021 | Accepted: 19 January 2022
DOI: 10.1111/2041-210X.13819

Methods in Ecology and Evolution
BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance

Atte Moilanen[1,2] | Pauli Lehtinen[1] | Ilmari Kohonen[1] | Joel Jalkanen[1] | Elina A. Virtanen[1,3] | Heini Kujala[1]

[1]Finnish Natural History Museum, University of Helsinki, Helsinki, Finland
[2]Department of Geosciences and Geography, University of Helsinki, Helsinki, Finland
[3]Marine Research Centre, Finnish Environment Institute, Helsinki, Finland

Correspondence
Atte Moilanen
Email: atte.moilanen@helsinki.fi

Funding information
Academy of Finland, Grant/Award Number: 312559; BioDiversa project FutureWeb/Academy of Finland, Grant/Award Number: 326343; Koneen Säätiö, Grant/Award Number: 201803179; Ympäristöministeriö, Grant/Award Number: project MetZo-III

Handling Editor: Robert Freckleton

Abstract

1. Spatial (conservation) prioritization integrates data on the distributions of biodiversity, costs and threats. It produces spatial priority maps that can support ecologically well-informed land use planning in general, including applications in environmental impact avoidance outside protected areas. Here we describe novel methods that significantly increase the utility of spatial priority ranking in large analyses and with interactive planning.
2. Methodologically, we describe a novel algorithm for implementing spatial priority ranking, novel alternatives for balancing between biodiversity features, fast tiled FFT transforms for connectivity calculations based on dispersal kernels, and a novel analysis output, the flexibility map.
3. Marking by N the number of landscape elements with data, the new prioritization algorithm has time scaling of less than $N\log_2 N$ instead of the N^2 of its predecessor. We illustrate feasible computation times with data up to billions of elements in size, implying capacity for global analysis at a resolution higher than 0.25 km^2, or close to 1-ha resolution for a continent.
4. The algorithmic improvements described here bring about improved capacity to implement decision support for real-world spatial conservation planning problems. The methods described here will be at the technical core of forthcoming software releases.

KEYWORDS
algorithm, ecological impact avoidance, land use planning, site selection, spatial conservation prioritization, spatial priority ranking, systematic conservation planning, zonation 5 release candidate

1 | INTRODUCTION

Spatial (conservation) prioritization is a tool used to identify priority areas for biodiversity conservation, included as a technical step within the framework of systematic conservation planning (Kukkala & Moilanen, 2013; Margules & Sarkar, 2007; McIntosh et al., 2017). Spatial prioritization integrates spatial data about the distributions of biodiversity features (species, habitat types, etc.), ecosystem services, land cost, human-induced pressures on biodiversity (threats), effects and costs of conservation interventions, and administrative restrictions such as land availability and ownership and the present protected area network. Kujala, Lahoz-Monfort, et al. (2018) and Kujala, Moilanen, and Gordon (2018) summarize how these different data types impact priorities. Specific uses of spatial prioritization include but are not limited to the following: (a) protected area network design, (b) planning for protected area network expansion, (c) protected area network evaluation, (d) the evaluation of impacts from development, (e) planning for ecological impact avoidance, (f) spatial planning for habitat restoration and/or maintenance, which can also inform biodiversity offsetting, (g) decision support for land use zoning, and (h) support for ecologically based land use planning in general.

With the increased availability of large high-resolution data and demand for their use in spatial planning (Wyborn & Evans, 2021), there is a need for conservation planning tools that can operate with very large biodiversity datasets with short processing times. It should be possible to use high-resolution data at native resolution, updates of large analysis sets should not be overly cumbersome computationally and interactive planning with stakeholders should be enabled. Here we present a new efficient optimization algorithm, which finds priority area solutions in large biodiversity datasets much faster than its predecessors. We describe this algorithm through the existing conservation planning software and concept Zonation.

Beginning from Moilanen et al. (2005), the Zonation family of software implements a set of methods for spatial prioritization, or spatial priority ranking on high-dimensional landscapes. Since then, these methods have been applied across the terrestrial, marine and freshwater realms, from local to global extents, for a broad variety of purposes, and on data ranging up to tens of thousands of biodiversity features and low hundreds of millions of grid cells with information. The approach converts general principles from ecology and conservation biology into computational form, including such as (a) more conservation coverage is better than less, (b) high local occurrence levels for features are preferable to low levels, (c) habitat quality and connectivity are both desirable, (d) there needs to be a sensible balance between features, implied by concepts such as complementarity, comprehensiveness, representativeness and adequacy (see Kukkala & Moilanen, 2013 for review), (e) solutions need to be cost/area-efficient (effective), and (f) minimizing loss maximizes what will remain (Lehtomäki & Moilanen, 2013; Moilanen et al., 2005, 2011). Maybe the single defining characteristic of a priority ranking is that the spatial solution is an *emergent property* of data and the generic rules implemented in the prioritization process. This is very different from the so-called target-based planning, in which a requirement (target) is *a priori* assigned for each biodiversity feature after which a minimum cost solution is sought for the optimization problem (Watts et al., 2009).

Methodologically, the present study includes the description of a new spatial priority ranking algorithm and new methods for balancing conservation resource-allocation between biodiversity features, and a new useful output, the flexibility map. A significantly improved capacity to analyse high-dimensional problems is demonstrated, which brings the ability to do on-the-fly analyses in co-creation workshops with stakeholders.

2 | MATERIALS AND METHODS

The following material can be understood with some familiarity about the two main Zonation outputs, the priority rank map and performance curves associated with it; please see Figures S1–S4.

To briefly introduce terminology, the *priority rank map* is a map in which all grid cells of the landscape are ranked in order of (conservation) priority as it emerges based on data and analysis settings (Figure S1). The directly linked feature-specific *performance curves* describe conservation coverage achievable for each feature for any selected top ranked fraction of the landscape (Figure S2), as determined by the rank map. The average performance curve summarizes information about mean (conservation) coverage achievable (Figure S3). The distribution of coverage across features can also be shown as a histogram at a selected fraction of the priority ranking (Figure S4). Detailed information can be extracted for specific areas of interest. For recent applications and additional references, please see, for example, Lehtomäki et al. (2019), Jalkanen, Toivonen, and Moilanen (2020) and Virtanen et al. (2018), which also describes the Finnish marine data used here. Notably, Zonation analysis serves both targeting of conservation (high priority areas) and targeting of *ecological impact avoidance* (low-priority areas, see Kareksela et al., 2013). Table 1 summarizes notation and simple relationships used throughout the methods.

2.1 | Novel priority ranking algorithm

The traditional Zonation meta-algorithm has operated so that ranking (a) starts from the full landscape, that is, all grid cells with data. Then, iteratively, (b) marginal loss of biodiversity caused by the loss of each site is evaluated for all remaining grid cells, (c) those cells leading to lowest losses are removed from the remaining landscape, giving them their rank, (d) what is remaining for features is updated (Moilanen et al., 2005, 2011). Steps (b)–(d) are iterated until nothing remains and the priority ranking is complete. In the resulting priority rank map, priority is defined by the removal order, with the most important grid cells for the (conservation) objective being removed and ranked last.

While easy to understand, this meta-algorithm has the disadvantage of N^2 time scaling, which in our experience starts to become an issue when the effective dimension (i.e. number of spatial elements with data), N, of the landscape goes into the high tens of millions of grid cells or above. Here we introduce a novel ranking algorithm that produces effectively the same result in terms of utility for conservation planning, but has advantageous computational properties that allow orders of magnitude larger analysis

TABLE 1 Symbols

Symbol	Explanation
i	Index for grid cell
j	Index for feature (species, habitat type, etc.)
w_j	Weight of feature j
o_{ij}	Original occurrence level of feature j in grid cell i in the input matrix (raster layer for the feature)
p_{ij}	Is p_{ij}, the fraction of occurrences of feature j in cell i, $p_{ij} = o_{ij}/\sum_i o_{ij}$
r_{ij}	Fraction of remaining occurrences of feature j in cell i, $r_{ij} = p_{ij}/r_j$
r_j	Fraction of occurrences remaining for feature j at a specific stage of the ranking
p-norm	Generalized expression for length of vector in multidimensional space. The p-norm of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of length n is defined as $\|x\|_p = \left(\sum_{i=1}^{n} \|x_i\|^p\right)^{1/p}$
M_i, M_{iP}, M_{iN}	Aggregate marginal loss from the removal of grid cell i from under 'conservation'. $M_i = M_{iP} - M_{iN}$, in which M_{iP} is aggregate marginal loss for positively weighted features (species) and M_{iN} is aggregate marginal loss for negatively weighted features (opposing factors, opportunity costs, etc.)

and fast interactive use. We call the new algorithm 'iterative conditional sort'.

Heuristically put, the iterative conditional sort is about finding a rank order of grid cells, in which the marginal loss for cells increases (or does not decrease) through the ranking. The complication is that the order of cells itself influences marginal losses via changes in feature coverages. Put in a simplified manner, (a) an order is proposed, (b) this ranking is evaluated conditional on the given order, which produces the so-called conditional marginal loss values for cells, and (c) cells are reordered according to this conditional marginal loss measure. (d) The cycle of conditional evaluation and reordering (re-ranking) is iterated until the ordering achieves convergence according to an error measure derived from the ranking. This development in the meta-algorithm is purely technical and aims at computational efficiency rather than at a solution that has different ecological characteristics.

The ranking algorithm below is a meta-algorithm, because an integral component—the marginal loss calculation—has multiple alternatives (sub-algorithms) that can be inserted into the meta-algorithm (Section 2.3).

New Zonation prioritization meta-algorithm: Iterative conditional sort

PART I: Initialization

1. Set iteration index $t = 0$. Calculate aggregate marginal loss M_i for all sites i. As a starting point, we set $M_i^{(0)}$ to the weighted range-size rarity of the grid cell (Section 2.3; Veach et al., 2017).
2. Sort sites i into ascending order based on their $M_i^{(0)}$-values that were just calculated: this produces the first sorted vector of site indexes, marked here by $S^{(0)}$. Set $t = t + 1$.

Technically, it is key that the sort can be implemented using a very fast sort algorithm with time complexity of $N\log_2 N$ or less, included in the C++ standard library. The next step is a conditional sort operation that is iterated until sufficient convergence. Here, conditional means that marginal values for cells are recalculated conditional on the specific order proposed by vector $S^{(t)}$.

PART II: Iteration

We mark by $i(s, t)$ the index of the site at position s in vector $S^{(t)}$, with $s = 1, 2, \ldots, N_c$. (N_c = number of cells with data.) For convenience, mark by $s(i, t)$ position of site i in $S^{(t)}$.

3. Generate a new vector of M_i-values that are now conditional on the present ordering, $M^{(t)}[S^{(t-1)}]$. First, set feature-specific fraction of occurrences remaining, r_j, to that in the full landscape for each feature j. Then, go through the full vector $S^{(t-1)}$ in increasing order of rank:

FOR $s = 1, 2, 3, \ldots, N_c$, DO

3.1 For each rank position s, take the index of the site $k = i(s, t - 1)$, evaluate conditional marginal loss for grid cell k (Section 2.3) and mark this quantity m_k.
3.2 Set $M_k^{(t)}[S^{(t-1)}] = m_k$.
3.3 Link in the occurrence levels of features in grid cells. Reduce feature-specific occurrences following hypothetical loss of site k, set $r_j = r_j - p_{jk}$, for all features j occurring in cell k. Values in vector $M^{(t)}$ thus become dependent on the order of sites in the order proposed by the previous iteration, $S^{(t-1)}$.

PART III: Check for convergence

4. We check for convergence by investigating the mean error, $Err(t)$, of the conditional marginal loss vector, $M^{(t)}$. The condition for no error is that marginal losses have been ordered into a non-decreasing manner. Hence, we accumulate the mean error of the ranking through all cells s, comparing marginal losses for pairs of cells, marked k and n, in sequential positions s and $s + 1$ in the ranking:

$$Err\,(t) = \frac{1}{N_c - 1} \sum_{s=1}^{N_c - 1} \begin{cases} 0, & \text{if } M_n^{(t)} \geq M_k^{(t)} \\ \dfrac{M_k^{(t)} - M_n^{(t)}}{M_{kP}^{(t)} + M_{kN}^{(t)}}, & \text{if } M_n^{(t)} < M_k^{(t)} \end{cases}.$$

Error is only accumulated for pairs of grid cells that are not ordered correctly according to the non-decreasing criterion. The sum of marginal loss components for positively and negatively weighted layers is used to scale the magnitude of cell-specific error observed.

5. We used a default requirement that $Err(t)$ should be less than 0.01%. If the sort has converged to this degree, end the iterative sort and move to outputting prioritization results. If the sort has not converged, then produce the new sorted rank vector $S^{(t)}$ by sorting sites into ascending order of the vector $M^{(t)}$, set $t = t + 1$, and return to step (3).

Heuristically put, each iteration cells with marginal losses that are high (low) given the position of the cell in the ranking move up (down) in the ranking. Over several iterations, the priority ranking reorganizes so that discrepancies become reduced, cells start to move less in the ranking, and the order stabilizes according to the convergence criterion.

2.2 | Novel options for balancing between (biodiversity) features

The method used to aggregate the marginal loss of conservation value across features is an integral part of spatial prioritization (in Section 3 of the meta-algorithm above). It is this mechanism that maintains the balance between features through ranking, or in the terminology of systematic conservation planning, complementarity (Kukkala & Moilanen, 2013; Williams, 2001).

There is no single unique way that a priority ranking should necessarily be developed in terms of the balance between features. The two main methods for maintaining balance in Zonation-style spatial prioritization have been core-area analysis (Moilanen et al., 2005) and the additive benefit function, which utilizes feature-specific species–area curves (Moilanen, 2007; Moilanen et al., 2011). Of these, the latter may find rankings with higher mean coverage across features, but with the cost of lowered coverage for the worst-off features. Here we introduce three new and useful methods for the aggregation of marginal loss.

Novel p-norm variants: *CAZ1*, *CAZ2*, and *CAZP*.

We introduce a family of novel alternatives for the marginal loss rule based on a *p*-norm of the weighted fraction of feature distribution remaining in the cell being evaluated. The *p*-norm is a basic mathematical expression for a generalized length measure of a vector (see e.g. Horn & Johnson, 1990). Here, it is used to aggregate the marginal loss of biodiversity across features when a grid cell is hypothetically lost. Technically, the *p*-norm of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of length n is defined as $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$.

Here, the elements of vector $\mathbf{x}$ of the generic norm expression need to be replaced with application-specific elements, feature-specific fraction remaining in the cell, r_{ij}, with the index running across features j. With feature weights included, the expression we used for marginal loss of cell i becomes $M_i = \left(\sum_{j=1}^{n} \left| w_j r_{ij} \right|^p \right)^{\frac{1}{p}}$. Absolute values of weights are used due to how positively weighted and negatively weighted features are combined in analysis, described in the next section.

Certain variants of the *p*-norm are well-known quantities (see Horn & Johnson, 1990). The so-called *L*1 norm, with $p = 1$, is simply the sum of the elements of the vector, also called the 'city block' distance. Here, the L1 norm is tagged CAZ1, $M_i = \sum_j (w_j r_{ij})$, which is the weighted range-size rarity calculation common in spatial ecology (Veach et al., 2017).

Having $p = \infty$ produces the so-called infinity norm, which is technically the same as picking the maximum element of the vector containing elements $w_j r_{ij}$. This is the same as previous core-area analysis (Moilanen et al., 2005), here called by CAZMAX. The *L*2 norm (here CAZ2), with $p = 2$, is the straight-line distance or *Euclidean* distance, technically the square root of the sum of squares of the elements.

For completeness, there is the option CAZP, in which the parameter p can be chosen ($p \geq 1$). The higher the value of parameter p, the more emphasis is given to the individual feature(s) with the highest weighted fractions of their remaining distributions in the grid cell ($w_j r_{ij}$). As they are special cases, optimized implementations were used for the *L*1, *L*2 and infinity norms.

2.3 | Treatment of negatively versus positively weighted features

We also introduce modified equations for the treatment of positively and negatively weighted features: clarified mathematical symmetry is introduced for features that oppose conservation (opportunity costs, invasive species, etc.) and ecologically desirable features (species, habitats, ecosystem services, etc.). The following formulations apply irrespective of the marginal loss method used.

For positively weighted features the aim is to maintain them to the top ranks of the prioritization efficiently and in a balanced manner. Hence, the basic quantity of interest is fraction remaining in the focal cell (Table 1), $r_{ij} = p_{ij}/r_j$. For negatively weighted features, the aim is to remove them in the low ranks, likewise efficiently and in a balanced manner. Hence, to introduce this symmetry, comparison is not to remaining r_j but to that removed already, $p_{ij}/(1 - r_j + p_{ij})$. In the context of reducing opportunity costs, the interpretation would be fractional addition to opportunity. The total marginal loss for cell i is then calculated as $M_i = M_{iP} - M_{iN}$, in which M_{iP} and M_{iN} are the marginal loss aggregates for positively and negatively weighted features, respectively. When calculating M_{iN} for negatively weighted features, absolute values of weights $|w_j|$ are used so that roots need not be taken of negative quantities; the negative is simply moved out in front, as $-M_{iN}$.

A final complication with positively and negatively weighted features is when direct costs (of conservation) are used in prioritization, aiming at cost-efficiency. Best areas for conservation have high biodiversity, low opposing factors and low direct cost of conservation. Best areas for other uses are the opposite. This is achieved by using direct cost to modify aggregate marginal loss as follows: $M_i = M_{iP}/c_i - c_i M_{iN}$, in which c_i is the normalized conservation cost of cell i, that is, original cost divided by mean cost across the landscape. Normalization is used to avoid implicit change in the relative weights of positively and negatively weighted features.

2.4 | Flexibility map

The Zonation priority ranking orders the grid cells of the landscape linearly in a rank from 0 to 1. Here we describe an additional result, the flexibility map, which helps with the interpretation of the ranking. Heuristically expressed, the flexibility map investigates how much could each grid cell's rank value move up or down the ranking without meaningful loss in the quality of the ranking. The relevance of this map is easy to understand in the context of data that has large areas of equal values for features. If, for example, 30% of the grid cells are identical in feature occurrences, then (barring other factors such as connectivity) these cells can be ordered in any order without loss of performance of the ranking. On the other hand, some highly ranked areas may be unique in the sense that their ranking could not be decreased without loss of quality of the ranking at large. Overall, the motivation for calculating the flexibility map is to identify areas for which alternatives most likely exist and areas that cannot really be replaced without loss. Conceptually, this calculation is related to the concept irreplaceability, but we do not use that term here, because technically the present calculation is nothing like the irreplaceability calculations found in spatial prioritization literature (Kukkala & Moilanen, 2013).

We calculate flexibility based on the converged priority ranking and the M_{iP} and M_{iN} components of the conditional marginal loss vector. Table 2 illustrates flexibility together with the ranking.

1. First, we set a threshold to how much suboptimality is allowed for the position of a grid cell: we used maximum error of $\varepsilon = 2\%$ for illustration.
2. Then each grid cell i is moved iteratively both up and down the priority ranking while investigating how much out of place the cell would be in the new hypothetical position.

2.1. The feasibility of the move is verified by comparison between (a) the conditional marginal loss of the focal cell i in the hypothetical position of the ranking and (b) the marginal loss of the cell that was originally ranked to that position. If these numbers differ by less than ε%, then the move is allowed and a position further out is tested. This operation results in two quantities: rank_up(ε, i) and rank_down(ε, i).

TABLE 2 Illustration of priority ranking and flexibility with an extremely simplified case of a landscape of five areas, two features (both equally weighted) and the CAZ1 measure. Input occurrence levels for features are normalized and an unconditional marginal loss is calculated for each area (grid cell), which produces the initial ranking. After convergence, areas have been ranked in the order of increasing conditional marginal loss, the calculation of which utilizes feature-specific information about the fraction of distribution remaining through the ranking. Flexibility calculation is shown for one area only, #2. It turns out to have high error if moved to ranks 1, 2 or 5 in the ranking, but at position 3 it would have zero error, which is trivially observed as areas #2 and #4 have identical occurrence levels for features. Assuming an allowable error level of 2%, only areas #2 and #4 have flexibility, they could each be moved 1/5 (20%) in the ranking. None of the other areas could be moved without the hypothetical marginal loss being over 2% out of place

1. Initial state

	Occurrence levels for features		Normalized occurrence levels		Unconditional ranking (low 1 to high 5)	
Area number (ID)	#1	#2	#1	#2	Marginal loss	Rank
1	3	5	0.3	0.5	0.8	5
2	2	2	0.2	0.2	0.4	3–4 (tied)
3	0	1	0	0.1	0.1	1
4	2	2	0.2	0.2	0.4	3–4 (tied)
5	3	0	0.3	0.0	0.3	2

2. State following converged iteration

		Remaining for feature (curves)			Flexibility for area #2 (4th in the ranking)		
Rank order	Converged rank order (area IDs)	#1	#2	Converged conditional marginal loss (CAZ1)	Hypothetical marginal loss	Error fraction; %	Move acceptable
1 (low)	3	1.0	1.0	0.1/1.0 = 0.1	0.4	(0.4–0.1)/0.1; 300%	No
2	5	1.0	0.9	0.3/1.0 = 0.3	0.422	(0.422–0.3)/0.3; 41%	No
3	4	0.7	0.9	0.2/0.7 + 0.2/0.9 = 0.51	0.51	(0.51–0.51)/0.51; 0%	Yes; 20% down
4	2	0.5	0.7	0.2/0.5 + 0.2/0.7 = 0.69	Focal cell, no error	0%	NA
5 (high)	1	0.3	0.5	0.3/0.3 + 0.5/0.5 = 2	0.2/0.3 + 0.2/0.5 = 1.07	(2–1.07)/2; 47%	No
Final		0.0	0.0				

2.2. We say the flexibility, F_i, of grid cell i is the length of the interval inside the cell could be moved without breaking the error level allowed F_i = rank_up(ε, i) − rank_down(ε, i).

The flexibility measure scales between zero, for a cell that cannot move at all without breaking the error level ε, and 1.0, for a cell that could travel up and down the entire length of the ranking. Returning to the hypothetical example above, all cells in the 30% area with equal feature occurrence levels could freely travel inside that 30% block without loss of quality, meaning that those cells would be identified with a flexibility of at least 30%. Flexibility can of course be trivially investigated in one direction only, up or down, if that is preferable for the analysis need.

A key in the evaluation of each hypothetical move is that different grid cells have different features in them. Hence, the marginal loss for a cell in a hypothetical position is evaluated using the feature occurrence levels of the cell itself and coverage levels at the hypothetical position. The calculation is a time-efficient one pass calculation when positions for rank_up(ε, i) and rank_down(ε, i) are identified using some computationally efficient search method.

2.5 | Tiled FFT computations for fast connectivity calculations

Several connectivity methods of Zonation have been based on radially symmetric, declining by distance dispersal kernels, commonly used in metapopulation biology (see Lehtomäki & Moilanen, 2013 for references). Here we tested the performance of connectivity calculation via tiled Fast Fourier Transforms (FFT) for large landscapes. The use of tiling is based on the following observations. (a) Species typically have limited dispersal distances, which are very short compared to the length of a continent or even a country. (b) The far tail of the dispersal kernel becomes so thin it can be truncated, to, for example, five standard deviations for a two-dimensional normal distribution, without effectively any loss of ecological relevance. Consequently, connectivity can be calculated on small overlapping tiles. Doing so is computationally advantageous compared to always doing full-landscape connectivity calculations in which all grid cells are connected to all other grid cells irrespective of landscape size and the truncated width of the dispersal kernel. As a further time saving device allowed by tiling, if the feature does not occur in the tile at all, computation can be skipped completely, as connectivity is zero.

2.6 | Memory-saving devices, quantization of input feature distributions and run length coding

It is well known that ecological observations have inaccuracies and that predictions of statistical species distribution models have significant errors. On the other hand, statistical models output predictions with high precision, commonly with 16 decimal places. Consequently, we have implemented and tested a memory saving device based on quantization of data to lower numerical resolution.

Typically, computers store numbers as single-precision floating point values (6–9 significant digits) or double-precision numbers (15–17 significant digits). These data types require 4 and 8 bytes of data for storage, respectively. If numerical resolution is lowered to 2 or 1 bytes (16 or 8 bits), 50% or 75% of memory is saved compared to single-precision floats. Effectively, 8-bit quantization takes the range of occurrence levels of an input layer and divides that range into 2^8 = 256 bins into which inputs are assigned. The use of 256 bins corresponds to a maximum quantization error per pixel of 0. 5 × 1/256 = 0.5 × 0.004 = 0.195%. This is order(s) of magnitude less than a typical expected modelling error. For comparison, thresholding corresponds to 1-bit quantization with a maximum error of 50% per pixel, and thresholded data have been commonly used in spatial analysis during the past decades. We tested the hypothesis that 8-bit quantization should not influence a priority ranking to any significant degree.

We paired quantization with a standard lossless data compression technique from computer science, run length coding, which codes sequences of identical numbers as data value and count instead of repeating the original value, which massively compresses sparse data. Memory usages reported below include the effects of both quantization and run length coding.

2.7 | Data for illustrative examples

We tested the computational properties of the new algorithm using two empirical datasets. The first data are for the Finnish marine areas (Virtanen et al., 2018), for which species and habitat distributions were available at 20 m native resolution as well as aggregated to 40 and 100 m grid cell resolution. These datasets have 204,501,900, 51,447,742 and 8,179,875 grid cells with data, respectively. The matrix size of the native resolution data is 24,100 × 38,076, which is large enough that Zonation 4 cannot perform connectivity transforms on it. The marine data were also resampled to 10 and 5 m resolutions to test operations on extremely high-dimensional data; the effective numbers of spatial elements for these data are approximately 818 million and 3.27 billion.

The second data are for Greater Hunter, New South Wales, Australia. These include distribution maps for 504 species (amphibians, birds, mammals, plants and reptiles) and 20 threatened ecological communities at 100 m resolution (Kujala et al., 2015). The data were further resampled to 200 and 50 m resolutions, with the three sets having 603,370, 2,413,196 and 9,652,784 grid cells with data, when going from the lowest to highest resolution.

We illustrate the new flexibility map feature using a dataset well suited for the purpose: suitability of urban areas for 10 taxa in the Helsinki metropolitan area, Finland (Jalkanen, Vierikko, &

Moilanen, 2020). A notable feature of the dataset is that ecologically low-value impermeable surfaces (asphalt; about 30% of the urban area) were included in suitability maps, to maintain full coverage of the metropolitan area.

The identities of these datasets make little difference for testing. From the perspective of computation, what matters is the matrix size, the effective number of elements and the number of data layers.

3 | RESULTS

We investigate the convergence of the new meta-algorithm and the speed and memory usage of the new computational methods versus Zonation 4. Speed and memory usage are important, because computational load limits the size of analysis that is feasible to implement, with relevance to high-resolution national, continental and global analyses.

3.1 | Meta-algorithm convergence

We tested the convergence of the new meta-algorithm by applying Zonation 4 (old meta-algorithm) and the proposed methods on the same datasets (low-resolution versions of the Finnish marine and Greater Hunter datasets), using those marginal loss rules that are available in both, ABF and CAZMAX. It was verified that both the old and new meta-algorithms find effectively the same result: the rank correlation between solutions from alternative algorithms varied from 98.18% to 99.97% for the 8-bit quantized occurrence data and from 99.99% to 99.9998% with 16-bit quantized data. Given the typical use of a full priority ranking, this is close enough given that (a) the meta-algorithms are completely different, (b) ties in data are resolved differently, (c) Zonation 4 was using an acceleration factor of approximately 0.06% of pixel count (that many pixels were removed and ranked in one iteration), (d) present analyses were run on quantized (compressed) pixel-level data, (e) Zonation 4 uses a derivative-based approximation in ABF, whereas the present implementation does not (see Moilanen 2007).

3.2 | Performance, speed and memory usage of the novel meta-algorithm

Figure 1 shows the performance of the new prioritization meta-algorithm and tiled FFT computations as a function of problem size. As expected, computation time increases close to linearly as a function of feature count (Figure 1a). The scaling of computation time as a function of effective number of grid cells is lower than $N\log_2 N$ (Figure 1b), which is a major improvement from the N^2 for the old version of the meta-algorithm. FFT-based kernel connectivity calculations run in seconds or minutes per layer (Figure 1c), which is at least an order of magnitude faster than before. Also, the transforms can work on landscapes in the order of billions of elements where Zonation 4 was capped at 1 billion elements for the input matrix size.

It was found that the speed difference between the present and past implementation of the prioritization algorithm increased with problem size, as expected. For example, the ratio of total computation times was 23.0 for the 100 m marine data and 136.6 for the 20 m marine data, which was the largest data that could be computed using Zonation 4. Overall, depending on data and using the current convergence settings, the total computation time used by the proposed methods ranged from 0.5% to 4.5% of that used by Zonation 4.

In addition to computational speed, memory usage may limit the size of a problem that can be analysed. Memory usage is case specific, depending on the average coverage of features across the landscape. Of the present data, the Greater Hunter data have statistical distribution models for species with relatively widespread occurrences predicted. With the 100-m Greater Hunter data, memory usage of proposed methods was 37% of that used by Zonation 4. With the 40-m marine data, in which species are less widespread, the memory demand of proposed methods was approximately 8.2% of Zonation 4.

3.3 | New marginal loss rules

Testing the new marginal loss rules, we confirmed that average performance is negatively correlated with the performance of the worst-off features: marginal loss rules that perform best on average (ABF & CAZ1) do less well in the tails than CAZ2 and CAZMAX (Figure 2). The new methods, CAZ1 and CAZ2, do well with this data, with relatively minor reduction in mean performance for significant gains in the lower tail of performance.

3.4 | Flexibility map

Figure 3 illustrates the flexibility map. Figure 3a shows the priority rank, in which a North–South gradient can be seen through areas with largely impermeable surfaces. These areas are equal in terms of biodiversity. Hence, the gradient emerges only because areas must by definition be ordered in a ranking. Figure 3b, the flexibility map calculated at 2% maximum error level, shows that the ranking in these areas is indeed very flexible, that is, cell ranks could be altered with hardly any impact on the quality of the ranking. Figure 3c,d shows flexibility up (3c) or down (3d) in ranks.

4 | DISCUSSION

This study describes major algorithmic developments in the subfield of conservation biology, spatial prioritization. These include a new priority ranking algorithm, novel ways of balancing trade-off between many features, more sophisticated kernel-based connectivity

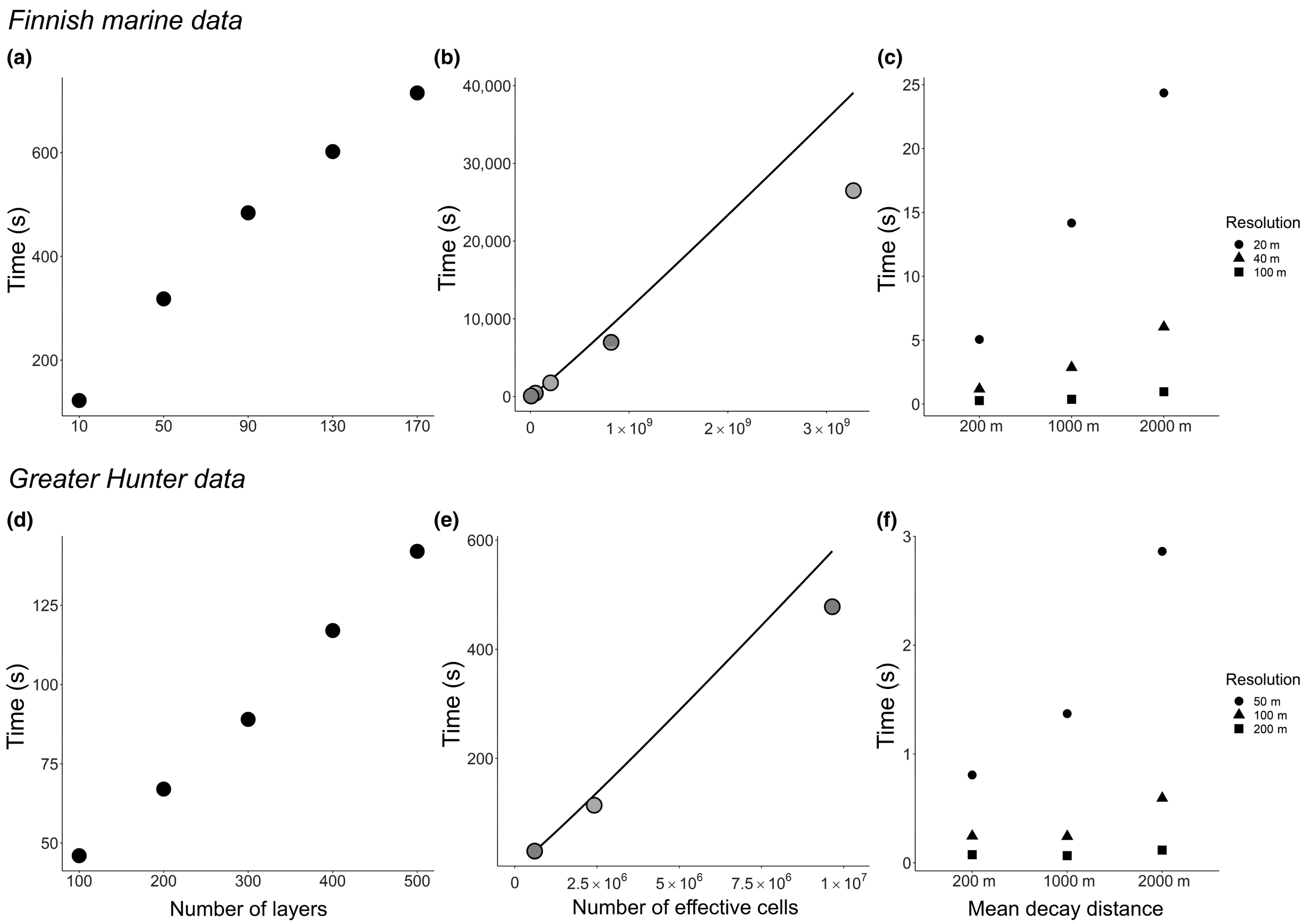

FIGURE 1 Algorithm performance using subsets or versions of the Finnish marine (a–c) and greater hunter (d–f) datasets. (a) and (d), scaling of computation time for full analysis as a function of feature layers, mean of five randomly selected sets of layers from the full data. (B) and (E), scaling of computation time used in the priority ranking, as a function of the effective number of grid cells with data in the landscape, N. A $N\log_2 N$ function is plotted for comparison, scaled to start from the same starting point. (c) and (f), scaling of computation time used for the tiled FFT transform of one layer. As a consequence of the tiled transform structure, computation time increases with the truncated width of the dispersal kernel

calculations, memory saving techniques and the flexibility map. The motivation for this work arose from the observation that high-resolution (e.g. 1 ha) national scale analyses started to become impossible, when the effective data size went up to 50–100 million grid cells with data (e.g. Virtanen et al., 2018). Large data imply lots of data to add to and to find errors from, inevitably leading to many repeats of potentially complicated analysis sets including, for example, variants with and without costs, with and without connectivity, for major taxa together and separately, etc. Clearly, being able to do analyses with much smaller computational resources would provide real operational advantages.

The proposed methods provide several advantages compared to methods previously available (Lehtomäki & Moilanen, 2013; Moilanen et al., 2005, 2011). (a) Analysis will be possible at a much higher spatial resolution than before, which links more directly to on the ground planning and reduces need to lose information via spatial aggregation, (b) development of analysis sets on very large problems becomes much more feasible than before, (c) the computational speed of the new algorithm may support interactive planning in working groups even for moderately large problems (Figure 1), and (d) the novel marginal loss rules described here. (e) Also, the new flexibility map facilitates interpretation of the strength of the priority pattern, allowing identification where the ranking is relatively flexible and where it is not. Effectively, these advances translate into major improvements in real-world planning capacity.

With respect to the flexibility map (Figure 3), the following interpretations would be useful commonly. If the area has high rank and little flexibility down, the area is certain to be good for conservation. If the area has low rank and little flexibility up, the area is a good candidate for impact avoidance—but with the assumption that underlying data are good.

With respect to the new marginal loss rules, differences between their results would depend on the dataset, namely the nestedness hierarchy of the features (species): are there many environments in analysis and do richness and rarity correlate strongly or not? Nevertheless, based on present results, CAZ1 and CAZ2 show good performance compared to the previously available ABF and CAZMAX, when looking at the trade-off between mean

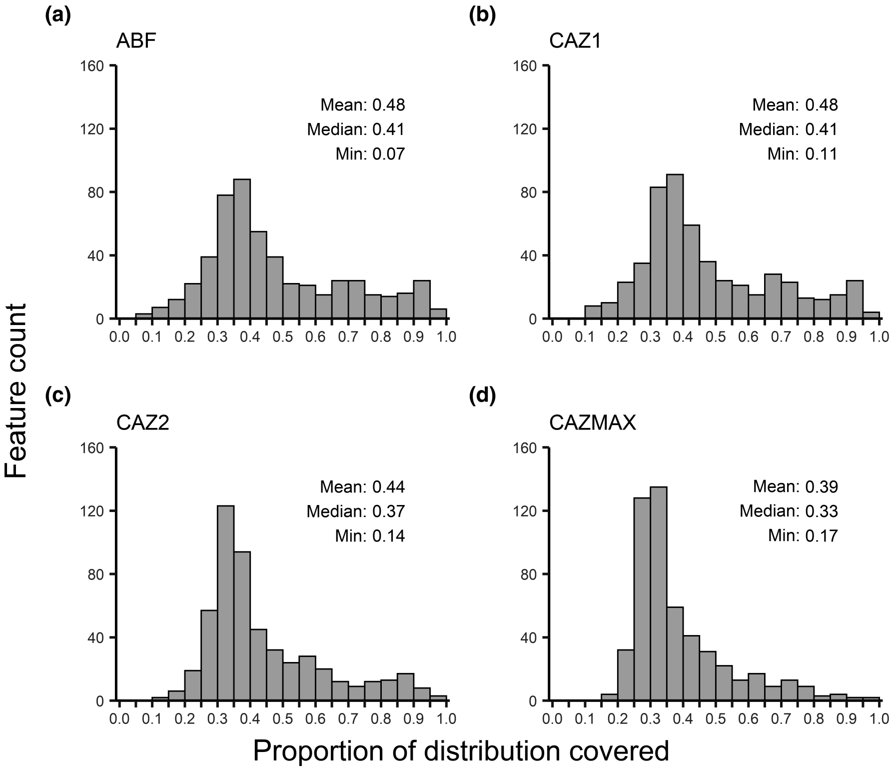

FIGURE 2 Illustrating differences between marginal loss rules by histograms of feature coverage. The histograms show counts of species at different levels of coverage inside the highest ranked 25% of the greater hunter study area. (a) ABF, (b) CAZ1, (c) CAZ2 and (d) CAZMAX. The mean, median and minimum coverage levels have been marked into each panel

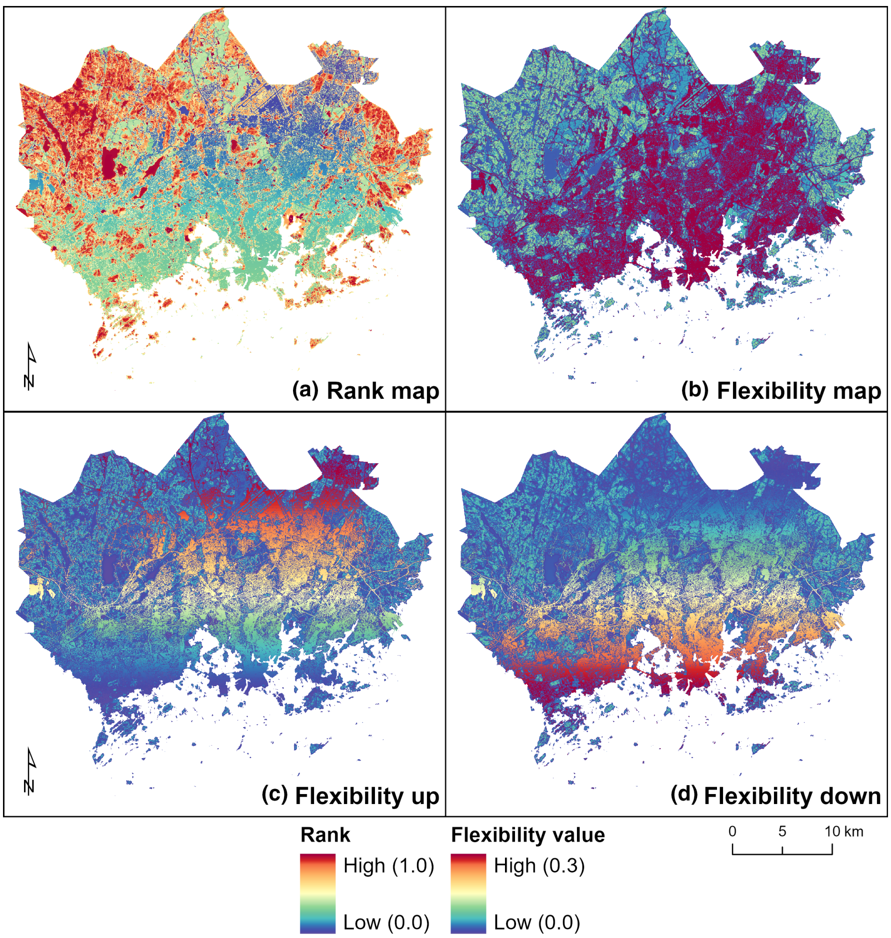

FIGURE 3 Illustrating flexibility with an analysis from the Helsinki capital district in Finland (Jalkanen, Vierikko, & Moilanen, 2020). This area has a large, mostly built, low value area, which shows high flexibility. (a) Priority rank map, (b) total flexibility map, (c) flexibility up, (d) flexibility down

coverage and performance in the lower tail of the coverage distribution (Figure 2).

We showed that the time used in analysis scales approximately linearly as a function of feature count. The time used by the new ranking algorithm scales as faster than $N\log_2 N$, as function of effective number of elements with data (N), compared to N^2 in Zonation 4. This makes a major difference when data size rises to tens of millions, hundreds of millions or billions of grid cells with data. For example, our tests show a run time of 333 min for an analysis with 177 feature layers, an effective landscape size of 818 million grid cells inside a matrix of dimension 48,200 × 76,152 (3.7 billion elements), and with connectivity calculations up to a 2 km distance applied on all layers. Dimensionally, this analysis corresponds to high-resolution continental-scale analysis and would have been impossible using previous implementations of spatial prioritization (in Zonation 4). The new algorithm together with technical solutions enables analyses of 10–50× the size that has been conveniently possible before (Virtanen et al., 2018). Analysis limitations have recently been observed by Wyborn and Evans (2021), who observe that many global analyses are not fully useful in local decision-making because of heavy data aggregation during analysis.

Because of a compressed main data structure and tiled FFT transforms, the memory usage of the proposed methods is with large problems in the order of 10% of that of Zonation 4, which indicates that the same computer will run a significantly larger analysis than would have been possible before. The difference in memory usage is explained by the computational techniques used, namely, data compression using run length coding and quantization of occurrence data. These savings in memory usage suggest that continental or global analyses, in which species typically occur in small parts of the area, would compress massively compared to past implementations.

To conclude, the present developments improve significantly the capacity of conservation science to address high-dimensional spatial (conservation) planning problems using balanced priority ranking. Methods described here are to be released in a forthcoming software release, which will also include many further developments of less methodological nature.

ACKNOWLEDGEMENTS

We gratefully thank the Koneen Säätiö, grant #201803179. A.M. and H.K. also thank the Strategic Research Council project #312559, IBC-Carbon and A.M. the Finnish Ministry of Environment (project MetZo-III) and the BioDiversa Belmont Forum project FutureWeb (Academy of Finland decision #326343) for support.

CONFLICT OF INTERESTS

Authors have no conflict of interest.

AUTHORS' CONTRIBUTIONS

Vision and major algorithms A.M. and P.L.; funding A.M. and H.K.; implementation P.L.; analyses and testing I.K., P.L., J.J., H.K. and E.A.V.; data E.A.V., H.K. and J.J.; manuscript preparation, all authors.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/2041-210X.13819.

DATA AVAILABILITY STATEMENT

The Zonation 5 software release candidate executables used here (Windows and Linux), documentation and example data are available for download from a data repository https://doi.org/10.5281/zenodo.5899003 (Moilanen et al., 2022).

ORCID

Atte Moilanen https://orcid.org/0000-0002-2194-0267
Joel Jalkanen https://orcid.org/0000-0002-6170-9916
Elina A. Virtanen https://orcid.org/0000-0001-9702-6677
Heini Kujala https://orcid.org/0000-0001-9772-3202

REFERENCES

Horn, R. A., & Johnson, C. R. (1990). *Norms for vectors and matrices. Chapter 5 in matrix analysis*. Cambridge University Press.

Jalkanen, J., Toivonen, T., & Moilanen, A. (2020). Identification of ecological networks for land-use planning with spatial conservation prioritization. *Landscape Ecology*, *35*, 353–371.

Jalkanen, J., Vierikko, K., & Moilanen, A. (2020). Spatial prioritization for urban biodiversity quality using biotope maps and expert opinion. *Urban Forestry & Urban Greening*, *49*, 126586.

Kareksela, S., Moilanen, A., Tuominen, S., & Kotiaho, J. S. (2013). Use of inverse spatial conservation prioritization to avoid biological diversity loss outside protected areas. *Conservation Biology*, *27*, 1294–1303.

Kujala, H., Lahoz-Monfort, J. J., Elith, J., & Moilanen, A. (2018). Not all data are equal: Influence of data type and amount in spatial conservation prioritisation. *Methods in Ecology and Evolution*, *9*, 2249–2261. https://doi.org/10.1111/2041-210X.13084

Kujala, H., Moilanen, A., & Gordon, A. (2018). Spatial characteristics of species distributions as drivers in conservation prioritization. *Methods in Ecology and Evolution*, *9*, 1121–1132. https://doi.org/10.1111/2041-210X.12939

Kujala, H., Whitehead, A. L., Morris, W. K., & Wintle, B. A. (2015). Towards strategic offsetting of biodiversity loss using spatial prioritization concepts and tools: A case study on mining impacts in Australia. *Biological Conservation*, *192*, 513–521. https://doi.org/10.1016/j.biocon.2015.08.017

Kukkala, A. S., & Moilanen, A. (2013). Core concepts of spatial prioritisation in systematic conservation planning. *Biological Reviews of the Cambridge Philosophical Society*, *88*, 443–464.

Lehtomäki, J., Kusumoto, B., Shiono, T., Tanaka, T., Kubota, Y., & Moilanen, A. (2019). Spatial conservation prioritization for the east Asian islands: A balanced representation of multi-taxon biogeography in a protected area network. *Diversity and Distributions*, *25*, 414–429. https://doi.org/10.1111/ddi.12869

Lehtomäki, J., & Moilanen, A. (2013). Methods and workflow for spatial conservation prioritization using zonation. *Environmental Modelling & Software*, *47*, 128–137. https://doi.org/10.1016/j.envsoft.2013.05.001

Lehtomäki, J., Moilanen, A., Toivonen, T., & Leathwick, J. (2016). *Running a zonation planning project*. isbn: 978-951-51-1922-3. University of Helsinki.

Margules, C. R., & Sarkar, S. (2007). *Systematic conservation planning* (1st ed.). Cambridge University Press.

McIntosh, E. J., Pressey, R. L., Lloyd, S., Smith, R. J., & Grenyer, R. (2017). The impact of systematic conservation planning. *Annual Review of Environment and Resources*, *42*, 677–697.

Moilanen, A. (2007). Landscape Zonation, benefit functions and target-based planning: Unifying reserve selection strategies. *Biological Conservation*, *134*, 571–579.

Moilanen, A., Anderson, B. J., Eigenbrod, F., Heinemeyer, A., Roy, D. B., Gillings, S., Armsworth, P. R., Gaston, K. J., & Thomas, C. (2011). Balancing alternative land uses in conservation prioritization. *Ecological Applications*, *21*, 1419–1426. https://doi.org/10.1890/10-1865.1

Moilanen, A., Franco, A. M. A., Early, R. I., Fox, R., Wintle, B., & Thomas, C. D. (2005). Prioritizing multiple-use landscapes for conservation: Methods for large multi-species planning problems. *Proceedings of the Royal Society B: Biological Sciences*, *272*, 1885–1891. https://doi.org/10.1098/rspb.2005.3164

Moilanen, A., Lehtinen, P., Kohonen, I., Jalkanen, J., Virtanen, E. A., & Kujala, H. (2022). Zonation 5 v1.0rc2 (release candidate 2) software upload. *Zenodo*. https://doi.org/10.5281/zenodo.5899003

Veach, V. L., Di Minin, E., Montesino Pouzols, F., & Moilanen, A. (2017). Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. *Diversity and Distributions*, *23*, 715–726.

Virtanen, E. A., Viitasalo, M., Lappalainen, J., & Moilanen, A. (2018). Evaluation, gap analysis, and potential expansion of the Finnish marine protected area network. *Frontiers in Marine Science*, *5*, 402.

Watts, M. E., Ball, I. R., Stewart, R. S., Klein, C. J., Wilson, K., Steinback, C., Lourival, R., Kircher, L., & Possingham, H. P. (2009). Marxan with zones: Software for optimal conservation based land- and sea-use zoning. *Environmental Modelling and Software*, *24*, 1513–1521.

Williams, P. (2001). Complementarity. In S. A. Levin (Ed.), *Encyclopedia of biodiversity* (pp. 813–829). Academic Press.

Wyborn, C., & Evans, M. C. (2021). Conservation needs to break free from global priority mapping. *Nature Ecology & Evolution*, *5*, 1322–1324.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Moilanen, A., Lehtinen, P., Kohonen, I., Jalkanen, J., Virtanen, E. A., & Kujala, H. (2022). Novel methods for spatial prioritization with applications in conservation, land use planning and ecological impact avoidance. *Methods in Ecology and Evolution*, 13, 1062–1072. https://doi.org/10.1111/2041-210X.13819