# What data to use for forest conservation planning? A comparison of coarse open and detailed proprietary forest inventory data in Finland

Joona Lehtomäki1,2\* Sakari Tuominen3 Tuuli Toivonen1 and Antti Leinonen4

1 Department of Biosciences, P.O. Box 65 (Viikinkaari 1), FI-00014 University of Helsinki, Finland  
2 Finnish Environment Institute, Natural Environment Centre, P.O. Box 140 (Mechelininkatu 34a), FI- 00251 Helsinki, Finland  
3 Finnish Forest Research Institute (Metsäntutkimuslaitos), Vantaa, Finland  
4 The Finnish Forest Centre (Suomen Metsäkeskus), Kajaani, Finland  
\* Corresponding author

**Contact:**  
joona.lehtomaki@helsinki.fi, tel. +358-9-191-57714  
sakari.tuominen@metla.fi  
tuuli.toivonen@helsinki.fi  
antti.leinonen@metsakeskus.fi

**Journal:**  
PLoS ONE

**Type of paper:**  
Original research paper

**Short title:**  
Open vs proprietary data in forest conservation

**Version:**  
v0.4.9

**Manuscript statistics:**  
Abstract length = 272  
Main text length = 6720

Figure captions length = 410  
Figures = 6  
Tables = 1  
References = 93

# Abstract

The boreal region is facing intensifying resource extraction pressure, but the lack of comprehensive biodiversity data makes operative forest conservation planning difficult. Many countries in the region have implemented forest inventory schemes that are making extensive and up-to-date forest databases increasingly available. Many detailed regional inventory databases, however, remain proprietary. Here, we investigate how well different types of open and proprietary forest inventory data sets in Finland suit the purpose of conservation prioritization. We also explore how much priorities are affected by using the less accurate, but open data. Firstly, we constructed a set of indices for forest conservation value based on quantitative information commonly found in forest inventories, such as the maturity of the trees, tree species composition, and site fertility. Secondly, using these data and accounting for connectivity between forest types, we investigated the patterns in conservation priority. We did the prioritization using Zonation, a method and software for spatial conservation prioritization and validated the prioritizations by comparing them to known areas of high conservation value. We show that the overall pattern of high priority area remains relatively robust across different data sources and analysis options. However, coarse data miss much of the fine-scale variation in forest structures and more detailed data is needed to be able to account for more small-scaled features of conservation value. These results underline the importance of making detailed inventory data publicly available, because basing conservation land-use decisions on too coarse data may entail serious risk of omission and commission errors. Finally, we discuss how the prioritization methodology we used could be integrated into operative forest management in especially in countries in the boreal zone.

**Keywords**: boreal forests; conservation planning; forest conservation management; open data; spatial conservation prioritization; Zonation software

# 1. Introduction

## 1.1 Multitude of objectives for conservation prioritization

Biodiversity conservation deals with multifaceted and complex problems [1] that call for inter- or transdisciplinary research and decision-making [1–3]. Several different kinds of data are typically required [4], such as spatial data on species distributions, habitats and ecosystem services [5,6], costs associated with conservation actions [7], the structure and representativeness of the existing reserve network [8], and increasingly information the present and future state of dynamic environments [9,10] and anthropogenic threats [11].

Spatial conservation prioritization is a form of conservation assessment primarily interested in when, where, and how should conservation action be taken in order to achieve conservation goals [12,13]. Conservation prioritization problems have been extensively studied conceptually and mathematically over the years [14] and consequently many software methods for solving a wide array of problems have also been published [15–19]. Here we are interested in how well different types of forest inventory data are suitable as input for spatial conservation prioritization, which largely depends on whether the data captures the occurrence of conservation value accurately enough. Furthermore, on-ground conservation decision are almost always tied to a relatively fine spatial scale which implies that the data used for conservation prioritization should also have resolution relevant for the prioritization problem at hand [6,20].

## 1.2 Informative conservation decision-making depends on available data

A plethora of studies has been published regarding the occurrence of biodiversity. New technological advancements such as relatively cheap and very accurate remote sensors have led to a formidable increase in the available biodiversity data [21]. However, in most regions of the world primary biodiversity data for conservation decision-making still remains scarce [22,23] and biased [24,25]. For most species and most parts of the world we simply do not have sufficient data [23,26] and even when we do, it is not necessarily accessible. Sharing data is especially desirable from the decision-making point of view because of the many benefits it entails, such as enabling integrative and synthesizing science [27], enabling exploration of new topics not envisioned by the data originators [28], and providing more verifiable research for policymakers [29]. Many public and private organizations collect and maintain research and monitoring databases that could be valuable for conservation decision-making, but remain unavailable because of lacking data-policies or technical barriers for data sharing. Often there are good reasons for withholding the data, such as detailed location data on endangered species or confidential information concerning the privacy of individuals [28,30,31]. Restricting access to such information is not only an ethical obligation, but also often a legal one. Much of the potentially useful biodiversity data thus remains of restricted availability, but decisions still have to be made [23,32]. With great potential for better informed decision-making, open access to relevant data is crucial for addressing increasingly complex conservation issues the world is facing [22,31,33–35].

Effective conservation planning should ideally be done simultaneously with general land-use and natural resource use planning [6], which further emphasizes the need to be able to synthesize and utilize data from various sources. Data for land use and natural resources use planning may also be useful for conservation planning, assuming that they act as surrogates for biodiversity features of conservation interest. The upside is that resources allocated for collecting these types of data usually exceed those allocated for conservation-related data collection. For example, countries with an active forest sector typically have high-resolution, national, forest inventory systems (NFIs) in place [36,37]. In addition to NFIs, many other public and private operators collect detailed forest inventory data for their own operational planning often at the national or regional scale. Recently, especially governments and public research institutions have started opening up their databases. For example, the Finnish Forest Research Institute has quite opened up their multi-source national forest inventory database (<http://www.metla.fi/ohjelma/vmi/vmi-moni-en.htm>).

## 1.3 Forest inventory data in spatial conservation prioritization and validation

Forest inventory data has historically been collected to assess the productive functions of forests [38], but especially large-scale NFIs are increasingly being used for monitoring forest biodiversity [39]. Following the classification by Corona et al. [39], biodiversity indicators estimated from forest inventory data can be classified into two categories: (i) compositional indicators directly measuring biodiversity, and (ii) structural indicators based on key structural features (e.g. variability in tree size and the amount of dead wood) acting as correlates or surrogates for biodiversity [39]. The latter approach largely relies on the assumption that broad structural and tree species diversity provides more habitats for different forest species [39–41]. It also requires that the features can be reliably estimated from forest inventory data [36,42]. The approach based on structural indicators also has many desirable qualities from the perspective of spatial conservation prioritization. First, as structural components are comparatively easy to measure, data about them are commonly included in forest inventories [38,39]. Second, the effects different forest management options have on structural features and thus on biodiversity can be easily assessed [43], which enables comparisons between different management scenarios. Third, forest inventories are typically repeated periodically, making it possible to monitor changing conditions [39]. Fourth, since forest inventory data still are primarily collected for forest management and planning purposes, conservation prioritization based on forest inventory data can be more easily understood by forestry practitioners. Finally, the data, and thus the results of conservation prioritization analyses, are produced at a resolution directly relevant for operative planning.

Validating the results of a conservation prioritization analysis is an important, but often overlooked part of the whole prioritization process (**REF**). In other words, maps and other results of prioritization assessments are often produced assuming that the input data is sensible quality and thus the priorities reflect on-ground reality adequately [44]. VA prioritization can be validated for example by comparison to additional local-scale data about distributions of species and habitats. Validating the results is even more important when relying on surrogate data such as structural forest inventory data. At the structural level, one can use as crude validation data the locations of known conservation value, such as existing protected areas, which should on average be more valuable than the surrounding (economically managed) landscape.

## 1.4 Aims and scope

Here, we develop a set of conservation prioritization analyses based on freely available and proprietary forest inventory data with a varying degree of detail across the province of South Savonia, Finland. We use the conservation prioritization software Zonation to develop complementarity-based priority maps while also accounting for connectivity in the solutions. With this approach, we are studying the following questions:

1) Can conservation prioritization analysis based on forest inventory data capture conservation value in boreal managed forest landscapes?

2) How well does freely available coarse forest inventory data perform compared to more detailed proprietary stand-based inventory data?

3) Given the differences in the data reliability and prioritization results, under what kind of planning circumstances is open but coarse inventory data sufficient for informative conservation decision-making?

We limit our attention to the effects that different data sources have on the quality of spatial prioritization, and we acknowledge that the computational analysis described here is just one part of a full conservation planning process [15,45]. While the results will be case-specific to a certain degree, the procedure itself is applicable to other countries that have similar forest inventory data available. The results should also be applicable to countries with a similar forest management history and current forest and conservation management needs. We have done the work in collaboration with the Finnish Forest Center and the results have already been utilized in the implementation of the METSO forest biodiversity programme.

To encourage other scientists and practitioners to build upon the work presented here, we also make available the full analysis implementation (**REF**) and the code necessary to produce the results from the prioritization analyses. While the proprietary data we have used cannot be shared because of privacy issues (however, see section 2.3.2), we have made all the stages of the implementation openly available for examination and re-use.

# 2. Material and Methods

## 2.1 Study area

The study area covers the region of Southern Savonia located in Southeastern part of Finland (Figure 1). South Savonia is one of 13 regional administrative units of Finnish Forest Centre. The size of the region is ca. 13990 km2 and it is characterized by a large number of lakes and fragmented waterways, which cover ca. 25% of the total area. Of the land area, approximately 88 % is forestry land that can further be divided into mineral soils (79%) and mires (21%). The south boreal vegetation zone covers the whole region and forests are mostly dominated by the Scots pine (*Pinus sylvestris*) and the Norway spruce (*Picea abies*), mixed with varying amounts of broadleaf trees. Land ownership is highly fragmented with private forest owners being the largest group (77.3%) followed by private companies (11.5%) and the state (6.2%) [46]. Most of the forestry land is under silvicultural management with only 2.5% strictly protected. This number is the same as the average for forestry land in Southern Finland (2.5%).

Whereas private forest land has several operators working there (including Finnish Forest Centre), the state-owned land is managed by a single organization, Metsähallitus, which is further divided into two independent departments: the Forestry Department manages the Finnish state production forests and the Natural Heritage Services manage forests outside of commercial operation, including protected areas.

## 2.2 Study design

Fig 2 presents design of our study and Table 1 the data sets used in the analysis. To address the main objectives, we 1) we acquired coarse and detailed forestry inventory data from Southern Savonia, 2) calculated comparable surrogate indices of conservation value out of these data, 3) carried out six different conservation prioritizations using three different input data sets and testing the influence of connectivity transformations, and 4) compared all prioritization results to areas with known high conservation value. We were interested only in forest land on mineral soils.

## 2.3 Datasets

### 2.3.1 Coarse data

The coarse data used in this study was based on the multi-source national forest inventory (MS-NFI) developed maintained by the Finnish Forest Research Institute (Metla). The MS-NFI method employs satellite images, digital maps and field measurements to estimate thematic digital maps about structural features of the forest across Finland at a spatial resolution of 20 m. MS-NFI data collection covers all land-use classes and ownership categories throughout the country [37,47,48]. The final data product contains over 40 forest variables in the form of thematic maps, including, e.g., the volumes by tree species and timber assortments, stand mean variables, the biomass by tree species groups and tree compartments and forest site type characteristics [e.g. 37,48,49]. In Finland, the MS-NFI is being used mostly for regional level forestry planning, but it has also been used for large-scale conservation prioritization studies [20,49,50].

The MS-NFI data has been publicly available since late 2012, the thematic maps can be viewed through a web portal, and the rasters can be downloaded through a file service [51]. The conservation value indexes used for the prioritization (see 2.4) require that information on both the average diameter and the volume are available for each tree species group. The standard MS-NFI rasters include only one estimate for average diameter over all tree species groups. In order to calculate estimates of average diameter for each tree species group, stand level variables were derived from the MS-NFI by the way of automatic stand delineation.

We did the stand delineation was carried out in the study area by automatic segmentation of the MS-NFI forest maps of the 10th NFI. As the input data for the segmentation, we used the thematic map layers on stand mean height and volumes of the tree species groups: pine, spruce, birch and other broadleaved trees. The segmentation was carried out using a modified implementation of the “segmentation with directed trees” algorithm by Nagendra & Goldberg [52]. The algorithm sis based on using the local edge gradient for linking individual pixels into larger spatially continuous units, i.e. segments. The automatic segmentation process is guided by parameters such as heterogeneity allowed within the segments and the desired minimum size of the segments [53]. Here the desired size of segments was approx. 1-2 ha. We calculated the stand level variables as average values of the individual pixels within each segment and the variables per tree species by weighting the pixel level variables by the volumes of individual tree species.

### 2.3.2 Detailed data

With “detailed”, we refer to detailed data for stands or forestry compartments. The data are produced by a combination of direct field inventories and a representative plot-based sampling system. Nowadays, inventory data is updated also using remote sensing data (LiDAR). These data are collated to provide very fine scale information for forest management planning [47] by different authorities and forestry organizations depending on land tenure. For this study, we used data from two authorities operating in the study region: Finnish Forest Centre (FFC) on private land and Metsähallitus (Finnish Forest and Park Service) Natural Heritage Services (NHS) on public land.

FFC inventories the forest stands are inventoried only on need-basis or when forestry operation take place, some of the inventory data can be relatively old and thus does not represent the current state of the forest very well. To account for this, we only used data gathered in year 2000 or after. After the filtering, the data available from FFC covered ~44% of the land area. Another additional source of information we used was spatial data on the planned forestry operations that contain information on planned operations such as thinnings and clear-cuts. We used these data to discount the value of forest areas that are planned to go through forest operations of varying degree (SI 1.3). The forest inventory data gathered and managed by FFC is not freely available as the Finnish Personal Data Act restricts the distribution of the data at a resolution that allows linkage of the data to properties of individual forest owners. It is possible, however, to get access to the data for research purposes [54].

NHS has similar inventory system in place on public land. The NHS information system update the database every yearly to simulate the growth of forests, and consequently we no filtering was needed. We received detailed stand-based data from NHS after signing a research collaboration agreement, as these data are not freely available for public use. We were unable the get any data from regions governed by Metsähallitus Forestry for reasons unknown to us. Detailed data from Metsähallitus NHS covers ~2.4% of the land area in South Savonia.

### 2.3.3 Data for validation

We used three different data sets for validating the prioritization results: spatial delineations of 1) established protected area network, 2) woodland key-habitats, and 3) recently acquired protected areas (Table 1).

NHS maintains the data on established protected areas and a spatial database is publically available. Protected areas also cover mires, but for validation, we used only protected areas on mineral soils (~1.9% of the whole landscape).

Woodland key-habitats (WKH) are a conservation instrument designed for maintaining landscape-level biodiversity in production forests by delineating and preserving small habitat patches of especially high conservation value [55]. The concept is in use in many Fennoscandian and Baltic countries and while their effectiveness as a conservation measure varies depending on the country and definition [55–57], WKHs seem to be hotspots for dead wood dependent and red-listed species, and for species richness in general [55]. Because of potential privacy issues, the exact spatial locations of WKHs are not public information, but the data is available for research use.

Recently acquired protected areas are related to the ongoing forest biodiversity conservation programme METSO that is an ongoing effort to halt the decline of forest biodiversity by year 2016 [58]. Individual forest owners can make offer their forest property to be protected, and if the particular offer fulfills given scientific selection criteria, it is admitted into the programme through either a permanent or temporary (10 years) conservation contract. Forest owner then receives a tax-free compensation based on the economic value of the growing stock and timber [59]. The sites selected in METSO are ecologically valuable than average Finnish forest containing more dead wood as well as many red-listed species [60]. We used only areas with permanent conservation contracts for validation, as the conservation effectiveness of temporary or fixed-term contracts is questionable [60]. This data is not publicly available before becomes integrated into the main protected areas database, but it can be accessed for research use.

## 2.4 Calculating conservation value indices from original data

We reclassified the original forestry data (both coarse and detailed) into four tree species groups: pine, spruce, birch, or other broadleaved (Table S1). We calculated an index of conservation value per pixel for each of the tree species groups in each of the data sources separately. This index measures an expert-opinion based view on how the average diameter and the volume of the growing stock relate to ecological features desirable for conservation. We transformed the average diameter of the growing stock per tree species group by a sigmoidal benefit function (see SI 1.2 and Figure S1) and then multiplied the transformed value by the volume of the growing stock. A similar approach has been used earlier in large-scale conservation prioritization [20,49] and in species-oriented prioritization [50].

All available data sets do have had information on site fertility class, which is often also associated with the formation of specific forest microhabitats. Therefore we further created two input data sets based on the coarse data: One with just the 4 index rasters (“Coarse”), one with the 4 index rasters each divided into 5 site fertility classes (“Coarse with classes”, Figure 2). We also hypothesized that prioritzation based on the more detailed - and more precise - inventory data would outperform those based on the coarser data (“Detailed with classes”)(Table S1). Note that input data set “Detailed with” is actually a combination of coarse and detailed as the detailed data only covered ~46% of the landscape (Table 1). Both “coarse” and “coarse with classes” are completely based on MS-NFI and thus publicly available data.

We converted all detailed vector data to rasters of the same resolution (20 m) and extent (SI 1.3) as the MS-NFI. For computational reasons, we aggregated the data to 60 x 60 meters pixel size using ArcGIS [61]. We wanted to retain as high a resolution as possible because conservation prioritization analyses should be carried out at a spatial scale that is informative about ecological components (e.g connectivity and average size of habitat patches) and relevant at the scale of operative planning [20]. For calculating conservation value indices at this resolution, we used custom-made geospatial scripts based on Python [62] bindings to GDAL [63].

## 2.5 Prioritizing locations for conservation

For the spatial prioritization, we used Zonation [16,64] version 4.0 [65]. It operates with a set of input rasters that describe the occurrence levels of biodiversity features across the landscape; in our case the features were the index rasters of forest conservation value (see 2.4). Zonation starts with looking at the study area and the proceeds by iteratively removing the least valuable pixels simultaneously accounting for the occurrence of features in pixels, the remaining occurrence of each feature across the landscape, and connectivity. In the end, Zonation has ranked the whole landscape according to its conservation priority. We encourage the reader to refer to a large body of existing literature for the conceptual background (**REF**), the exact operational principles (**REF**), and the different applications of Zonation (**REF**).

Following best-practices for constructing Zonation runs [15], we started from the simplest possible configurations, enabling more complex features one at a time. This way, it is possible to test for the exact effects each component of the analysis introduces and the sensitivity of the results for different parameter values. After testing with several different combinations, we set up two runs for each input data sets (“Coarse”, “Coarse with classes”, and “Detailed with classes”): one with and one without connectivity. Thus, we completed six different analysis runs in total (R1-R6 in Figure 2).

All six runs R1-R6 shared certain Zonation configuration options. We used the additive benefit function mode in Zonation [66], because it is appropriate when dealing with habitat data that acts as surrogate for biodiversity at large [49]. The weight that each individual feature received (Table S2) was based on expert opinion, and the weights reflect subjective importance given to particular tree species groups and site fertility classes (see SI 1.3).

Runs R1-R6 also had differences (Figure 2). The number of biodiversity features (the index rasters) varied from four in R1 and R2 to twenty in R3-R6. Runs based on the detailed data (R5 and R6) used additional information about planned forestry operations (see 2.3.2). Technically, we implemented this in Zonation using the data as a condition layer, where local quality (as measured by the values in the index rasters) was reduced at locations with forestry operations [65]. Runs R1, R3, and R5 are so called “local” variants in the sense that they do not include any connectivity transformations. Runs R2, R4, and R6 on the other hand account for connectivity between different forest types (see SI 1.2 and Table S3). We used the so-called matrix-connectivity feature of Zonation [49,65], in which connectivity is shared between multiple partially similar environments (SI 1.3). The spatial scale of the connectivity transformation effect is in Zonation controlled by a feature-specific parameter (α), which is derived from the scale of landscape use of each species of community occupying a habitat type [15,20,49]. We used a value of α (0.001), which corresponds to an average dispersal distance of 2.0 kilometers in a negative exponential dispersal kernel. See Lehtomäki et al. [49] and Sirkiä et al. [50] for further discussion and references about the distances chosen. We also tested the sensitivity of results replicating analysis with scales of 0.2 and 4.0 km, but these did not change the qualitative interpretation of results significantly. See Arponen et al. [20] further discusses the role of the spatial scale.

## 2.6 Comparison and validation of analysis runs

We examined the spatial patterns of rank priorities at different spatial resolutions and between different priority ranges. We did comparisons between (i) runs based on the same input dataset but different analysis settings (i.e. the effect of connectivity) and (ii) between different input datasets analyzed using the same settings (i.e. the effect of the input data). We performed all comparisons using the standard Zonation outputs and data described in section 2.4. We used R (version 3.1.0) statistical language [67] and the zonator R-package [68].

Visual examination of the priority rank maps should give an initial idea how well the different runs - and hence the different input datasets - converge especially in terms of high and low priorities. We also compare the spatial overlap between analyses by calculating Jaccard coefficients (the intersection of two sets divided by the union of those sets) for different priority ranges. In other words, we divide each rank raster into 10 equal range bins and compare each bin to every other. This way we can compare for example the spatial overlap of the best 10% of the landscape in runs R1 and R3, but also for example the worst 10% in R1 with the best 10% in R3.

It is possible to load the priority rank map of a Zonation solution and evaluate the same prioritization using different input features or Zonation options. This procedure allows examination of how much performance is lost when the analysis criteria and evaluation criteria differ [15,65]. We used this technique to evaluate how much lower the representation levels are for the input features based on the detailed data (R5) when the priority ranking is taken from analyses based on the coarser data. Assuming that the detailed data also is more precise, we can then answer the question of how much feature representation do we risk losing if forced to rely on the coarser data.

We examined how well different prioritization runs were able to identify forest regions with high conservation value. We did this by simply examining the priority distributions within areas covered by each of the validation data sets (see 2.3.3).

# 3. Results

## 3.1 Spatial patterns of rank priority

Overall, the spatial pattern of priority was roughly consistent across non-spatial runs accounting for local quality only (Figure 3A, 3C, and 3E). A major concentration of high-priority areas was identified in the southwestern corner of the study area. Classifying the coarse input dataset according to the site fertility classification (R3, see 2.1 and Figure 2) had only a minor effect of distributing the top priorities more equally across the study area (Figure 3A and 3C). Zonation tries to retain a balanced representation of all features throughout the analysis and therefore introducing more classes (i.e. features) will produce more even spatial priority patterns unless the most valuable features are spatially aggregated. R5, which is based on the more detailed data, produced a different priority pattern (Figure 3E). Top priorities distribute even more equally over the study area (marginal plots in Figure 3E). Regions of high-concentrations of top-priority areas also partially shift towards the northeastern part of the study area. This shift is at least partly explained by the fact that the more detailed data gives higher value to the two large national parks in the northeastern region.

Runs including connectivity between forest types (R2, R4, R6) display very similar rank priority patterns compared to their non-spatial counterparts. Regions with higher density of top-priority areas (Figure 3B, 3D, and 3F) receive higher overall priority because of the connectivity effect, which is evident from the marginal plots in Figure 3B, 3D, and 3F.

## 3.2 Spatial overlap of priority ranges

Comparing rank priority ranges between solutions shows the effect of using the site fertility classification. Figure 4A displays an asymmetrical pattern of overlap between priority ranges in R1 and R3. Large areas in R3 receive slightly lower priorities than in R1, which is balanced by a small set of areas having significantly higher ranks in R3 than in R1. There is little overlap between high priorities in R1 and low priorities in R3 (upper-left part of panel 4A) whereas the inverse is different: there is some overlap with relatively high priorities in R3 and low priorities in R1 (lower-right part of panel 4A). The classification of the data causes these overlapping patterns in priority range. More specifically, some soil fertility classes (most notably the herb-rich and xeric soil fertility types) are rarer than others and consequently receive more weight in Zonation analysis. This is because Zonation will give more priority to features that are rare to begin with.

The best and worst 10% of the priorities have the largest spatial overlaps in all comparisons. Since data classification is the only difference between R1 and R3, their overall similarity is larger which also explains the higher overlap of the best and worst 10% of priorities (Figure 4A). The overlap is smaller for the best and worst 10% of priorities between R1/R5 and R1/R3, but still those overlaps are higher than for the rest of the priority ranges. In other words, the best and the worst areas are more similar between all the analyses even if the underlying input datasets are different.

Figure 4D-4F show the spatial overlaps between runs that account for connectivity. Patterns are similar to the patterns in the non-spatial (i.e. not accounting for connectivity, Figure 4A, 4B, and 4C) runs with the difference that the patterns are smoother and more aggregated in all comparisons (Figure 4D-4F). Comparisons between runs based on the same input data set with and without accounting for connectivity (Figure 4G-4I) show a strong overlap between the same priority ranges in different runs. The overlap tends to increase when moving towards the highest or lowest priority areas of the study area. This reaffirms that connectivity as defined in this study has an effect only on a local scale.

## 3.3 Feature representation

Loading the priority rank order from the runs based on coarse input data (R1 and R3) revealed differences in performance. Figure 5 shows the overall performance, i.e. how much of the initial representation levels from the detailed data can be covered by protecting a given fraction of the landscape. Figure 5A shows that on average, priority rankings R1 and R3 perform much worse than R5. For example, protecting the best 10% of the landscape using the ranking from R5 would cover on average approximately 54% of the original distributions of all features from the detailed input dataset. In comparison, solutions R1 and R3 would cover on average only ~15% and ~16% of the features in the detailed data, respectively (Figure 5A). This difference is even more pronounced when examining the solutions that use additional site fertility classes. For example, the best 10% of the landscape covers ~93% of features in herb-rich sites, whereas solutions R1 and R3 only achieve a coverage of ~15% and ~14%, respectively (Figure 5B). For every other site fertility class except for mesic, the performance of R5 is superior to that of R1 and R3. The performance levels of runs that account for connectivity (R2, R4, R6) are omitted here, because they are very similar to those of R1, R3, and R5.

## 3.4 Comparison to spatial validation data

Protected areas have relatively high median priorities in all runs R1-R6 (Figure 6). R5 and R6 have the highest median priorities (~0.85 and ~0.90), followed by R1 and R2 (~0.71 and ~0.69) (REF). Woodland key-habitats also have quite high median priorities in solutions R5 and R6 (both ~0.69), but the distribution of priorities is not as skewed as with protected areas. For R3 and R4, WKHs have a median priority of ~0.48, and the median values are even lower for R1 and R2 (~0.42 and ~0.41). Locations admitted to the METSO programme receive the highest median priorities values in R5 and R6 (both ~0.82). R1 and R2 have a median priority value similar to those of protected areas (~0.72 and ~0.70), as do R3 and R4 (~0.68 and ~0.65). In all cases, the difference between a runs with and without connectivity is small, except for in the case of protected areas. Overall solutions R5 and R6 perform better than the others, demonstrating the utility of using detailed data from proprietary on-the-ground forest inventories.

# 4. Discussion

## 4.1 Can forest inventory data be used to identify valuable areas for conservation?

Our results demonstrate that 1) inventory data collected primarily for operational forest planning is informative for spatial conservation prioritization, and 2) openly available remote-sensing based data performs reasonably well for large mature forest areas, but fails to detect valuable sites of smaller size. Therefore, if the spatial prioritization includes objectives for detecting small scale biodiversity feature occurrences such as the WKHs, a more detailed input data are needed.

On the scale of the whole study area, priority patterns between runs based on the coarse and on the detailed data are relatively similar with at least three key differences. First, analyses based on the coarser data give higher priority to a large area at the southwestern part of the province. This is because the MS-NFI data has high estimated values for birch and other deciduous trees in the region, which also has a high incidence of fertile soils. Deciduous trees and fertile soil types are less common than other tree species and soil types. They furthermore have higher weights assigned in the Zonation analysis due to relatively high associated biodiversity values (see S1.2 for further explanation). Second, analyses based on the more detailed data give existing large protected areas even much higher priorities. This is most probably because compared to coarse data, the detailed data available from within protected areas describes more accurately the mature stands within the PAs. Third, since the detailed data has information also on the occurrence of small but valuable forest (e.g. herb-rich sites or mature deciduous trees) that is not correctly represented in the coarse data, the high-priority sites are more evenly distributed over the whole study area (see the marginal plots in Figure 3).

Of the three validation data sets, woodland key-habitats have the smallest average size per site and the most fine-grained structural features important for biodiversity. The coarse data is simply unable to pick up such features. This is not surprising as the coarse data we are using (MS-NFI) is known to have low statistical precision for small area estimates [39,69]. Of course, when available, information about WHKs can be included in the prioritization process itself. We did not do so here, because that would have excluded the use of WHK data as an external validation source.

Extent and resolution are important factors in analyses that account for connectivity. The small effect of connectivity has on the priority rank distributions of the validation data sets may appear surprising, since the effect of connectivity is quite pronounced over larger areas (Figure 3). However, even when combined the validation data sets cover only a small fraction of the total landscape (2.5%, Table 1) and the mean decay distance for dispersal we used (2 km, see 2.5) is relatively large compared to the average size of sites in the validation data. For these reasons, accounting for connectivity actually decreases the median priority for all other validation data sets except the PAs, which are larger and thus by definition better connected internally.

The validation procedure we have used relies on few key assumptions. First, we assume that the indices we have constructed truly reflect conservation value. While we have not validated the indexes against actual species occurrence data, features we have emphasized in the construction of the index are important for biodiversity in the Finnish boreal forest (see e.g. [49,50], S1.2, and S1.3). Second, we assume that the validation data sets actually describe locations of high conservation value, and that they should therefore receive higher than average priority in spatial prioritization analyses. Protected areas have traditionally been established on less productive soils [70,71] and they usually do not represent the full spectrum of species or habitats in any given region. However, being set aside from the prevailing forest management regimes will over time lead to a less even forest structure [72], thereby accumulating important resources such as dead-wood [73]. METSO-sites are on average smaller than many of the existing PAs, but because of the stringent selection criteria and on-ground evaluation of each site, their ecological quality is high and studies have shown that they do indeed have higher species richness and rarity than their surrounding areas [60]. WKHs are scattered more evenly over the landscape and according to a recent meta-analysis [55] they contain elevated amounts of critical resources (dead-wood, etc.) that support a comparatively large number of species. However, the average size of a WKH site is small (0.67 ha in Finland [74]), meaning that their capability to support populations in the long run is questionable.

## 4.2 Trade-offs between different data and prioritization objectives

Conservation scientists, managers, and practitioners are often faced with tight schedules and limited budgets, and thus have to decide whether it is worth the time and money to try to collect more data [75,76]. Collecting more data also includes spending time and money on trying to gain access to more detailed data that is not openly available. Conservation prioritization based on incomplete data runs the risk of commission and omission errors, selecting sites that are not valuable in reality or missing sites that are [24]. According to our results, the spatial overlap of the top priorities is fairly consistent between the analysis based on coarse and the detailed data (Figure 4) and the same applies to the lowest priorities. More importantly, the top priorities of any of the analyses do not much overlap with the low priorities in any other run. If they did, using coarse data as basis for prioritization would produce wildly different and often incorrect results.

While coarse data reveals broad priority patterns correctly, we found that the less abundant biodiversity features such as herb-rich and xeric forest types are not identified well (Figure 5). For example, if we are interested in the top 10% of the landscape, prioritization based on coarse data with classes captures only half of the representation of biodiversity features that can be achieved if using detailed data. Even if the top priority locations have a large overlap spatially, using the coarser data misses much of the occurrences of herb-rich sites and woodland key habitats.

The differences between the analyses based on the coarse and coarse with classes input datasets are particularly interesting, because a simple classification scheme that improves the performance of the results would be easy to implement in practice. The inclusion of the classification does slightly improve the performance for rarer classes (Figure 5) so everything else being equal, an ecologically justified classification of the data can improve the results.

Including connectivity in the analysis raises the priority of regions that have high quality sites at high densities, thus identifying regions where metapopulations might be able to persist. This is particularly important for many threatened forest species that suffer from habitat loss and fragmentation [77–79]. However, emphasizing connectivity will happen at the expense of individual high-quality sites that are relatively isolated [20,80]. Increasing the priority of medium-quality and well-connected forests will lower the priority of other locally similar sites and possibly even poorly connected high-quality sites (Figure 3). Including connectivity will also emphasize large, overall high-quality areas such as protected areas (Figure 6).

## 4.3 Opening up forest inventory data is an opportunity for integrated forest and conservation planning in the Boreal zone

The circumpolar boreal forest is the second largest biome in the world [81]. Countries in the boreal zone have traditionally utilized their forest-based natural resources extensively, which has led to changes in forest structure, species composition, habitat diversity, and large-scale disturbance dynamics [82–86]. While it is not the most species-rich or threatened biome on the planet [87,88], there are still many reasons for increasing conservation efforts in the boreal zone. First, boreal forests host a great number of highly specialized species that are dependent on resources such as dead wood [89–91]. Many of these species have become endangered because of intensive forestry practices. Second, because of their large extent and biomass, boreal forests have a major role in carbon sequestration and climate change adaption [81,92]. Third, many parts of boreal zone, especially in the Russian Federation and Canada, remain inaccessible presenting an opportunity to protect large tracts of relatively intact forest [70,93].

Open forest inventory data has a major role in conservation planning and decision-making in the boreal region. It enables equal access to the best available data, it makes the supporting scientific analysis more transparent, and it enhances the repeatability of the whole conservation planning process [28]. Repeatability is especially important for applied research supporting decision-making, because underlying objectives may change, old data is updated, and new information can accumulate rapidly. Transparency and repeatability are also important for the process of translating regional plans into local conservation action: whereas regional plans incorporate important factors such connectivity and the representativeness of the protected area network as a whole, local action can be understood as individual management actions that sometimes unfortunately are poorly linked to regional planning [93]. Plugging into regional and local forest planning through the use of forest inventory data presents new opportunities for conservation prioritization especially in countries of the boreal zone which already have sophisticated forest planning and inventory systems in place.

In summary, we have shown that coarse, NFI-derived data works reasonably well in the identification of broad spatial conservation priorities, but we also found that more detailed inventory data is needed to capture the structural attributes at the local-scale. While it is encouraging to see that inventory data is becoming more openly available, conservation research and decision-making would benefit from more open data policies especially in government organizations. The approach we have taken in this work builds upon previously published work [20,49,50] and methodology [15]. Here we make all analysis implementations (see S1.1) and data (where possible, see 2.3) available to enable others to adapt the approach for their own uses. The approach described here is being used in the implementation of the Finnish national forest conservation programme, and we continue our efforts to improve the approach.

# 5. Acknowledgments

# References

1. Game ET, Meijaard E, Sheil D, MacDonald-Madden E (2014) Conservation in a wicked complex world; challenges and solutions. Conserv Lett 7: 271–277. doi:10.1111/conl.12050.This.

2. Reyers B, Roux DJ, Cowling RM, Ginsburg AE, Nel JL, et al. (2010) Conservation Planning as a Transdisciplinary Process. Conserv Biol 24: 957–965. doi:10.1111/j.1523-1739.2010.01497.x.

3. Pooley SP, Mendelsohn JA, Milner-Gulland EJ (2013) Hunting Down the Chimera of Multiple Disciplinarity in Conservation Science. Conserv Biol 28: 22–32. doi:10.1111/cobi.12183.

4. Keane A (2013) Unusual data in conservation science: searching for validation. Anim Conserv 10: 604–605. doi:10.1111/acv.12091.

5. Ferrier S, Drielsma M (2010) Synthesis of pattern and process in biodiversity conservation assessment: a flexible whole-landscape modelling framework. Divers Distrib 16: 386–402. doi:10.1111/j.1472-4642.2010.00657.x.

6. Ferrier S, Wintle BA (2009) Quantitative Approaches to Spatial Conservation Prioritization: Matching the Solution to the Need. In: Moilanen A, Wilson KA, Possingham HP, editors. Spatial Conservation Prioritization: Quantitative Methods & Computational Tools. Oxford: Oxford University Press. p. 304.

7. Naidoo R, Balmford A, Ferraro PJ, Polasky S, Ricketts TH, et al. (2006) Integrating economic costs into conservation planning. Trends Ecol Evol 21: 681–687.

8. Branquart E, Verheyen K, Latham J (2008) Selection criteria of protected forest areas in Europe: The theory and the real world. Biol Conserv 141: 2795–2806. doi:10.1016/j.biocon.2008.08.015.

9. Scheller RM, Mladenoff DJ (2007) An ecological classification of forest landscape simulation models: tools and strategies for understanding broad-scale forested ecosystems. Landsc Ecol 22: 491–505.

10. Kujala H, Moilanen A, Araújo MB, Cabeza M (2013) Conservation planning with uncertain climate change projections. PLoS One 8: e53315. doi:10.1371/journal.pone.0053315.

11. Visconti P, Pressey RL, Segan DB, Wintle BA (2010) Conservation planning with dynamic threats: The role of spatial design and priority setting for species’ persistence. Biol Conserv 143: 756–767. doi:10.1016/j.biocon.2009.12.018.

12. Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson KA (2007) Conservation planning in a changing world. Trends Ecol Evol 22: 583–592.

13. Wilson KA, Underwood EC, Morrison SA, Klausmeyer KR, Murdoch WW, et al. (2007) Conserving Biodiversity Efficiently: What to Do, Where, and When. PLoS Biol 5: 12.

14. Moilanen A, Possingham HP, Polasky S (2009) A Mathematical Classification of Conseravation Prioritization Problems. In: Moilanen A, Wilson K, Possingham HP, editors. Spatial Conservation Prioritization: Quantitative Methods & Computational Tools. Oxford: Oxford University Press. pp. 28–42.

15. Lehtomäki J, Moilanen A (2013) Methods and workflow for spatial conservation prioritization using Zonation. Environ Model Softw 47: 128–137. doi:10.1016/j.envsoft.2013.05.001.

16. Moilanen A, Kujala H, Leathwick JR (2009) The Zonation framework and software for conservation prioritization. In: Moilanen A, Wilson KH, Possingham HP, editors. Spatial Conservation Prioritization. Oxford University Press. pp. 196–210.

17. Pressey RL, Watts ME, Barrett TW, Ridges MJ (2009) The C-Plan Conservation Planning System: Origins, Applications and Possible Futures. In: Moilanen A, Wilson K, Possingham HP, editors. Spatial Conservation Prioritization: Quantitative Methods & Computational Tools. Oxford University Press. pp. 211–234.

18. Ciarleglio M, Barnes JW, Sarkar S (2009) ConsNet: new software for the selection of conservation area networks with spatial and multi-criteria analyses. Ecography (Cop) 32: 205–209. doi:10.1111/j.1600-0587.2008.05721.x.

19. Possingham HP, Ball IR, Andelman SJ (2000) Mathematical methods for identifying representative reserve networks. In: Ferson S, Burgman MA, editors. Quantitative methods for conservation biology. New York: Springer-Verlag. pp. 291–305.

20. Arponen A, Lehtomäki J, Leppänen J, Tomppo E, Moilanen A (2012) Effects of connectivity and spatial resolution of analyses on conservation prioritization across large extents. Conserv Biol 26: 294–304. doi:10.1111/j.1523-1739.2011.01814.x.

21. Pettorelli N, Laurance WF, O’Brien TG, Wegmann M, Nagendra H, et al. (2014) Satellite remote sensing for applied ecologists: Opportunities and challenges. J Appl Ecol 44: early view. doi:10.1111/1365-2664.12261.

22. Maeda EE, Torres JA (2012) Open environmental data in developing countries: who benefits? Ambio 41: 410–412. doi:10.1007/s13280-012-0283-4.

23. Pressey RL (2004) Conservation planning and biodiversity: assembling the best data for the job. Conserv Biol 18: 1677–1681. doi:10.1111/j.1523-1739.2004.00434.x.

24. Rondinini C, Wilson K a, Boitani L, Grantham H, Possingham HP (2006) Tradeoffs of different types of species occurrence data for use in systematic conservation planning. Ecol Lett 9: 1136–1145. doi:10.1111/j.1461-0248.2006.00970.x.

25. Boitani L, Maiorano L, Baisero D, Falcucci A, Visconti P, et al. (2011) What spatial data do we need to develop global mammal conservation strategies? Philos Trans R Soc B Biol Sci 366: 2623–2632.

26. Grantham HS, Pressey RL, Wells JA, Beattie AJ (2010) Effectiveness of Biodiversity Surrogates for Conservation Planning: Different Measures of Effectiveness Generate a Kaleidoscope of Variation. PLoS One 5: 1–12. doi:10.1371/journal.pone.0011430.

27. Carpenter SR, Armbrust EV, Arzberger PW, Iii FSC, Elser JJ, et al. (2009) Accelerate Synthesis in Ecology and Environmental Sciences. Bioscience 59: 699–701. doi:10.1525/bio.2009.59.8.11.

28. Uhlir PF, Schröder P (2007) Open data for global science. Data Sci J 6: 36–53.

29. Wolkovich EM, Regetz J, O’Connor MI (2012) Advances in global change research require open science by individual researchers. Glob Chang Biol 18: 2102–2110. doi:10.1111/j.1365-2486.2012.02693.x.

30. Egloff W, Patterson DJ, Agosti D, Hagedorn G (2014) Open exchange of scientific knowledge and European copyright: The case of biodiversity information. Zookeys 135: 109–135. doi:10.3897/zookeys.414.7717.

31. Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, et al. (2004) An International Framework to Promote Access to Data. Science (80- ) 303: 1777–1778. doi:10.1126/science.1095958.

32. Bode M, Wilson KA, Brooks TM, Turner WR, Mittermeier RA, et al. (2008) Cost-effective global conservation spending is robust to taxonomic group. Proc Natl Acad Sci U S A 105: 6498–6501. doi:10.1073/pnas.0710705105.

33. Reichman OJ, Jones MB, Schildhauer MP (2011) Challenges and Opportunities of Open Data in Ecology. Science (80- ) 331: 703–705. doi:10.1126/science.1197962.

34. Pullin AS, Salafsky N (2010) Save the whales? Save the rainforest? Save the data! Conserv Biol 24: 915–917. doi:10.1111/j.1523-1739.2010.01537.x.

35. Huang X, Hawkins B a., Lei F, Miller GL, Favret C, et al. (2012) Willing and unwilling to share primary biodiversity data: results and implications of an international survey. Conserv Lett 5: 399–406. doi:10.1111/j.1755-263X.2012.00259.x.

36. Chirici G, Winter S, McRoberts RE (2011) National Forest Inventories: Contributions to Forest Biodiversity Assessments. Dordrecht: Springer.

37. Tomppo EO, Haakana M, Kaitila M, Peräsaari J (2008) Multi-source national forest inventory – methods and applications. Springer.

38. Chirici G, Mcroberts RE, Winter S, Bertini R, Bra U, et al. (2012) National Forest Inventory Contributions to Forest Biodiversity Monitoring. For Sci 58: 257–268.

39. Corona P, Chirici G, Mcroberts RE, Winter S, Barbati A (2011) Contribution of large-scale forest inventories to biodiversity assessment and monitoring. For Ecol Manage 262: 2061–2069. doi:10.1016/j.foreco.2011.08.044.

40. Winter S, Chirici G, Mcroberts E, Hauk E, Tomppo E (2008) Possibilities for harmonizing national forest inventory data for use in forest biodiversity assessments. Forestry 81. doi:10.1093/forestry/cpm042.

41. Kallio AM, Hänninen R, Vainikainen N, Luque S (2008) Biodiversity value and the optimal location of forest conservation sites in Southern Finland. Ecol Econ 67: 232–243. doi:10.1016/j.ecolecon.2008.05.005.

42. Tomppo E, Gschwantner T, Lawrence M, McRoberts RE (2010) National Forest Inventories Pathways for Common Reporting. Media: 612.

43. McElhinny C, Gibbons P, Brack C, Bauhus J (2005) Forest and woodland stand structural complexity: Its definition and measurement. For Ecol Manage 218: 1–24. doi:10.1016/j.foreco.2005.08.034.

44. Langford WT, Gordon A, Bastin L, Bekessy SA, White MD, et al. (2011) Raising the bar for systematic conservation planning. Trends Ecol Evol 26: 634–640. doi:10.1016/j.tree.2011.08.001.

45. Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, et al. (2013) Predicting species distributions for conservation decisions. Ecol Lett 16: 1424–1435. doi:10.1111/ele.12189.

46. Finnish Forest Research Institute (2013) Finnish Statistical Yearbook of Forestry: 450.

47. Tuominen S, Balazs A, Korhonen KT, Muinonen E (2014) NFI plots as complementary reference data in forest inventory based on airborne laser scanning and aerial photography in Finland. Silva Fenn 48: 1–20. doi:10.14214/sf.983.

48. Tomppo EO (2006) The Finnish National Forest Inventory. In: Kangas A, Maltamo M, editors. Forest inventory: Methodology and applications. Dordrecht: Springer. pp. 179–194.

49. Lehtomäki J, Tomppo E, Kuokkanen P, Hanski I, Moilanen A (2009) Applying spatial conservation prioritization software and high-resolution GIS data to a national-scale study in forest conservation. For Ecol Manage 258: 2439–2449. doi:10.1016/j.foreco.2009.08.026.

50. Sirkiä S, Lehtomäki J, Lindén H, Tomppo E, Moilanen A (2012) Defining spatial priorities for capercaillie Tetrao urogallus lekking landscape conservation in south-central Finland. Wildlife Biol 18: 337–353.

51. Finnish Forest Research Institute (2014) Multi-source National Forest Inventory (MS-NFI). Available: http://www.metla.fi/ohjelma/vmi/vmi-moni-en.htm. Accessed 4 May 2014.

52. Nagendra PM, Goldberg M (1980) Image Segmentation with Directed Trees. IEEE Trans Pattern Anal Mach Interligence 1: 185–191.

53. Pekkarinen A (2002) Image segment-based spectral features in the estimation of timber volume. Remote Sens Environ 82: 349–359. doi:10.1016/S0034-4257(02)00052-4.

54. Mönkkönen M, Juutinen A, Mazziotta A, Miettinen K, Podkopaev D, et al. (2014) Spatially dynamic forest management to sustain biodiversity and economic returns. J Environ Manage 134C: 80–89. doi:10.1016/j.jenvman.2013.12.021.

55. Timonen J, Gustafsson L, Kotiaho JS, Mönkkönen M (2011) Hotspots in cold climate: Conservation value of woodland key habitats in boreal forests. Biol Conserv 144: 2061–2067. doi:http://dx.doi.org/10.1016/j.biocon.2011.02.016.

56. Aune K, Jonsson BG, Moen J, Gunnar B (2005) Isolation and edge effects among woodland key habitats in Sweden: Is forest policy promoting fragmentation? Biol Conserv 124: 89–95. doi:10.1016/j.biocon.2005.01.015.

57. Pykälä J, Heikkinen RK, Toivonen H, Jääskeläinen K (2006) Importance of Forest Act habitats for epiphytic lichens in Finnish managed forests. For Ecol Manage 223: 84–92. doi:10.1016/j.foreco.2005.10.059.

58. Finnish Government (2008) Government Resolution on the Forest Biodiversity Programme for Southern Finland 2008-2016 (METSO): 15. Available: http://www.mmm.fi/attachments/metsat/5yckfcmWR/METSOResolution2008-2016\_ENGL.pdf. Accessed 3 May 2014.

59. Korhonen K, Hujala T, Kurttila M (2013) Diffusion of voluntary protection among family forest owners: Decision process and success factors. For Policy Econ 26: 82–90. doi:10.1016/j.forpol.2012.08.010.

60. Siitonen J, Penttilä R, Ihalainen A (2012) METSO-ohjelman uusien pysyvien ja määräaikaisten suojelualueiden ekologinen laatu Uudenmaan alueella. Metsätieteen Aikakausik: 259–284.

61. ESRI (2014) ArcGIS Desktop, version 10.2.1. Redlands, CA: Environmental Systems Research Institute.

62. Python Development Team (2014) Python Language Reference, version 2.7. Python Software Foundation.

63. GDAL Development Team (2014) GDAL - Geospatial Data Abstraction Library, version 1.10.1. Open Source Geospatial Foundation.

64. Moilanen A, Franco AMA, Early RI, Fox R, Wintle BA, et al. (2005) Prioritizing multiple-use landscapes for conservation: methods for large multi-species planning problems. Proc R Soc B Biol Sci 272: 1885–1891. doi:10.1098/rspb.2005.3164.

65. Moilanen A, Meller L, Leppänen J, Montesino-Pouzols F, Kujala H, et al. (2014) Zonation spatial conservation planning framework and software V4.0, User manual. Helsinki.

66. Moilanen A (2007) Landscape Zonation, benefit functions and target-based planning: unifying reserve selection strategies. Biol Conserv 134: 571–579.

67. R Core Team (2014) R: A Language and Environment for Statistical Computing, version 3.1.0. Vienna, Austria: R Foundation for Statistical Computing.

68. Lehtomäki J (2014) zonator: Utilities for Zonation spatial conservation prioritization software. R package version 0.3.9. Helsinki.

69. Tomppo EO (2006) The Finnish multi-source National Forest Inventory-small area estimation and map production. In: Kangas A, Maltamo M, editors. Forest inventory: Methodology and Applications. Dordrecht: Springer. pp. 195–224.

70. Elbakidze M, Angelstam PK, Sobolev N, Degerman E, Andersson K, et al. (2013) Protected area as an indicator of ecological sustainability? A century of development in Europe’s boreal forest. Ambio 42: 201–214. doi:10.1007/s13280-012-0375-1.

71. Scott JM, Davis FW, McGhie RG, Wright RG, Groves C, et al. (2001) Nature reserves: Do they capture the full range of America’s biological diversity? Ecol Appl 11: 999–1007. doi:10.1890/1051-0761(2001)011[0999:NRDTCT]2.0.CO;2.

72. Kuuluvainen T, Tahvonen O, Aakala T (2012) Even-aged and uneven-aged forest management in boreal Fennoscandia: a review. Ambio 41: 720–737. doi:10.1007/s13280-012-0289-y.

73. Siitonen J, Martikainen P, Punttila P, Rauh J (2000) Coarse woody debris and stand characteristics in mature managed and old-growth boreal mesic forests in southern Finland. For Ecol Manage 128: 211–225. doi:10.1016/S0378-1127(99)00148-6.

74. Timonen J, Siitonen J, Gustafsson L, Kotiaho JS, Stokland JN, et al. (2010) Woodland key habitats in northern Europe: concepts, inventory and protection. Scand J For Res 25: 309–324. doi:10.1080/02827581.2010.497160.

75. Grantham HS, Wilson K a, Moilanen A, Rebelo T, Possingham HP (2009) Delaying conservation actions for improved knowledge: how long should we wait? Ecol Lett 12: 293–301. doi:10.1111/j.1461-0248.2009.01287.x.

76. Grantham HS, Moilanen A, Wilson KA, Pressey RL, Rebelo TG, et al. (2008) Diminishing return on investment for biodiversity data in conservation planning. Conserv Lett 1: 190–198. doi:10.1111/j.1755-263X.2008.00029.x.

77. Nordén J, Penttilä R, Siitonen J, Tomppo E, Ovaskainen O (2013) Specialist species of wood-inhabiting fungi struggle while generalists thrive in fragmented boreal forests. J Ecol 101: 701–712. doi:10.1111/1365-2745.12085.

78. Henle K, Davies KF, Kleyer M, Margules C, Settele J (2004) Predictors of species sensitivity to fragmentation. Biodivers Conserv 13: 207–251. doi:10.1023/B:BIOC.0000004319.91643.9e.

79. Ranius T, Kindvall O (2006) Extinction risk of wood-living model species in forest landscapes as related to forest history and conservation strategy. Landsc Ecol 21: 687–698. doi:10.1007/s10980-005-5222-3.

80. Hodgson JA, Thomas CD, Wintle BA, Moilanen A (2009) Climate change, connectivity and conservation decision making: back to basics. J Appl Ecol 46: 964–969. doi:10.1111/j.1365-2664.2009.01695.x.

81. Bradshaw CJ, Warkentin IG, Sodhi NS (2009) Urgent preservation of boreal carbon stocks and biodiversity. Trends Ecol Evol 24: 541–548. doi:10.1016/j.tree.2009.03.019.

82. Kuuluvainen T, Grenfell R (2012) Natural disturbance emulation in boreal forest ecosystem management - theories, strategies, and a comparison with conventional even-aged management. Can J For Res 1203: 1185–1203. doi:10.1139/X2012-064.

83. Mönkkönen M (1999) Managing Nordic boreal forest landscapes for biodiversity: ecological and economic perspectives. Biodivers Conserv 8: 85–99.

84. Stephens SL, Burrows N, Buyantuyev A, Gray RW, Keane RE, et al. (2014) Temperate and boreal forest mega-fires: characteristics and challenges. Front Ecol Environ 12: 115–122. doi:10.1890/120332.

85. Halme P, Allen KA, Auninš A, Bradshaw RHW, Brumelis G, et al. (2013) Challenges of ecological restoration: Lessons from forests in northern Europe. Biol Conserv 167: 248–256. doi:10.1016/j.biocon.2013.08.029.

86. Esseen P, Ehnström B, Ericson L, Sjöberg K (1997) Boreal forests. Ecol Bull 46: 16–47.

87. Mace GM, Masundire H, Baillie JEM (2005) Chapter 4: Biodiversity. In: Hassan R, Scholes RJ, Ash N, editors. Ecosystems and human well-being: current state and trends, Volume 1. Washington D.C.: Island Press. pp. 79–122.

88. Puumalainen J, Kennedy P, Folving S (2003) Monitoring forest biodiversity: a European perspective with reference to temperate and boreal forest zone. J Environ Manage 67: 5–14. doi:10.1016/S0301-4797(02)00183-4.

89. Esseen P, Ehnström B, Ericson L, Sjöberg K (1992) Boreal forests - the focal habitats of Fennoscandia. In: Hansson L, editor. Ecological Principles of Nature Conservation. London: Elsevier. pp. 252–325.

90. Hanski I (2000) Extinction debt and species credit in boreal forests: modelling the consequences of different approaches to biodiversity conservation. Ann Zool Fennici 37: 271–280.

91. Martikainen P, Siitonen J, Punttila P, Kaila L, Rauh J (2000) Species richness of Coleoptera in mature managed and old-growth boreal forests in southern Finland. Biol Conserv 94: 199–209.

92. Moen J, Rist L, Bishop K, Chapin III FS, Ellison D, et al. (2014) Eye on the Taiga: Removing global policy impediments to safeguard the boreal forest. Conserv Lett. doi:10.1111/conl.12098.

93. Andrew ME, Wulder MA, Coops NC (2012) Identification of de facto protected areas in boreal Canada. Biol Conserv 146: 97–107. doi:10.1016/j.biocon.2011.11.029.

# Tables

**Table 1** The spatial datasets used in the study. Column Use indicates whether the dataset is used as input for the analyses or in validation.

| **Dataset** | **Type** | **Use** | **Coverage (%)** | **Source\*** | **Availability\*\*** |
| --- | --- | --- | --- | --- | --- |
| Multi-source National Forest Inventory Data | Raster | Analysis | 100.00 | FRI | 1 |
| Stand-based forest inventory data | Vector | Analysis | 46.36 | FC, NHS | 2/3 |
| Protected areas | Vector | Validation | 1.87 | FPS | 2 |
| Woodland key-habitats | Vector | Validation | 0.50 | FC | 3 |
| METSO-deals | Vector | Validation | 0.13 | CEDTE | 2 |

\* FRI = Finnish Forest Research Institute, FC = Finnish Forest Centre, NHS = Metsähallitus (Finnish Forest and Park Service) Natural Heritage Services, CEDTE = Centre for Economic Development, Transport and the Environment

\*\* 1 = freely available, 2 = available for research purposes upon request under lax conditions, 3 = available for research purposes upon request under strict conditions.

# Figures legends

**Figure 1.** Location of Southern Savonia in Finland and Northern Europe. Map in ETRS89 / ETRS-LAEA coordinate reference system.

**Figure 2.** Schematic of the flow of analysis. The figure is divided into three sections. The first describes how the index rasters (see 2.5 for description) were combined to produce the input datasets used in the prioritization analysis. The second summarizes the main characteristics of spatial conservation prioritization analysis done using Zonation. The third section summarizes how the priority rank maps produced by the six different Zonation runs were analyzed.

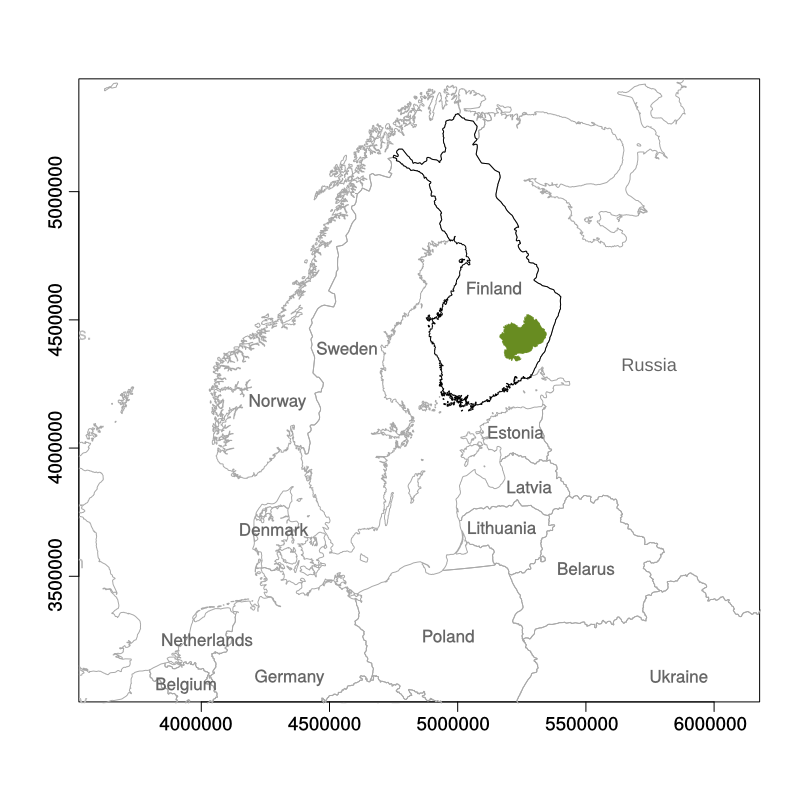
**Figure 3.** Priority rank maps for runs R1 (A), R2 (B), R3 (C), R4 (D), R5 (E), and R6 (F) (Figure2). The marginal plots on top and on the left side of each panel show the count of cells in the top 10% of the landscape (with highest priority) along both latitudinal and longitudinal gradients. (Note different scales on the y-axes of marginal plots.) The insets expand the priority pattern from a selected smaller area.

**Figure 4.** Spatial overlap of 10% intervals of priority classes between selected pairs of analyses, measured by the Jaccard coefficient. An overlap of 1.0 indicates complete match, whereas overlap of 0.0 means absence of overlap. Panels A-F show comparisons between runs based on different input data sets. Panels G-I show comparisons between analyses that used the same input data, but with and without connectivity. Note that the scale is different for panels A-F and G-I.

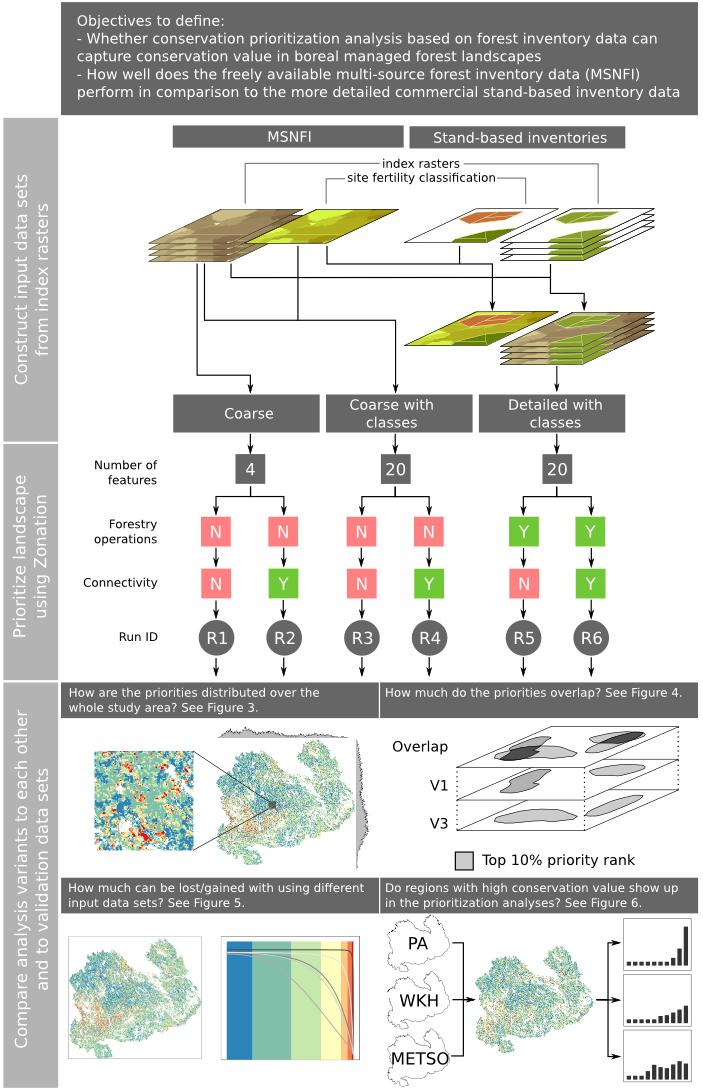
**Figure 5.** The performance of solutions based on coarser data measured by their ability to cover features in the detailed data. The performance curves show for each site fertility class the mean occurrence levels of biodiversity features in the detailed data. The solid curves are for R5, which uses detailed data. The dotted (R1) and dashed (R3) represent coarse data solutions, and show how much representation of the detailed – and presumably more accurate – data would be lost if the prioritization was based on coarser data. The same comparison between R2, R4, and R6 produced very similar results (not shown).

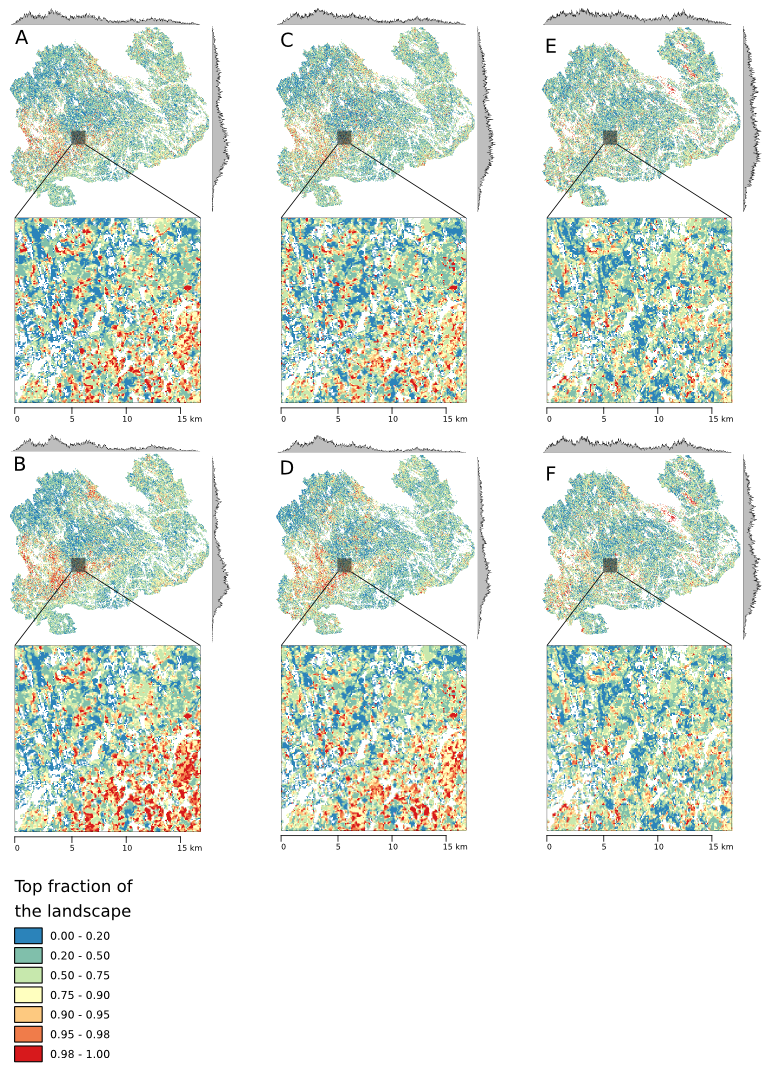
**Figure 6.** Priority rank maps evaluated against independent spatial validation data (Table 1). The first row corresponds to protected areas, the second to woodland key habitats, and the third to made METSO-deals. The columns in each panel show the difference between variants with (left) and without connectivity (right). All spatial validation data should on average have higher conservation value than the surrounding forests, which mostly have a history of economically motivated management. Red vertical line corresponds to the median value.

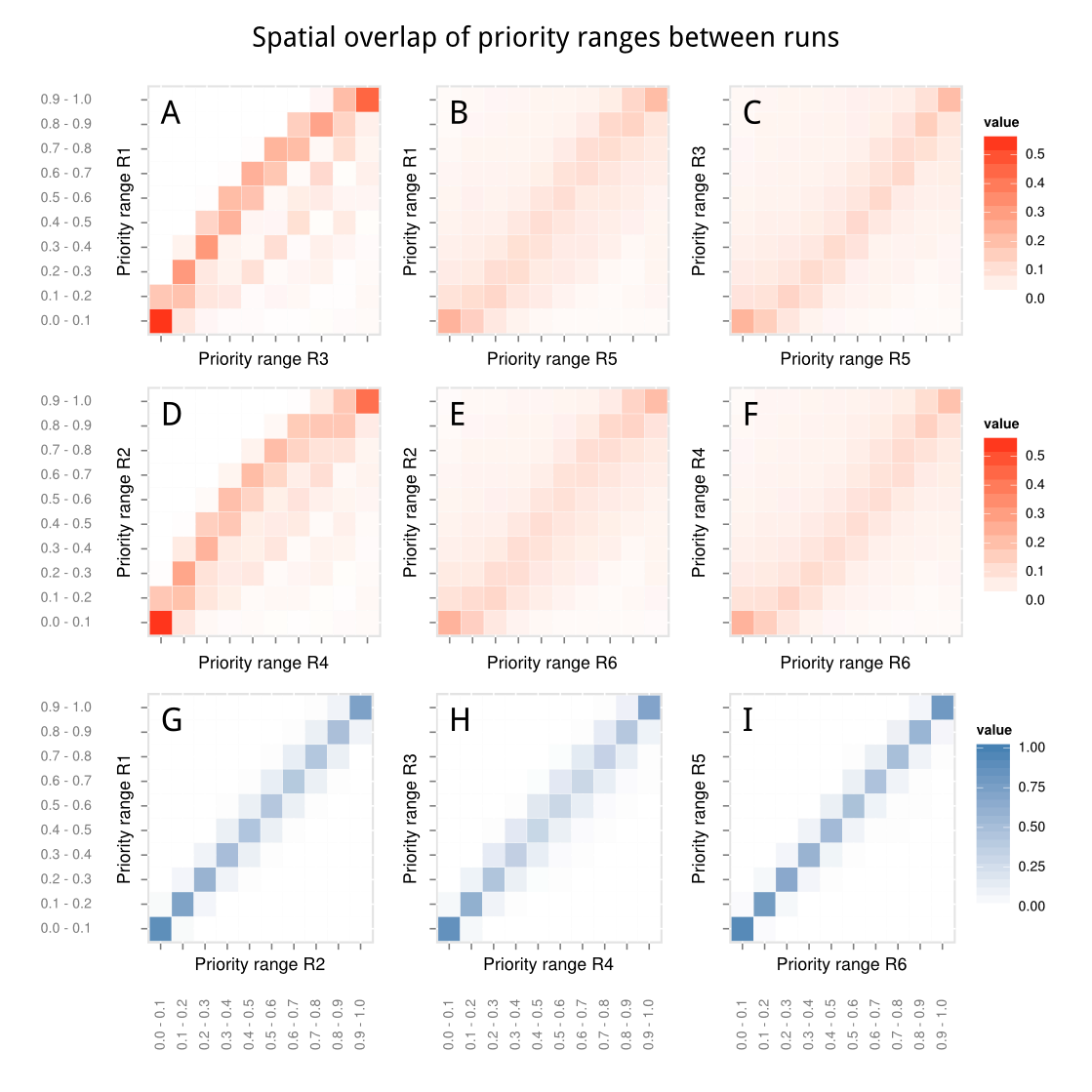
# Figures



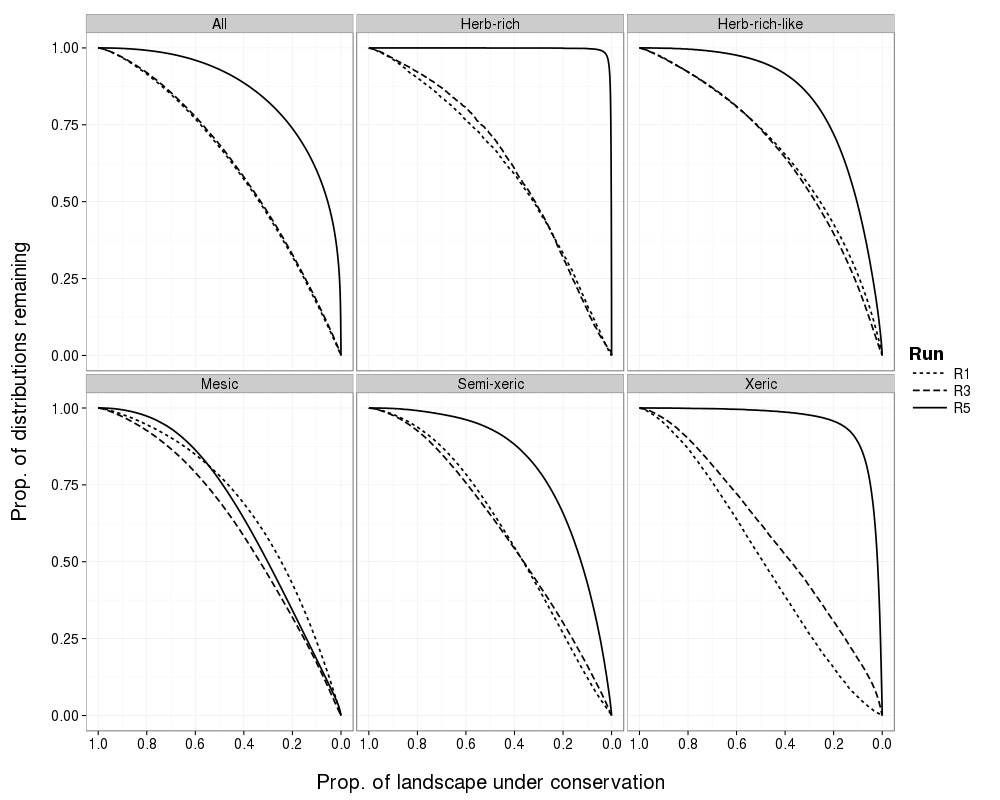
**Figure 1**

 **Figure 2**

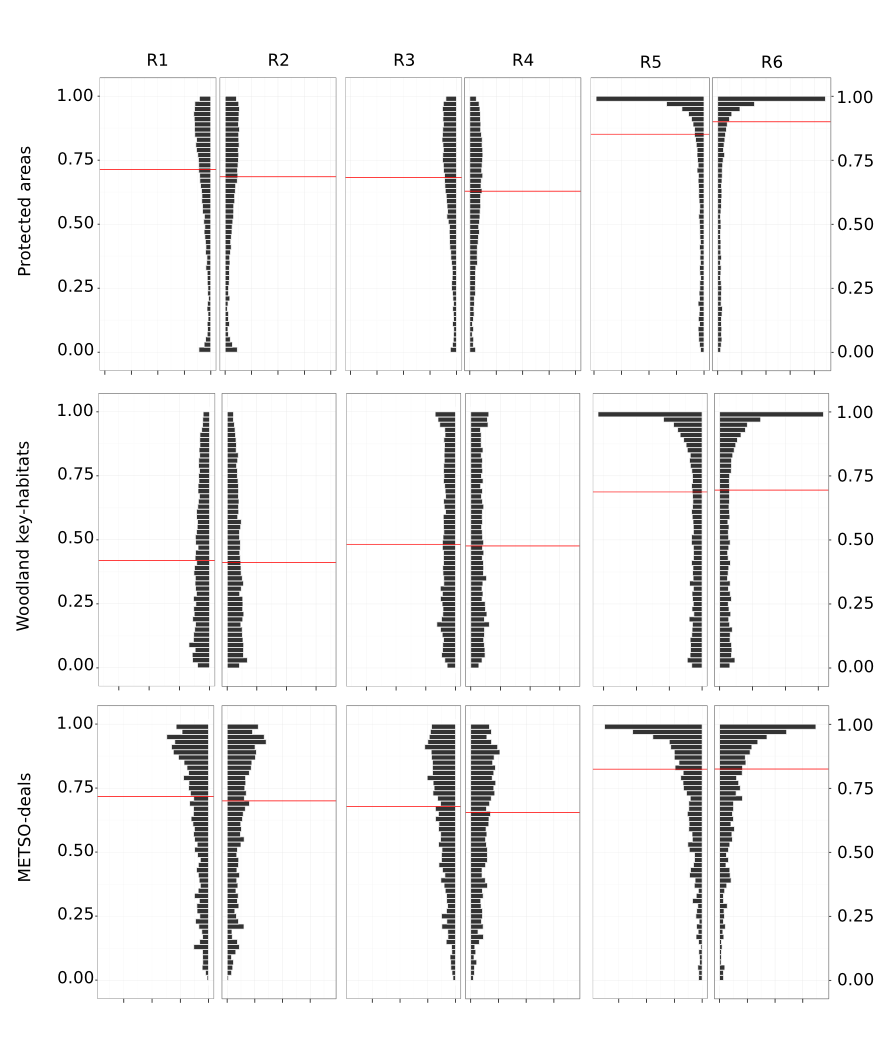
**Figure 3**



**Figure 4**



**Figure 5**



**Figure 6**

# Supplementary material

## S1.1 Data, results, and implementation of data pre-processing and Zonation analysis

The following online repositories contain the data, such results that can be shared, and the implementation of various steps of the study in the form of source code or configuration files. Each repository has its own licensing information.

**Data is still missing, should be deposited in Dryad etc.**

1. https://github.com/jlehtoma/validityms - The sources of this manuscript including text, figures and all code needed to analyze the results.
2. https://github.com/jlehtoma/zsetup-esmk/ - Zonation configuration files.
3. https://github.com/cbig/zonator - zonator: Utilities for the Zonation spatial conservation prioritization software.

## S1.2 Expert elicitation

Figure S1: The benefit functions used to scale the perceived, expert opinion based conservation value (y-axis) to structural characteristics of the forest (x-axis). These functions are specific to tree species.

* Weights and connectivity coefficients in the connectivity matrix were developed for the more detailed classification (count of features 20). Weights and connectivity coefficients were averaged for R1 and R2, which only have 4 features (CHECK).

## S1.3 Data pre-processing

### 1.3.1 Segmentation

The segmentation was carried out using a modified implementation of the “segmentation with directed trees” algorithm by Nagendra & Goldberg . The “segmentation with directed trees” aims at detecting regions without using absolute thresholds (Nagendra & Goldberg 1980). The method is based on using an edge image. The algorithm starts by dividing an image to edge and plateau pixels by computing an edge gradient value for each image pixel based on its 3x3 pixel neighborhood. If the edge gradient value was higher than a user-defined gradient threshold value (i.e. indicating high local variation), the pixel was an edge pixel, otherwise a plateau pixel. In the following phases, 1.) the edge pixels linked to the direction of maximum positive edge gradient, and 2.) the plateau pixels are linked to the pixels of the same plateau. A more detailed description of the segmentation algorithm can be found in Pekkarinen (2002).

Thus, the automatic segmentation process is guided by the local heterogeneity of the input data pixels. In the case when larger segments than produced by initial segmentation are required, separate region merging algorithm can be applied. The region-merging algorithm is based on t-ratio threshold (Hagner 1990) and it is guided by parameters such as the desired minimum size of the final segments and the similarity or dissimilarity of the segments (measured by t-ratio and a user-defined threshold value). Here the size of the output segments was in the range of desired segment size, approx. 1-2 ha, and separate region merging phased was not required.

The stand level variables variables were calculated as average values of the individual pixels within each segment. The variables per tree species were calculated by weighting the pixel level variables by the volumes of individual tree species.

## S1.4 Zonation analysis setup:

* Table S1: Input data features and their weights
  + Include categorical variable for one of [“coarse”, “coarse with classes”, “Detailed”]
* Table S2: The connectivity matrix that specifies the degree to which different forest types assist each other's connectivity.

## Supplementary references

Narendra, P. M., and M. Goldberg. 1980. Image Segmentation with Directed Trees. IEEE Transactions on Pattern Analysis and Machine Interligence **1**:185–191.

Hagner, O. 1990. Computer aided forest stand delineation and inventory based on satellite remote sensing. In: Proceedings of SNS/IUFRO workshop in Umeå 26–28.12.1990: The usability of remote sensing for forest inventory and planning. Swedish University of Agriculture Sciences. Remote Sensing Laboratory. Umeå.

Pekkarinen, A. 2002. Image segment-based spectral features in the estimation of timber volume. Remote Sensing of Environment **82**:349–359.