Title:

Is forest inventory data informative for conservation prioritization?

1. Intro

1.1 Data availability and available data

- Reliable and extensive data on species occurrence is rarely available.
- In most Fennoscandian countries (Finland included) there is a wealth of forest inventory data available and the data is relatively accurate
- Detailed inventory data is often in restricted use, but more and more publicly funded data are becoming openly available (e.g. Metla's MSNFI data).
  - open data policies have a huge potential for conservation planning
- Using forest inventory data as a basis for conservation prioritization effectively surrogate use for the primary forest biodiversity data (species and habitats occurrence data) → how well do inventory data work?
- Setting the reference for surrogacy is complicated, but we have a fairly good understanding what is valuable in boreal forests from conservation point of view

1.2 Spatial conservation prioritization

- Conceptual and methodological ways of defining priorities for given conservation actions (conservation, restoration etc).
- Importantly, conservation of *highest* priorities is only one aspect; equally important is where the *lowest* conservation priorities are → these are areas for alternative land use.
- Forest management has a long history of multi-objective planning so detailed conservation prioritization should fit into existing customs well.
- Priorities given by scientific analyses often not verified but rather taken at face value → multitude of available forest inventory data also opens up possibilities to assess and validate prioritization results.

1.3 Conservation planning in boreal forests

- Major ecoregion with large geographical extent and important role for example in carbon storage and sequestration.
- While not the most species-rich or threatened ecoregion on the planet, there is still good opportunity to protect/manage large continuous areas of forest
- Much conservation research has been done in boreal forest context on various fields → most impact from implementation point of view on work done in general forest management planning
- Regional conservation programmes (e.g. METSO) often must reconcile varying interests and objectives of forest owners, but again much of the general context is forestry management
- Forestry is mostly driven by market economics and conservation has also been shifting into more voluntary-based models emphasizing the role multi-use and ecosystem services → need for fine-grained spatial plannign

1.4 Aims and scope

- Main questions of the work are:

  1) Can conservation prioritization analysis based on forest inventory data capture conservation value in boreal managed forest landscapes?
  2) How well does freely available multi-source forest inventory data perform compared to more detailed commercial stand-based inventory data?
  3) What are the requirements for data imposed by the conservation objectives (e.g inclusion of connectivity and complementarity)?

- Additional the work presented aims at addressing the following practical issues:

  1) Ways to integrate conservation planning into operative forestry management and planning
  2) Making use of the available data considering the existing ownership restriction
  3) Assessing what kind of (forestry) data would be useful for conservation planning

1.4 Aims and scope

- In this work, we approach the aforementioned questions by a case-study of conservation prioritization in the regional forest center of South-Savo, Finland.
- Work described was done in the context of METSO – the forest biodiversity conservation programme for Southern Finland – with the aim of putting the results into practice.

2. Material and methods

2.1 Data for prioritization and comparison

- Data used must be available across the entire study area.
- Source of the data matters as different data sets have different levels of uncertainty etc.
- Prioritization data sets (more description in Table 1):

  1) Stand-based forest inventory data from South-Savo Forest Center
  2) Stand-based forest inventory data from Metsähallitus Natural Heritage Service
  3) Segmented multisource National Forest Inventory from Finnish Forest Research Institute

Availability (could be incorporated into Table1):

1 – Available for research purposes upon request, strict conditions

2 – Available for research purposes upon request, loose conditions

3 – Freely and openly available for anyone (still have to figure out whether this applies to the segmented version)

Table 1. Forest inventory data sets used for conservation prioritization.

Notes:

- NOT EVEN STARTED

Dataset	Data type	Production method	Variables	Source
Forest inventory data	Vector	Inventory	…	MK
Forest inventory data	Vector	Inventory (auto-increment)	…	MHLP
Multi-source National Forest Inventory Data	Raster	kNN, remote-sensed data with field plot calibration	…	

2.1 Data for prioritization and comparison

Table 2. Spatial validation data sets used.

Notes:

- NOT EVEN STARTED

Dataset	Code	Analysis	Owner
Existing protected areas	PA	DP, RC (?)	MH
Woodland key-habitats	WKH	DP, RC (?)	MK
METSO deals (permanent)	MD	DP, RC (?)	ELY

2.2 The ecological model of conservation value

2.3 Spatial conservation prioritization

- Freely

2.4 Comparing prioritization results A

- Between data sources
- Spatial overlaps between different fractions of different solutions can be used to determine “consistently valuable locations”. When analysis setup changes it is not necessarily expected that the top priorities stay in same spatial locations, but often they will because:
  - Of patterns in the underlying data.
  - More complicated variants are typically built on top of simpler variants and will thus contain many of the same analysis features.

2.4 Comparing prioritization results B

- In relation to spatial validation data

Table 3. Zonation analysis variants used in the study. See Figure 1 for graphical representation of the analysis setup. ID and variant name are used in the text and in other figures to refer to a specific variant. Weight indicates that features used in the variant have been weighted according to expert opinion (see table SX for the exact weights). Condition indicates whether to features used in the analysis have been modified using information on the forestry operation planned since the data were collected. Internal connectivity refers to short-range connectivity between different forest types (see table SX for the exact values) implemented as Zonation matrix-connectivity. Interaction connectivity can refer either to connectivity to woodland key habitats (WKH) (variant 3) or protected areas (variant 4).

Notes:

- If these are the variants used then weights and condition are in principle redundant as all variants actually have them.
- Not very good as is right now, needs re-thinking

ID	Variant name	Weights	Condition	Internal connectivity	Interaction connectivity
1	Local quality	x	x		
2	Matrix quality	x	x	x	
3	WKHs	x	x	x	x

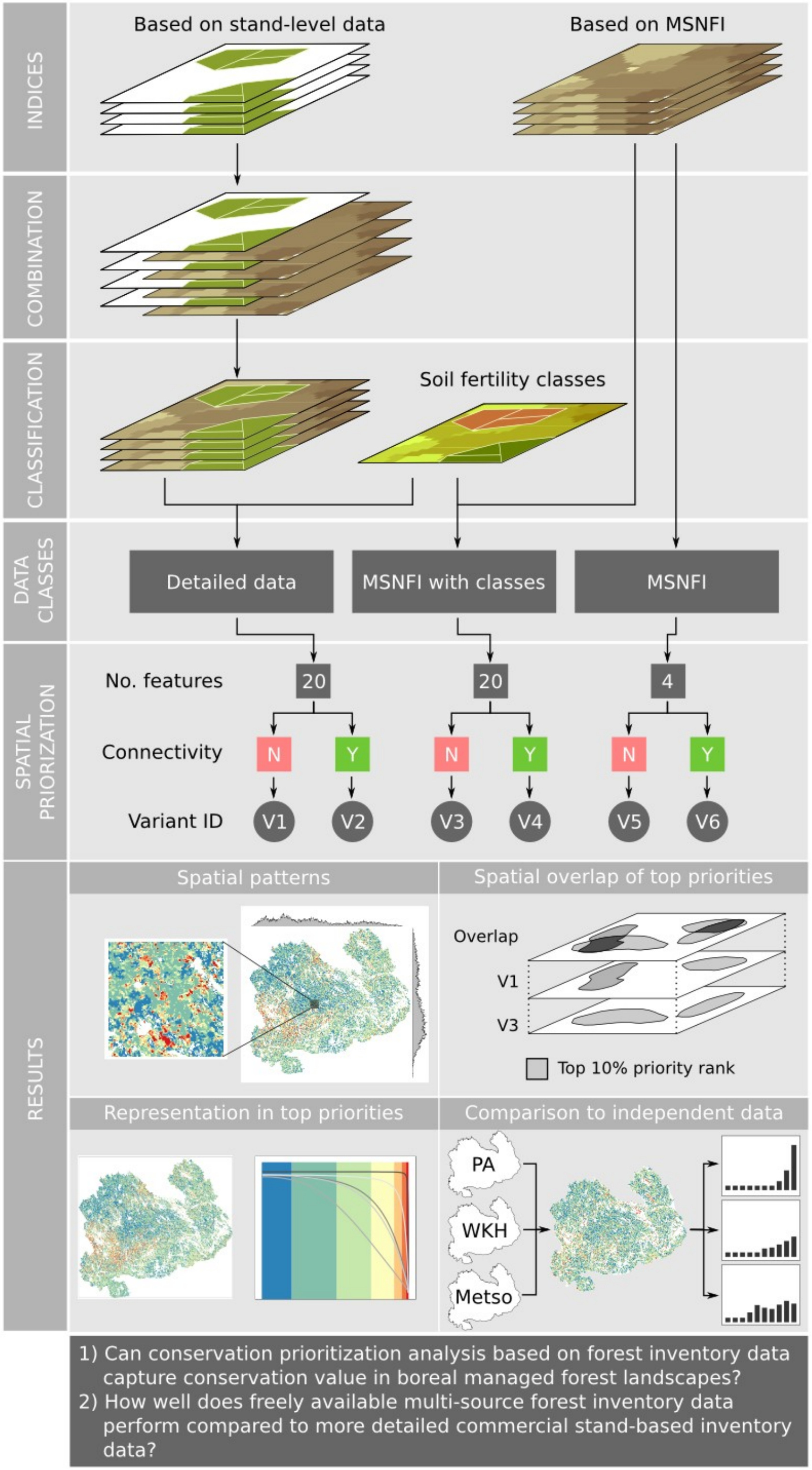


Figure 1. Schematics of the prioritization analysis and how the results are analyzed.

Notes:

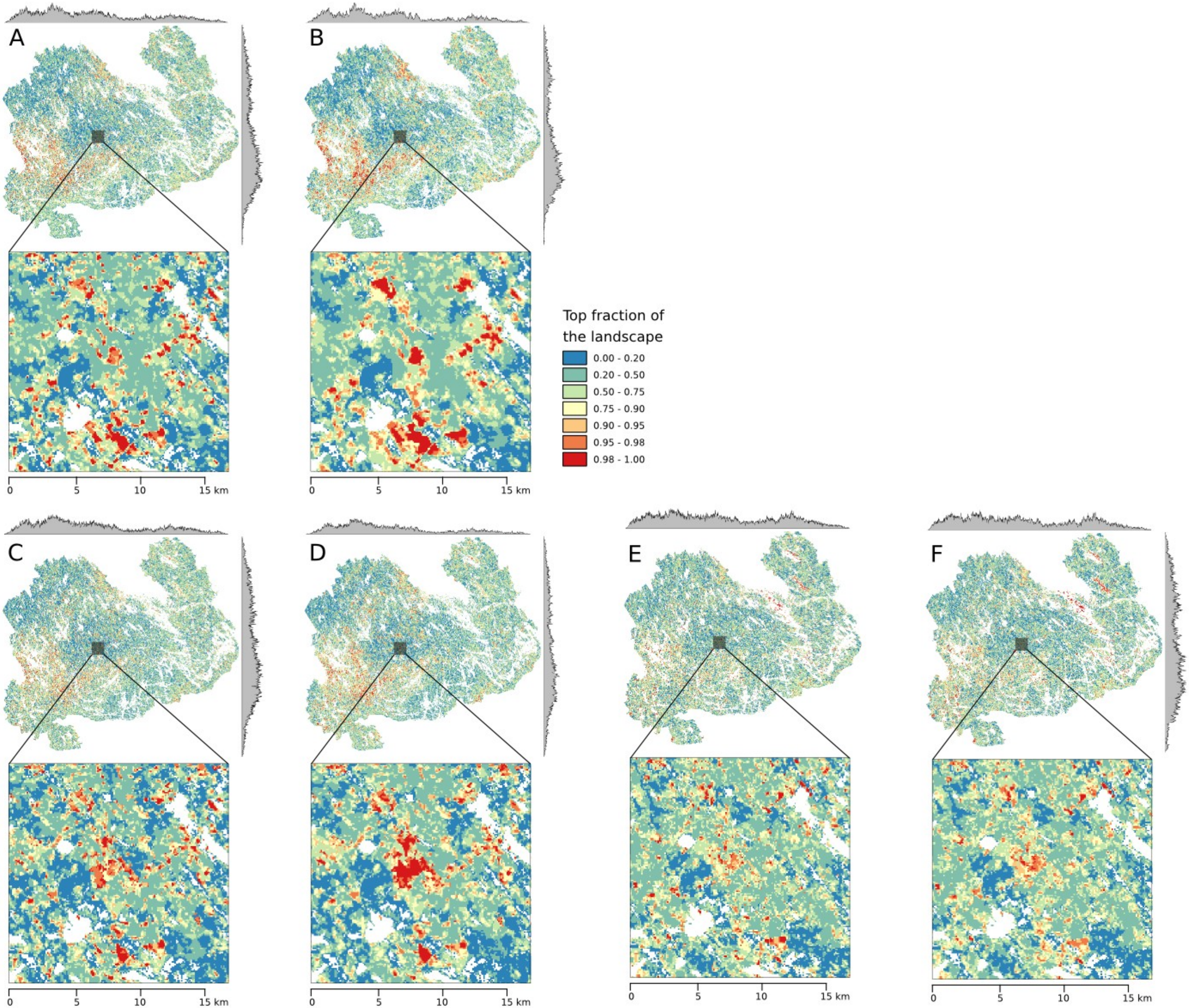
- First 3 sections need to be condensed
- There is nothing in the figure to point to how the indices were constructed. How do deal with this? References to already published papers
- There are now arrows pointing to to results, would it need some?
- Make explicit references to actual figures in results?
- Results has two components 1) differences between different data classes, and 2) differences between different variants. This is not accounted for in the current Figure, should it?

3. Results

3.1 Spatial patterns of rank priorities

- Following descriptions apply only to variant 1.
- MSNFI-only
  - Spatial pattern of top-priorities is more equally spread over the landscape
  - Highest priorities are located in areas with mature forest with relatively large volumes on (coniferous) trees
  - Lowest priorities can be found in ...
- Detailed data
  - Spatial pattern more concentrated to particular locations
  - Highest priorities are located in areas with high amount other deciduous tree species on fertile soils. Also very dry (xeric) soil types come out as high priorities (although this was not intentional).
  - Lowest priorities are found in areas with low overall tree volume (including peatlands) and on mid-fertility soil type classes.
- Difference in patterns

Figure 2. Rank priority maps.


A
B
C
D
E
F
Top fraction of the landscape
0.00 - 0.20
0.20 - 0.50
0.50 - 0.75
0.75 - 0.90
0.90 - 0.95
0.95 - 0.98
0.98 - 1.00
0
5
10
15 km

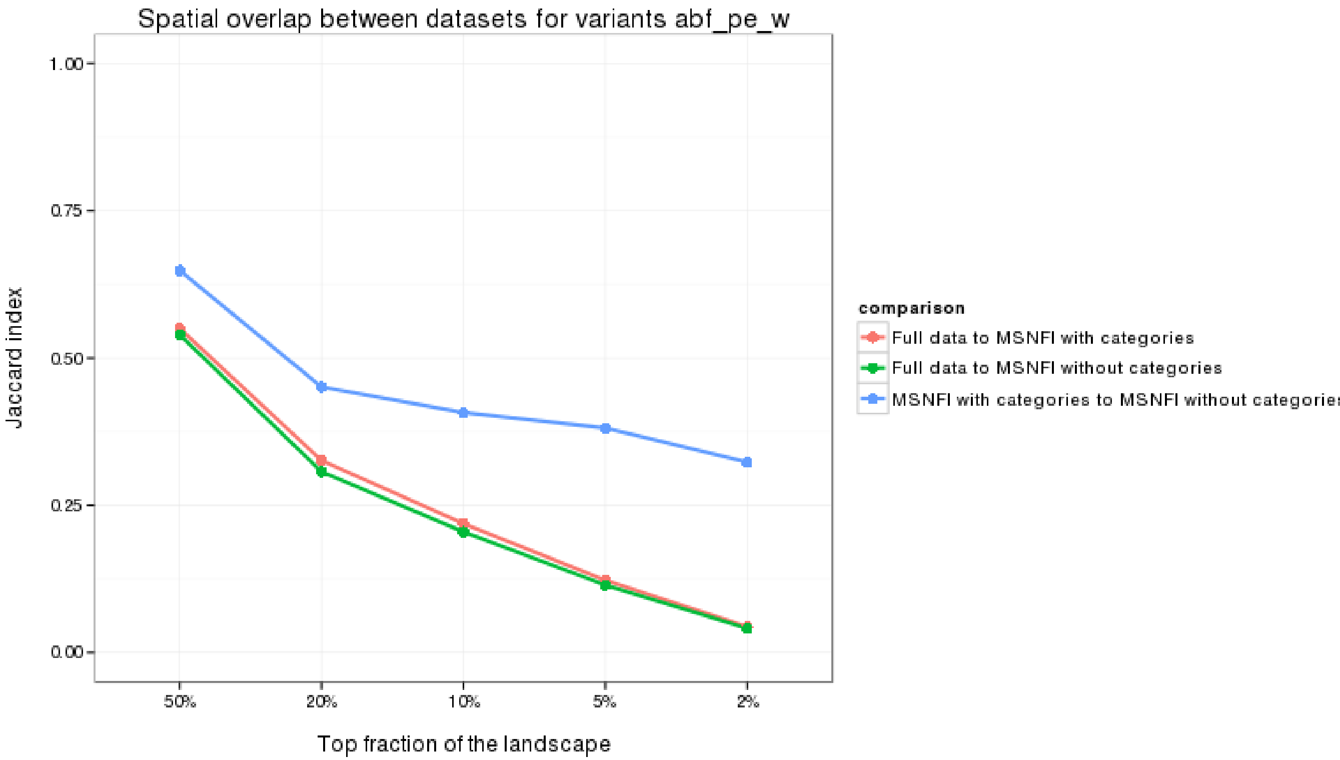
3.2 Spatial overlap of top fractions

- Spatial overlap between the top fractions of the different solutions can be studied along two dimensions: 1) data sources, and 2) variants.
- Overlaps measured by the Jaccard coefficient (1 = the same set, 0 = completely dissimilar set) tend to be greater among analysis based on different data sources (msnfi-only vs. more detailed data) when the top fraction if larger. In other words, the absolute top priorities are in different locations (which is expected as the more detailed data is, well, more detailed).
- Between variants and different data sources, including connectivity (and hence spatial aggregation) increases the spatial overlap of top fractions.

Figure 3. The spatial overlap as defined by Jaccard similarity.

Notes:

- Generally the more compact the solution is (i.e. the more connectivity it has OR the more clumped the data are) the higher the average representation is for a given fraction of the landscape.


Spatial overlap between datasets for variants abf_pe_w
Jaccard index
1.00
0.75
0.50
0.25
0.00
50%
20%
10%
5%
2%
Top fraction of the landscape
comparison
Full data to MSNFI with categories
Full data to MSNFI without categories
MSNFI with categories to MSNFI without categories

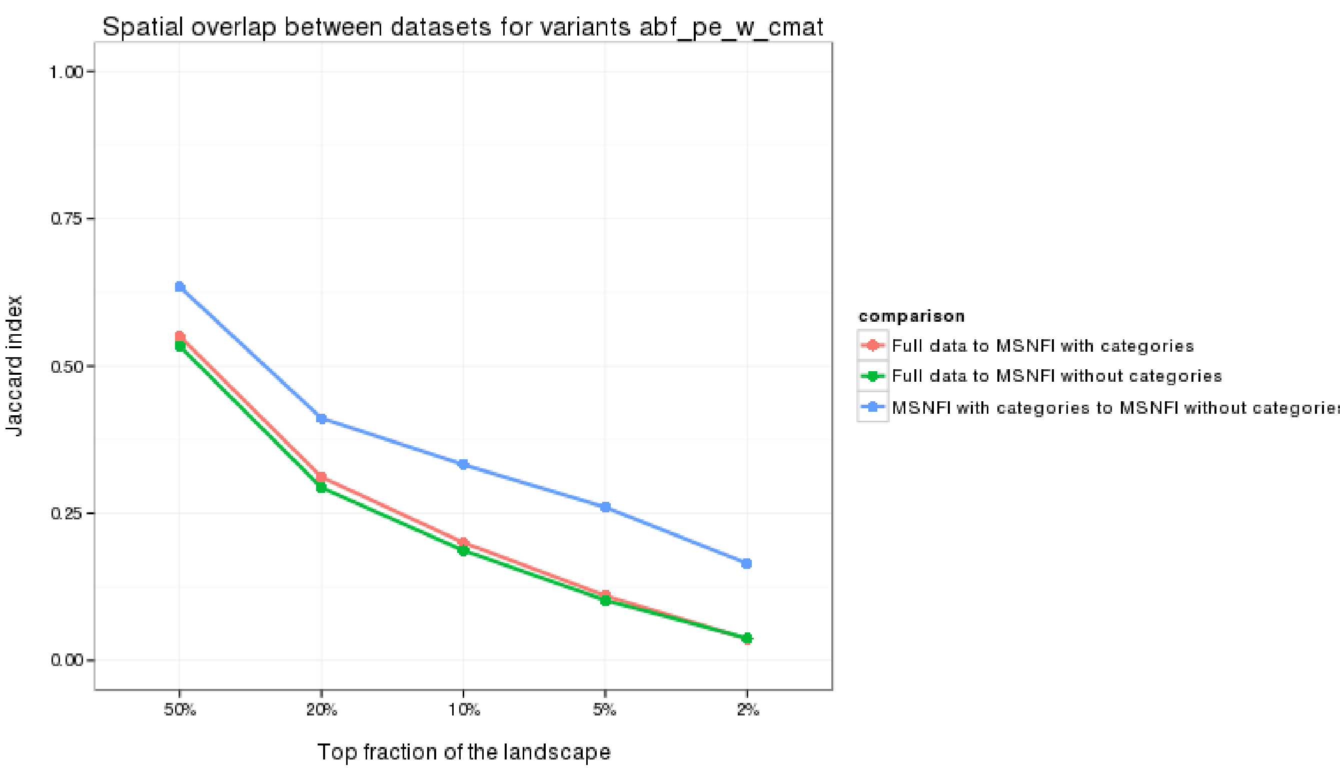
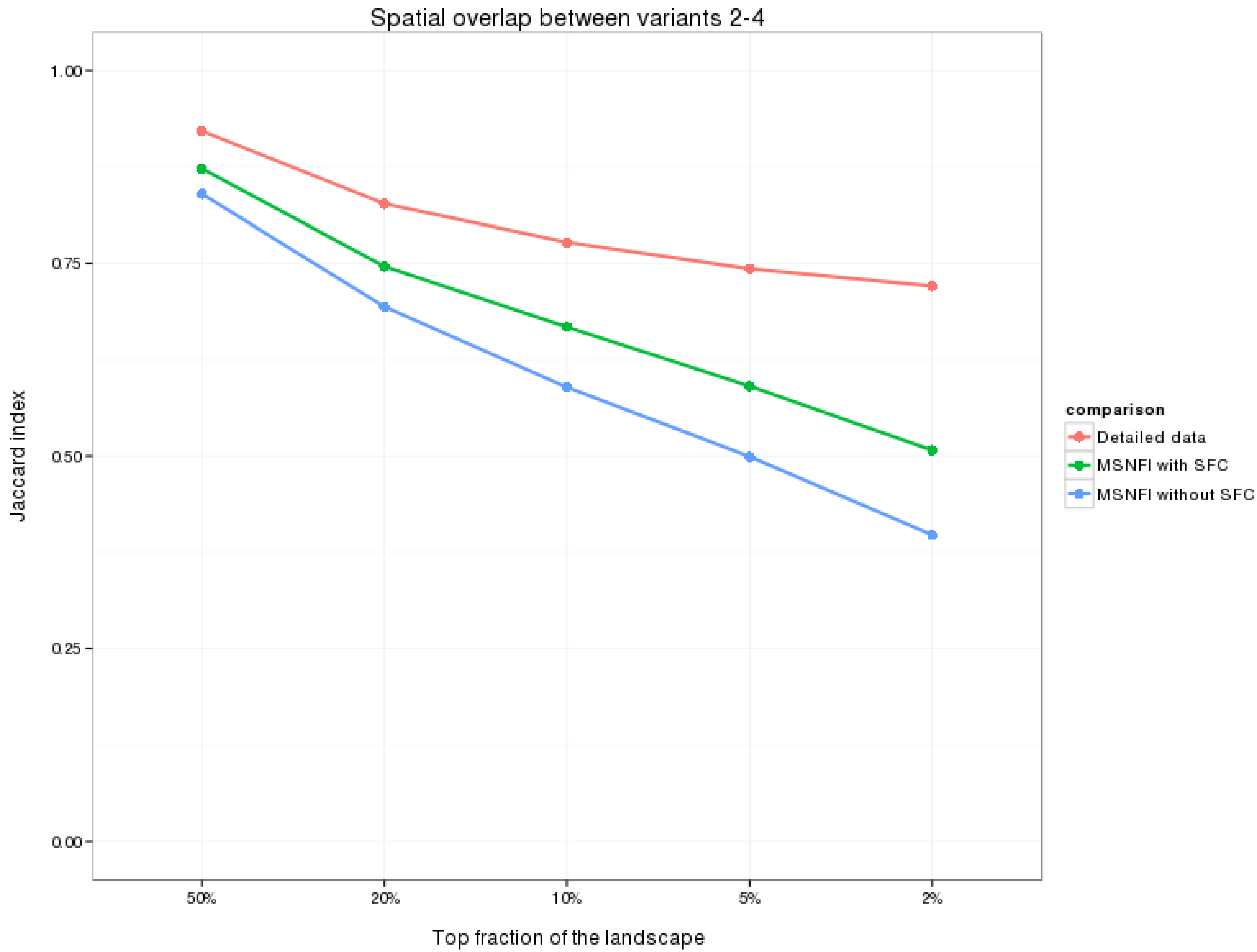

Spatial overlap between datasets for variants abf_pe_w_cmat
Jaccard index
1.00
0.75
0.50
0.25
0.00
50%
20%
10%
5%
2%
Top fraction of the landscape
comparison
Full data to MSNFI with categories
Full data to MSNFI without categories
MSNFI with categories to MSNFI without categories

Figure 3 EXTRA. The spatial overlap of top fractions between the variants (without and with connectivity).

Notes:

-

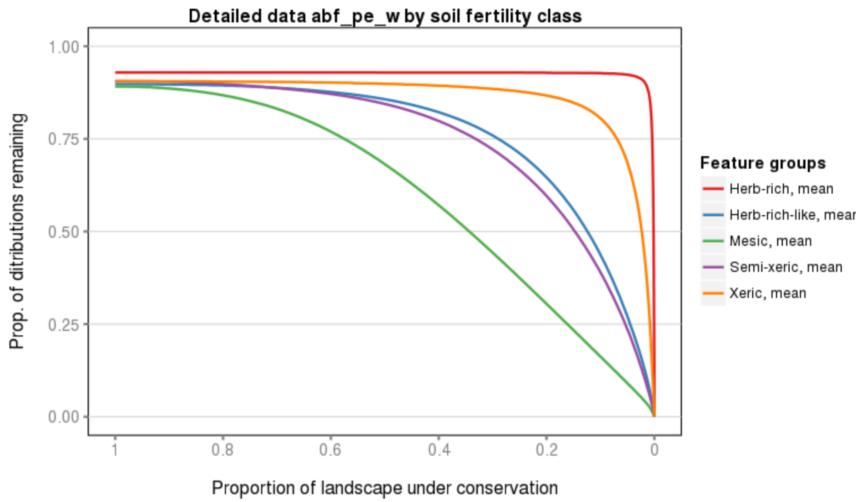


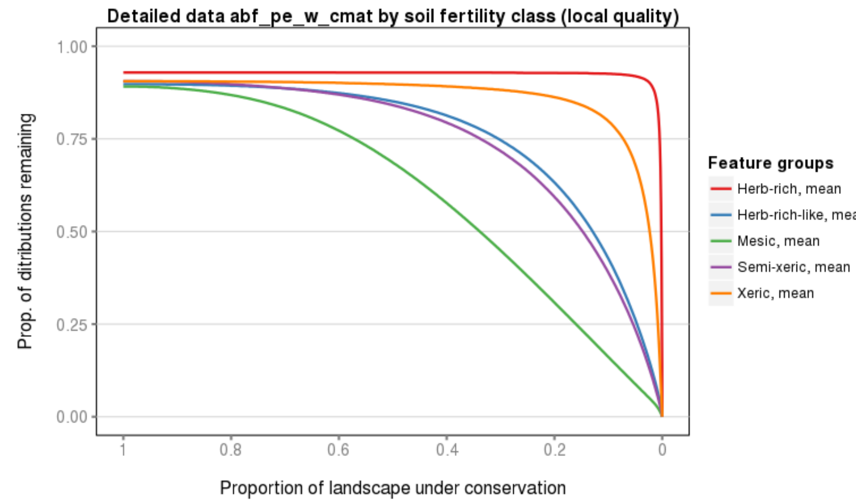
Spatial overlap between variants 2-4
1.00
0.75
0.50
0.25
0.00
Jaccard index
50%
20%
10%
5%
2%
Top fraction of the landscape
comparison
Detailed data
MSNFI with SFC
MSNFI without SFC

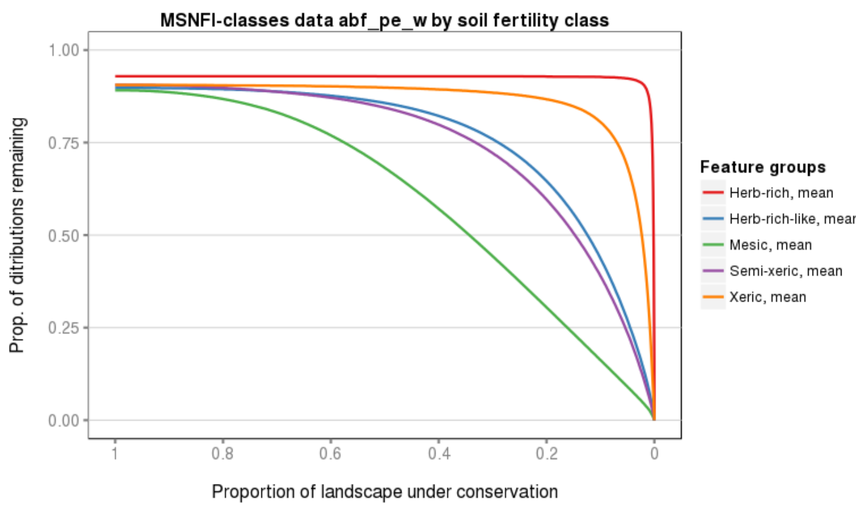
3.3 Feature representation

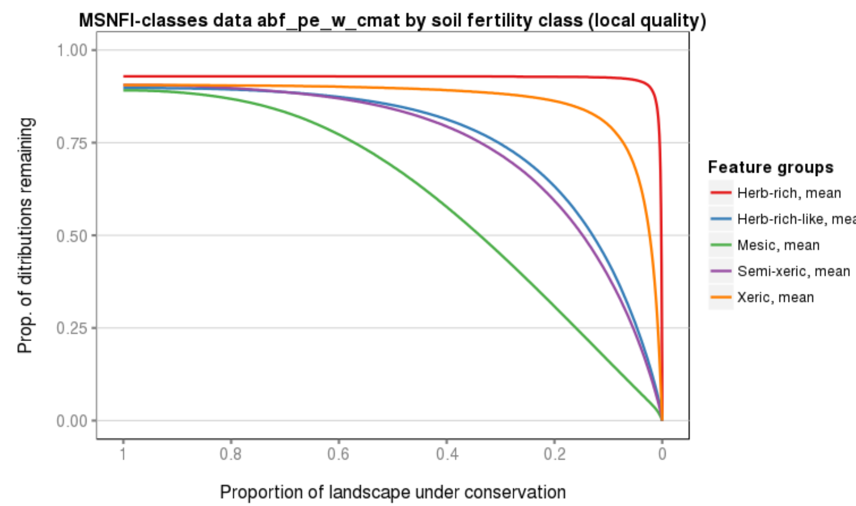
- Within data sources, variants with connectivity have higher on average feature representation, but this is largely due to the fact that many of the features *are* connectivity transformed duplicates of the original feature stack.
- However, looking at the representation level of the non-transformed features ("local quality") stack we see a slight drop in overall performance. The drop is not very large as the transformed and non-transformed features are highly correlated. (THIS HAS NOT BEEN IMPLEMENTED YET)
- msnfi-only results tend to have slightly lower representation in selected top fractions as the data varies more smoothly across the landscape.

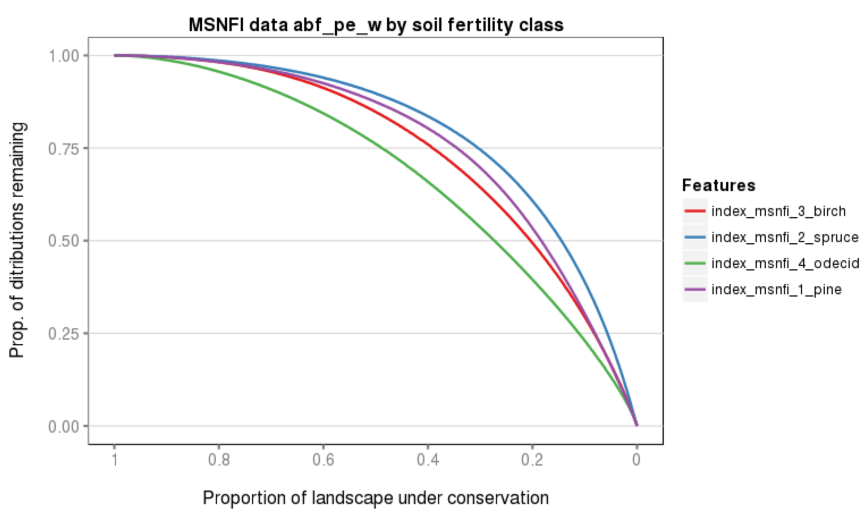
Figure 4A. Feature group representation levels (group by soil fertility class where applicable) for all variants.


Detailed data abf_pe_w by soil fertility class
Prop. of ditributions remaining
1.00
0.75
0.50
0.25
0.00
1
0.8
0.6
0.4
0.2
0
Proportion of landscape under conservation
Feature groups
Herb-rich, mean
Herb-rich-like, mean
Mesic, mean
Semi-xeric, mean
Xeric, mean


Detailed data abf_pe_w_cmat by soil fertility class (local quality)
Prop. of ditributions remaining
1.00
0.75
0.50
0.25
0.00
1
0.8
0.6
0.4
0.2
0
Proportion of landscape under conservation
Feature groups
Herb-rich, mean
Herb-rich-like, mean
Mesic, mean
Semi-xeric, mean
Xeric, mean


MSNFI-classes data abf_pe_w by soil fertility class
Prop. of ditributions remaining
1.00
0.75
0.50
0.25
0.00
1
0.8
0.6
0.4
0.2
0
Proportion of landscape under conservation
Feature groups
Herb-rich, mean
Herb-rich-like, mean
Mesic, mean
Semi-xeric, mean
Xeric, mean


MSNFI-classes data abf_pe_w_cmat by soil fertility class (local quality)
Prop. of ditributions remaining
1.00
0.75
0.50
0.25
0.00
1
0.8
0.6
0.4
0.2
0
Proportion of landscape under conservation
Feature groups
Herb-rich, mean
Herb-rich-like, mean
Mesic, mean
Semi-xeric, mean
Xeric, mean


MSNFI data abf_pe_w by soil fertility class
Prop. of ditributions remaining
1.00
0.75
0.50
0.25
0.00
1
0.8
0.6
0.4
0.2
0
Proportion of landscape under conservation
Features
index_msnfi_3_birch
index_msnfi_2_spruce
index_msnfi_4_odecid
index_msnfi_1_pine

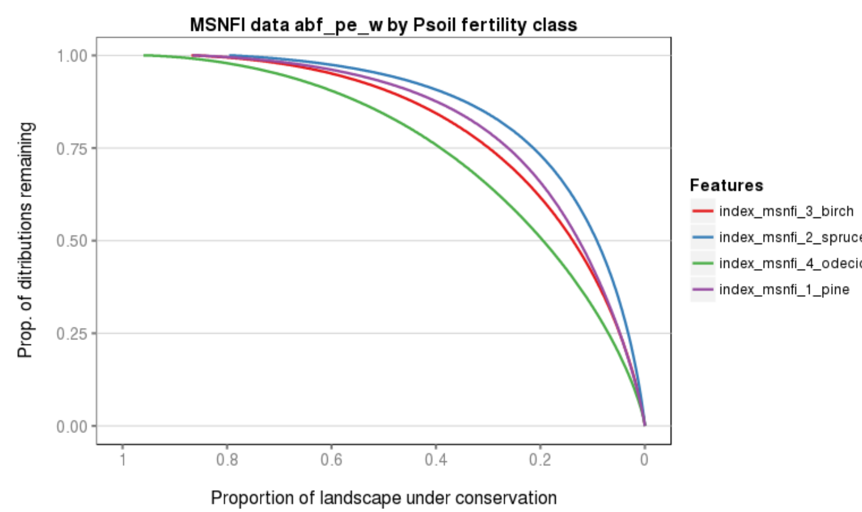
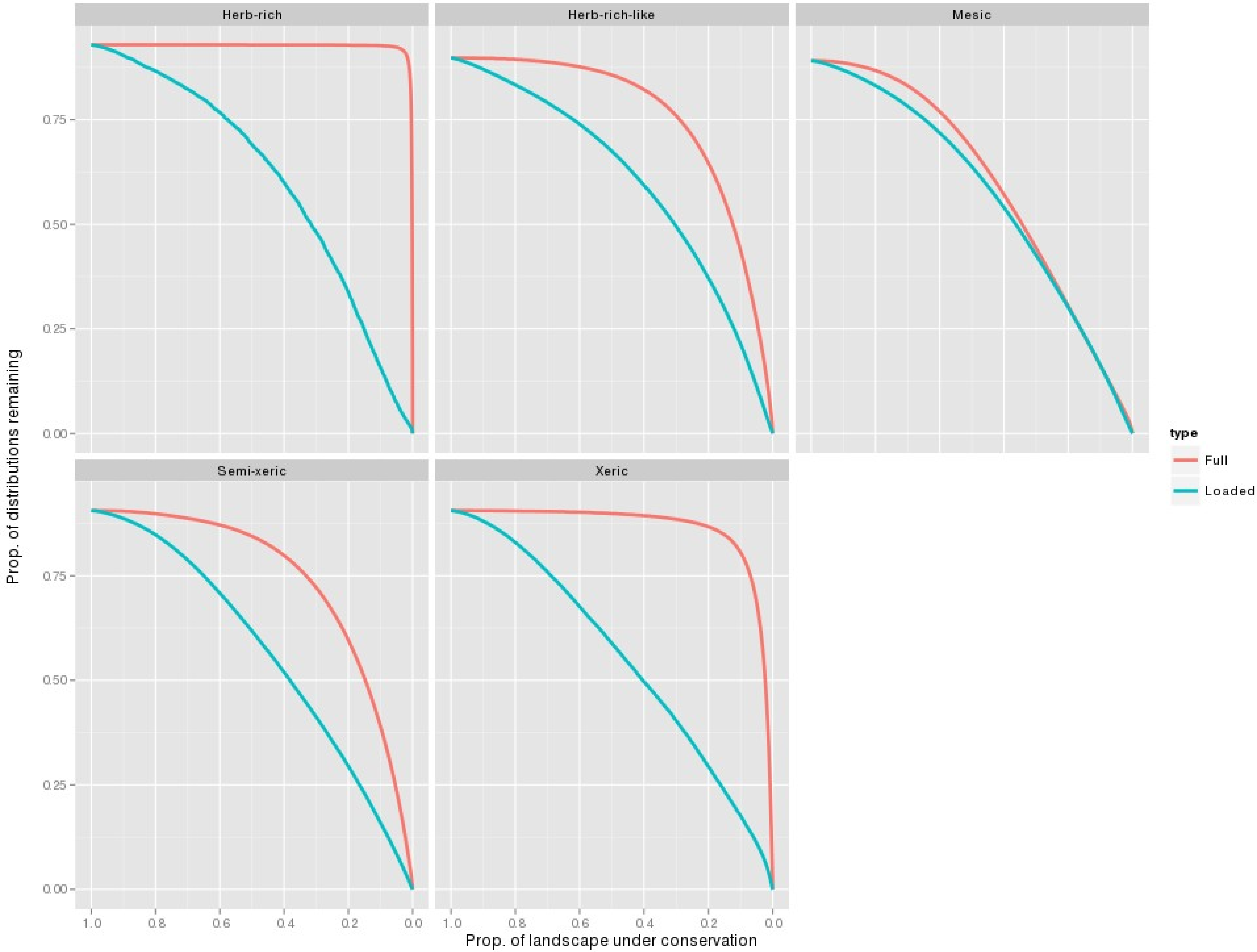

MSNFI data abf_pe_w by Psoil fertility class
Prop. of ditributions remaining
1.00
0.75
0.50
0.25
0.00
1
0.8
0.6
0.4
0.2
0
Proportion of landscape under conservation
Features
index_msnfi_3_birch
index_msnfi_2_spruce
index_msnfi_4_odecid
index_msnfi_1_pine

Figure 4B. How much feature representation is lost if the coarser MSNFI with classes is used as a basis for ranking, but the feature data used comes from the detailed data?

Notes:

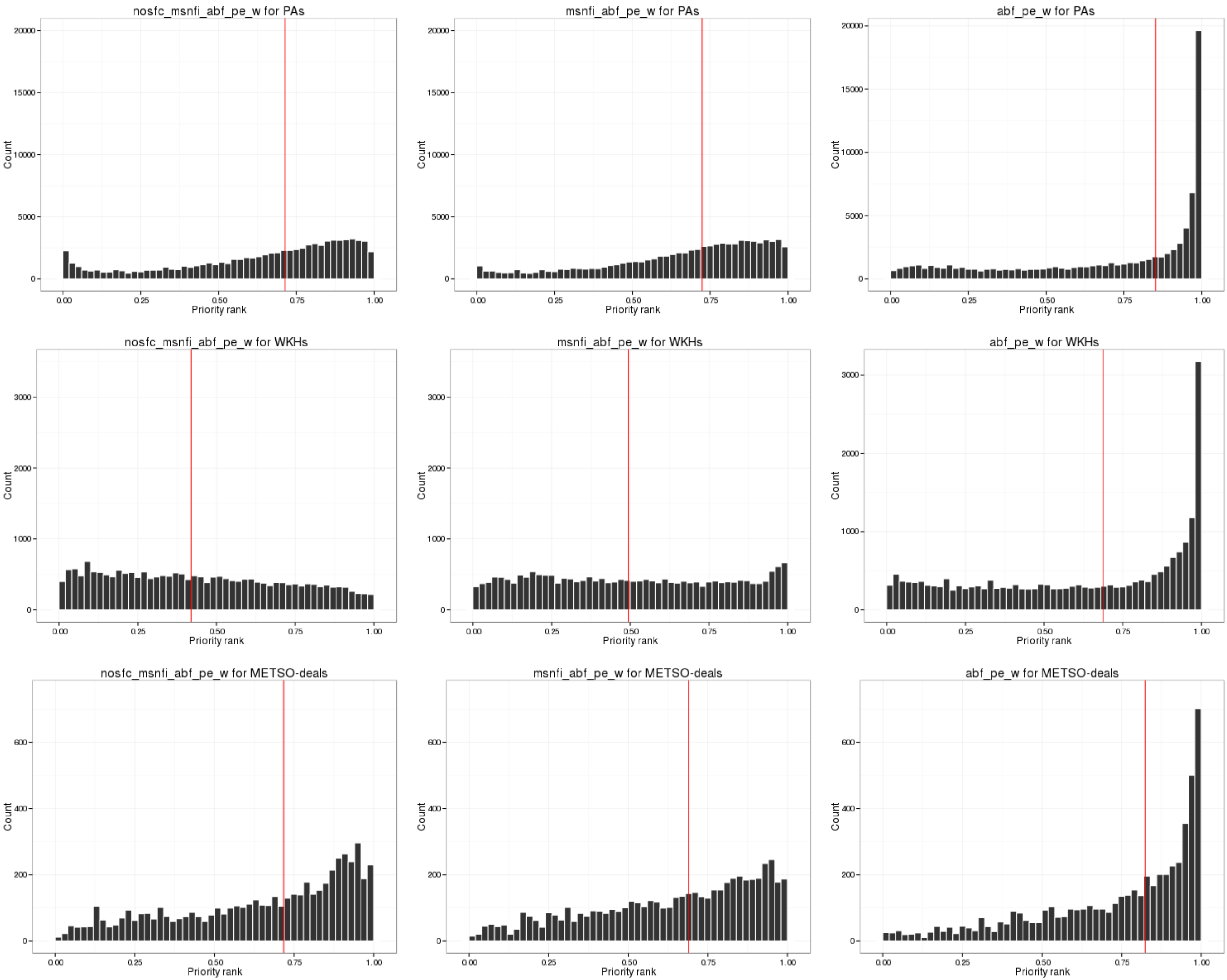
-


Herb-rich
Herb-rich-like
Mesic
Semi-xeric
Xeric
0.00
0.25
0.50
0.75
1.0
0.8
0.6
0.4
0.2
0.0
Prop. of distributions remaining
Prop. of landscape under conservation
type
Full
Loaded

3.4 Comparison to spatial validation data

- Between the different variants for the detailed data source, all areas contained in the spatial validation data sets come out with larger than average priority with a priority rank distribute strongly skewed to the left.
  - Especially protected areas come out consistently with a high average priority
  - When considering only the local quality, also the done METSO-deals have a high mean priority.
  - Woodland-key habitats come out with high priority only after connectivity to them is included (this is circular, cannot be used).
- Between the different variants for the msnfi-only data source the outcome is different
  - Protected areas still come out with higher than average priority, but skewness of the histogram of rank priorities is not as pronounced.
  - Woodland key habitats have an average priority that is not different from averages.
  - Done METSO-deals also have higher than average priority.

Figure 5. The average priority rank and the distribution of rank priorities of the landscape within the independent spatial validation data. First column corresponds to protected areas, second to woodland key habitats, and the third one to made METSO-deals. Each of the spatial validation data is assumed to have on average higher conservation value than the surrounding managed forest.



nosfc_msnfi_abf_pe_w for PAs
msnfi_abf_pe_w for PAs
abf_pe_w for PAs
nosfc_msnfi_abf_pe_w for WKHs
msnfi_abf_pe_w for WKHs
abf_pe_w for WKHs
nosfc_msnfi_abf_pe_w for METSO-deals
msnfi_abf_pe_w for METSO-deals
abf_pe_w for METSO-deals
Count
Priority rank
0.00
0.25
0.50
0.75
1.00
0
5000
10000
15000
20000
1000
2000
3000
200
400
600

4. Discussion

Main messages

- Freely available national forest inventory data can be used to inform practitioners and decision-makers on the spatial location on conservation priorities, but only when the scale is coarse and when the qualitative characteristics of the ecological model emphasize certain things (i.e. volume) → more suitable when reference is PAs
- More detailed stand-based forest inventory data will almost always outperform the coarser MSNFI-only data and is a requirement if planning requires small scale and detailed habitat characteristics information.
- For large-scale planning (and depending on the ecological model used) MSNFI might be suitable.
- Increasing the complexity of the analysis and including several important ecological factors such as connectivity might introduce trade-offs in which case there is no single best solution, but rather a suite of solutions.
- Current forest planning operation mostly consider the occurrence of biodiversity in small, contained areas →
- Primary objective is to increase the area (even modest increases can lead to large gains if placed right) [how to show this with current work?]

Caveats

- Known issues with MSNFI
  - May be imprecise on the scale of single pixels
  - Averages distributions
  - Does not pick up e.g. other deciduous tree species
  - May systematically categorize certain soil fertility classes wrongly
- Ecological model is not "fitted" with real ecological data, thus it only conveys information on our subjective views attached to conservation value.

Future aspects

- Other forestry data available that we didn't use (e.g. costs, ownership information)
- Habitat suitability indices (Tikkanen, Kouki et al.) would form a stronger foundation for the ecological model, but also require a more strict definition of which species we're talking about (which is not necessarily bad)
- Simulating forest dynamics fairly well understood (e.g. Mönkkönen et al. 2014) and could be included in the ecological model with moderate amount of work

Supplementary material

Figure SX. Graphical representation of the feature setup in the Zonation analysis.

- Weights
- Aggregate weights for combinations
- Feature duplication
- Matrix connectivity
- Interaction connectivity

Issues

1) Utility of different data and validation of the prioritization results need to be welded together firmly → cannot come out us too separate issues (this outline still doesn't have a very strong emphasis on data issues)

2) Should the validation results actually be statistically tested (sampling from distribution)