Spatially Based Approach of the Fennoscandian Shield and the Borealic Uplands (Illies ecoregion 20 and 22)

Ulrika Beier and Erik Degerman
National Board of Fisheries
Institute of Freshwater Research
SE-178 93 Drottningholm

2004-02-24

Contents

Abstract 3
Introduction 4
1. Fish based typology 5
1.1 Data selection 5
1.2 Description of the calibration dataset 5
1.3 Methods, data and tests 8
PCA-correlation 9
Decorana 10
Hierarchical clustering 11
1.4 Resulting typology and decision model 12
1.5 Fish-based typology applied to calibration sites 12
2 Screening of metrics 15
2.1 Calibration data for the screening of metrics 15
3 Selection of metrics 17
3.1 Potential metrics 17
3.2 Screening of potential metrics 18
Fish type “Trout dominated” 19
Fish type “Trout with others” 21
Fish type “Lake trout” 22
Fish type “Salmon” 23
Fish type “Sea trout” 24
4 Discussion 25
4.1 Too few sites with high impact 25
4.2 The use of a combined impact variable 26
4.3 Impact assessment improvment 27
4.4 Reclassification of species 27
4.5 Distribution of modelling metrics within calibration sites 28
4.6 The connectivity issue 28
5 References 30

Abstract

It was decided within the FAME project that the spatially based approach for assessing the ecological status of running waters conducted by the Swedish partner should concern Illies' ecoregion 20, "Borealic uplands" and 22, "Fenno-Scandian shield". "Borealic uplands" cover the Swedish mountain range and the whole of Norway, whereas the "Fenno-Scandian shield" covers the lowlands of northern Sweden and the whole of Finland. Out of 623 sites in the national data set, 279 sites were from ecoregions 20 (n=41) and 22 (n=238). Due to the few sites in ecoregion 20 it was decided to produce a joint typology for ecoregions 20 and 22. In ecoregions 20 and 22, impacted sites are few, and especially heavily impacted sites are very rare in the data set.

The dataset was not initially split up into one part used for analyses and one part for validation due to that it could be foreseen that some of the fish types would be represented by few sites in the dataset. The calibration dataset was selected primarily according to the rules suggested at the meeting in Lyon, i.e. human impact should be less than 3 on five critical impact variables: "Connectivity segment", "Hydrological regime site", "Morphological condition site", "Toxic-Acidification site", "Nutrients-organic input site". The variable "Connectivity segment" was after the meeting in Portugal replaced by the variable "Connectivity multiscale". In addition to this, we added criteria for calibration data that "Introduction of fish site" and "Impact of stocking site" must be 1, to avoid including localities with stocked salmon, eel and trout. These criteria for calibration data resulted in 209 sites, which corresponds to 75% of all sites in ecoregion 20 and 22. A single fishing occasion was randomly selected for the fish typology analyses. The general scores for the variables describing human impact was low in the dataset for ecoregions 20 and 22.

Methods used to classify fish types were Principal component analysis (PCA), Detrended correspondance analysis (Decorana) and Hierarchical clustering. Seven fish types were distinguished, described as: 1: Subalpine lake in/outlets -Arctic char, 2: Inland brook - Trout dominated, 3: Inland streams - Trout with others, 4: Lake in/outlets - Lake-trout, 4: Lenthic - Non-salmonid 5: Large coastal waters - Salmon (and sea trout), 7: Small coastal waters - Sea-trout. A simple decision model was produced to designate sites to fish types using historical data on sea trout, salmon and lake trout, Huet zonation and expert judgement regarding the occurrence of Arctic char. The non-salmonid type found was rare in ecoregions 20/22 and all of these sites could be designated to the Sea trout type using the decision model including historical data.

The correlation of FAME list of metrics was screened against the mean of the five selected impact variables. 14 additional metrics were defined using information from fish lengths. These were added to address the question of length/age, which may be useful in species poor systems. Due to lack of highly and moderately impacted sites in the dataset, it was for none of the seven fish types possible to produce a model to determine ecological status based on the response of any of the tested metrics. The reasons for failure to develop a model for ecoregions 20/22 were mostly due to the low occurrence of impacted sites. It can also be discussed that combining the impact of the five impact variables into a mean impact may result in effects of one variable counteracting effects of another. Additionally, the impact variables used might not reflect all human impact on fish in ecoregions 20/22 or the classification of species into the used metrics may be less relevant in these species poor systems.

Introduction

This is a presentation of an attempt to produce a spatially based approach for assessing the ecological status of running waters using electric fishing data. The work is based on the data present in the FAME database FIDES. Primarily a reference data set was selected. From this a classification of rivers was done using fish. For each of the resulting “fish types” different metrics proposed by the FAME consortium were tested.

It was decided that the spatial approach conducted by the Swedish partner should concern Illies’ ecoregion 20, “Borealic uplands” and 22, “Fenno-Scandian shield”. “Borealic uplands” cover the Swedish mountain range and the whole of Norway, whereas the “Fenno-Scandian shield” covers the lowlands of northern Sweden and the whole of Finland.

The southern part of Sweden is a part of ecoregion 14, “Central plains”. This was the area in the national data set with most of the impacted sites. In ecoregions 20 and 22, impacted sites are few, and especially heavily impacted sites are very rare in the data set. Out of 623 sites in the national data set, 279 sites were from ecoregions 20 (n=41) and 22 (n=238). Due to the few sites in ecoregion 20 it was decided to produce a joint typology for ecoregions 20 and 22.

1. Fish based typology

1.1 Data selection

In spite of the advised methods for the spatially based approach (Böhmer and Schmutz 2003), the dataset was not initially split up into one part used for analyses and one part for validation. The reason for that was that it could be foreseen that some of the fish types would be represented by few sites in the dataset, and that it was therefore preferred to validate the results later using another dataset from the Swedish electrofishing register.

The calibration dataset was selected primarily according to the rules suggested at the meeting in Lyon, i.e. human impact should be less than 3 on five critical impact variables. These were; Connectivity segment, Hydrological regime site, Morphological condition site, Toxic-Acidification site, Nutrients-organic input site (Beier et al. 2002). For ecoregion 20 and 22 the criteria could have been set more strictly, e.g. human impact 1, but to comply with the FAME consortium it was decided to follow the recommendations from the meeting in Lyon. After the meeting in Portugal also the impact variables Connectivity_segment and Connectivity_river were joined in a combined new variable, Connectivity_multiscale. This was used in the analyses.

For the Swedish data it was decided that “Introduction of fish site” and “Impact of stocking site” must be 1. This was to avoid including localities with stocked salmon, eel and trout that would cause incorrect evaluations. Such stocking used to be rather common, especially in coastal streams, but is today diminishing except for large-scale reintroductions of salmon.

This selection resulted in 209 sites, which corresponds to 75% of all sites in ecoregion 20 and 22. From each calibration site where multiple fishing occasions were present, a single fishing occasion was randomly selected for the fish typology analyses.

The general scores for the variables describing human impact was low in the dataset for ecoregions 20 and 22. The number of sites with a mean of 1 of the five impact variables (‘reference sites’) were 20 (49%) in ecoregion 20, and 109 (46%) in ecoregion 22.

1.2 Description of the calibration dataset

No introduced species or hybrids occurred in the calibration dataset. In total 23 taxa (20 species) were caught (Table 1). Salmo trutta was the most common species and occurred at almost 90% of the calibration fishing occasions. Trout has been divided into three ecological forms; *Salmo trutta trutta* (sea trout), *Salmo trutta lacustris* (lake-migrating trout) and *Salmo trutta fario* (stream-living trout). The dominance of trout in the data set is to a great extent due to the selection of sites for electric fishing, i.e. wadable sections with hard bottom. The second most frequent taxa was *Cottus spp.* (*Cottus gobio* and *C. poecilopus*), present at 63% of the calibration fishing occasions. Burbot and minnow were the only two further species that were present at more than 20% of the calibration fishing occasions. Eleven species were found at less than 5% of the occasions, and out of these six species were found at only a single occasion. One locality (SE0520) lacked fish due to high altitude and small size, and was omitted in analyses.

Table 1. Species occurrence at the calibration fishing occasions (n= 209 sites) used in the ordination, factor analysis and clustering to obtain fish types.

		Species	***Species occurrence***	***Per cent of all fishing occasions***
1		Salmo trutta	188	89,8
	1a	*Salmo trutta fario*	136	65
2		Cottus gobio	99	47,4
3		Lota lota	73	34,9
4		Phoxinus phoxinus	58	27,8
	1b	*Salmo trutta trutta*	45	21,5
5		Cottus poecilopus	33	15,8
6		Esox lucius	30	14,4
7		Lampetra planeri	29	13,9
8		Salmo salar	27	12,9
9		Thymallus thymallus	24	11,5
10		Rutilus rutilus	10	4,8
	1c	*Salmo trutta lacustris*	7	3,3
11		Perca fluviatilis	5	2,4
12		Lampetra fluviatilis	2	1
13		Gymnocephalus cernuus	2	1
14		Salvelinus alpinus	2	1
15		Alburnus alburnus	1	0,5
16		Leuciscus leuciscus	1	0,5
17		Gasterosteus aculeatus	1	0,5
18		Pungitius pungitius	1	0,5
19		Barbatula barbatula	1	0,5
20		Coregonus albula	1	0,5

The Swedish data were originally selected from a larger dataset to be as representative as possible for the whole of Sweden. Sites which were well known, e.g. had been electrofished several times, were preferred in the selection process. These criteria apply to the ecoregions 20 and 22 as well. In the dataset, there is an overweight for wadable, smaller streams, which however possibly reflects the relative amount present in the area. It seems that the sites are reasonably evenly spread out, and present in 19 different rivers or streams in ecoregion 20, and in 92 rivers in ecoregion 22 (Fig. 1).

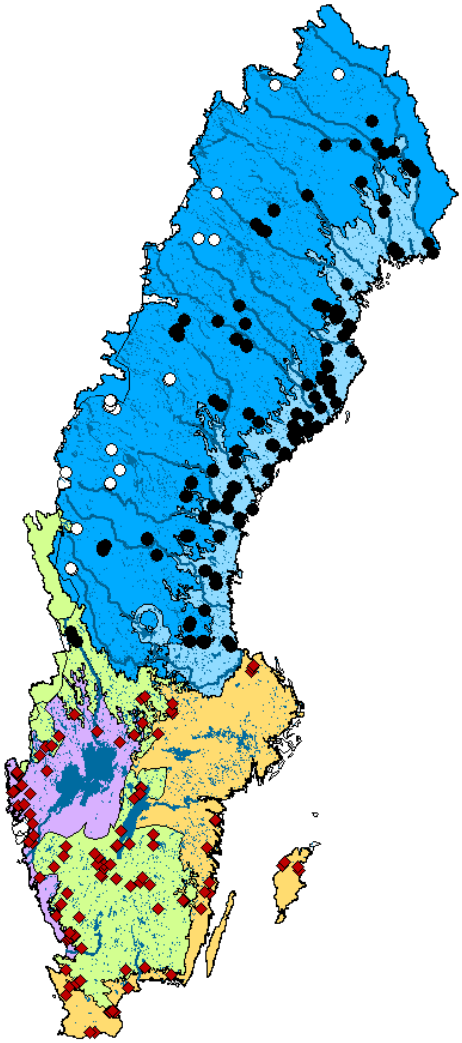


Figure 1. Calibration sites in the Swedish FIDES dataset. White rings belong to Illies' ecoregion 20 (Borealic uplands), black rings to ecoregion 22 (Fennoscandian shield) and diamonds to ecoregion 14 (Central plains). Differently coloured areas represent five delimited subecoregions: 1. Northern highland (dark blue), 2. Northern lowland (light blue), 3. Southern highland (green), 4. Southern lowland (orange) and 5. West coast (purple) (Beier et al. 2002).

1.3 Methods, data and tests

Used methods were Principal component analysis (PCA), Detrended correspondance analysis (Decorana) and Hierarchical clustering. Tests were performed using SPSS (version 11.5.1) and Cap (Community Analysis Package, 2.2).

Decorana addresses some problems with Correspondance analysis. However, Decorana is sensitive to data sets with samples that differ in their composition or distribution to the rest of the data set. Such samples should be removed. Therefore only species occurring on more than one site was used in the analyses.

Hierarchical clustering was done using Wards method, which is also termed minimum variance. At each iteration all possible pairs of groups are compared and the two groups chosen for fusion are those that will produce a group with the lowest variance. As distance measure Euclidean distance was used on presence/absence data.

PCA-correlation

Seven factors with eigenvalues above 1 were found explaining 66% of the variation among sites. The first factor distinguished between sites with stream resident trout (alone) and sites with sea trout and salmon (Figure 2). The second separated sites with lake trout from sites with percids and pike, i.e. non-salmonid slow flowing streams. The third factor separated sites with sea trout from those with stream resident. The fourth factor again separated species living in stronger current from those living in slow flowing waters.

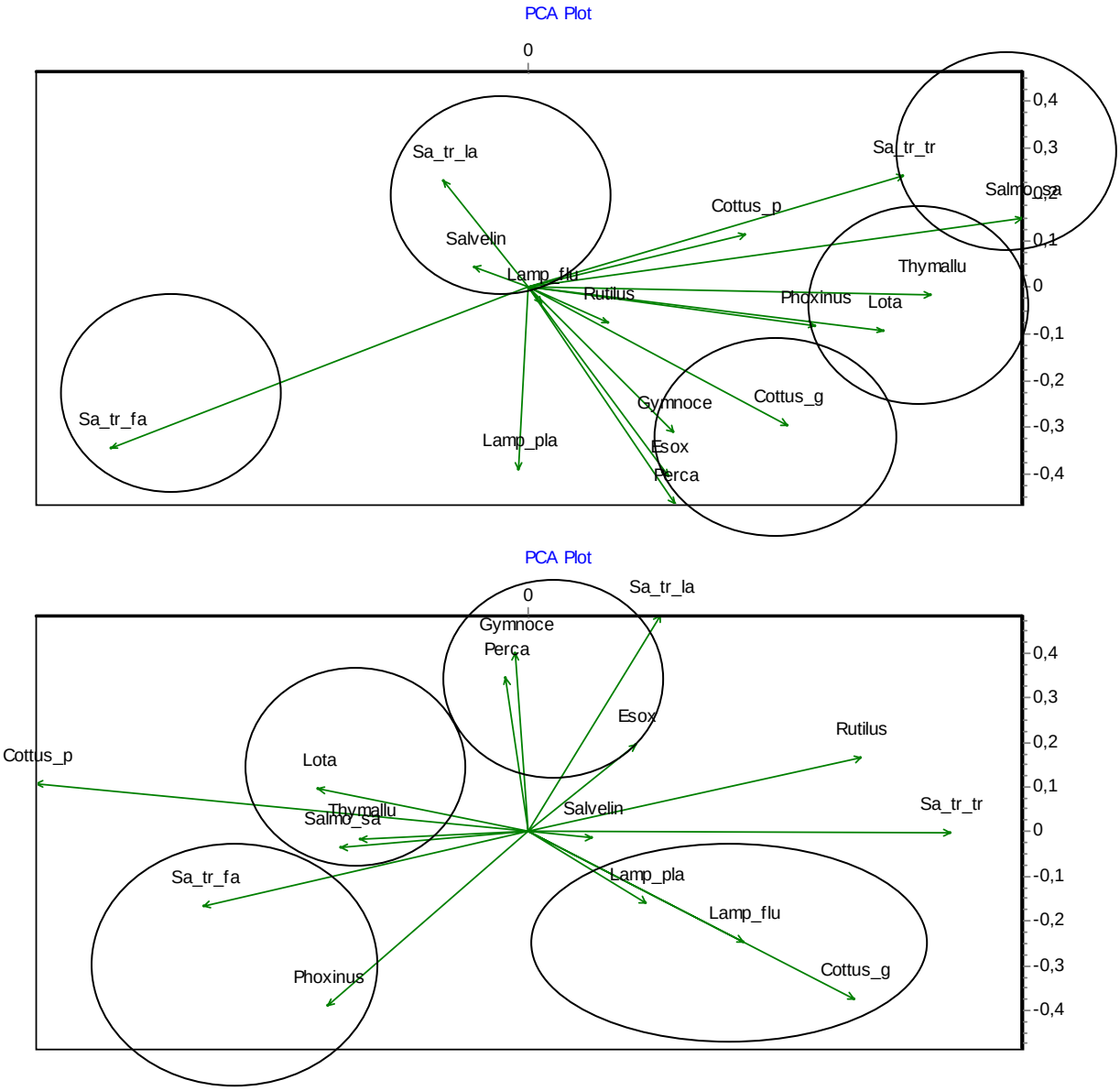


Figure 2. The first four PCA factors (sites not shown). Factor 1 and 2 above, 3 and 4 below.

Decorana

As stated Decorana was performed using the reduced data set, i.e. without low frequency species. The four axis had eigenvalues from 0,99 to 0,32. The first axis distinguished sites with lake trout from other sites, whereas the second factor separated the Arctic char from other species (Figure 3). The third factor separated sites with stream-resident trout (and occasionally brook trout) from sites with sea migrating salmonids (including grayling which is sea migrating in the northern part of the Gulf of Bothnia). The fourth factor separated sites with species in slow flowing waters (e.g. roach, ruffe, perch, pike) from such with more rheophilic species as minnow, brook lamprey and alpine bullhead.

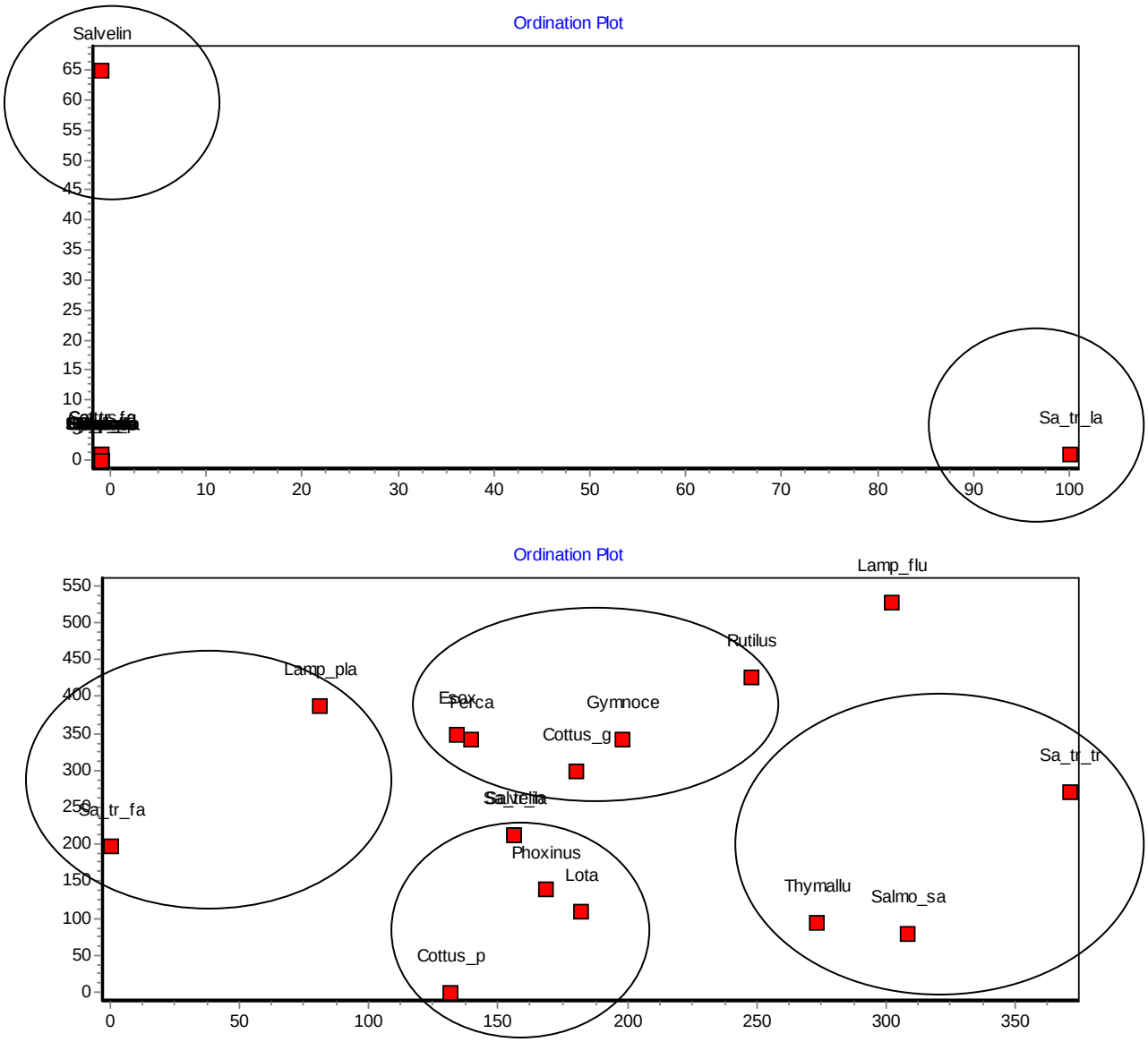


Figure 3. Axis 1 and 2 (above) and axis 3 versus axis 4 (below) on Detrended correspondance analysis (Decorana) on species occurrence.

Hierarchical clustering

Seven groups that separated the data set well were identified. As with the other methods applied, lake trout and Arctic char formed distinct groups (Figure 4). Also river lamprey formed an own group, but has previously been associated with trout, bullhead and brook lamprey.

Stream-living trout was found together with bullhead, minnow and brook lamprey. Sea trout formed an own group, but often co-occurred with salmon (and burbot, grayling, and alpine bullhead). As above species preferring low water velocity were grouped together; pike, perch, roach and ruffe.

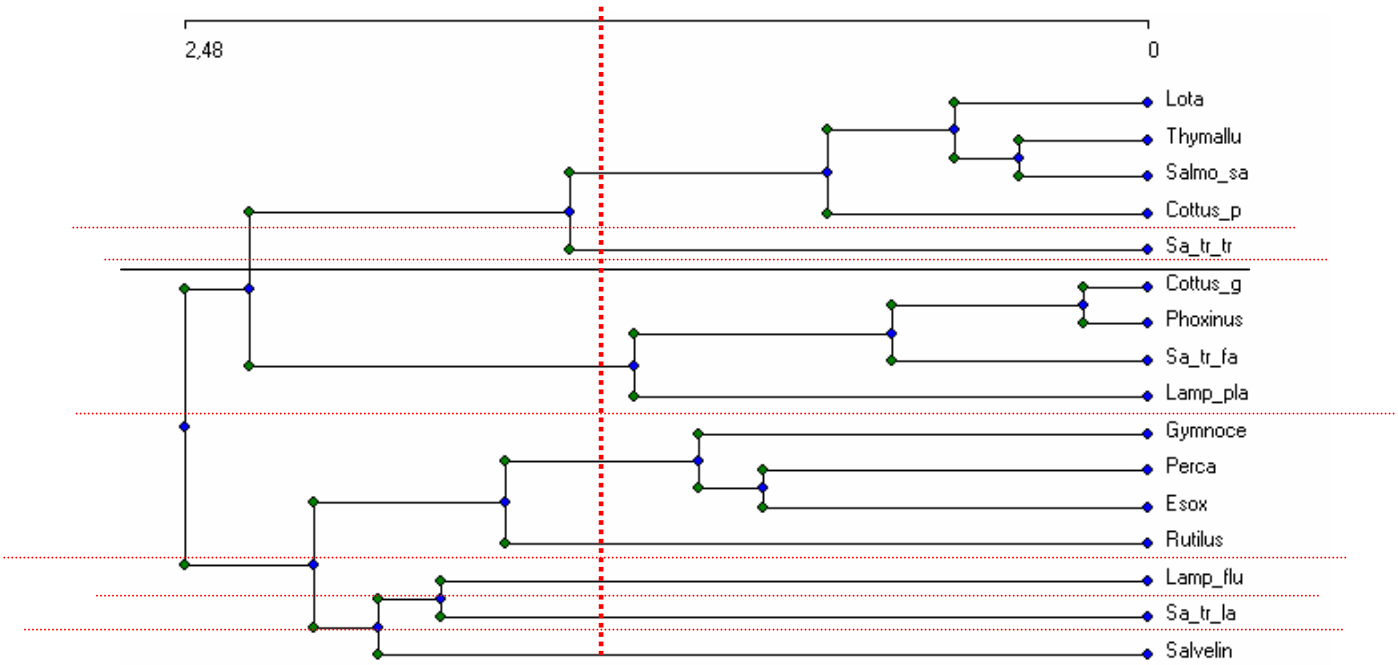


Figure 4. Wards agglomerative cluster using Euclidean distance.

1.4 Resulting typology and decision model

From the results above it was suggested that seven species assemblages could be distinguished (Table 3).

Table 3. Suggestion for fish-based typology for ecoregions 20 and 22.

No.	Water type	Fish type
1	Subalpine lake in/outlets	Arctic char
2	Inland brook	Trout dominated
3	Inland streams	Trout with others
4	Lake in/outlets	Lake-trout
5	Lenthic	Non-salmonid
6	Large coastal waters	Salmon (and sea trout)
7	Small coastal waters	Sea-trout

For allocation of a specific site to the correct typological unit (t.u.) a model has been suggested to allow identification of correct t.u. for each new site.

1. From historical data fish types “Sea trout”, “Salmon” and “Lake trout” are classified.
2. For the reminder of sites a division is made between salmonid (trout, grayling, barbel zones) and non-salmonid waters (=bream zone according to Huet).
3. Fish type “Arctic char” is separated from other salmonid waters by expert judgement*.
4. Remaining salmonid waters are classified from stream order, where stream order 1-3 = “trout dominated”, and higher stream order “trout with others”.

*Arctic char is found at altitudes above 375 m.a.s.l. in ecoregion 20. It is lake-migrating and subalpine lakes with present or historical presence of Arctic char is a prerequisite.

1.5 Fish-based typology applied to calibration sites

Especially “Sea trout”, “Salmon”, “Trout dominated” and “Trout with others” were well represented. “Arctic char” was only represented with two sites (Table 4). More sites can be added from the Swedish electrofishing register in the future.

In this classification no non-salmonid sites were identified. The sites where non-salmonids dominated, as seen in clusters and ordinations, were later classified as “Sea trout” from historical data. Historical data was not used in the analyses above and this would explain the discrepancy between the ordination/clustering and outcome of the final typology.

Table 4. Distribution of calibration sites (n=209) in the resulting typological units based on fish.

Fish type	Frequency of sites	Percent of sites
1-Arctic char	2	1%
2-Trout dominated	66	31,6%
3-Trout with others	66	31,6%
4-Lake-trout	8	3,8%
5-Non-salmonid	0	0%
6-Salmon	35	16,7%
7-Sea trout	32	15,3%

Tables 5-8 are descriptive showing the outcome of the fish-based typology with respect to stream characteristics and classification of human inpact.

Table 5. Average of environmental variables for the different fish typological groups.

FISHTYPE	Altitude (m)	Air temp. (mean)	Gradient slope (pro mille)	Wetted width (m)
1 Arctic char	434	-0,5	53,3	12,9
2 Trout dominated	346	2,1	20,5	5,3
3 Trout with others	326	1,9	17,9	20,4
4 Lake trout	750	0,5	21,2	6,9
6 Sea trout	48	3,2	17,7	6,4
7 Salmon	117	1,2	3,8	131

Table 6. Size of catchment (km^2) for sites in the different fish typological groups.

Label	Fishtype no	<10	<100	<1000	<10000	>10000	Total
Arctic char	1	1	1	0	0	0	2
Trout dominated	2	9	53	4	0	0	66
Trout with others	3	11	42	5	6	2	66
Lake-trout	4	4	4	0	0	0	8
Sea-trout	6	10	16	3	3	0	32
Salmon	7	0	0	0	21	14	35
Total		35	116	12	30	16	208

Table 7. Average of impact variables for the different fish typological units.

Label	Fishtype no	Conn_river	Land_use_seg	Urban_seg	Rip_z_seg	Conn_seg
Arctic char	1	3,0	1	1	1	1
Trout dominated	2	4,9	1	1	1,2	1,1
Trout with others	3	5,0	1	1,1	1,2	1
Lake-trout	4	5,0	1	1	1,1	1
Sea-trout	6	1,6	1,1	2,0	1,4	1
Salmon	7	1,6	1,1	2,2	1,1	1

Label	Fishtype no	Sedim_seg	Hydrol_site	Morph_site	Tox_site	Nutr_site
Arctic char	1	1	1	1	1	1
Trout dominated	2	1,2	1	1,1	1,2	1
Trout with others	3	1,2	1	1	1,2	1,1
Lake-trout	4	1	1	1	1,4	1
Sea-trout	6	1	1,1	1,1	1	1,4
Salmon	7	1,1	1	1	1,1	1

Table 8. Occurrence of fish species (and groups) at sites in the different typological units.

	Arctic char	Trout dom.	Trout & oth.	Lake trout	Sea trout	Salmon
Sea-trout	0	0	2,4	0	**88,2**	56
Bullhead	0	26,5	76,2	0	**70,6**	68
Roach	0	4,1	2,4	0	**14,7**	0
Dace	0	0	0	0	**2,9**	0
R. Lamprey	0	1	0	0	**2,9**	0
Bleak	0	0	0	0	**2,9**	0
Salmon	0	0	0	0	5,9	**100**
Minnow	0	24,5	23,8	0	23,5	**64**
A. Bullhead	24,5	10,2	26,5	0	0	**56**
Grayling	0	3,1	9,5	0	14,7	**48**
Th. Stickleback	0	0	0	0	0	**4,2**
Ni. Stickleback	0	0	0	0	0	**4**
Stone loach	0	0	0	0	0	**0,4**
Burbot	0	26,5	30,9	0	35,3	84
Pike	0	12,2	16,7	0	14,7	20
Lake-trout	0	0	0	**100**	0	0
Trout	0	**98**	90,5	0	2,9	4
B. Lamprey	0	10,2	**31**	0	11,8	8
Perch	0	0	**7,1**	0	2,9	4
Ruffe	0	0	**2,4**	0	0	0,4
Cisco	0	0	**2,4**	0	0	0
Arctic char	**100**	0	2,4	0	0	0

2 Screening of metrics

2.1 Calibration data for the screening of metrics

For the selection and screening of potential metrics, both calibration and impacted sites must be present and all 279 sites were chosen initially. It was decided to select the first fishing occasion at each site for the analyses after the fish typology had been carried out, to obtain the largest possible impact. The reason for this is that effects of acidification, being one of the major impacts in the country, were more pronounced earlier on before liming had been carried out. We are perfectly aware that this was not correct in a statistical sense. However, to be able to proceed with the screening and selection of metrics, and classification of status, this was done to obtain a dataset with a slightly greater span in impact variables.

As several sites were represented with several fishing occasions it was decided to use the first fishing occasion at each site where length measurements were available. The reason taking this into account was that preliminary analyses pointed towards the need for some metric based on length data.

Sites where the sum of impact variables changed between fishing occasion (n=65) were allowed to be represented by two fishing occasions, i.e. these two fishing occasions were not treated as replicates. If yet another changes in the sum of impact variables had occurred at the site a third fishing occasion was allowed to enter the calibration data set. This was rare (n=2). Finally, the calibration data set consisted of 333 fishing occasions from 265 sites (Table 9).

Although additional fishing occasions have been added there was still a very low proportion of impacted sites in the data set (Table 10).

Table 9. Number of sites in different impact classes (sum of the five impact variables) depending on selection criteria for the calibration data set. Data set 1=Only one fishing occasion per site in Ecoregion 20-22, Data set 2=Addition of fishing occasions where the sum of impact changed at another date.

	Data set	1 (n=265)	2 (n=333)
Sum of impact	5	**130**	**154**
	6	**56**	**78**
	7	**41**	**57**
	8	**23**	**27**
	9	**11**	**12**
	10	**2**	**3**
	11	**2**	**2**

Table 10. Number and distribution of sites in the different fish types and the five impact variables in calibration data set 2.

Arctic char	Impact	Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=4)	1	4	4	4	2	4
	2				1	
	3				1	

Trout dominated		Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=104)	1	95	100	99	64	101
	2		2	5	18	3
	3	9	2		16	
	4				5	
	5				1	

Trout & others		Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=106)	1	89	106	97	64	101
	2	5		9	20	5
	3	8			14	
	4	1			7	
	5	3			1	

Lake trout		Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=16)	1	16	16	16	6	14
	2				6	1
	3				2	1
	4				2	

Salmon		Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=49)	1	52	46	47	47	36
	2	1	2	6	3	15
	3		5		3	2
	4					
	5	1				

Sea trout		Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=33)	1	33	31	30	19	23
	2		2	3	8	9
	3				5	1
	4				1	
	5					

All fish types		Connectivity	Hydrological	Morphological	Toxic	Nutrients
(n=333)	1	306	318	307	214	292
	2	5	8	26	60	37
	3	17	7		42	4
	4	1			15	
	5	4			2	

3 Selection of metrics

3.1 Potential metrics

The FAME consortium produced a list of a large number of metrics to be tested against the five impact variables, for each fish type (Böhmer and Schmutz 2003). In addition to these metrics, 14 other metrics were defined using information from length measurements. The new metrics were developed to address the question of length/age, with potential size ranges of migrating species as indicators, which may be useful in species poor systems. The list of additional metrics was as follows:

- No of lengths
- % of lengths gt 150 (150 mm)
- No of salmonid inds (*Salmo* species)
- No of salmonid inds gt 150
- % of salmonid inds gt 150
- No of migrating sp inds (*Anguilla anguilla, Salmo salar, Salmo trutta trutta, Salmo trutta lacustris*)
- No of migrating inds gt 150
- % of migrating sp gt 150
- No of long migrating sp inds (*Anguilla anguilla, Salmo salar, Salmo trutta trutta*)
- No of long migrating sp gt 150
- % of long migrating sp gt 150
- No of S t l inds
- No of S t l gt 150
- % of S t l gt 150

It can be argued that several of the listed metrics are probably internally correlated, and it could be sufficient to use only for example the % metrics from the list. However, at this stage it was chosen to include all to investigated which of the above that were associated with the original metrics also.

3.2 Screening of potential metrics

As a simple way to exclude metrics with no relationship to the impact variables it was suggested by Böhmer and Schmutz (2003) that the Spearman Rank correlation procedure could used as it is robust and not does not require parametric data or linear relationships.

The “sum of impact” values and the “mean of five impacts” were calculated and a bivariate correlation matrix was established between this, each of the five impact variables and the potential metrics for each fish type. Metrics with the highest correlation (both positive and negative) were selected for further screening.

After this initial screening using Spearman Rank correlation each candidate metric was plotted;

1/ using box-plots (non-parametric) to compare the metric at reference sites with impact sites.
2/ using box-plots to study the distribution of metric values for categories of impact of the five impact variables.
3/scatterplots of metrics against sum of impact variables.

Candidate metrics should be chosen to address different ecological guilds (habitat, feeding, migration, rheophily, reproduction habitat, life span) and the four WFD criterions (richness, tolerance, taxonomic composition, reproduction) for each fish type.

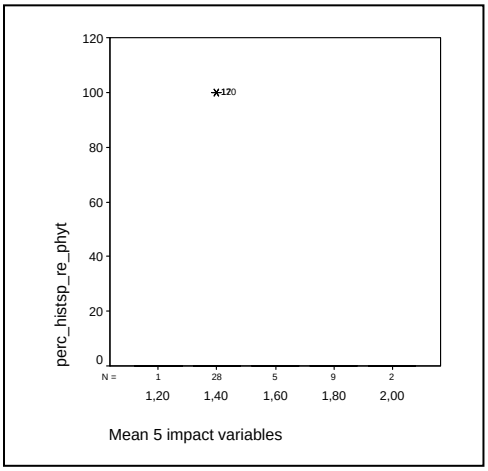
For the fish type “Arctic char” to few sites were included and this fish type was omitted in the analysis.

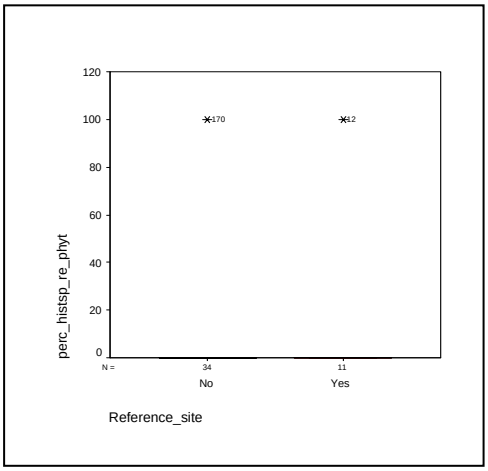
Fish type "Trout dominated"

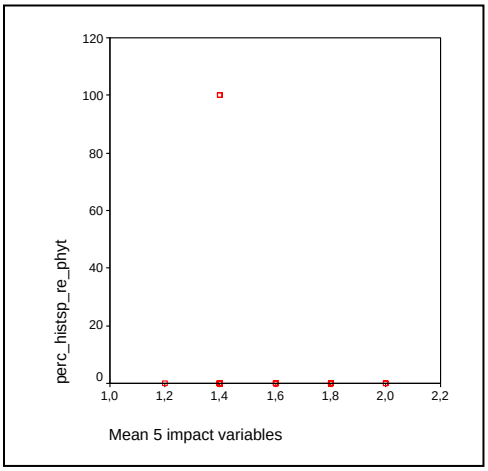
For the fish type "Trout dominated" only weak associations existed between the impact variable (mean of five impacts) and the potential metrics (Table 11). Obviously, to few impacted sites were included in the data set. The Spearman rank correlation r of –0,21 and 0,23 were to low to indicate any meaningful correlation between metrics and impact. The two highest correlated metrics were displayed in a box-plot against "mean of five impacts", and also against reference-impacted sites. Finally, a scatterplot was produced between metrics and "mean of five impacts" (Figure 5).

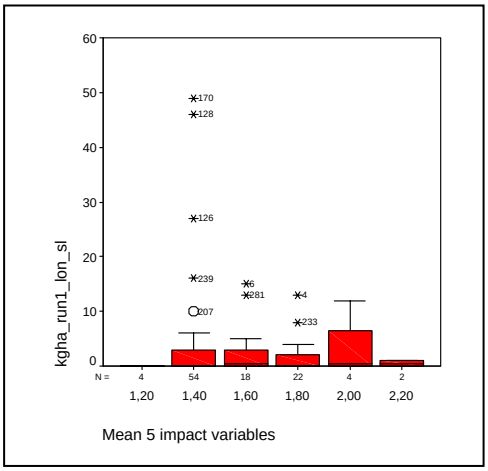
Table 11. Bivariate Spearman rank correlation matrix between metrics and mean of five impact variables for fish type "Trout dominated". Only the potential metrics with the strongest correlations (positive and negative) are shown.

Metrics	Spearman
perc_histsp_re_phyt	-0,21088
Presence_0plus_Sal_far	-0,20793
perc_0plus_Sal_far	-0,17586
density_0plus_Sal_far	-0,17572
perc_histsp_lon_sl	0,171923
kg_ha_Lon_sl	0,193186
n_ha_Lon_sl	0,196545
kgha_run1_lon_sl	0,233831









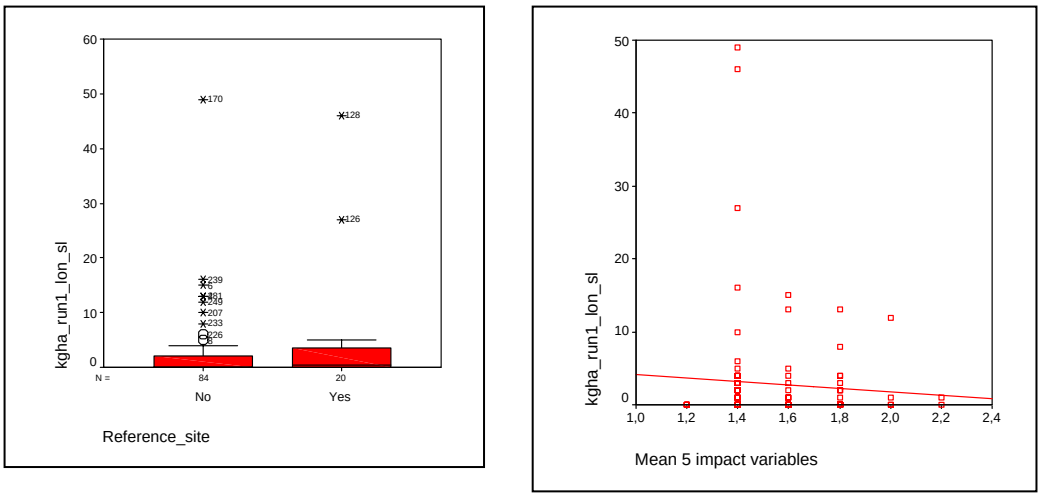


Figure 5. Box-plots and scatterplot between the two most correlated metrics and "mean of five impacts" and reference sites and impacted sites, respectively.

Unfortunately, the lack of heavily impacted sites makes further efforts meaningless. It has been discussed to include impacted sites from Ecoregion 14. However, it was decided to follow the original plan. Thus, it was concluded that it was not possible to produce a model for this fish type with the present data set.

Fish type "Trout with others"

For the fish type "Trout with others" also few impacted sites were present in the material. Only weak associations existed between the impact variable (mean of five impacts) and the potential metrics (Table 12). Plots were produced for the three most correlated metrics (Figure 6). As for the preceding fish type it was not meaningful to carry on the work.

Table 12. Bivariate Spearman rank correlation matrix between metrics and "mean of five impacst" variables for fish type "Trout with others". Only the potential metrics with the strongest correlations (positive and negative) are shown.

Metrics	Spearman
perc_histsp_hab_b	-0,43544
perc_sp_hist	-0,40166
perc_histsp_fe_insev	-0,39204
perc_histsp_intol	-0,39124
perc_histsp_lon_sl	-0,38898
perc_histsp_re_lith	-0,3756
perc_histsp_hab_rh	-0,33713

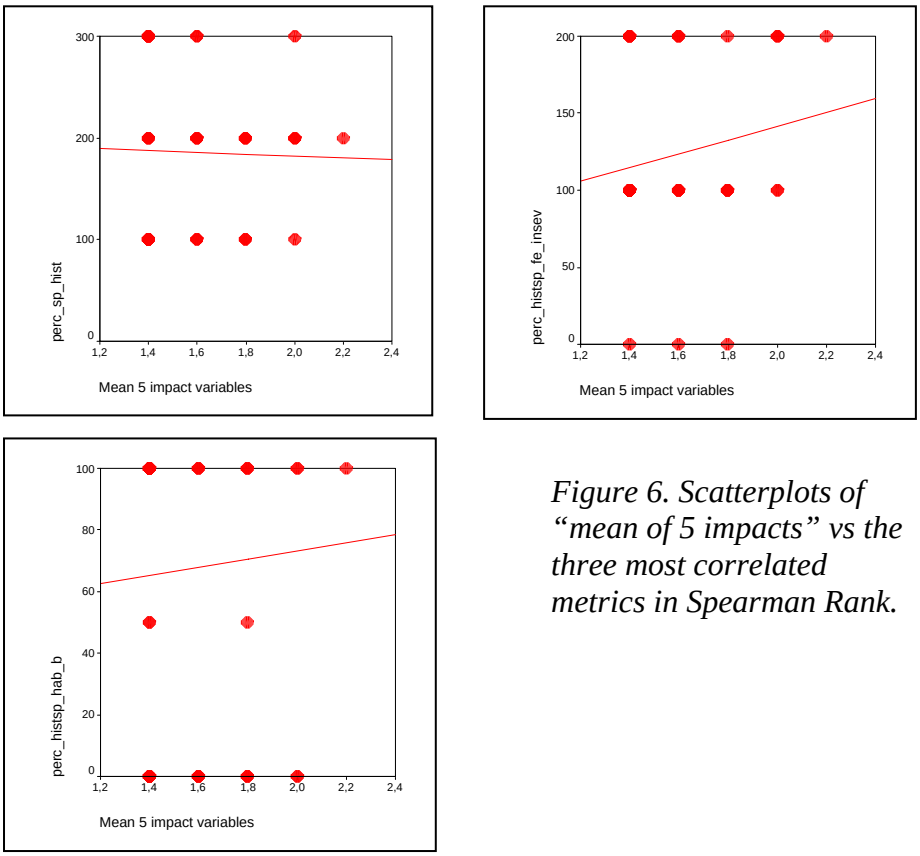


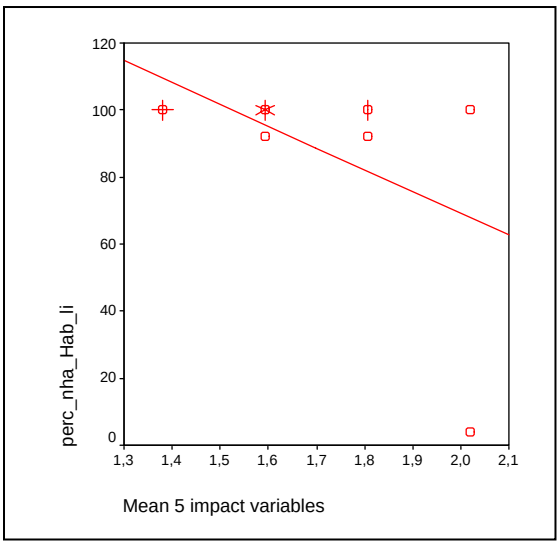
Figure 6. Scatterplots of "mean of 5 impacts" vs the three most correlated metrics in Spearman Rank.

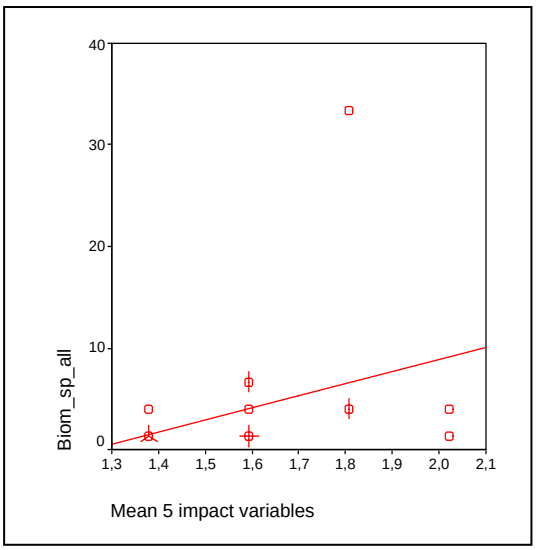
Fish type "Lake trout"

For the fish type "Lake trout" slightly higher correlations were found than for the two preceding fish types (Table 13). However, scatterplots revealed that the situation was similar to the other fish types (Figure 7), i.e. it was not possible to derive an index from the present data.

Table 13. Bivariate Spearman rank correlation matrix between metrics and "mean of five impacts" variables for fish type "Lake trout". Only the potential metrics with the strongest correlations (positive and negative) are shown.

Metrics	Spearman
perc_nha_Hab_li	-0,43539
perc_nha_Hab_wc	-0,43539
perc_sp_Hab_li	-0,43539
perc_sp_Hab_wc	-0,43539
Perc_sp_native	-0,4143
biom_run1_alien	0,414304
perc_nha_Hab_b	0,435393
perc_sp_Hab_b	0,435393
Biom_sp_all	0,446944
biom_run1_all	0,499408





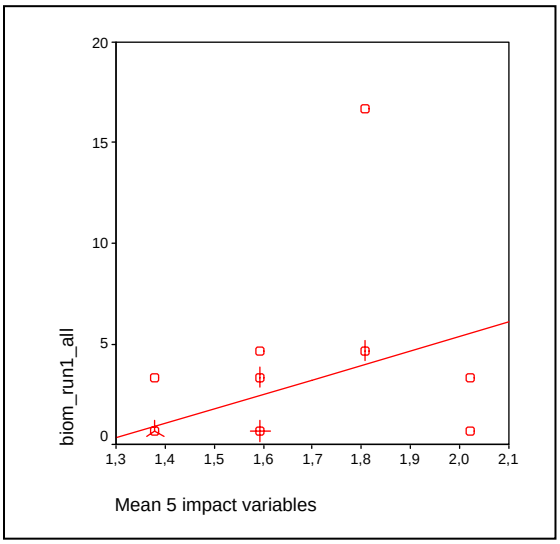


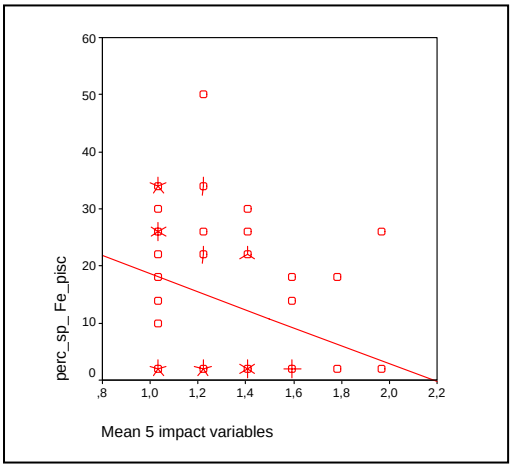
Figure 7. Scatterplots of "mean of 5 impacts" vs the three most correlated metrics in Spearman Rank.

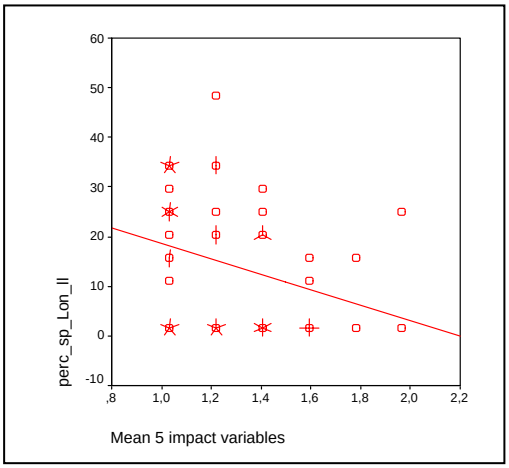
Fish type "Salmon"

For the fish type "Salmon" only weak associations existed between the impact variables and the potential metrics (Table 14). Correspondingly, the plots showed that the small amount of impacted sites made the result ambiguous (Figure 8).

Table 14. Bivariate Spearman rank correlation matrix between metrics and "mean of five impacts" variables for fish type "Salmon". Only the potential metrics with the strongest correlations (positive and negative) are shown.

Metrics	Spearman
perc_sp_ Fe_pisc	-0,3683
perc_sp_Lon_ll	-0,3683
perc_sp_Hab_b	-0,36767
n_sp_Hab_b	-0,33544
n_sp_hist	-0,33091
n_sp_ Fe_pisc	-0,32255
n_sp_Lon_ll	-0,32255
perc_histsp_fe_pisc	-0,32255
perc_histsp_lon_ll	-0,32255
n_ha_Re_lith	0,343847
Density_sp_all	0,347514
Density_sp_native	0,347514
n_ha_Hab_rh	0,355219
perc_sp_Hab_wc	0,367669





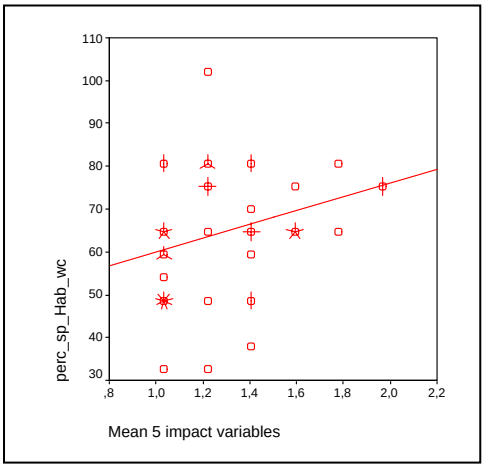


Figure 8. Scatterplots of "mean of five impacts" vs the three most correlated metrics in Spearman Rank.

Fish type "Sea trout"

For the fish type "Sea trout" some strong associations were found between the impact variable and the potential metrics (Table 15). The quantities of long-migrating species or individuals were important metrics in this fish type (Figure 9). As for the other fish types the amount of impacted sites was too low for further model development.

Table 15. Bivariate Spearman rank correlation matrix between metrics and mean of five impact variables for fish type "Sea trout". Only the potential metrics with the strongest correlations (positive and negative) are shown.

Metrics	Spearman
perc_sp_Mi_long	-0,65308
perc_nha_Mi_long	-0,58166
perc_kgha_Mi_long	-0,57162
n_ha_Mi_long	-0,4477
density_0plus_Sal_lac	-0,44602
perc_kgha_Hab_wc	-0,42156
perc_nha_Hab_wc	-0,41447
perc_sp_Hab_wc	-0,40432
n_sp_Lon_sl	0,401557
perc_sp_Hab_b	0,404322
perc_nha_Hab_b	0,416693
perc_kgha_Hab_b	0,421558
perc_histsp_lon_sl	0,429311
perc_sp_Lon_sl	0,441978

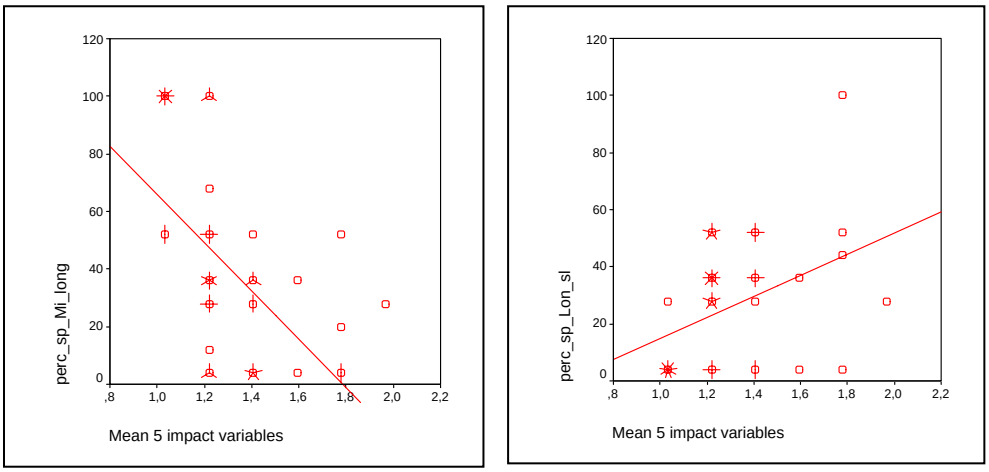


Figure 9. Scatterplots of "mean of 5 impacts" vs the two most correlated metrics in Spearman Rank.

4 Discussion

The effort failed to produce a model to assess ecological status based on a spatial approach in ecoregion 20 and 22. Below we discuss the major causes to this and propose how a repeated effort could be successful.

4.1 Too few sites with high impact

There is an obvious need to extend the dataset to include more fishing occasions with higher environmental impact. One of the major causes to the skewness towards low-impact sites is that many of the sites are fished as a part of monitoring of salmonid stocks. Thereby the data set is biased towards “interesting” populations where environmental improvement projects, e.g. liming to counteract acidification, are carried out.

As a consequence of the few sites with high impact the data set was imbalanced. A single value at the impacted site influences the outcome of the Spearman Rank correlation, i.e. chance has to large freedom to act.

In figure 10, it is shown that one impacted site may determine the whole outcome of a correlation study. In the upper figure the value of the metric is 75, if by chance the metric would have been 55 (perhaps well within natural variation) as in the lower figure, the outcome of the correlation study would have been completely different. Again, the screening procedure would function well in a balanced data set. Hence, the lack of impaired sites is really the problem.

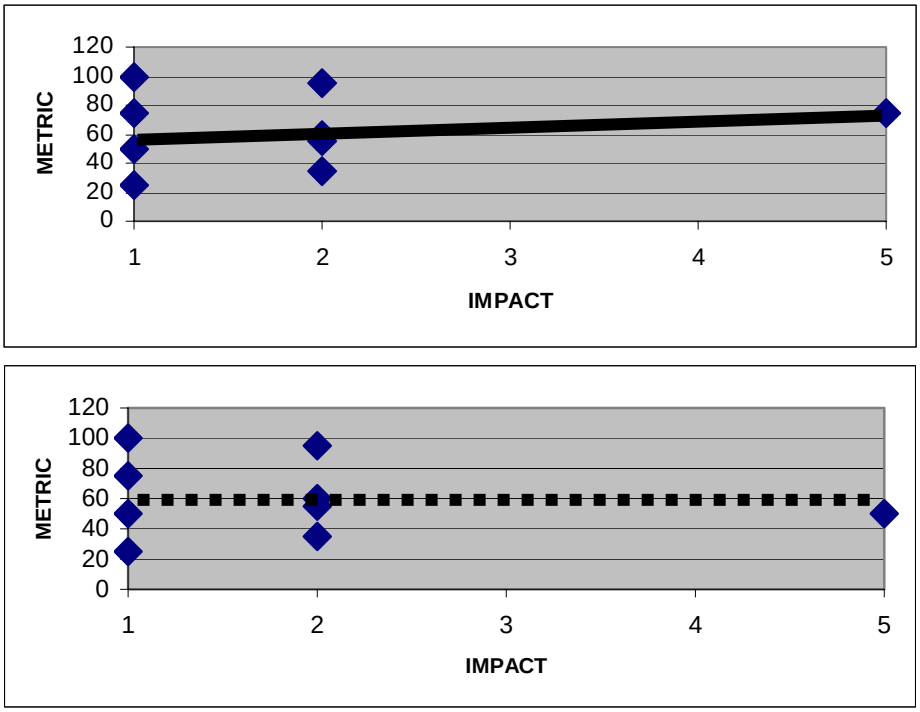


Figure 10. Illustration of how the effect of the value of a metric for a single impacted site may influence the outcome trend of the impact.

4.2 The use of a combined impact variable

Using the sum (or mean) of impact variables can in certain cases result in a scenario where one impact variable is counteracted by another. If for instance a site is under heavy nutrient load (impact value=5) while the other impact variables all have value 1 this will result in an average of 1.8 and a sum of impact of 9. In spite of this relatively low value the fish fauna may be severely disturbed. But at a site where the nutrient load is low (impact=1), but the other four impact variables have a value of 2, the combined impact will be average 1.8 and sum of impact 9, i.e. precisely the same as with the disturbed site above.

One way of circumventing this problem is to use maximum impact of any impact variable, which results in more sound correlation between metrics and impact (Figure 11).

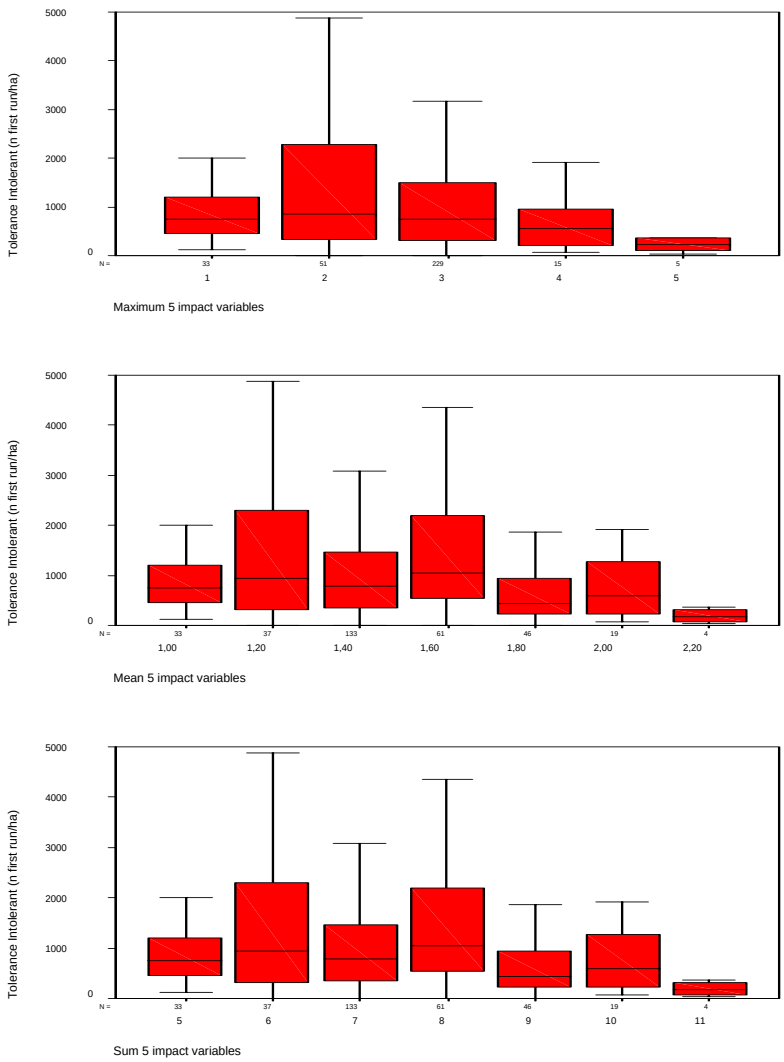


Figure 11. Number of intolerant individuals (first run) versus a combined impact variable calculated as maximum impact, mean impact and sum of impact.

4.3 Impact assessment improvment

Recent work has shown that certain kinds of impact have not been included, although the effect on biota is large. The majority of sites are situated in forest areas where forestry is generally carried out. The potential effect is diffuse with possible increased sediment load, loss of large woody debris (LWD), and deterioration of riparian forest. It is strongly felt that the diffuse effect of forestry might be underestimated. Degerman et al. (2004) compared sites with different amount of LWD in forest streams in Sweden. It could be shown that sites without LWD had only half of the abundance of trout as compared to sites with more than 4 pieces of LWD per 100 m^2. Also growth of trout, size of the largest trout and occurrence of brook lamprey differed significantly. It was argued that loss of LWD, mainly due to forestry, has a major impact on Swedish forest streams. If so, it might be that the reference sites chosen were indeed not unaffected by forestry.

4.4 Reclassification of species

In the investigated waters the number of species was low. The classification of species unfortunately, and quite understandingly, only resulted in a species grouping with regard to tolerance into two groups; tolerant and intolerant. In this classification Cottus spp. was classified as intolerant, which is correct in a European perspective. However, along a degradation gradient in ecoregion 20-22 we normally have salmonids, followed by cottids and then cyprinids. Due to the lack of severely impacted sites this fine distinction between species became troublesome. The suggested solution, however, is to add more impacted sites, and not to alter the common classification.

Furthermore, the metrics used in the FAME project are designed to cover a wide range of habitats with a wide range of species assemblages, which are often more species rich compared to ecoregions 20 and 22. The fewer number of species is the reason why metric values calculated for data from these two ecoregions often result in zero values. Table 16 shows that for the six fish types found in the dataset, four metrics from the final modelling approach are almost all zero values, among them all three metrics increasing with impact. In further developing a system for the ecoregions 20 and 22, care would have to be taken to choose metrics in a way to detect differences in impact, and not just species richness.

Table 16. Percentage of zero values in the dataset from ecoregions 20 and 22 (North Sweden) for the ten metrics used in the FAME modelling approach, as well as for Salmo trutta fario for comparison. Values ≥90 are in bold.

Fish type	Arctic char	Trout with others	Trout dominated	Lake trout	Salmon	Sea trout
n	4	104	106	16	53	50
% zero values						
n_ha_Fe_insev	50	3,8	3,8	0	0	0
n_ha_Fe_omni	***100***	***96,2***	***99,1***	***100***	***94,3***	***92***
n_ha_Re_phyt	***100***	***90,4***	***92,5***	***100***	83	***90***
n_sp_Hab_b	25	44,2	49,1	81,3	1,9	34
n_sp_Hab_rh	25	2,9	1,9	81,3	0	0
n_sp_Mi_long	***100***	***99***	***100***	***100***	11,3	22
n_sp_Mi_potad	25	74	73,6	6,3	30,2	54
perc_nha_Re_lith	0	0	0,9	0	0	0
perc_sp_Intol	0	3,8	3,8	0	0	0
perc_sp_Tol	***100***	***94,2***	***98,1***	***100***	***90,6***	***90***
S. trutta fario (n/ha)	50	8,7	5,7	100	88,7	82

4.5 Distribution of modelling metrics within calibration sites

The distribution of metrics for the calibration sites, within fish types, would give an indication of how wide the distribution might have been in more affected sites. Figure 12 shows that among the ten metrics, which were the outcome of the modelling approach, three were almost all zero values (being the three metrics which increased with impact), and one contained almost only 100% values (percentage of n per ha of lithophilic species) in the data from ecoregion 20 and 22. Furthermore, among the other six metrics there was only one (n per ha of insectivorous species) not based on the number of species. These facts strongly indicate that the modelling approach may give a lower precision in classification for ecoregions that generally have relatively few species caught at each site.

4.6 The connectivity issue

The metrics perc_sp_Mi_long, perc_nha_Mi_long, and perc_kgha_Mi_long were one of the few metrics that seemed useful for the fish type Sea trout. To show that long migrating species indicate the effects of human impact on natural connectivity is important from the point of view of the WFD. Impact on river connectivity is one of the major impacts found in ecoregions 20 and 22, where the major part of Sweden's electricity from hydropower is produced. The impact of Connectivity_segment was not well represented in the data from ecoregions 20 and 22. The variable Connectivity_river was responsible for almost all the variation in the combined Connectivity_multiscale. It is clear that for the fish types Salmon and Sea trout, especially the percentage variables (n per ha and biomass per ha) representing long migratory fish, decline with the degree of impact on river connectivity (Figure 13). For the Sea trout type, no long migratory fish is present when the degree of impact is greater than 2.

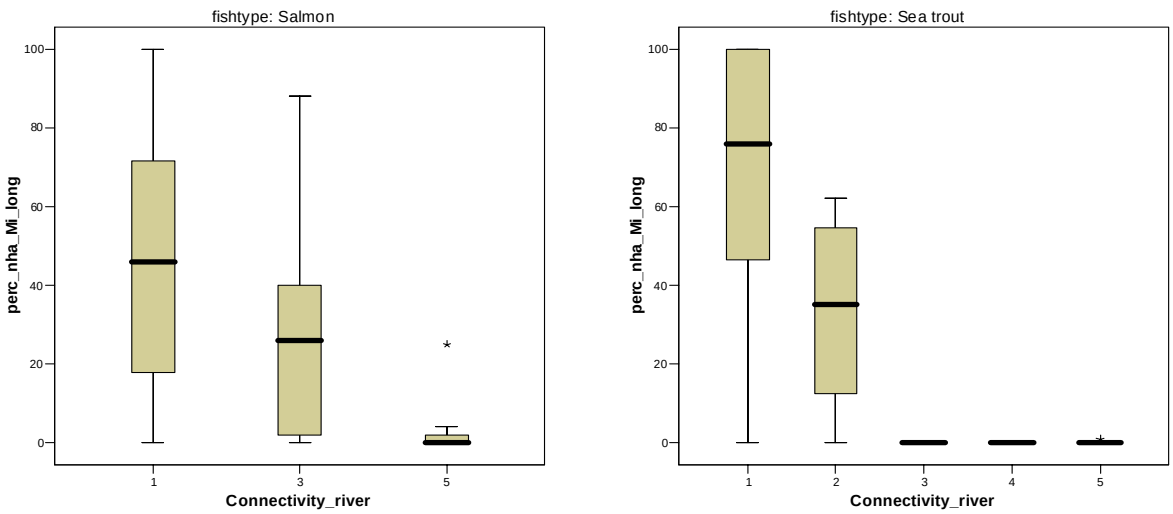


*Figure 13. Boxplots for percentage (n per ha) of long migrating species in two different fish types including long migratory species, for different impact classes of river connectivity. The data is from ecoregion 20 and 22. For fishtype Salmon, number of cases is for connectivity river impact class **1**, 55; **3**, 18; **5**, 7; for fishtype Sea trout, **1**, 59; **2**, 7; **3**, 7; **4**, 3; **5**, 6.*

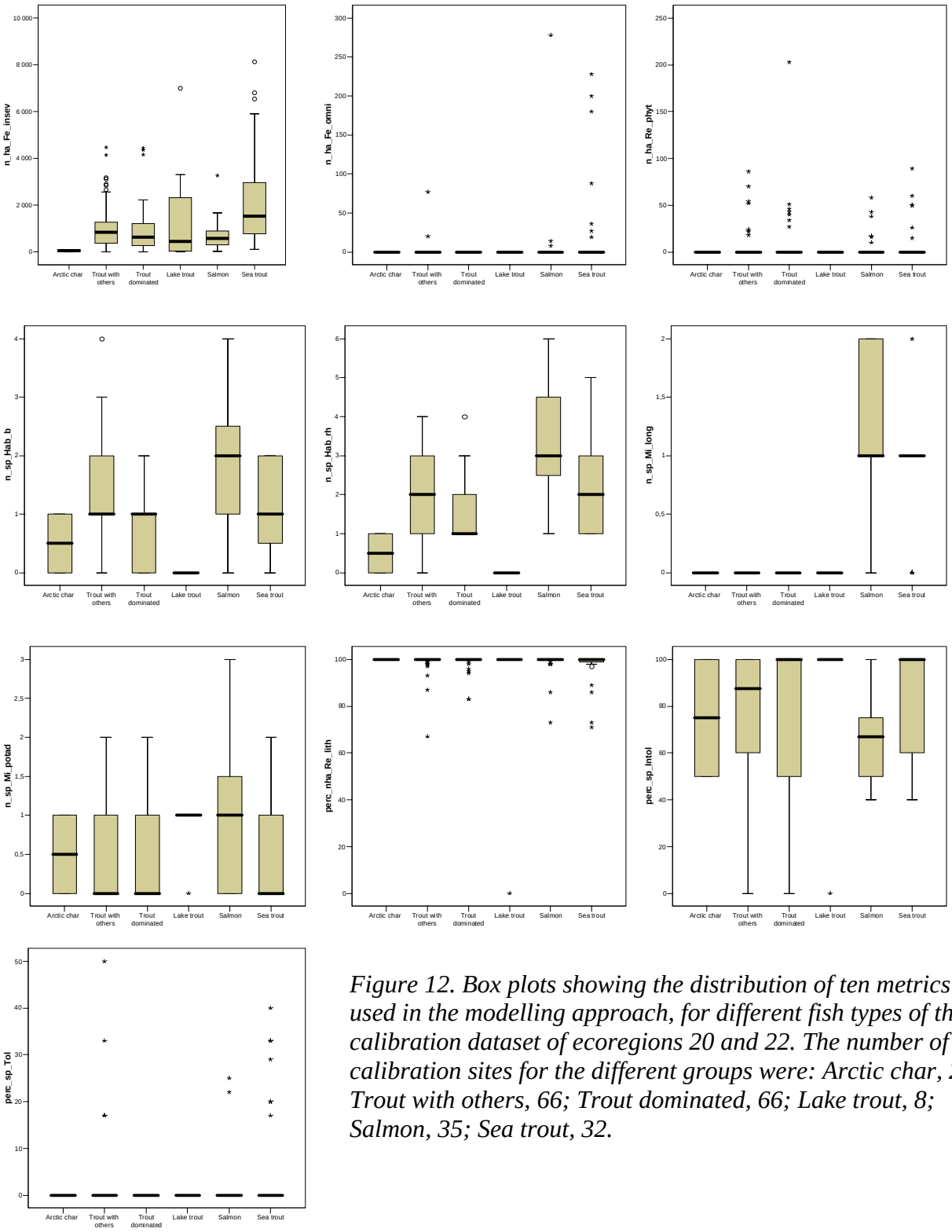


Figure 12. Box plots showing the distribution of ten metrics used in the modelling approach, for different fish types of the calibration dataset of ecoregions 20 and 22. The number of calibration sites for the different groups were: Arctic char, 2; Trout with others, 66; Trout dominated, 66; Lake trout, 8; Salmon, 35; Sea trout, 32.

5 References

Beier, U., Degerman, E., Bergquist, B. and Holmgren, K. 2002. DElimitation of Swedish Ichthyological REgions (DESIRE). Working paper, contribution to the project report ”Net of references and surface water stations”, National Board of Fisheries, Swedish Environmental Protection Agency and University of Agricultural Sciences.

Beier, U., Degerman, E., and Wirlöf, H. 2002. Data input to the ACCESS-2000© database FIDES (Fish Database of European Streams). The FAME project.

Degerman, E., Sers, B., Törnblom, J. & P. Angelstam, 2004. Large woody debris and brown trout in small forest streams – towards targets for assessment and management of riparian landscapes. In press: Ecological Bulletins 51, 29.

Böhmer, J. and Schmutz, S. 2003. Guidelines for spatially-based approach. Version 2.0, 7 July 2003. Working paper, the FAME project.