



Project no.: 0044096
Project acronym: EFI+

Improvement and spatial extension of the European Fish Index

Instrument: STREP
 Thematic Priority: Scientific Support to Policies (SSP) - POLICIES-1.5

D 3.3, Task 3.2. – Evaluation of existing European Fish Index

Due date of deliverable: 31.12.2007
 Actual submission date: 14.02.2008

Start date of project: 01.01.2007

Duration: 24 Month

Organisation name of lead contractor for this deliverable:
 CEMAGREF (HYAX),
 3275 Route de Cézanne, CS 40061,
 13182 AIX EN PROVENCE Cedex 5,
 FRANCE

Revision [draft 1]

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Workpackage 3.2:

Evaluation of Existing European Fish Index (M9-M12)



Authors: Pierre BADY* & Didier PONT*

* CEMAGREF (HYAX), 3275 Route de Cézanne, CS 40061, 13182 AIX EN PROVENCE Cedex 5, FRANCE

Created: 08/11/2007

Revision: 09/02/2008

Version: 3.1

Abstract:

This document contains some comments on the existing European Fish Index (EFI). Our purpose is to complete the substantial study proposed by Quataert et al. (2007) on performance evaluation of the index to correctly classify river sites at the European level and propose some necessary adaptations for the development of the news index (EFI+).

In first section, we show that shape patterns of rarefaction curves fluctuate in function of sites and metrics. Consequently, we can't only propose empirical rules to reduce the potential bias induced by the heterogeneous sampling effort. Best alternative should consist to select the sites with individual number superior to 100 or 150. However, for the development of EFI+ and after our first evaluation of database, an objective of 50/100 individuals appears to be more realistic to conserve an adequate sample size of calibration dataset.

In second section, our investigation highlights the link between the metric values and regional classification proposed by Reyjol et al. (2007) and that metric values appear depend to river size. Accordingly, we systematically recommend the integration or/and test of biogeographical effect and river size for each individual metrics in the future developments.

In third section, we illustrate that estimation of EFI error is a complex process and show that the multimetric approach doesn't favour the computation of prediction error and confidence interval. However, the computation of confidence interval can be envisaged for single metric with non-parametric procedure (Efron & Tibshirani 1993, Davison & Hinkley 1997).

In the last section, the comparisons of metrics modelling with the main statistical methods (generalized linear model, generalized additive model, partial least squares regression and neural networks) show that the modelling approaches provide similar results. High fitting is generally associated with low prediction capacity. An efficient procedure must be a compromise between model complexity and prediction error (Hastie et al. 2001).

Keywords:

EFI, index variability, sampling effort, biogeographical effect, predictive error, model comparisons, sample size

Contents:

1- Introduction.....	4
1.1- Dataset definition.....	4
1.2- Metric definition.....	6
2- Sampling Effort.....	6
2.1- Motivation.....	6
2.2- Methods.....	7
2.2.1- Data.....	7
2.2.2- Outcome definition.....	8
2.2.3- Simulation procedure.....	8
2.3- Results.....	10
2.4- Discussion.....	14
3- Geographical Effect.....	15
3.1- Motivation.....	15
3.2- Methods.....	16
3.2.1- Data.....	16
3.2.2- Statistical Analyses.....	17
3.3- Results.....	18
3.4- Discussion.....	21
4- Error associated with EFI.....	22
4.1- Motivation.....	22
4.2- Material and Methods.....	22
4.3- Results.....	23
4.4- Discussion.....	25
5- Comparison of metrics modelling.....	26
5.1- Motivation.....	26
5.2- Material and methods.....	27
5.2.1- Dataset.....	27
5.2.2- Outcome definition.....	27
5.2.3- Statistical Analyses.....	27
5.3- Results.....	29
5.3.1- Continuous data.....	29
5.3.2- Count data.....	30
5.3.3- Proportion data.....	30
5.4- Discussion.....	31
6- References.....	34
Annexes A: Description of the dataset used in the section 2.....	39

1 - Introduction

The establishing of the limits associated with assessments tools is an essential step for the users as well as the future developments. In FAME project (Schmutz et al. 2007, <http://fame.boku.ac.at/>), first evaluations of the existing EFI were mainly based on the global index. In this way, Quataert et al. (2004, 2007) proposed a very substantial exploration of performance of European Fish Index (Pont et al. 2004, 2006 and 2007) in terms of the detection capacity (sensitivity and specificity, impacted and non impacted predictive value) and consistency. The authors stressed on the difficulties to establish correct measures of pressures and appropriate calibration dataset. However, the authors shown that the discriminative capacity of EFI between reference and disturbed sites was good. The specificity and sensitivity were close to 80%. The separate analyses by country, river-group by component did not major inconsistencies (Quataert et al. 2004). They observed that the relation with existing national or regional methods is relatively poor, except in France.

The objectives of our study consist in completing the checking list of existing European Fish Index (EFI). Firstly, we propose to examine of the relationship between the sampling effort and the values of the metrics. Indeed, several authors (Cao et al. 2002, Hughes et al. 2002, Hughes & Herlihy 2007, Reynolds et al. 2003) showed that the low sampling effort could bias the estimation of the functional structure of fish assemblage. Secondly, we show how metrics values can be affected by the geographical regions (e.g. Reyjol et al. 2007) and by river size. Thirdly, we explore the variability structure of the EFI in function of the pressure index based on hydromorphological and chemical conditions. The last section presents a performance comparison of several modelling approaches of the metric values. Moreover, we also present the difficulties to compute an estimation of error interval associated with EFI values. For each section, we will propose some recommendations for the future development of new European Fish Index (EFI+).

1.1- Dataset definition

The whole of works proposed in this document are based on the initial dataset exploited in the FAME project, which contained 5238 sites among 1849 European rivers (Figure 1, Beier et al. 2007). The sampling sites are spatially distributed among 11 European countries (without the Greece data): Austria (581), Belgium (956), France (804), Germany (156),

Lithuania (219), Poland (118), Portugal (103), Spain (247), Sweden (417), The Netherlands (654), United Kingdom (983).

The FIDES database contains information about the fish assemblage, pressure level and environmental condition such as slope, altitude, and temperature for each site. This enables the computation of predicted values for each functional metrics (Pont et al. 2006, 2007). The pressure levels are based on four groups of pressure established to describe the status of the levels of nutrients, toxic, morphological and hydrological impacts (specific pressure). These variables were classically coded in 5 classes: (1) high, (2) good, (3) moderate, (4) poor and (5) bad (more information in <http://fame.boku.ac.at/>). Concerning the definition for the reference sites (quasi-reference), we use the definition proposed in FAME project. Reference sites are considered as minimally impacted sites (i.e. sites assigned to the class 1 or 2 of the 4 pressure classes). The reference dataset also contains 1611 reference sites: Austria (320/581), Belgium (45/956), France (405/804), Germany (13/156), Lithuania (97/219), Poland (60/118), Portugal (39/103), Spain (112/247), Sweden (280/417), The Netherlands (0/654), United Kingdom (240/983).

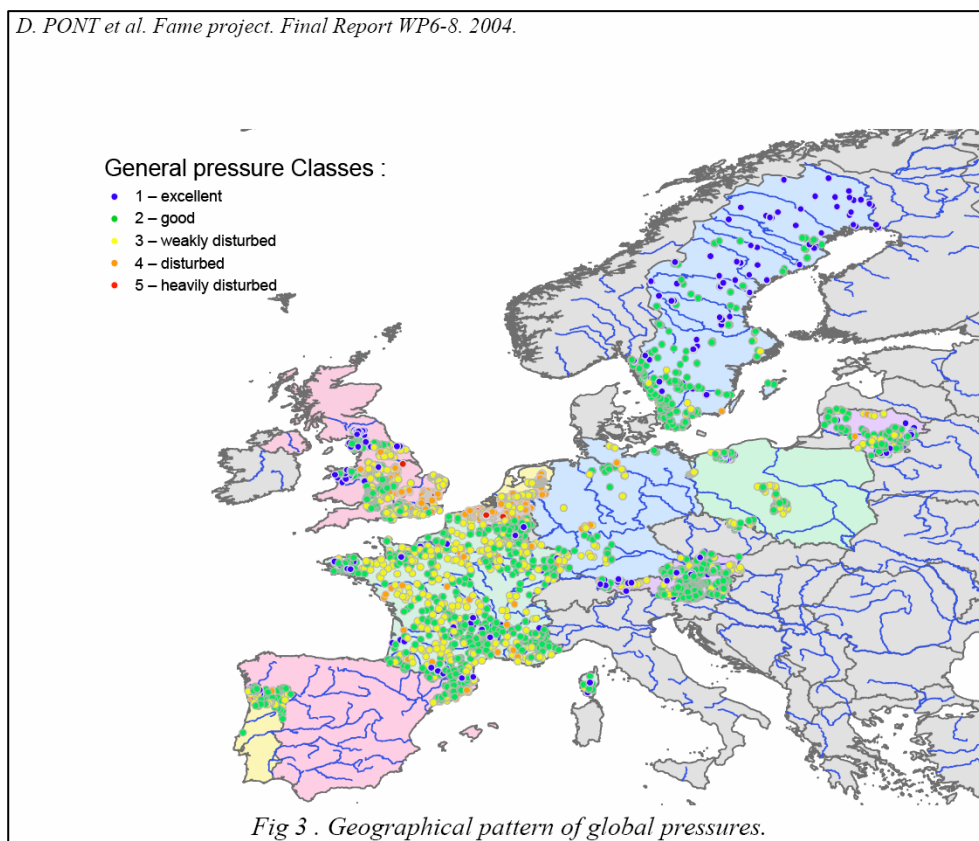


Figure 1. Localisations of the sampling sites from FIDES database (<http://fame.boku.ac.at/>).

The evaluation must be considered internal because it is based on the same dataset used to the construct the fish index. As a result, it involves potential bias for error rate and risks

underestimating (Quataert et al. 2007). However, the choice of this dataset was constrained by the availability of EFI+ database, which was in construction (year 2007).

1.2- Metric definition

The calculation of the European Fish Index involves the use of a list of 10 potential metrics reflecting different aspects of fish assemblage integrity (i.e. tolerance guilds, habitat guilds, trophic guilds, see table 1), taxonomic richness and individual abundance (Pont et al. 2006, 2007). More details on the choice of the metrics are available in the documentation of FAME project (<http://fame.boku.ac.at/>) and in Pont et al. (2006, 2007).

Table 1. Description of the metrics used to compute the European Fish Index (Pont et al, 2006).

code	type	guild	description
Ni_INSE	density	insectivorous	Density of insectivorous species
Ni_OMNI	density	omnivorous	Density of omnivorous species
Ni_PHYT	density	phytophilous	Density of phytophilous species
Ns_BENT	number	benthic	Number of benthic species
Ns_RHEO	number	rheophilic	Number of rheophilic species
Ns_LONG	number	long migratory	Number of long migratory species
Ns_POTA	number	potamodrous	Number of potamodrous species
perNi_LITH	percentange	lithophilous	Relative Abundance of lithophilous species
perNs_INT	percentange	intolerant	Relative number of intolerant species
perNs_TOLE	percentange	tolerant	Relative number of tolerant species

2- Sampling Effort

2.1- Motivation

As all indices of biotic integrity, the estimation of EFI requires an accurate evaluation of fish distribution, assemblage structure and assessment of the stream biotic integrity (e.g. Cao et al. 2002, Hughes et al. 2002, Hughes & Herlihy 2007, Maret & Herlihy 2007, Reynolds et al. 2003). The computation of EFI directly depends to specie number (richness) and individual abundance. For example, the estimation of relative abundance of lithophilous species is linked to the species abundance. In the same manner, the relative number of intolerant species is based on the measure of richness. Therefore, the accuracy of the measure of richness or individual abundance influences the accuracy of metric values as well as EFI values. Indeed, the effects of sampling effort on the functional structure of species assemblage have been largely documented in the literature (May 1981, Palmer 1990, Moreno & Halfpeter 2001, Bremmer et al. 2003, Diaz & Cabido 2001, Petchey & Gaston 2001, Magurran 1988). For

example, these works showed that the sampling effort such as the individual number, sampling area could affect the accuracy of the richness estimation. As a result, if the sampling effort is not substantial, there is a strong risk to underestimate the richness, which involves a bias in the sampling of fish assemblage.

Presently, concerning the European Fish index, the authors chose to consider sites with individual number superior to 30 to limit the potential bias in the estimation of the functional structure of fish assemblage in the calibration dataset (Fame documentation). With regard to the literature, the assumption of the authors appear justify, but they did not propose a clear analysis of the relationship between the accuracy of metric/EFI values and the sampling effort. Consequently, the objective of this section is to determine how the sampling effort (e.g. number of individual) can modify the evaluation of the functional metric values. A good accuracy of the measures as species richness, abundances, etc. are essential to obtain an acceptable estimation of metrics values. Hence, we propose to use fish assemblage data from 12 European rivers from FAME database to build rarefaction curves and to assess the number of caught fish required to estimate (i.e. closeness to the maximal value) the metrics values linked to the richness.

2.2- Methods

2.2.1- Data

To evaluate the potential effect of the sampling effort on metric estimation, we select 12 sites from FIDES database (<http://fame.boku.ac.at/>). The observed numbers of individuals of the sites are superior to 500 individuals and their main environmental characteristics are described in the table 2 and in the annexe A.

Table 2. Description of the sampling site. The terms Richness and Abundance correspond to species number and individual number respectively.

River name	country	code	FIDES code	Slope (m/km)	Temperature (°C)	Size of basin (km ²)	Richness	Abundance
Tirolerbach	AT	S1	ATTIBA5000	7	9	<10	4	537
Mare	FR	S2	FR06340022	8	12	<10	6	658
FINKENBACH	DE	S3	DEHE3011	16	9.5	<100	3	889
Selves	FR	S4	FR05121001	20	8.75	<100	3	719
Tees	UK	S5	UK20382	5.22	8.5	<100	12	4264
Sinaize	FR	S6	FR04180035	2	11.5	<1000	21	774
Minija	LT	S7	LTM10001	2.94	5.9	<1000	7	665
Piesting	AT	S8	ATPIES6900	4	9.3	<1000	3	1426
Doubs	FR	S9	FR06390206	0.6	10.5	<10000	26	1107
Mur	AT	S10	ATMUR_857	4.3	5.1	<10000	3	812
Thaya	AT	S11	ATTHAY155	0.3	8.9	>10000	29	3794
Mur	AT	S12	ATMUR_206	1.4	9	>10000	6	1389

2.2.2- Outcome definition

We focus only on the count metrics linked to the richness, because your simulation procedure is not adapted to the metric based on abundance or density of individuals. The procedure propose in the next section can't integrate the variation of fished area. Indeed, we suppose that the fished area is fixed for a given sample. This assumption can be strongly dubious in the cases of the density estimation. The metrics is described in the table 1 of the section 1.2.

2.2.3- Simulation procedure

To evaluate the effect of the sampling effort, we proposed to simulate an increase in the number of caught fish and to compute rarefaction curves by a Monte-Carlo procedure (see e.g. Magurran 1988, Cao et al. 2002). Species accumulation and rarefaction curves are frequently used to estimate the expected number of species, to estimate minimum sample size required to characterize the communities, to minimize the sampling costs and as a means of data standardisation prior to among-sites comparisons (Gotelli & Colwell 2001).

The procedure is relatively simple. For a given sampling effort, we build theoretical assemblage based on multinomial distribution and randomly draw without replacement (50 repetitions). To rebuild the Fish assemblage, we postulate that the proportion of each species follows a binomial distribution. We define p_j as the proportion of the individual of the j^{th} species in the assemblage. The estimation of p_j was given by:

$$\hat{p}_j = \frac{n_j}{n} \quad (1)$$

where n_j and n correspond to the number of individual for the j^{th} species and the total number of the individual. Under the above hypothesis, we can suppose that the assemblage based on k species y_1, y_2, \dots, y_k follows multinomial distribution. The estimation of the probabilities for multinomial distribution (p_1, p_2, \dots, p_k) is defined as:

$$(\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k) = \left(\frac{n_1}{n}, \frac{n_2}{n}, \dots, \frac{n_k}{n} \right) \quad (2)$$

Function of probability is obtained as follow:

$$P(Y_1 = y_1, Y_2 = y_2, \dots, Y_k = y_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} \cdot p_2^{n_2} \dots p_k^{n_k} \quad (3)$$

where $\sum_{i=1}^k n_i = n$ and $\sum_{i=1}^k p_i = 1$. We can compute means and variances as follow:

$$E[Y] = (n \cdot p_1, n \cdot p_2, \dots, n \cdot p_3) \tag{4}$$

$$V[Y] = \begin{cases} M_{ij} = -n \cdot p_i \cdot p_j, i \neq j \\ M_{ii} = n \cdot p_i \cdot (1 - p_i) \end{cases} \tag{5}$$

The resulting 50 simulated values for a given metric are averaged to provide a mean and variance (see e.g. Cao et al. 2001). According to the simulation procedure, we can't evaluate the accuracy of the metric based on the abundance: density of insectivorous species, density of omnivorous species and density of phytophilous species, relative Abundance of lithophilous species (see the section called *Outcome definition*).

Figure 2. Simulation procedure.

```

Algorithm:

For i in 0 to 400 individuals {
    1) Repeat 50 times {
        1.1) Computation of the theoretical assemblage based on
            multinomial distribution for ith individuals
        2.1) Metrics Computation
    }
    2) Computation of the mean and the variance
}
    
```

To evaluate the effect of the sampling effort in function of the number of caught fish, we use the tools proposed in Hellmann & Fowler (1999). These authors defined three measures called Bias, Precision and accuracy. The bias evaluates the distance between the true values and the simulated values for a given number of individuals (Saporta 2006, Manly 2007):

$$bias(\hat{S}) = \left[\left(\sum_{i=1}^n \hat{S}_i \right) / n \right] - S \tag{6}$$

The precision corresponds to the Monte Carlo variance obtained for each estimator. It is defined as follow:

$$var(\hat{S}) = \left[\sum_{i=1}^n \hat{S}_i^2 - \left(\sum_{i=1}^n \hat{S}_i \right) / n \right] / (n - 1) \tag{7}$$

These values enable the computation of confidence intervals for each estimator (Manly 2007). All the routines necessary for calculating EFI, performing simulations and computing

accuracy and modelling precision are implemented in the R software (R Development Core Team 2007).

2.3- Results

The QQplot representation illustrates the good coherence between the theoretical distribution and the observed distribution (Figure 2). This means that the choice of multinomial distribution to describe the organisation of the fish assemblage is an acceptable assumption.

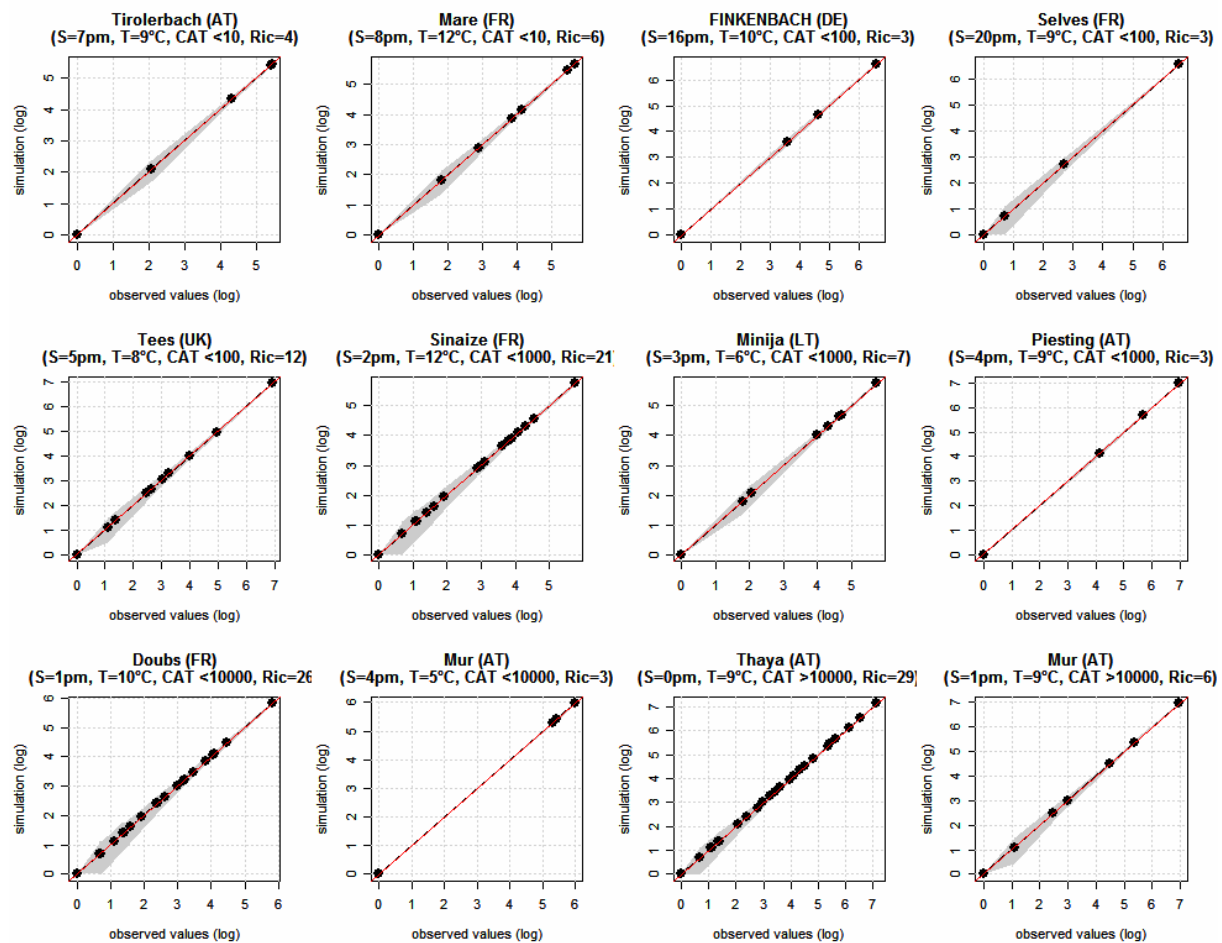


Figure 3. QQplot representation of simulated (1000 repetitions) and observed values after log-transformation. The red line correspond to the line $y = x$. The grey polygons give the confidence interval.

The bias of an estimator is the difference between the expected value of the estimator and the true species richness value. Positive and Negative bias correspond to the overestimation and underestimation of the true species richness respectively. The overestimation is only observed for the metrics based on percentage (e.g. Relative number of intolerant species, Relative number of tolerant species, Table 3). We observe that the relationship between the expected values and the true values are very complex. From figures 4, 5 and 6, it easily follows that the

shape patterns of rarefaction curves fluctuate in function of the sites and the metrics. However, we note that the sites (s6, s9 and s11) with high richness and high entropy (Shannon's indices available in annexe A1) are generally characterised by more important bias and strong Monte-Carlo variance (figures 5 and 6).

Table 3. Ratios between the simulated and observed values of the metrics associated with richness for 30, 50 and 100 simulated individuals. The bold values correspond to the values with $0.80 \leq \left(\frac{\sum_{i=1}^n \hat{S}_i}{n} \right) / S \leq 1.20$. The terms x was used to indicate the values of observed metric equal to zero.

River name	code	Ind. Nb	Ns_BENT	Ns_RHEO	Ns_LONG	Ns_POTA	perNS_INT0	perNs_TOLE
Tirolerbach	S1	30	0.38	0.82	x	0.99	0.93	x
		50	0.58	0.85	x	1.00	0.95	x
		100	0.74	0.96	x	1.00	0.96	x
Mare	S2	30	0.53	0.77	x	0.56	1.32	x
		50	0.67	0.84	x	0.63	1.27	x
		100	0.83	0.92	x	0.81	1.11	x
FINKENBACH	S3	30	0.85	0.87	x	0.66	1.00	x
		50	0.93	0.96	x	0.88	1.00	x
		100	0.99	0.99	x	1.00	1.00	x
Selves	S4	30	0.42	0.52	x	0.04	2.22	x
		50	0.62	0.53	x	0.06	1.97	x
		100	0.90	0.67	x	0.14	1.70	x
Tees	S5	30	0.77	0.48	0.52	0.23	0.67	1.31
		50	0.75	0.52	0.59	0.36	0.66	1.17
		100	0.77	0.64	0.61	0.53	0.75	1.19
Sinaize	S6	30	0.38	0.61	0.48	0.57	0.54	0.81
		50	0.48	0.73	0.82	0.64	0.58	0.85
		100	0.61	0.80	0.98	0.78	0.67	0.84
Minija	S7	30	0.92	0.75	x	0.12	0.62	x
		50	0.99	0.78	x	0.32	0.67	x
		100	1.00	0.84	x	0.48	0.79	x
Piesting	S8	30	x	0.93	x	0.85	1.13	x
		50	x	0.95	x	0.96	1.05	x
		100	x	1.00	x	1.00	1.02	x
Doubs	S9	30	0.19	0.31	0.00	0.27	0.56	1.36
		50	0.25	0.34	0.06	0.42	0.68	1.31
		100	0.39	0.46	0.10	0.51	0.84	1.25
Mur	S10	30	x	1.00	x	1.00	1.00	x
		50	x	1.00	x	1.00	1.00	x
		100	x	1.00	x	1.00	1.00	x
Thaya	S11	30	0.36	0.36	0.00	0.46	0.74	1.32
		50	0.40	0.44	0.02	0.57	0.58	1.20
		100	0.48	0.53	0.04	0.68	0.75	1.05
Mur	S12	30	0.12	0.68	x	0.76	1.69	x
		50	0.40	0.76	x	0.83	1.53	x
		100	0.54	0.88	x	0.89	1.35	x

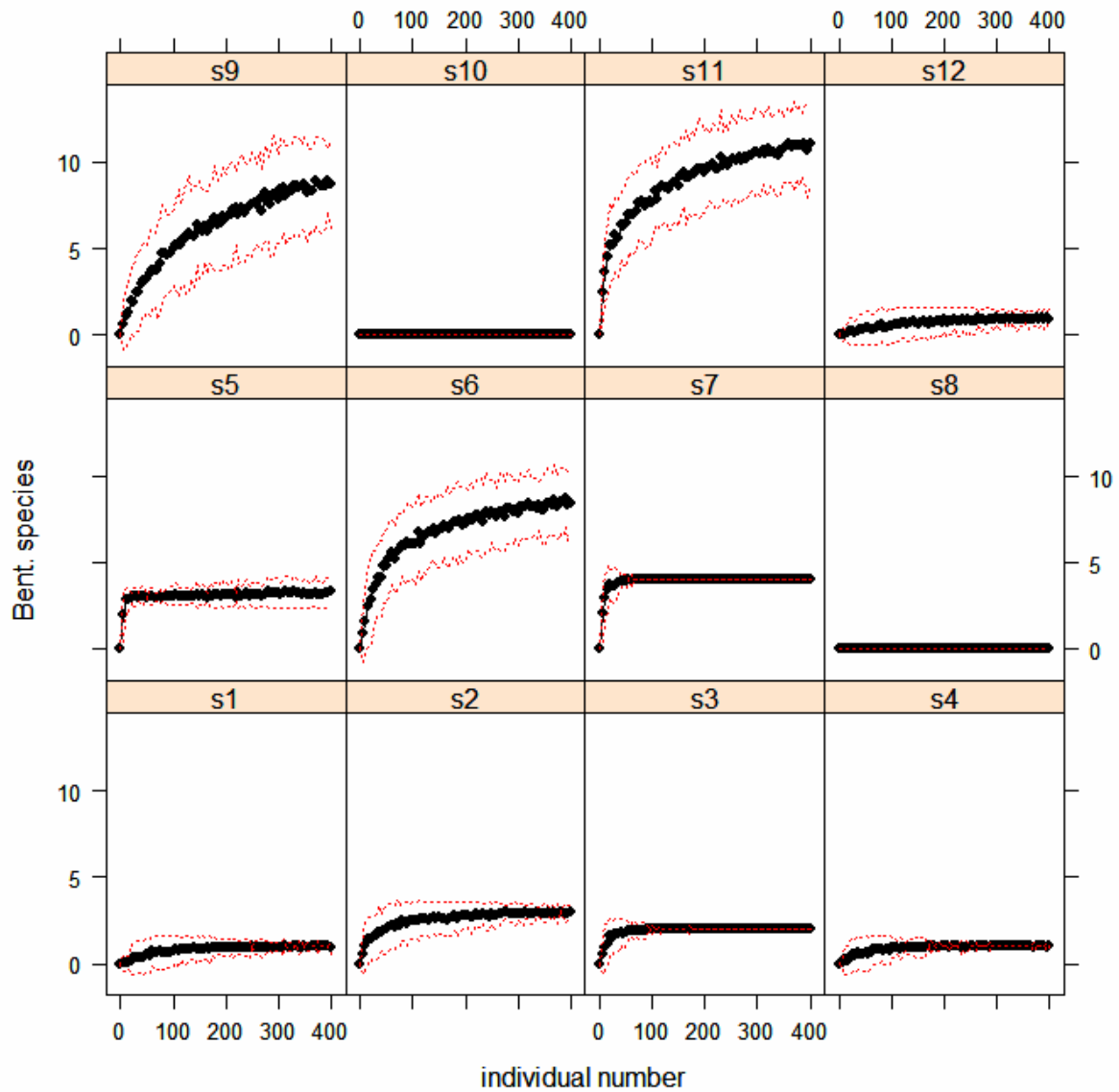


Figure 4. Rarefaction curves of the percentage of intolerant species (y). The black curves correspond to the simulated number of potamodrous species in function of the individual number ($\hat{S} = f(\text{individual number})$). The red dotted lines represent 95% confidence interval ($\hat{S} \pm \varepsilon_{1-0.05} \sqrt{\text{var}(\hat{S})}$) where $\hat{S} = \sum_{i=1}^n \hat{S}_i / n$).

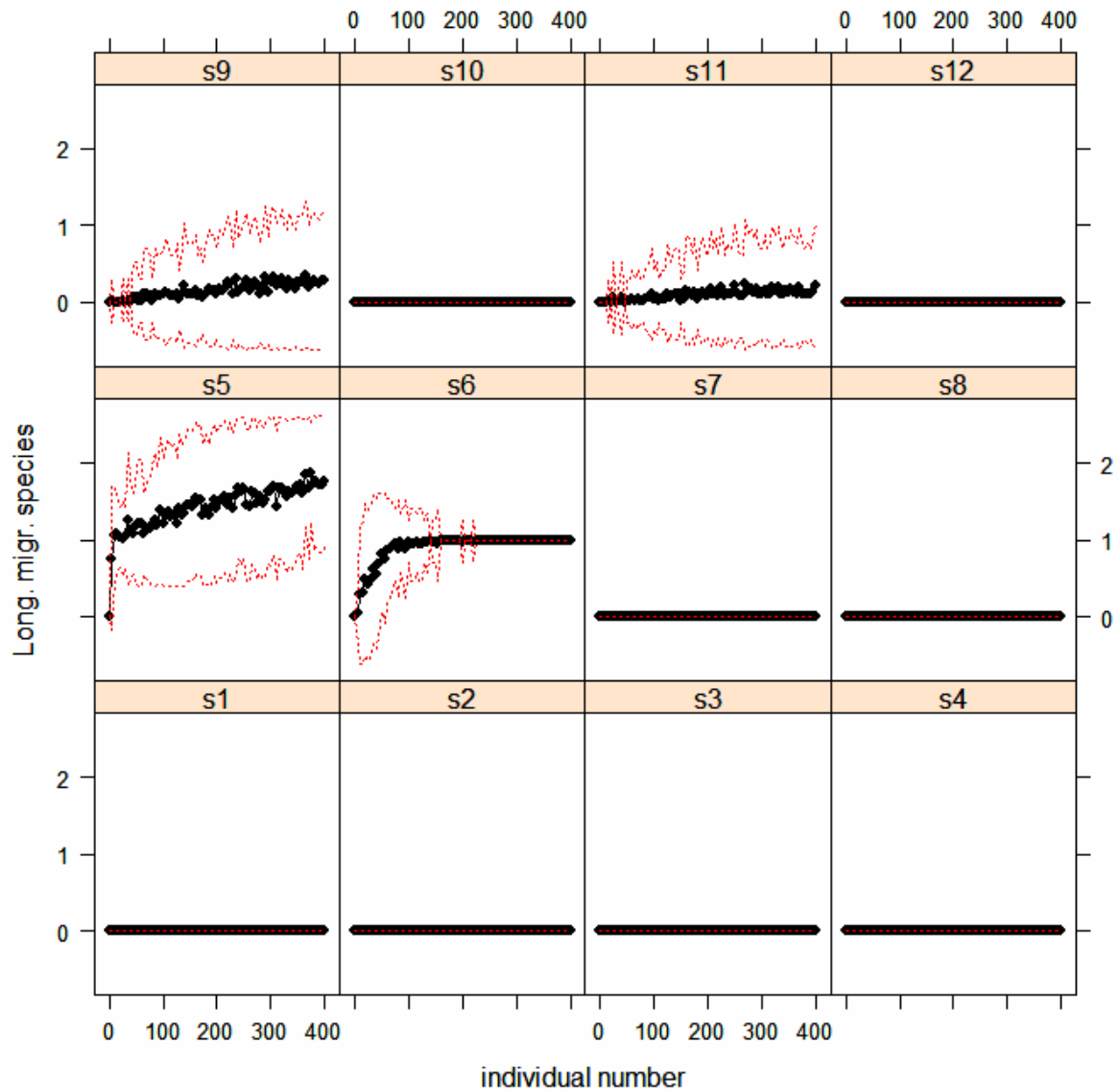


Figure 5. Rarefaction curves of the percentage of intolerant species (y). The black curves correspond to the simulated number of long migratory species in function of the individual number ($\hat{S} = f(\text{individual number})$). The red dotted lines represent 95% confidence interval ($\hat{S} \pm \varepsilon_{1-0.05} \sqrt{\text{var}(\hat{S})}$ where $\hat{S} = \sum_{i=1}^n \hat{S}_i / n$).

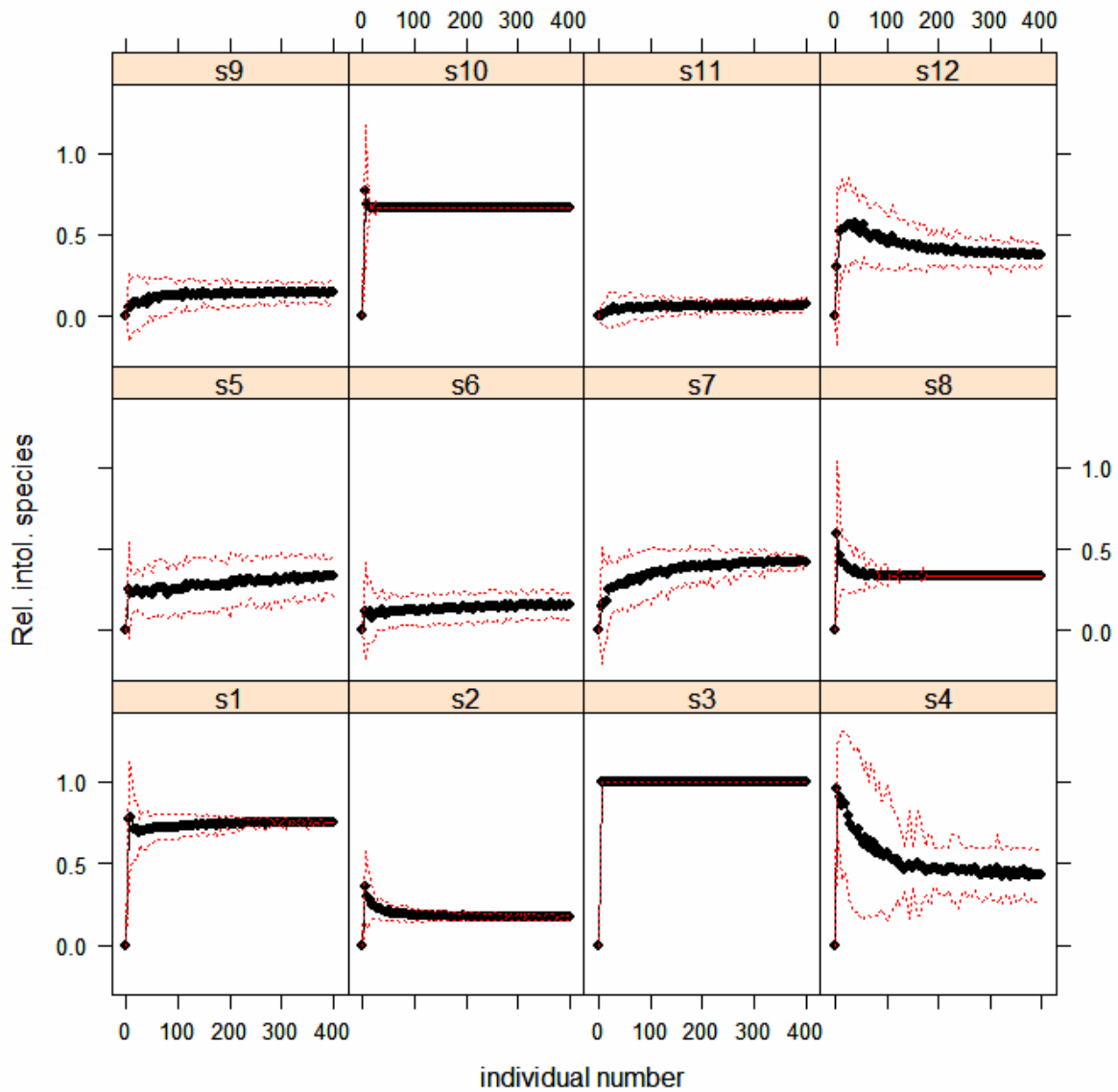


Figure 6. Rarefaction curves of the percentage of intolerant species (y). The black curves correspond to the simulated percentage of intolerant species in function of the individual number ($\hat{S} = f(\text{individual number})$). The red dotted lines represent 95% confidence interval ($\hat{S} \pm \varepsilon_{1-0.05} \sqrt{\text{var}(\hat{S})}$) where $\hat{S} = \sum_{i=1}^n \hat{S}_i / n$.

2.4- Discussion

The observations of rarefaction curves for each sites suggests that there is no common shape associated with the relationship between number of individuals and values of metrics. The relations between the expected values and the true values are complex. In fact, we observe that the accuracy of the estimators appears dependent upon the properties of the assemblage including total number of species, the number of rare species and the pattern of species distribution. Gotelli & Graves (1996) and Gotelli & Colwell (2001) suggested that greater evenness of the relative abundance distribution is combining with the steeper

rarefaction curve. These results are only partially congruent with their propositions. In our study, the sites with high richness, high entropy and high evenness (e.g. sites s6, s9, s11) present low potential estimation of the true values of the metrics (annexe A1, figures 4 and 5). Yet, we observe that the sites with Low richness and low entropy are associated with low potential estimation of the true values of the metrics. The site 4 contains one very dominant specie and two secondary others species with low probability of capture. If the number of individual is small, it's therefore very difficult to catch the rare species (Magurran 1988). In the both case, we need a lot of individuals to obtain a correct estimation of the assemblage structure.

The definition of the calibration dataset must include sampling effort information, but the definition of the efficient rules is a complex process. Sampling effort is a complex a combination of correlated variables such as the sampling area (see Magurran 1988, Cao et al. 2001, Cao et al. 2002), individual number (Thompson et al. 2003). For these reasons, we can't only propose empirical rules to reduce the potential bias induced by the heterogeneous sampling effort among the records. Hence, the best alternative should consist to select the sites with individual number superior to 100 or 150 (see figures 4, 5 and 6). For the development of EFI+ and according our first evaluation, an objective of 50/100 individuals appears to be more realistic to conserve an adequate sample size of calibration dataset.

3- Geographical Effect

3.1- Motivation

Over the years, the reflections of the statistician, scientist and practitioner on the spatial patterns of data are increasing more and more in ecological study (Legendre et al. 2002, Schabenberger & Gotway 2005, Dormann et al. 2007). The analyses of spatial characteristic integrate several levels of observation such as autocorrelation, geographical cluster effect. For example, Legendre et al. (2002) define spatial dependence and spatial autocorrelation as follow:

1. Spatial dependence implies that the response variable is spatially structured because it depends upon explanatory variables (e.g. temperature, precipitation) that are themselves spatially structured by their own generating process.
2. In Spatial autocorrelation, the value of response variable (y) for a given site is assumed to results from some dynamic process within the variable y itself. This refers

to the lack of independence among error components of field data, as a function of geographic distance among site

In this study, we focus only on the potential effect of spatial dependence on the metric values. This includes the exploration of the difference between Northern-Western with Central-Eastern and Southern (Mediterranean) Europe. Hence, to answer this question, we propose to test the effect of biogeographical regions proposed by Reyjol et al. (2007, figure7). These regions correspond approximately to grouping of river groups used in the computation of the EFI (Pont et al. 2006). This variable takes in account the difference between the regional specie pools among the European River and, indirectly, geographical localisation. The regionalisation gives a first approximation of the spatial effect on the metric value. We complete this section with the test of river size effect on the metric values.

3.2- Methods

3.2.1- Data

The dataset contained the 1611 references sites from FIDES database (see section 1.1, Beier et al. 2007). First, to test biogeographical effects, we use the variables “ichtyoregion” (BR region proposed by Reyjol et al. 2007, figure 7). Second, we examine relationship between metric values and river size effect described by the variable Catchments sizes. This variable is defined in five classes : <10 km², <100 km², <1000 km², <10000 km² and >10000 km². This variable is used in the computation of two metrics (relative Abundance of lithophilous species and number of potamodrous species). The selected outcomes (section 1.2) correspond to the score after the transformation in probability. The distributions of these values are statistically more exploitable (quasi-normal and without limited between 0 and 1).

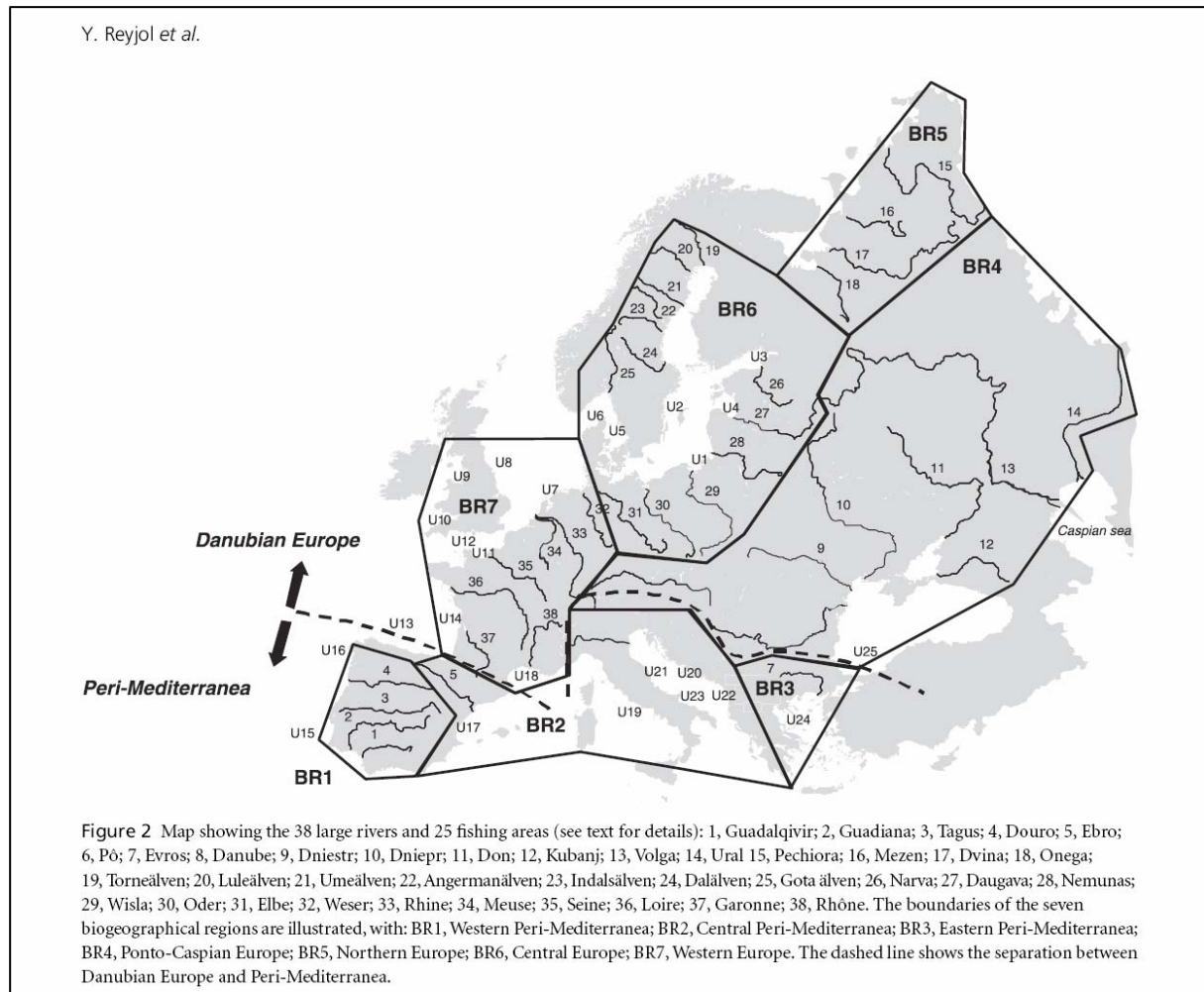


Figure 7. Biogeographical regions according to Reyjol et al (2007).

3.2.2- Statistical Analyses

The classical method to test the groups effects (e.g. regions, country and catchment size) is based on linear model, analyse of variance (ANOVA). An important characteristic of this method is that it requires the equality of variance across the groups (for example see Sokal & Rohlf 1969). Yet, Levene's tests (Fox 2002, Greene 2002) showed that this assumption was supported only for two metrics with the variable catchment size (Table 5). Then to limit the effect of the heteroskedasticity, we use Robust Huber-White-Sandwich procedure to estimate covariance matrix of the model (Zeileis 2004, Greene 2002). We modify the element $\hat{\Omega} = \text{diag}(\omega_1, \dots, \omega_n)$ in the computation of the variance of the parameters $\hat{\beta}(\text{var}[\hat{\beta}] = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \hat{\Omega} \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1})$ where $\omega_i = \frac{\hat{u}_i^2}{(1 - h_i)^2}$, \hat{u}_i correspond to the residuals between the estimated and observed data and $h_i = \mathbf{H}_{ii}$ is the diagonal of hat matrix

(MacKimmon & White 1985, Greene 2002, Zeileis 2004). The use of the new covariance matrix enables the correction of the values obtained by t-test associated with the model parameters and Wald test to compare nested models. Tests de WALD with robust estimation of the VCOV Matrices is based on the comparison between Null model (Model 0: without the group variable, e.g. biogeographical groups) and the complete model (Model 1: with group variable). This procedure is equivalent to ANOVA.

$$\text{Model 0: } y = \mu \tag{8}$$

$$\text{Model 1: } \begin{cases} y = \mu + \gamma_{region} \\ \text{or } y = \mu + \gamma_{catchment\ size} \end{cases} \tag{9}$$

In this study, the interactions are not considered because their constructions are not systematically possible. All the routines necessary for calculating metric values and EFI were implemented in the R software (R Development Core Team 2007). The linear models and estimation of the covariance matrix are performed with the packages **lmtest** (Zeileis & Hothorn 2002) and **sandwich** (Zeileis 2004).

3.3- Results

Levene’s tests show that the hypothesis of the equality of variance across groups is rejected in the majority of cases except for the variable Ns_BENT and Ns_RHEO (tables 4 and 5). Consequently, these results justify the use of the correction of covariance matrices by robust procedure to analyse the differences among means of reference sites between BR regions and catchments size.

Table 4. Summary of the Levene’s tests with the variable ‘BR’. The table contains coefficients, degree of freedom and F-values and additionally gives p-values.

Variable	Res df	df	F	P(>F)
Ni_INSE	1606	4	8.1633	< 0.0001
Ni_OMNI	1606	4	5.1212	0.0004
Ni_PHYT	1606	4	18.787	< 0.0001
Ns_BENT	1606	4	4.9486	0.0006
Ns_RHEO	1606	4	6.2627	< 0.0001
Ns_LONG	1606	4	88.55	< 0.0001
Ns_POTA	1606	4	4.077	0.0027
perNi_LITH	1606	4	6.5103	< 0.0001
perNS_INT0	1606	4	30.808	< 0.0001
perNs_TOLE	1606	4	18.685	< 0.0001

Table 5. Summary of the Levene's tests with the variable CAT. The table contains coefficients, degree of freedom and F-values and additionally gives p-values.

Variable	Res df	df	F	P(>F)
Ni_INSE	1606	4	16.665	< 0.0001
Ni_OMNI	1606	4	28.569	< 0.0001
Ni_PHYT	1606	4	37.209	< 0.0001
Ns_BENT	1606	4	0.8032	0.5231
Ns_RHEO	1606	4	1.138	0.3369
Ns_LONG	1606	4	6.7059	< 0.0001
Ns_POTA	1606	4	11.858	< 0.0001
perNi_LITH	1606	4	45.7	< 0.0001
perNS_INT0	1606	4	7.1748	< 0.0001
perNs_TOLE	1606	4	14.622	< 0.0001

Concerning the biogeographical variable (regions 'BR'), the null hypotheses of the Wald's test are systematically rejected at the 5% significance level for all the metrics (Table 6). These results suggest that the biogeographical region could be important variable to add to models. Nevertheless, we observe that the distribution of the metrics in proportion (persNs_LITH, perNs_INT0, perNs_TOLE) are very asymmetric and include many outliers (Figure 8).

Table 6. Summary of the Wald's tests for the variable 'BR' based on robust covariance matrix estimators. The table contains coefficients, degree of freedom and F-values and additionally gives p-values.

Variable	Res df	df	F	P(>F)
Ni_INSE	1606	4	2.9161	0.02033
Ni_OMNI	1606	4	4.5799	0.0011
Ni_PHYT	1606	4	4.6275	0.001
Ns_BENT	1606	4	39.701	< 0.0001
Ns_RHEO	1606	4	18.569	< 0.0001
Ns_LONG	1606	4	97.243	< 0.0001
Ns_POTA	1606	4	7.8975	<0.0001
perNi_LITH	1606	4	5.5865	0.0002
perNS_INT0	1606	4	10.789	< 0.0001
perNs_TOLE	1606	4	20.532	< 0.0001

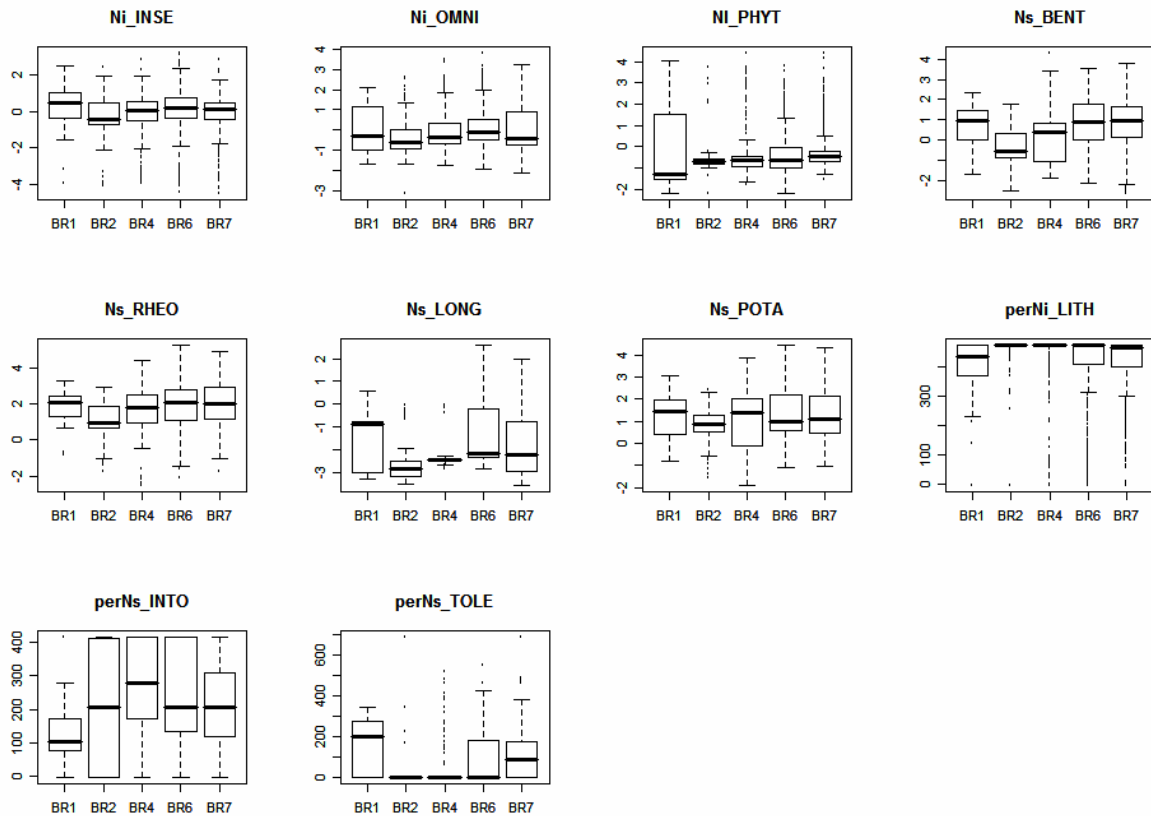


Figure 8. Boxplot representations of the scores each metric in function of the regions BR.

As with the biogeographical variables, the averages of metric values among the catchment size groups are significantly different (Table 7). These results confirm that the variable catchments size is important variable in the modelling of metric. For several metrics, the averages trend to rise with increase of the catchment size. However, despite the integration of this variable in models (Pont et al. 2006, 2007), the means among groups are significantly different for the metrics perNi_LITH and Ns_Pota. This could indicate some potential problems in the adjustment of these two models. For the metric “Ns_POTA”, the problem appears to be relatively important. Indeed, we clearly observe that the increase of metrics values is associated with the increase of catchments size (mean by catchments size = 0.97, 1.21, 1.52, 1.54, 1.85, Figure 9). Concerning the metric “perNi_LITH”, the distributions are very asymmetric and the intra-group variances are very different (Figure 9). In this extreme condition, the correction of the covariance matrix and Wald’s test are probably not efficiency.

Table 7. Summary of the Wald's tests for the variable "catchment size" based on robust covariance matrix estimators. The table contains coefficients, degree of freedom and F-values and additionally gives p-values.

Variable	Res df	df	F	P(>F)
Ni_INSE	1606	4	6.7404	< 0.0001
Ni_OMNI	1606	4	8.9909	< 0.0001
Ni_PHYT	1606	4	4.8811	0.0007
Ns_BENT	1606	4	50.694	< 0.0001
Ns_RHEO	1606	4	55.118	< 0.0001
Ns_LONG	1606	4	9.2055	< 0.0001
Ns_POTA	1606	4	13.344	< 0.0001
perNi_LITH	1606	4	31.867	< 0.0001
perNS_INT0	1606	4	61.166	< 0.0001
perNs_TOLE	1606	4	21.35	< 0.0001

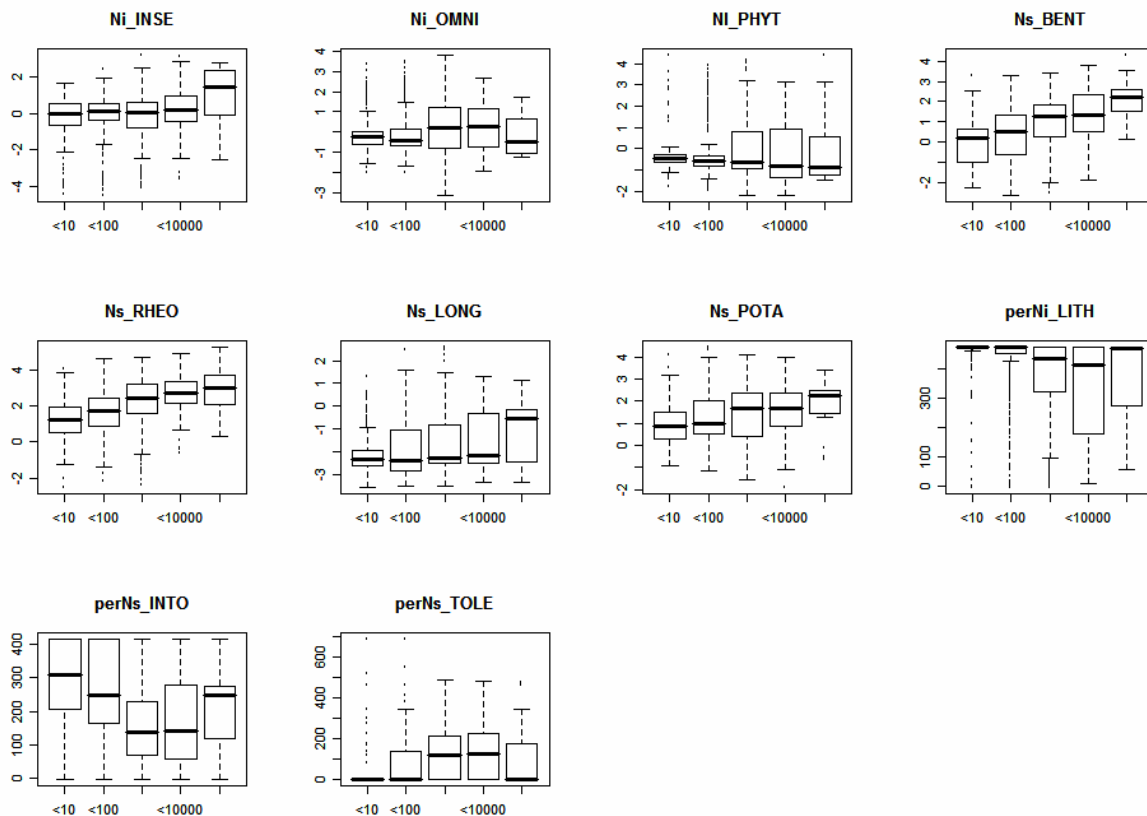


Figure 9. Boxplot representations of the scores each metric in function of catch size.

3.4- Discussion

The main result of this study concerns the significant effect of catchment size on the metric values (Table 7). We observe that individual scores are not independent to the river size. For example, the relative number of intolerant species decreases when catchment size increases. In contrast, the metrics based on number of rheophilic species, number of benthic species and number of potamodrous species rise with increase in the catchment size.

However, Pont et al. (2006, figure 3) observed that the distribution of the index score didn't appear to be affected by catchment area. The steps of metric aggregation and probability transformation tend to compensate this effect and to stabilize the average response of the score. This could be explains these results. As regards the biogeographical effect, we show that all metric are significant affected by our geographical variable. Distribution patterns of individual score are not independent to the localisation. These results are not totally surprising because the bioregion used in Reyjol et al. (2007) integrate complex spatial patterns: It's a mixture of countries/geographical effects and regional specie pools (partial corrected by the variable "river group" in EFI). To conclude, we recommend to integrate or/and to test systematically, for each individual metrics, the biogeographical effect and river size in the future development of EFI+.

4- Error associated with EFI

4.1- Motivation

In a decisional context, the evaluation and quantification of the error associated with the assessment tools are essential information for the end-user (e.g. Bardos 2002, Bernier et al. 2000). For this reason, the objective of this section is to present 1) how we can compute an error, confidence and/or prediction interval for the EFI values and 2) how we can estimate the variability associated with EFI in function of the global pressure index (Pont et al. 2006, 2007).

4.2- Material and Methods

We use the EFI values of 5238 sites from FIDES database (description in section 1.1, Beier et al. 2007) and four pressures variables to describe the status of the levels of nutrients, toxic, morphological and hydrological impacts (specific pressure). Pressure variables are classically coded in 5 classes: (1) high, (2) good, (3) moderate, (4) poor and (5) bad (more information in the FAME documentation, <http://fame.boku.ac.at/>). A global pressure index is obtained by the sum of these scores, which ranged between 4 and 20 (see Pont et al 2006).

To assess the variability and approximate error associated with EFI, linear model is used to explore the relationship between global pressure index and the index values (see Pont et al. 2006). This model is defined as:

$$y_{efi} = \alpha + \beta \cdot x_{pressure} \quad (10)$$

Where α , β , y_{efi} and $x_{pressure}$ correspond to the intercept, the slope, EFI values and values of global pressure index respectively. Accordingly, Confidence and prediction intervals were obtained as follow:

$$\hat{y}_x \pm \hat{\sigma}(\hat{y}_x) t_{1-\alpha, n-p} \tag{11}$$

Where \hat{y}_x , $\hat{\sigma}(\hat{y}_x)$ and $t_{1-\alpha, n-p}$ correspond to the estimated values, standard deviation associated with estimated values and the theoretical values from Student distribution (α = error type I ; $n - p$ = degree of freedom). The both intervals differ from the computation of the variance (more mathematical detail in Greene 2002 and Saporta 2006). The Estimation of the variance of the estimated values \hat{y}_x for confidence interval is:

$$\sigma^2(\hat{y}_x) = \sigma^2 \mathbf{X}'_x (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_x \text{ where } \hat{\sigma}^2 = \frac{1}{n - p - 1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{12}$$

and the estimation of the variance of the estimated values \hat{y}_x for confidence prediction is:

$$\sigma^2(\hat{y}_x) = \sigma^2 \left(1 + \mathbf{X}'_x (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_x \right) \tag{13}$$

The calculations are completed by the estimation of the first and last quantiles and quartiles by bootstrap procedure (Efron & Tibshirani 1993, Davison & Hinkley 1997). Yet, we remind that the computation of the classical confidence and prediction intervals did not take into account the error associated with individual metric models. Consequently, we tend to underestimate the confidence and prediction intervals. For lack of a better solution, we shall use these values to approximate error associated with EFI.

4.3- Results

The adjustment to EFI values on the global pressure index is satisfactory with a R-squared equal to 0.4748 and F-statistic highly significant (F=4734, d =1 and 5236, P<0.0001). QQ-plot representation show that Pearson residuals followed quasi-normal distribution (Figure 10). Graphic of Residuals against Fitted values show that there is a slight problem of heteroskedasticity. The error appears to be less important for the small fitted. The parameters of the regression model (table 3) also are highly significant (P< 0.0001). Nevertheless, the significant status of the parameters is not surprising with regard to the large number of observations.

Table 8. Summary statistics of the fitted linear model $efi=f(\text{pressure index})$. The table gives coefficients, standard errors, etc. and additionally gives p-values.

parameters	Estimate	Std. Error	t-value	P (> t)
(Intercept)	0.842	0.005	169.6	< 0.0001***
Pressure Index	-0.034	0.005	-68.8	< 0.0001***

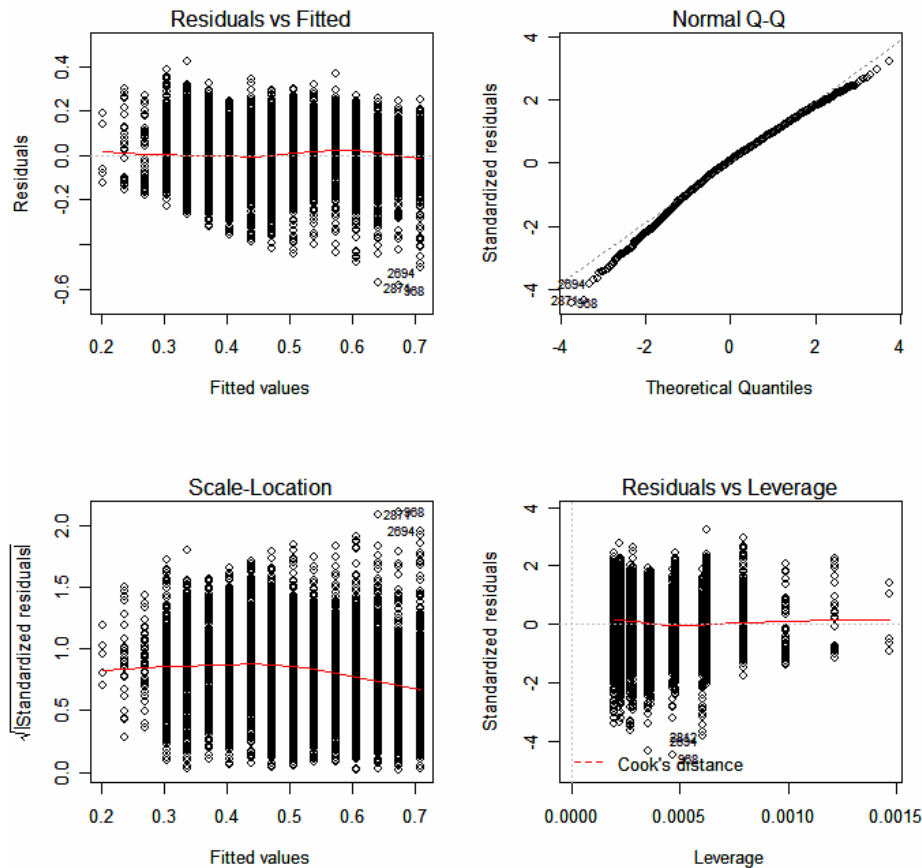


Figure 10. Graphical representation associated with the regression to EFI values on global pressure index. The first graphic corresponds to residuals in function of the fitted values. The second shows the QQ-plot representation of standardised residuals against normal theoretical quantiles. The third correspond to the representation of the square root of standardized residuals against the fitted values. The last graphic plots the leverage against the standardised residuals to detect the potential influent points.

We observe a significant relationship among EFI values and the pressure level. The index decrease is associated with the increase of the pressure level. However, the illustration of quantiles values in Figure 11 (black dotted line) shows that the variability is high. The strong value of crude classical estimation of the error (grey polygon, Figure 11) suggests that the

predicted error is strong. The average approximation error is in the order of ± 0.26 units. This corresponds more or less to one class¹.

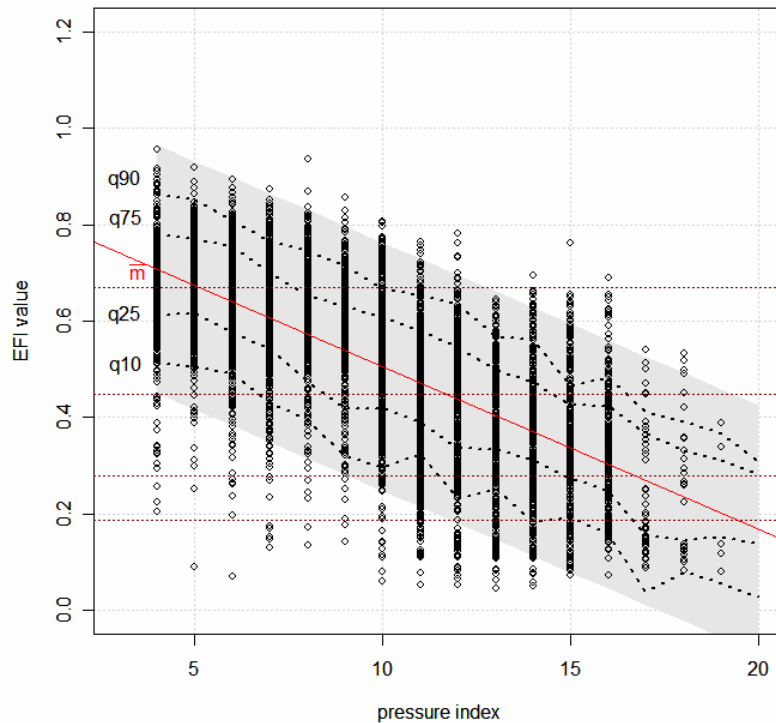


Figure 11. Relationship between EFI values and Global pressure index (black circles). Terms q_{10} , q_{25} , q_{75} , q_{90} and \bar{m} correspond to the first quantile, first quartile, third quartile, last quantile and predicted values respectively (black dotted lines and red line). Grey polygon corresponds to the predictive interval associated with the linear model. The dark red dotted lines correspond to the thresholds associated with the five water quality classes (see FAME Project).

4.4- Discussion

Our study confirms that the estimation of EFI error is a complex process and show that the multimetric approach doesn't favour the computation of prediction error and confidence interval. Indeed, a large part of the metric is statistically not independent that involves serious difficulties to estimate the variance of EFI. The multi-metric index exploits the redundancy among metrics to stabilize the final score by their aggregation, but this involves that the

¹ Thresholds associated with the five water quality classes (FAME Project):

Status	Score	Description
1 - HIGH	[0.669 - 1.00]	Less than 1 % of reference sites are classified as disturbed sites
2 - GOOD	[0.449 - 0.669]	Impairment threshold.
3 - MODERATE	[0.279 - 0.449]	The risk for a reference site to be classified as a impaired site is 20%
4 - POOR	[0.187 - 0.279]	
5 - BAD	[0.00 - 0.187]	Less than 1 % of disturbed sites are classified as reference sites

metrics are often strong correlation. In addition, the statistical distribution of EFI values is unknown, which poses problems for the estimation of variance parameter. There are no simple solutions to compute a **true** error associated with EFI. Yet, the computation of confidence interval can be envisaged for single metric with non-parametric procedure (Efron & Tibshirani 1993, Davison & Hinkley 1997). Presently, we are testing some procedure based on Kernel Density Estimation (KDE, Scott 1992, Venables & Ripley 2002) to reduce the statistical difficulties.

The crude approximation based on linear model is biased and it underestimated the real error rate. However, we observed that the variability of EFI values were relatively strong. The results support the previous studies proposed by Quataert et al. (2007) and Pont et al. (2006). Indeed, we observe that the discriminative capacity between quasi-reference sites and disturbed sites was good, but the separation between the intermediate classes is less efficient. The main reason of the low performance appears to be the definition of the global pressure index. In the present version, we suppose that the pressures have additive effect on the fish assemblage. However, the accuracy of pressure index appears to be not sufficient to integrate the complex association between pressures. Consequently, the difference between the intensity levels and effect of the pressure could explain the high variability for a given values of global pressure index. In this context, the development of new measures of the pressure and an evaluation of the combining of pressure effect seems an acute point for the next index.

5- Comparison of metrics modelling

5.1- Motivation

A very dense bibliography shows that the comparison of the modelling strategies was a recurrent biological question (e.g. Austin 2002, Austin et al. 2006, Manel et al. 1999, Potts & Elith 2006). The biologist is often lost in view of the variety of statistical methods (e.g. (generalized) linear model, neural networks, etc.). Hence, to clarify the situation in the context of fish index development, this study proposes performance comparison of several methods to model the metrics. Modelling approaches are chose for their compatibility² with the construction of the European Fish Index in line with Pont et al. (2006, 2007). As a result, we don't use classification procedure such as partitioning by K-means or hierarchical divisive clustering (e.g. Hartigan 1975, Legendre & Legendre 1998, Gordon 1999) and tree regression approach (e.g. Bardos 2001, Faraway 2006, Venables & Ripley 1999). Explicative variables

² Metric estimation is based on a linear combination of environmental variables.

and model selections differ in function of the methods, that's why the results of this study must be interpreted with the greatest caution.

5.2- Material and methods

5.2.1- Dataset

The dataset is based on the references sites from FIDES database (<http://fame.boku.ac.at/>, Beier et al. 2007, Schmutz et al. 2007). It contains 1611 references sites: Austria (320/581), Belgium (45/956), France (405/804), Germany (13/156), Lithuania (97/219), Poland (60/118), Portugal (39/103), Spain (112/247), Sweden (280/417), The Netherlands (0/654), United Kingdom (240/983). In our study, we select only the observations using in Pont et al. (2006, 2007): the training and test sample contain 1000 and 304 **independent** reference sites respectively. Concerning the definition for the reference sites (quasi-reference), we use the definition proposed in FAME project. Reference sites are considered as minimally impacted sites (i.e. sites assigned to the class 1 or 2, see the FAME documentation).

5.2.2- Outcome definition

We mainly focus on three different types of variables: continuous, count and binary/proportion variables. As results, each variable was associated with particular distribution (e.g. Binomial, Poisson, etc...) and/or particular transformation (e.g. log-transformation for the count data). We select one variable by type (table 1, section 1.2): density of insectivorous species, number of benthic species and relative number of intolerant species.

5.2.3- Statistical Analyses

We propose to compare the results obtained by the four main family of model: generalized linear model, generalized additive model, Partial Least Squares Regression and Neural Network.

In Generalized Linear Model (GLM), each outcome Y is assumed to be generated from a distribution function in the exponential family and the mean μ of the distribution depends on the independent variables X . The GLM consists of three elements: A distribution function, from the exponential family, a linear predictor ($g(y) = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon$) and link function g ($E(Y) = \mu = g^{-1}(\eta)$). The parameters estimation is based on the maximum

likelihood (Nelder Wedderburn 1972, McCullagh & Nelder 1989). The linear model is a particular case defined by the *Gaussian* family and *identity* link.

Generalized Additive Model (GAM) replaces the parameter terms $\beta_i x_i$ of multiple regression with functions $f(x_i)$: $g(y) = \beta_0 + f(x_1) + \dots + f(x_m) + \varepsilon$. The functions $f(x_i)$ are arbitrary and often nonparametric (b-spline, polynomial, loess, fractional polynome, etc...). A stepwise procedure can be adapted to the GAM. More mathematical details are available in Hastie & Tibshirani (1989).

Partial Least Squares Regression (PLS) is a compromise between the multiple regression of Y on \mathbf{X} and the PCA of \mathbf{X} (Tenenhaus 1998). Wold et al. (1983) proposed a complete review on this analysis. The PLS algorithm is defined as follow:

1. Search k orthogonales component (T_k) with good explicative power and correlated to Y .
2. Regression of Y on T_k .
3. Expression of regression in fonction of \mathbf{X} .

A generalized version of the PLS (gpls, Marx 1996) can be adapted to model categorical (ex. binomial and multinomial distribution) and count data (Poisson distribution).

Neural Network (NNET) is learning method. It's based on *nonlinear weighted sum* and minimized the distance between the observed and expected data (ex. mean of quadratic error). More details were available in Abdi (1994), Ripley (1996) and Venables & Ripley (1999).

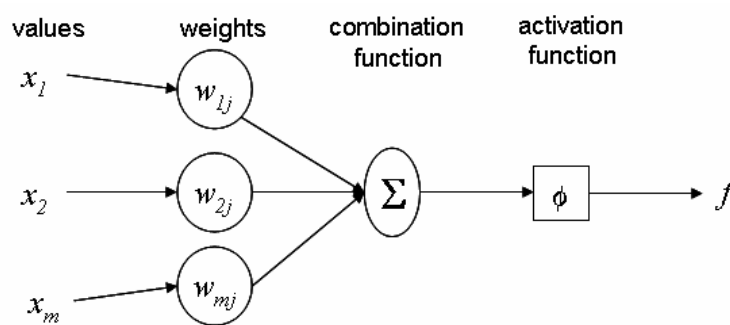


Figure 12. Structure of one unit in the hidden layer.

To compare models, we propose to use the tools using by Potts & Elith (2006) in a similar study:

- **RMSE** (Root Mean Square Error) provides indication on the divergence between the observed and predicted values.
- Pearson correlation (**CORP**) provides indication of similarities between the ranks of the observed and predicted values.

- **AVER** (Residuals distribution) provides indication on the goodness of fit.

For more detail concerning the statistical methods, we encourage people to consult the appropriate bibliography. All the routines necessary for computing models are implemented in the R software (R Development Core Team 2007).

5.3- Results

5.3.1- Continuous data

Continuous data are generally characterised by Gaussian distribution. We often use transformation as such square root, logarithmic to stabilize the variability of the outcome. To model continuous data, we propose to use four different approaches (table 9): Generalized linear model with stepwise procedure, general additive model with stepwise approach, partial least square regression with optimization of the component number and neural networks with optimization of the unit number in the hidden layer.

Table 9. Approaches to model continuous outcome.

type	f(y)	distribution	link	variable selection	number of component	units in the hidden layer	f(x)
glm	log(y+1)	gaussian	identity	stepwise	-	-	poly
gam	log(y+1)	gaussian	identity	stepwise	-	-	spline
pls	log(y+1)	gaussian	identity	-	1-7 (?)	-	-
nnet	log(y+1)	-	-	-	-	2-4 (?)	-

The performance analyses for the modelling of density of insectivorous or invertivorous Species (INSE) are presented in the table 10. We observe that the results were relatively similar among the procedures (see the values of RMSE and Pearson correlation for training and test samples).

Table 10. Performance analyses of the models. The terms RMSE, AVER, CORP and RMSE ratio correspond to Root Mean Square Error, averaged residuals, Pearson correlation and RMSE ratio between training and test samples respectively.

model type	transfor.	family	link	Calibration Data			Validation Data			
				RMSE	AVER	CORP	RMSE	AVER	CORP	RMSE ratio
glm1	log(y+1)	gaussian	identity	1.72	0.00	0.59	1.80	-0.16	0.59	0.96
gam1	log(y+1)	gaussian	identity	1.72	0.00	0.59	1.98	-0.19	0.58	0.87
pls	log(y+1)	gaussian	identity	1.86	0.00	0.49	1.93	-0.21	0.51	0.96
nnet	log(y+1)	-	-	1.63	0.00	0.64	1.87	-0.13	0.56	0.88

5.3.2- Count data

Count data can be characterised by Poisson distribution with logarithmic link or Gaussian distribution with identity link and logarithmic to stabilize the outcome variability. In this study, we evaluate six different approaches to model count data (table 11): Generalized linear model with stepwise procedure for Gaussian and Poisson distribution, general additive model with stepwise approach for Gaussian and Poisson distribution, partial least square regression with optimization of the component number and neural networks with optimization of the unit number in the hidden layer.

Table 11. Approaches to model count outcome.

type	f(y)	distribution	link	variable selection	number of component	units in the hidden layer	f(x)
glm	log(y+1)	gaussian	identity	stepwise	-	-	poly
gam	log(y+1)	gaussian	identity	stepwise	-	-	spline
glm	y	poisson	log	stepwise	-	-	poly
gam	y	poisson	log	stepwise	-	-	spline
pls	log(y+1)	gaussian	identity	-	1-7 (?)	-	-
nnet	log(y+1)	-	-	-	-	2-4 (?)	-

The performance analyses for the modelling of Number of Benthic Species (BENT) are presented in the table 12. As expected, we observe that the results are relatively similar among the procedures among GAM, GLM and PLS. The NNET appear to be more efficient than the other models in the prediction of the training sample (RMSE = 0.37) but their performance in the prediction of the test sample are comparable to the other model. Yet, very low differences among RMSEs complicate their interpretation.

Table 12. Performance analyses of the models. The terms RMSE, AVER, CORP and RMSE ratio correspond to Root Mean Square Error, averaged residuals, Pearson correlation and RMSE ratio between training and test samples respectively.

model type	transfor.	family	link	Calibration Data			Validation Data			
				RMSE	AVER	CORP	RMSE	AVER	CORP	RMSE ratio
glm1	log(y+1)	gaussian	identity	0.41	0.00	0.75	0.44	-0.01	0.69	0.93
glm2	y	poisson	log	0.42	0.08	0.75	0.46	0.08	0.69	0.91
gam1	log(y+1)	gaussian	identity	0.41	0.00	0.75	0.45	0.06	0.68	0.91
gam2	y	poisson	log	0.41	0.08	0.76	0.47	0.14	0.69	0.87
pls	log(y+1)	gaussian	identity	0.42	0.00	0.73	0.46	0.00	0.66	0.91
nnet	log(y+1)	-	-	0.37	0.00	0.80	0.47	0.01	0.65	0.79

5.3.3- Proportion data

Proportion data is associated with Binomial distribution with logit link or Gaussian distribution with identity link and square root of *arcsin*. We examine three different

approaches to model count data (table 13): Generalized linear model with stepwise procedure for Binomial distribution, general additive model with stepwise approach for Binomial and neural networks with optimization of the unit number in the hidden layer with outcome transformation by *arcsin* square root.

Table 13. Approaches to model binary or proportional outcome.

type	f(y)	distribution	link	variable selection	number of component	units in the hidden layer	f(x)
glm	y	binomial	logit	stepwise	-	-	poly
gam	y	binomial	logit	stepwise	-	-	spline
nnet	y	-	-	-	-	2-4 (?)	-

The performance analyses for the modelling of proportion of intolerant Species (INTOL) are presented in the table 14. Once again, we observe that the results are similar among the procedures among GAM, GLM. The NNET appear to be less efficient than other models in the prediction of the training sample. However, differences among RMSE are relatively low.

Table 14. Performance analyses of the models. The terms RMSE, AVER, CORP and RMSE ratio correspond to Root Mean Square Error, averaged residuals, Pearson correlation and RMSE ratio between training and test samples respectively.

model type	transfor.	family	link	Calibration Data			Validation Data			
				RMSE	AVER	CORP	RMSE	AVER	CORP	RMSE ratio
glm1	y	binomial	logit	0.24	-0.06	0.69	0.27	-0.08	0.66	0.92
gam1	y	binomial	logit	0.24	-0.05	0.69	0.27	-0.09	0.66	0.90
nnet	asin(sqrt(p))	-	-	0.27	0.07	0.65	0.29	0.08	0.62	0.93

5.4- Discussion

In this study, we focus on the raw predictive power of the analysis. Therefore, the estimation of the goodness of fit includes other criteria. For example, the linear models (LM) and generalized linear models (GLM) require a complete diagnostic of the residuals (e.g. autocorrelation, heteroskedasticity, etc...), examination of the influent observations (e.g. leverage effect with the diagonal elements of the hat matrix, influence measures with Cook's distance) and potential multicollinearity effect (e.g. variance inflated factor, VIF). Robust and/or resampling methods give some solutions to solve a part of problems associated with linear and generalized linear model (Davidson & Hinkley 1997, Fox 2002, Greene 2002, Marazzi 1993, Zeileis 2004). In Partial least squares regression, the choice of the component number and the representation of the canonical weights can be added to the above list. For neural networks, the choice of unit number in the hidden layer was a key step of the analysis.

Table 15. Comparison of the approaches to model the functional metrics. Additional terms were defined as follow: ??? with bootstrap procedure, * difficulties to chose of the adequate component number and ** difficulties to chose the adequate number of units in the hidden layer

characteristic	lm	glm	gam	pls	nnet
continuous	good	gaussian	gaussian	good	good
binary	poor	binomial	binomial	poor	good
proportion	with transfo	binomial	binomial	with transfo	with transfo
count	with transfo	poisson	poisson	with transfo	with transfo
easy use	good	fair	fair	fair	poor
Robutness of outliers	fair	fair	fair	fair	poor
Robutness of missing values	poor	poor	poor	fair	poor
natrual handling of data of "mixed" type	good	good	good	good	poor
monotones transformation of inputs	fair	fair	good	fair	poor
extraction of linear combinations	good	good	poor	good	pair
Interpretability	good	good	fair	fair	poor
predictive power	fair	fair	poor	fair*	good**
transportable in Excel	good	good	poor	good	poor
confidence interval	good	fair	fair	good	???
prediction interval	good	fair	fair	good	???

A comparison of the approaches is proposed in the table 15. The table shows that the choice of the modeling strategy is a complex step. High fitting is generally associated with low prediction capacity. An efficient procedure is a compromise between model complexity and prediction error (Hastie et al. 2001, Figure 2).

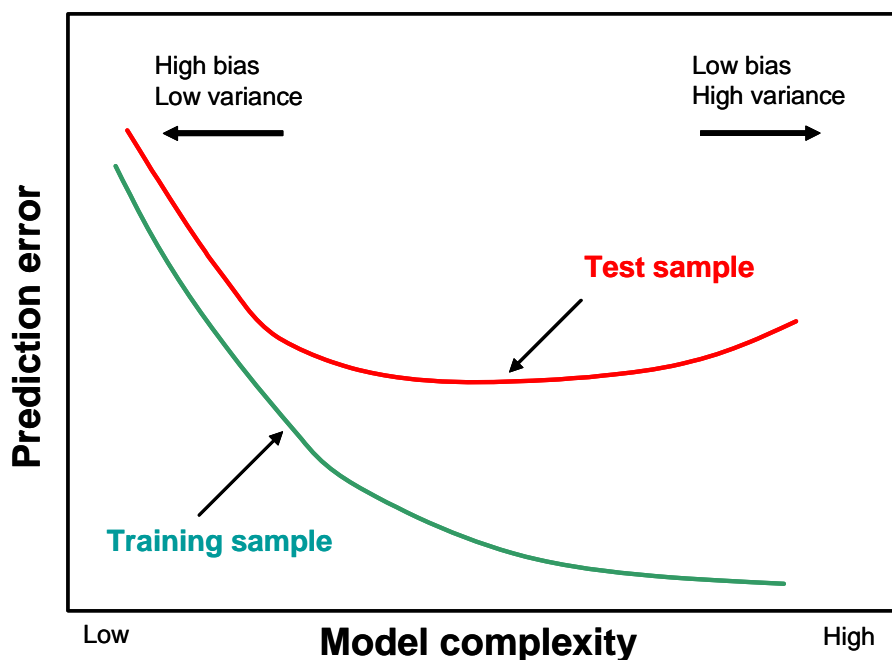


Figure 13. Relationship between the learning and prediction power of model in function of the complexity (Hastie et al. 2001 and Saporta 2006).

Neural network (NNET) are regularly using in ecological study since last some years. However, this method is generally good for prediction but bad for understanding (Faraway

2006, Saporta 2006) and the weights computed by the NNET are almost uninterpretable. Moreover, NNET introduce complex interactions that often do not reflect the reality (Faraway 2006). Venables and Ripley (1999) warn the users about the problem of overfitting and recommend the use of cross-validation and others resampling methods. In fact, without careful control, the NNET can easily overfit the data resulting in overoptimistic predictions (Faraway 2006, Saporta 2006). If the training sample is representative, the machine learning is very efficient. Nevertheless, generally, their performances are relatively lower than the others models in regions of little and no data. In the case of EFI development, it's clearly a limitation. The training sample is not representative of population³ of the European River. For example, there are not sites characterized by Strahler order superior to six in the calibration dataset (Figure 3).

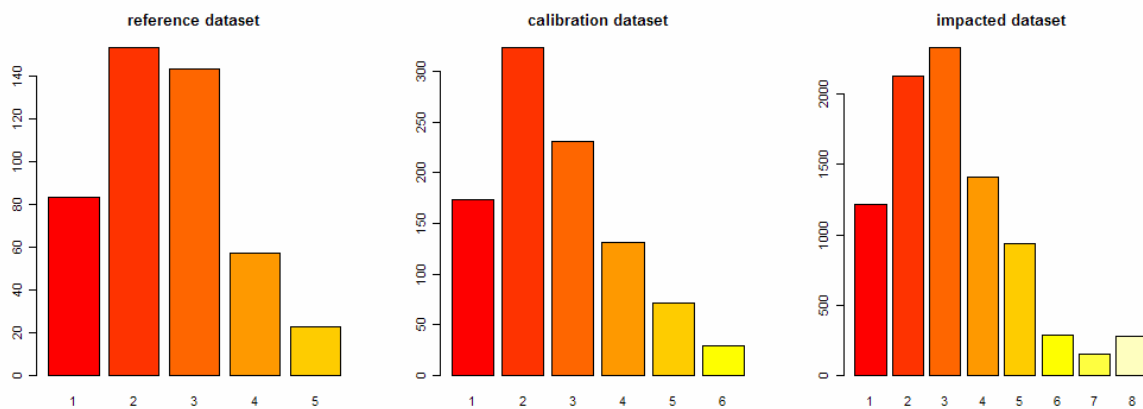


Figure 14. Distributions of sites in function of the Strahler order for reference, calibration and impacted datasets.

To conclude this section, we remind some statistical principles on modeling: all model are wrong, model comparison was complex processes, quality of the data is often more important than type of model and, according to McCullagh & Nelder (1989) and Austin et al. (2006), there is no absolute model.

³ “In statistics, a statistical population is a set of entities concerning which statistical inferences are to be drawn, often based on a random sample taken from the population”. **Wikipedia**, *The Free Encyclopedia*, Statistical population, http://en.wikipedia.org/wiki/Statistical_population.

6- References

- Abdi, H. (1994) A neural network primer. *Journal of Biological Systems*, **2**, 247-281.
- Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D. & Luoto, M. (2006) Evaluation of statistical models used for predicting plant species distributions: Role of artificial data and theory, *Ecological Modelling*, **199** (2), 197-216.
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecological Modelling*, **157**, 101-118.
- Bardos, M. (2001) *Analyse discriminante. Application au risque et scoring financier*. Dunod, Paris.
- Bernier, J., Parent, E. & Boreux, J.-J. (2000) *Statistique pour l'environnement traitement bayésien de l'incertitude* Editions TEC & DOC (Lavoisier), Paris.
- Beier, U., Degerman, E., Melcher, A., Rogers, C. & Wirlof, H. (2007) Processes of collating a European fisheries database to meet the objectives of the European Union Water Framework Directive. *Fisheries Management and Ecology*, **14**(6), 407-16.
- Bremner, J., Rogers, S.I. & Frid, C.L.J. (2003) Assessing functional diversity in marine benthic ecosystems: a comparison of approaches. *Marine Ecology Progress Series*, **254**, 11-25.
- Cao, Y., Larsen, D.P. & Hughes, R.M. (2001) Evaluating sampling sufficiency in fish assemblage surveys: a similarity-based approach. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 1782-93.
- Cao, Y., Larsen, D.P., Hughes, R.M., Angermeier, P.L. & Patton, T.M. (2002a) Sampling effort affects multivariate comparisons of stream assemblages. *Journal of the North American Benthological Society*, **21**(4), 701-14.
- Cao, Y., Williams, D.D. & Larsen, D.P. (2002b) Comparaison of ecological communities: the problem of sample representativeness. *Ecological Monographs*, **72**(1), 41-56.
- Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap Methods and Their Application*. Cambridge University Press.
- Diaz, S. & Cabido, M. (2001) Vive la difference: plant functional diversity matters to ecosystem processes. *Trends in Ecology & Evolution*, **16**(11), 646-55.
- Dormann C. F., McPherson J. M., Araújo M. B., Bivand R., Bolliger J., Carl G., Davies R. G., Hirzel A., Jetz W., Kissling W. D., Kühn I., Ohlemüller R., Peres-Neto R. P., Reineking B., Schröder B., Schurr F. M. & Wilson R. (2007) Methods to account for spatial

- autocorrelation in the analysis of species distributional data: a review, *Ecography*, **30**, 609-628.
- Efron, B. and Tibshirani, R. (1993) *An Introduction to the Bootstrap*. Chapman & Hall.
- Faraway, J. J. (2006) *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*, Chapman & Hall/CRC, New-York, 301 pages.
- Fox, J. (2002) *Applied Regression, Linear Models, and Related Methods*. Sage.
- Gordon, A. D. (1999) *Classification, Second Edition*, Chapman and Hall / CRC, London, pp.
- Gotelli, N.J. & Graves, G.R. (1996) *Null models in ecology* Smithsonian Institution Press, Washington.
- Gotelli, N.J. & Colwell, R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379-91.
- Greene (2002) *Econometric Analysis*, Prentice Hall, US, 5e International Ed, 1026 pp.
- Hartigan, J.A. (1975), *Clustering Algorithms*, New York: John Wiley & Sons, Inc.
- Hastie, Trevor, and Tibshirani, R. (1989) *Generalized Additive Models (with discussion)*, *Statistical Science*, **1**(3), 297-318.
- Hastie, T., Tibshirani, R. & Friedman J. (2001). *The Elements of Statistical Learning: Data mining, inference and prediction*, Springer-Verlag, New-York, 536 pages.
- Hellmann, J.J. & Fowler, G.W. (1999) Bias, precision and accuracy of four measures of species richness. *Ecological Applications*, **9**(3), 824-34.
- Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Kincaid, T.M., Reynolds, L. & Larsen, D.P. (1998) A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 1618-31.
- Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Instelmann, S.S., Corbett, S.C., Arbogast, M.C. & Hjort, R.C. (2002) Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. *North American Journal of Fisheries Management*, **22**, 1229-40.
- Hughes, R.M. & Herlihy, A.T. (2007) Electrofishing distance needed to estimate consistent index of biotic integrity (IBI) score in raftable Oregon rivers. *Transactions of the American Fisheries Society*, **136**, 135-41.
- Kestemont P. & Goffaux D. (2004) *Metric Selection and Sampling Procedures for FAME, FINAL REPORT, Development, Evaluation & Implementation of a Standardised Fish-based Assessment Method for the Ecological Status of European Rivers A Contribution to the Water Framework Directive*,
http://fame.boku.ac.at/downloads/D4_6_metrics_and_sampling_procedure.pdf

- Legendre P., Dale M. R. T., Fortin, M.-J., Gurevitch J., Hohn M. & Myers D. (2002) The consequences of spatial structure for the design and analysis of ecological study, *Ecography*, **25**, 601-615.
- Legendre, P. & Legendre, L. (1998) *Numerical Ecology*. Second English Edition Elsevier.
- MacKinnon JG, White H (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics*, **29**, 305–325.
- Magurran, A.E. (1988) *Ecological diversity and its measurement* Croom Helm Limited, London.
- Manel, S., Dias, J.-M. & Ormerod, S.J. (1999) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, **120**(2-3), 337-47.
- Manly, B.F.J. (2007) *Randomization and Monte Carlo methods in biology*, third edition, Chapman and Hall, London.
- Marazzi, A. (1993) *Algorithms, routines, and S functions for robust statistics*. Wadsworth & Brooks/Cole, Pacific Grove, CA.
- Maret, T.R., Ott, D.S. & Herlihy, A.T. (2007) Electrofishing effort required to estimate biotic condition in Southern Idaho rivers. *North American Journal of Fisheries Management*, **27**, 1041-52.
- Marx, B.D (1996) Iteratively reweighted partial least squares estimation for generalized linear regression. *Technometrics*, **38**(4): 374-381.
- May, R. (1981). *Patterns in multi-species communities*. In *Theoretical ecology: principles and applications*, second edition (ed B.S. publication), pp. 197-227, London.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*, second edition edn. Chapman & Hall/CRC, London.
- Moreno C. E. & Halffter G. (2001) On the measure of sampling effort used in species accumulation curves. *Journal of applied. Ecology*, **38**,487–90.
- Nelder, J; & Wedderburn R. (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society. Series A (General)*, **135**, 370-384.
- Palmer M. W. (1990) The estimation of species richness by extrapolation. *Ecology* **71**, 1195–1198.
- Petchey, O.L. & Gaston, K.J. (2002) Functional diversity (FD), species richness and community composition. *Ecology Letters*, **5**, 402-11.

- Pont D., Hugueny B., Roset N. & Rogers C. (2004) Analysing Reference conditions and Assessing degraded conditions, FINAL REPORT, Development, Evaluation & Implementation of a Standardised Fish-based Assessment Method for the Ecological Status of European Rivers A Contribution to the Water Framework Directive, http://fame.boku.ac.at/downloads/D9_13_Final%20Modelling%20Report.pdf
- Pont D., Hugueny B., Roset N. & Rogers C. (2004) Definition of an European Fish Based Index (EFI) Using The Reference Condition Modelling Approach, FINAL REPORT, Development, Evaluation & Implementation of a Standardised Fish-based Assessment Method for the Ecological Status of European Rivers A Contribution to the Water Framework Directive, http://fame.boku.ac.at/downloads/D9_13_Final%20Modelling%20Report.pdf
- Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., Rogers, C., Roset, N. & Schmutz, S. (2006) Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology*, **43**(1), 70-80.
- Pont, D., Hugueny, B. & Rogers, C. (2007) Development of a fish-based index for the assessment of river health in Europe: the European Fish Index. *Fisheries Management and Ecology*, **14**(6), 427-39.
- Potts, J. & Elith, J. (2006) Comparing species' abundance models. *Ecological Modelling* **199**, 153-163.
- Quataert P., Breine J. & Simoens I. (2004) Comparison of the European Fish Index with the Standardised European Model, the Spatially Based Models (eco-regional and European), and Existing Methods, FINAL REPORT, Development, Evaluation & Implementation of a Standardised Fish-based Assessment Method for the Ecological Status of European Rivers A Contribution to the Water Framework Directive, http://fame.boku.ac.at/downloads/D16_17_MethodComparison.pdf
- Quataert, P., Breine, J. & Simoens, I. (2007) Evaluation of the European Fish Index: false-positive and false-negative error rate to detect disturbance and consistency with alternative fish indices. *Fisheries Management and Ecology*, **14**(6), 465-72.
- R Development Core Team (2007). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Reyjol, Y., Hugueny, B., Pont, D., Bianco, P.G., Beier, U., Caiola, N., Casals, F., Cowx, I., Economou, A., Ferreira, T., Haidvogel, G., Noble, R., de Sostoa, A., Vigneron, T. &

- Virbickas, T. (2007) Patterns in species richness and endemism of European freshwater fish. *Global Ecology and Biogeography*, **16**(1), 65-75.
- Reynolds, L., Herlihy, A.T., Kaufmann, P.R., Gregory, S.V. & Hughes, R.M. (2003) Electrofishing effort requirements for assessing species richness and biotic integrity in Western Oregon streams. *North American Journal of Fisheries Management*, **23**, 450-61.
- Ripley, B. D. (1996) *Pattern Recognition and Neural Networks*. Cambridge.
- Saporta, G. (2006) *Probabilités, analyses de données et statistique*, Editions TECHNIP, Paris, 622 pages.
- Schabenberger, O. & Gotway, C.A. (2005) *Statistical methods: for spatial data analysis* Chapman & Hall/CRC, Florida.
- Schmutz, S., Cowx, I.G., Haidvogel, G. & Pont, D. (2007) Fish-based methods for assessing European running waters: a synthesis. *Fisheries Management and Ecology*, **14**(6), 369-80.
- Sokal, R.R. & Rohlf, F.J. (1969) *Biometry* Third edition. W.H. Freeman and Company, New-York.
- Scott, D. W. (1992) *Multivariate Density Estimation. Theory, Practice and Visualization*. New York: Wiley.
- Tenenhaus, M. (1998) *La régression PLS, théorie et pratique*, Editions TECHNIP, Paris, 254 pages.
- Thompson, G.G., Withers, P.C., Pianka, E.R. & Thompson, S.A. (2003) Assessing biodiversity with species accumulation curves; inventories of small reptiles by pit-trapping in Western Australia. *Austral Ecology*, **28**, 361-83.
- Venables W. N. & Ripley B. D. (1999) *Modern Applied Statistics with S-plus*, Third edition, Statistics and computing, Springer-Verlag, New York, pp. 501.
- Wold, S., Martens, M. & Wold, H. (1983) in *The Multivariate Calibration Problem in Chemistry by PLS*, ed. Ruhe A. & Kagstrom, B. , Springer, Heidelberg, pp. 286–289.
- Zeileis A, Hothorn T (2002). “Diagnostic Checking in Regression Relationships.” *R News*, **2**(3), 7–10. URL <http://CRAN.R-project.org/doc/Rnews/>.
- Zeileis A (2004). *Econometric Computing with HC and HAC Covariance Matrix Estimators*. *Journal of Statistical Software*, **11**(10), 1–17. URL <http://www.jstatsoft.org/v11/i10/>.

Annexes A: Description of the dataset used in the section 2

Species	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12
<i>Cottus gobio</i>	7	0	103	0	1000	1	74	0	1	0	0	0
<i>Oncorhynchus mykiss</i>	75	0	0	1	0	0	0	297	0	198	0	1055
<i>Salmo trutta fario</i>	233	235	750	704	11	2	7	1068	0	222	0	215
<i>Thymallus thymallus</i>	222	0	0	0	20	0	5	0	3	392	0	87
<i>Barbatula barbatula</i>	0	17	0	14	1000	37	109	0	2	0	0	11
<i>Barbus meridionalis</i>	0	5	0	0	0	0	0	0	0	0	0	0
<i>Gobio gobio</i>	0	63	0	0	3	60	54	0	3	0	0	0
<i>Leuciscus cephalus</i>	0	47	0	0	53	96	0	61	87	0	120	19
<i>Phoxinus phoxinus</i>	0	291	0	0	1000	313	313	0	334	0	0	0
<i>Lampetra planeri</i>	0	0	36	0	0	0	0	0	0	0	0	0
<i>Anguilla anguilla</i>	0	0	0	0	1000	21	0	0	1	0	2	0
<i>Esox lucius</i>	0	0	0	0	2	0	0	0	31	0	10	2
<i>Leuciscus leuciscus</i>	0	0	0	0	137	19	0	0	1	0	1	0
<i>Rutilus rutilus</i>	0	0	0	0	25	74	0	0	336	0	277	0
<i>Salmo salar</i>	0	0	0	0	13	0	0	0	0	0	0	0
<i>Alburnoides bipunctatus</i>	0	0	0	0	0	44	0	0	10	0	0	0
<i>Alburnus alburnus</i>	0	0	0	0	0	49	0	0	31	0	1	0
<i>Ameiurus melas</i>	0	0	0	0	0	4	0	0	13	0	0	0
<i>Barbus barbus</i>	0	0	0	0	0	17	0	0	10	0	227	0
<i>Carassius carassius</i>	0	0	0	0	0	6	0	0	1	0	0	0
<i>Chondrostoma nasus</i>	0	0	0	0	0	19	0	0	0	0	19	0
<i>Cyprinus carpio</i>	0	0	0	0	0	2	0	0	0	0	36	0
<i>Lepomis gibbosus</i>	0	0	0	0	0	2	0	0	46	0	0	0
<i>Lota lota</i>	0	0	0	0	0	2	0	0	6	0	0	0
<i>Perca fluviatilis</i>	0	0	0	0	0	1	0	0	58	0	30	0
<i>Rhodeus sericeus</i>	0	0	0	0	0	3	0	0	24	0	51	0
<i>Scardinius erythrophthalmus</i>	0	0	0	0	0	2	0	0	19	0	3	0
<i>Cobitis taenia</i>	0	0	0	0	0	0	103	0	0	0	0	0
<i>Abramis brama</i>	0	0	0	0	0	0	0	0	4	0	690	0
<i>Blicca bjoerkna</i>	0	0	0	0	0	0	0	0	57	0	1261	0
<i>Gymnocephalus cernuus</i>	0	0	0	0	0	0	0	0	3	0	25	0
<i>Sander lucioperca</i>	0	0	0	0	0	0	0	0	1	0	59	0
<i>Silurus glanis</i>	0	0	0	0	0	0	0	0	2	0	76	0
<i>Tinca tinca</i>	0	0	0	0	0	0	0	0	23	0	2	0
<i>Abramis ballerus</i>	0	0	0	0	0	0	0	0	0	0	212	0
<i>Abramis sapa</i>	0	0	0	0	0	0	0	0	0	0	89	0
<i>Aspius aspius</i>	0	0	0	0	0	0	0	0	0	0	25	0
<i>Carassius auratus</i>	0	0	0	0	0	0	0	0	0	0	457	0
<i>Gobio albipinnatus</i>	0	0	0	0	0	0	0	0	0	0	15	0
<i>Gymnocephalus schraetser</i>	0	0	0	0	0	0	0	0	0	0	7	0
<i>Leuciscus idus</i>	0	0	0	0	0	0	0	0	0	0	90	0
<i>Pseudorasbora parva</i>	0	0	0	0	0	0	0	0	0	0	3	0
<i>Sander volgensis</i>	0	0	0	0	0	0	0	0	0	0	2	0
<i>Vimba vimba</i>	0	0	0	0	0	0	0	0	0	0	1	0
Zingel zingel	0	0	0	0	0	0	0	0	0	0	3	0
richness	4	6	3	3	12	21	7	3	26	3	29	6
abundance	537	658	889	719	4264	774	665	1426	1107	812	3794	1389
H	1.53	1.84	0.75	0.15	2.34	2.99	2.12	0.98	3.01	1.52	3.17	1.12
J	0.21	0.25	0.10	0.02	0.32	0.41	0.29	0.14	0.42	0.21	0.44	0.16

Richness: number of species, Abundances: number of individuals, H: Shannon's index, J: Equitability index.