Annex 2.1.5.2 Part I REBECCA short contribution to the IC exercise for Rivers.
Part II Relationships between the Intercalibration Common Metric index (ICMi) and a Land Cover Pressure Index for the French invertebrates IC datasets.

REBECCA - Short contribution to the Intercalibration exercise
for the Central Baltic River GIG - Invertebrates

Part I
REBECCA short contribution to the IC exercise for Rivers (Buffagni, Erba, M. Dobiasova)

Part II
Relationships between the Intercalibration Common Metric index (ICMi) and a Land Cover Pressure Index for the French invertebrates IC datasets (WASSON, VILLENEUVE, MENGIN)

15th September 2006

REBECCA (Contract no: SSPI-CT-2003-502158) short contribution to the IC exercise for Rivers

Prepared by A. Buffagni*, S. Erba* and M. Dobiasova**
*CNR-IRSA Brugherio (IT)
**SHMU Bratislava (SK)

This report provides some results related to the investigation of the correlation between organic pollution and Intercalibration metrics. These results are derived from the REBECCA WP4 deliverable that will be published at the end of autumn 2006. Detailed information on the analyses run and on the data used can be found in this deliverable.

One of the aims of the project was to investigate the relationship between organic pollution and biological metrics. The first stage involved combining chemical parameters in order to provide a single abiotic descriptor to quantify organic pollution. Invertebrate intercalibration metrics were also calculated. ICMs and STAR_ICMi were thus correlated to the newly developed organic pollution descriptor.

Calculation of an abiotic organic descriptor index

The organic pollution descriptor was obtained by assigning a score to the observed concentration of single chemical parameters. Decreasing scores are assigned with reference to boundaries obtained as multiples of the 75^{th} percentile of values found in reference sites. Each chemical variable is assigned a score according to the observed concentration so that the result ranges between 0 and 1 (see Table 1).

Table 1. Criteria used for the definition of classes when assigning scores to chemical parameters. A = 75^{th} percentile of reference sites.

	Class I	Class II	Class III	Class IV	Class V
Boundary	<1*A	<2*A	<4*A	<8*A	>8*A
Score	1	0.5	0.25	0.125	0.0625

The scores obtained from each chemical parameter are averaged to obtain the final index value (site score). The variables considered for scoring are: O_2%, BOD_5, COD, N-NH_4, N-NO_3, NO_2, PO_4, TP, Chloride and *E. coli*. Not all variables were recorded for all samples considered so the final site score may be based on e.g. 5 out of the 10 variables (e.g. O_2%, BOD_5, N-NO_3, Chloride and *E. coli*). On average, more than 8 or 5 variables were available for Italy and Slovakia, respectively and thus used for each sample to derive the site score for the organic pollution indicator.

Calculation of determination coefficients between organic pressures and intercalibration metrics

The following tables present some examples of r^2 values between intercalibration metrics and the organic pollution descriptor from a selection of datasets provided for the REBECCA project. Data presented below were derived from Italian and Slovak monitoring networks and from AQEM Italian data.

Some examples from Italian data

Table 2 shows determination coefficients for Italian monitoring data (Italian Environment Agency from VENETO and EMILIA-ROMAGNA Regions, Northern Italy) R-M1, R-M2 and R-M3 all combined. The Iberian ASPT and the Number of selected families of ETD (Ephemeroptera, Trichoptera and Diptera) are the best performing metrics. The Pearson coefficients of these metrics are comparable to the number of EPT families and to the ICM indices. Notable coefficients are also obtained by the metric based on the abundance of selected taxa of Ephemeroptera, Plecoptera, Trichoptera and Diptera (Log_SelEPTD). In general, coefficients are not very high. This may be explained by the fact that only organic pollution was quantified in these examples and no information was available on other acting pressures. For comparison, the r^2 values obtained by the IBE index (official method in Italy before the WFD) are shown in the following tables.

Table 2. Pearson coefficients of determination in Italian data (ARPA data).

	Org_descriptor	
	r^2	P
ASPT	0.29	3E-13
Iberian ASPT	0.34	5E-10
No of EPT Families	0.30	1E-08
Total No of Families	0.02	0.225
1-GOLD	0.01	0.1632
Shannon Diversity	0.006	0.3206
Log_SelEPTD+1	0.24	6E-11
No of Selected Families of ETD / Total No of Families	0.35	5E-10
STAR_ICMi	0.32	2E-09
MedQual_ICMi	0.31	4E-09
IBE	0.13	0.0003

In table 3 determination coefficients are shown for Italian data from R-M1 from Southern Italy. These data were collected within the AQEM project (Buffagni et al., 2004). Total number of families was the poorest performing metric (lowest coefficient), while abundance-based metrics showed a good relationship with the organic pollution descriptor. The qualitative mediterranean index gave a similar performance to the STAR ICMi. The mediterrannean qualitative index comprises metrics that are primarily developed to detect organic pollution, while the STAR_ICMi also comprises metrics developed to detect habitat degradation. The mediterranenan qualitative index, therefore, would be expected to show a stronger relationship with the organic pollution descriptor than the STAR ICMi. It should be noted that the mediterranenan ICMi does not take abundance into account and is not considered to be fully WFD-compliant.

Table 31. Pearson coefficients of determination for the different options of deriving the organic pollution descriptors and selected biological metrics in Southern Italy (AQEM data; N=66).

	Org_descriptor	
	r^2	p
ASPT	0.61	7E-15
Iberian ASPT	0.63	2E-15
No of EPT Families	0.59	8E-14
Total No of Families	0.28	4E-06
1-GOLD	0.39	2E-08
Shannon Diversity	0.43	2E-09
log(SelEPTD+1)	0.60	2E-14
No of Selected Families of ETD / Total No of Families	0.72	2E-19

STAR_ICM index	0.67	5E-17
Med_ICM index	0.68	2E-17
IBE	0.40	1E-08

Both multimetric indices developed for intercalibration purposes gave a better response to organic pollution than the IBE index.

The relationship between the ICM indices and pressures, other than organic pollution, is also presented (Table). The other pressures considered are presented below:

1. Morphological alteration expressed in terms of Habitat Modification Score (HMS: Raven et al., 1997);
2. Land use expressed in terms of the Land Use Index (LUI), according to scores assigned in relation to the percentage of anthropic land use (Feld, 2004);
3. General degradation expressed in terms of Index of Fluvial Functioning (IFF: Siligardi et al., 2000).
4. Habitat Quality Assessment (HQA: Raven et al., 1997), that within this dataset was proved to be an indicator of general degradation (Balestrini et al., 2004)

Table 4. Pearson coefficients of determination for pressures other than organic pollution and ICM indices and IBE (AQEM data). N=66.

	r^2	p	r^2	p
	LUI_catch		HQA	
STAR_ICM index	0.01	0.347	0.51	1E-11
Med_ICM index	0.002	0.741	0.50	4E-11
IBE_EQR	9E-6	0.981	0.36	1E-7
	LUI_floodp		IFF	
STAR_ICM index	0.32	6E-7	0.44	1E-9
Med_ICM index	0.29	3E-6	0.38	3E-8
IBE_EQR	0.13	0.003	0.21	9E-5
	HMS		INT_press	
STAR_ICM index	0.09	0.013	0.42	4E-9
Med_ICM index	0.06	0.049	0.37	8E-8
IBE_EQR	0.02	0.317	0.19	0.0002

The different pressures that have been averaged and combined in order to obtain an integrated pressure index include: HMS, LUI, HQA (even if not always linked to general degradation), IFF and the combined organic pollution index. Results showed that the best performing index in relation to the combination of pressures is the STAR_ICMi. Low Pearson coefficients are found with HMS because sites with poor communities affected by organic pollution can have a very good morphology. The relationship between metrics and land use at the catchment level is not significant, while it is significant for land use in the floodplain. This seems to be confirmed by Wasson et al (contribution to REBECCA WP4 deliverable), who demonstrated that land use at the buffer level can show higher relationships with biota than that at catchment level. In the dataset presented here, organic pollution generally represents the strongest impact at sites and Pearson coefficients are higher between biological metrics and the organic pollution index than with integrated pressures.

Some examples from Slovak data

The relationship between selected biological metrics (ASPT, Saprobic Index and STAR_ICMi) and single parameters indicating organic pollution were analyzed by calculating Pearson coefficients. The selected biological metrics were predominantly intended to assess organic pollution (saprobic and ASPT) and are used in the Intercalibration exercise (STAR ICMi) over a large part of Europe. The first stage was to evaluate the performance of single metrics by calculating their correlations with single chemical parameter concentrations (Table 5).

Table 5. Pearson coefficients of determination between selected metrics and single chemical parameters.

	r^2	p	N		r^2	p	N
BOD5				TP			
Czech_saprobic	0.42	0	256	Czech_saprobic	0.15	7.2E-11	259
ASPT	0.27	8.5E-19	256	ASPT	0.18	1.4E-12	259
STAR_ICMi	0.32	3.5E-23	256	STAR_ICMi	0.19	8.9E-14	259
N-NH4				N-NO2			
Czech_saprobic	0.09	2.1E-06	252	Czech_saprobic	0.29	4.8E-20	253
ASPT	0.12	2.1E-08	252	ASPT	0.24	1.3E-16	253
STAR_ICMi	0.11	4E-08	252	STAR_ICMi	0.28	2.1E-19	253
N-NO3				P-PO4			
Czech_saprobic	0.24	8.5E-17	253	Czech_saprobic	0.03	0.26551	48
ASPTn	0.18	1.2E-12	253	ASPTn	0.06	0.09411	48
STAR_ICMi	0.24	5.2E-17	253	STAR_ICMi	0.08	0.05125	48

The highest coefficient of determination occured between the Czech saprobic index and BOD_5. Of the chemical parameter considered in general, BOD_5 seemed to show the best correlation with biology. Correlations between orthophosphate and selected metrics were not significant (note that data availability is scarce). The other chemical parameters do not show very high determination coefficients (r^2 always < 0.3), and gave a similar performance with all of the biological metrics. The Saprobic index performed quite well to detect organic pollution but, in most cases, the STAR_ICMi gave similar results. It should be noted that the saprobic index is based on specific taxa identification while ASPT and the STAR_ICMi are based on family level identification. Even if species level identification is very useful for the detection of particular impact types, in some cases family level identification can give similar results with much less taxonomic effort (Hewlett, 2000). Moreover it has to be emphasized that the ICMi is structured to detect general degradation, thus it is not expected to have as strong a relationship with organic pollution as the saprobic index, which is dedicated to detect this impact.

The second stage tested the combined chemical descriptor against the biological metrics. Table 6 reports the calculated r^2 values.

Table 6. Pearson coefficients of determination between selected biological metrics and the different options of combining chemical parameters in Slovak data.

	r^2	p	N
Org_descriptor			
Czech_saprobic	0.51	0	259
ASPT	0.33	3.9E-24	259
STAR _ICMi	0.42	0	259

The calculation of determination coefficients between combined chemical data and biological metrics confirm the strong performance of the Saprobic index. The second best performing metric is the STAR_ICMi, confirming the fact that combining information derived from different components of the community gives better results than considering only one aspect even when considering only one impact type (Brabec et al., 2004; Buffagni et al., 2004; Camargo et al., 2004; Ofenböck et al., 2004).

When assessing the response of biological metrics to chemical water quality (e.g. organic pollution), the approach of combining single chemical parameters into an integrated indicator gives better correlations than chemical variables considered individually.

REBECCA (Contract no: SSPI-CT-2003-502158)

Short contribution to the Intercalibration exercise for the Central Baltic River GIG - Invertebrates

Relationships between the Intercalibration Common Metric index (ICMi) and a Land Cover Pressure Index for the French invertebrates IC datasets.

Prepared by Jean-Gabriel WASSON, Bertrand VILLENEUVE and Nicolas MENGIN
CEMAGREF, Lyon, France

1 Introduction

This brief report provides some results related to the relationships between large scale anthropic pressures, evaluated through a land cover index, and the biological response of river invertebrates.

These results are derived from the REBECCA WP4 deliverable that will be published at the end of autumn 2006. Detailed information on the analyses run and on the data used will be found in this deliverable.

One of the aims of the project was to investigate the relationship between combined pressures and biological metrics. The basin and riparian corridors land cover was selected as the best indicator to evaluate large scale pressures, taking into account most of the driving forces leading to physico-chemical and hydro-morphological impacts at the river scale. The first stage was to derive a land cover pressure index in order to provide a single descriptor of large scale combined pressures. The invertebrate response was evaluated in using the Intercalibration Common Metric index (ICMi), especially developed for the purpose of the Intercalibration (IC) process.

2 Methods

In a first step, diagnostic models based upon Partial Least Squares (PLS) regressions were developed to evaluate the correlation between the ICMi and land cover categories. This kind of model allows to attribute to each land cover category a coefficient of impact upon the biological index.

Land cover was evaluated by the mean of the CORINE Land Cover (CLC) spatial database. For France, we used 4 categories of land cover variables

- Urban and artificial areas (variable : **artif**);
- High impact agriculture (variable : **agrii**), mainly arable land;
- Low impact agriculture (variable : **agrif**), mainly pastures;
- Forests and natural areas (variable : **espnat**).

PLS regressions were run between these four land cover variables and the ICMi in order to build an integrated land cover pressure index. For each site, the land cover was calculated for the whole basin of the site (catchment scale), and within a local riparian buffer (river corridor scale).

These GIS delineated buffers were 3 km long, with a total width varying from 100m for the low order streams (order 1 to 3 in Strahler’s ordination) to 1200 m for the large rivers (order 7). The corresponding land cover variables are denominated **artifbuf, agriibuf, agrifbuf** and **espnatbuf**.

The biological data are those used by France for the river invertebrates IC process, for the 3 GIGs (Alpine, Central Baltic and Mediterranean) : 490 sites from A1, A2, C1, C2, C3, C4, C6, M1, M4, corresponding to 1593 samples. Invertebrates were collected in using the field protocol of the normalized French IBGN method, but with quantitative laboratory evaluation allowing to calculate the ICM index.

3 Diagnostic PLS models linking ICMi to land cover categories

3.1 PLS Model at the catchment scale

A PLS regression between the average ICMi of the sites and the catchment land cover was run for the whole French IC dataset. The determination coefficient of the model is 15.6%. The normalized regression coefficients of land cover variables are presented in figure 1 and table 1.

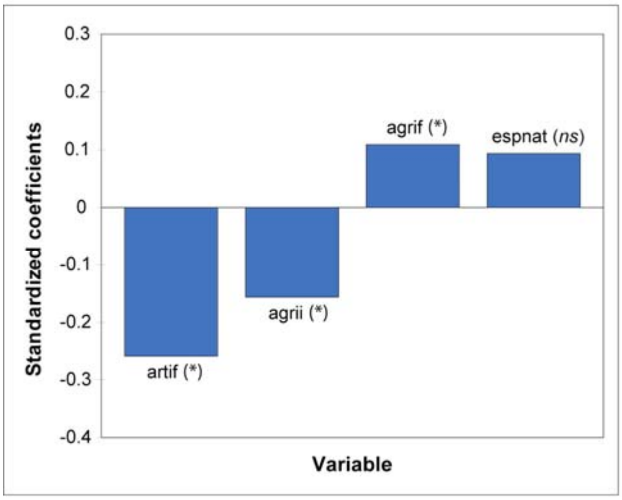

Figure 1. Standardized coefficients (and significative level : "*" p-value <= 0.05, "*ns*" non-significative) of the PLS regression between catchment land cover variables (artif, agrii, agrif, espnat) and ICMi; data from the French invertebrates IC dataset.

Variable	Coefficient
artif	-0.258
agrii	-0.156
agrif	0.109
espnat	0.094

Table 1. Values of the standardized coefficients of the PLS regression between catchment land cover variables (artif, agrii, agrif, espnat) and ICMi; data from the French invertebrates IC dataset.

3.2 PLS Model at the river corridor scale

Similarly, a PLS regression between the average ICMi of the sites and the river corridor land cover was run on the same dataset. The determination coefficient of the model is 10.8%. The normalized regression coefficients of land cover variables are presented in figure 2 and table 2.

Variable	Coefficient
artifbuf	-0.205
agriibuf	-0.140
agrifbuf	0.117
espnatbuf	0.090

Table 2. Values of the standardized coefficients of the PLS regression between river corridor land cover variables (artifbuf, agriibuf, agrifbuf, espnatbuf) and ICMi PLS Model; data from the French invertebrates IC dataset.

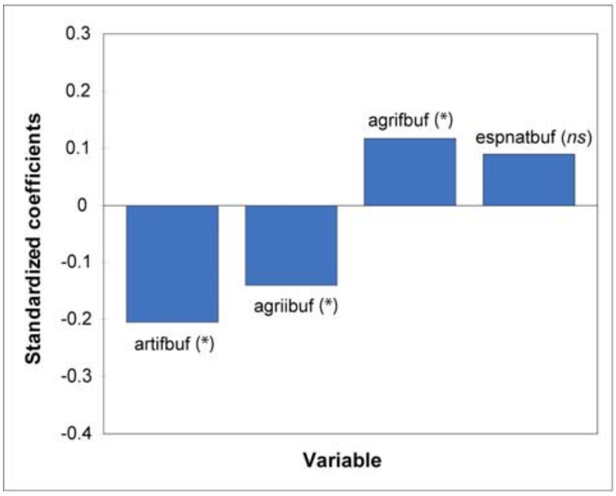

Figure 2. Standardized coefficients (and significative level : "*" p-value <= 0.05, "*ns*" non-significative) of the PLS regression between river corridor land cover variables (artifbuf, agriibuf, agrifbuf, espnatbuf) and ICMi; data from the French invertebrates IC dataset.

4 Range of variation of land cover variables

However, all the land cover variables do not vary in the same range : while the agricultural or natural areas can cover the entire basin, a small share of artificial areas yet represents a huge human pressure. The range of variation of each land cover category, at the catchment and corridor scale, is reported in the table 3. For instance, in the basins of the tested sites, the 95th percentile of the urban land cover (artif) is only 10%, while the 95th percentile of arable land (agrii) is 78%. The 95th percentiles of these distributions allow to determine a realistic maximum range of variation for each variable.

	artif	agrii	agrif	espnat	artifbuf	agriibuf	agrifbuf	espnatbuf
Minimum	*0%*	*0%*	*0%*	*0%*	*0%*	*0%*	*0%*	*0%*
95th percentile	*10%*	*78%*	*70%*	*94%*	*34%*	*93%*	*99%*	*95%*
Maximum	*47%*	*95%*	*97%*	*100%*	*97%*	*100%*	*100%*	*100%*

Table 3. Distribution of the land cover variables at the catchment and corridor scales in the French invertebrates IC dataset.

As a consequence, it is not relevant to make a simple linear combination of land cover variable to build a land cover pressure index: one percent of variation of urban land cover in the basin, representing 1/10th of the possible range of variation of this pressure, cannot be compared to one percent of variation of arable land. It seems thus logical to normalize each variable in the land cover index by its 95th percentile.

5 Calculation of a Land Cover Pressure Index

The results of the PLS models (§3) evidenced different senses and magnitudes of impact of the land cover variables upon the ICMi response. To build up an integrated Land Cover Pressure Index (denominated here **IPOS** according to its French acronym: *Indice de Pression d'Occupation du Sol*), the idea was to combine both previous results : the determination coefficients of the PLS models give an "impact factor" of the land cover variables, and the 95th percentiles of the distributions in the whole French dataset give a realistic range of variation of the same variables. We used then a linear combination of the land cover variables weighted by the PLS determination coefficients and normalized according to the 95th percentile of their distribution.

The general formula of the IPOS index can be written like in equation 1 :

$$IPOS = \sum_i (-coef_i * \frac{\mathrm{var}_i}{95^{th}\ p(\mathrm{var}_i)} *100) \qquad (1)$$

According to the previous results, the IPOS index for the basin land cover pressure ($IPOS_{bassin}$) is given by equation 2:

$$IPOS_{bas\sin} = 0.25*\frac{artif}{95^{th}\ p(artif)}*100 + 0.15*\frac{agrii}{95^{th}\ p(agrii)}*100 - 0.1*\frac{agrif}{95^{th}\ p(agrif)}*100 - 0.1*\frac{espnat}{95^{th}\ p(espnat)}*100 \qquad (2)$$

And for the river corridor land cover pressure, the IPOS index ($IPOS_{buffer}$) is given by equation 3:

$$IPOS_{buffer} = 0.20*\frac{artifbuf}{95^{th}\ p(artifbuf\ )}*100 + 0.15*\frac{agriibuf}{95^{th}\ p(agriibuf\ )}*100 - 0.1*\frac{agrifbuf}{95^{th}\ p(agrifbuf\ )}*100 - 0.1*\frac{espnatbuf}{95^{th}\ p(espnatbuf\ )}*100 \qquad (3)$$

Both indices can take positive and negative values because of PLS coefficients.

6 Relationships between the IPOS Index and Chemical pressures

Does this index represent correctly the pressures impacting river invertebrates? This question was addressed in analysing the correlation between the IPOS indices (basin and corridor) and chemical parameters characterising the samples of the French IC dataset; hydro-morphological pressure at the site scale were not available for this dataset. A Multiple PLS regression was realised between $IPOS_{bassin}$ and $IPOS_{buffer}$. and mean concentrations of selected chemicals (NO3, PO4, BOD5 and NH4) corresponding to the year preceding a sample date (11 months before, 1 after).

The correlations map allows to visualize on the first two components the correlations between IPOS and chemical parameters (figure 3). The determination coefficient is 35% with $IPOS_{basin}$, but only 5% with $IPOS_{buffer}$. The normalized regression coefficients of chemical variables are presented in figures 4 (A and B) for the IPOS index at the basin and corridor level. $IPOS_{bassin}$ is principally explained by NO3 and a combination of PO4, BOD5 and NH4. Conversely, $IPOS_{buffer}$ has no significant relationships with chemical variables.

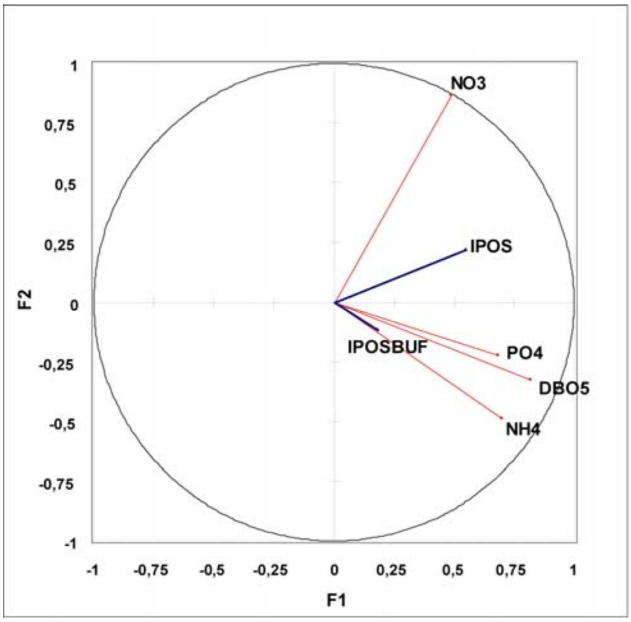

Figure 3. Correlation circle of multiple PLS regression between mean annual concentration of chemical parameters and the land cover pressure index for the basin (IPOS) and river corridor (IPOSBUF); data from the French invertebrates IC dataset.

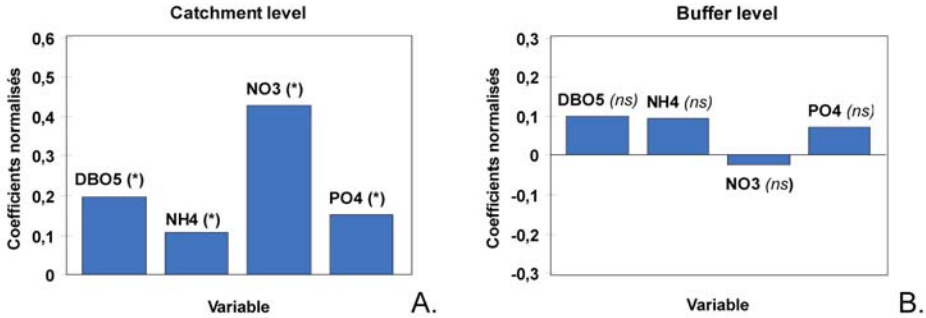

Figure 4. Standardized coefficients (and significative level : "*" p-value <= 0.05, "*ns*" non-significative) of the PLS regression between chemical variables and the IPOS index for the basin (A) and the river corridor (B); data from the French invertebrates IC dataset.

From these results, we can infer that the IPOS index at the basin scale is an indicator of the chemical pressures coming both from urban and agricultural areas, while the IPOS at the corridor level is not correlated to the pollution discharge, and indicates a local, physical pressure at the site scale.

7 Relationships between IPOS and ICMi

7.1 Linear regressions between ICMi and IPOS

There are a clear negative relationships between the average ICMi of the sites and the IPOS at both basin and riparian corridor scales (figure 5). The determination coefficient of the linear regression between average ICMi and $IPOS_{bassin}$ is 16%, and 11% between average ICMi and $IPOS_{buffer}$. In both cases, the average ICMi variability explained by the combined land cover index is equivalent to that explained by the PLS model with the 4 independent land cover variables.

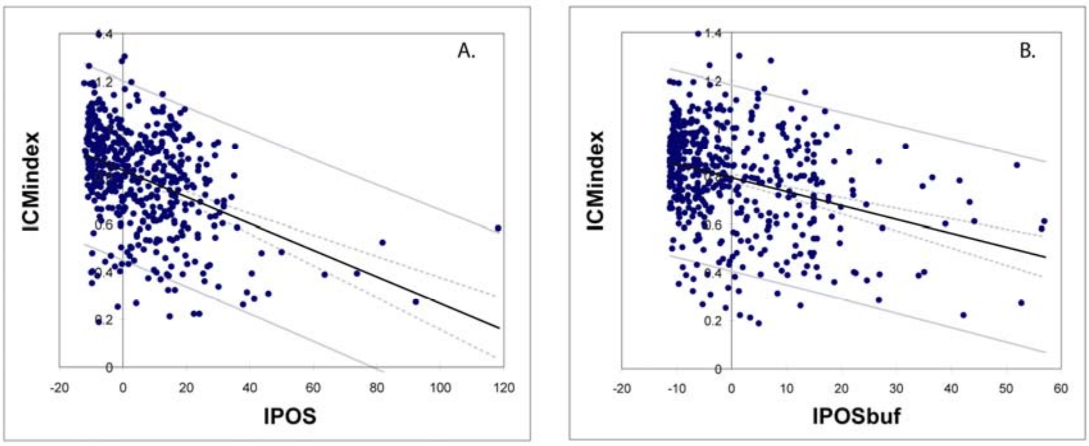

Figure 5. Linear regression between ICMi and land cover pressure index (IPOS) at the basin scale (A) and at the river corridor scale (B); data from the French invertebrates IC dataset.

7.2 Combined relationships at both catchment and river corridor scale

When combining the two land cover pressure indices, the value of the determination coefficient of the linear regression between average ICMi and ($IPOS_{bassin}$ + $IPOS_{buffer}$) is 18%. This represents only 2 % more than the $IPOS_{bassin}$ alone. These two levels of pressure are intercorrelated and their effects cannot be added.

However, the standardized coefficients of the PLS regression demonstrate that the effects of both indices are well balanced between the river corridor and the catchment level (figure 6).

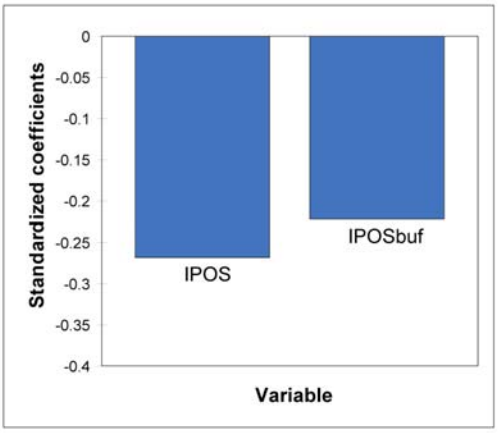

Figure 6. Standardized coefficients of regression between ICMi and land cover pressure indices at the catchment scale (IPOS) and river corridor level (IPOSbuf); data from the French invertebrates IC dataset.

Moreover, a complete PLS regression between average ICMi and ($IPOS_{basin}$ + $IPOS_{buffer}$ + NO_3 + NH_4 + PO_4 + DBO_5) give a determination coefficient of 21 %. This represent only 3% more than without chemical variables.

7.3 Relationships for the different IC Types

Linear regressions were realised between the average ICMi of the sites and the IPOS at the catchment and river corridor levels for each IC type, for the French Alpine and CB GIG datasets; determination coefficient are given in table 4.

	Level	
IC Type	Catchment	River corridor
R-A1	10%	21%
R-A2	20%	21%
R-C1	0.5%	12%
R-C2	16%	5%
R-C3	28%	25%
R-C4	8%	9%
R-C6	13%	6%

Table 4. Determination coefficient of linear regression between ICMi and IPOS at the catchment and river corridor levels by IC type

Except for the type R-C1, corresponding in France to the sandy “Landes” region mainly covered with pine forests, the correlations with the basin IPOS vary between 8% and 28% for the different types, with highest values in the mountainous regions (R-A2 and R-C3). The correlations with the river corridor IPOS vary between 5% and 25%, here also with higher values in the mountains.

8 Conclusions

The ICM index was derived to assess at the European scale the ecological status of the invertebrate quality element in rivers. Other works (see annexes 2.2.5.1 and 2.2.5.2) have demonstrated its response to various pressures including chemical and hydro-morphological ones evaluated at the site scale. We examined in France only the relationships between average ICMi calculated for each site and a land cover pressure index (IPOS) evaluated at the basin and at the river corridor scale.

The results presented here evidenced clear and significant negative relationships between average ICMi and IPOS evaluated at both basin and riparian corridor scale for the whole French IC dataset corresponding to the

Alpine and CB GIG types. Theses relationships vary according to the types, but are evident for all the types (except for the basin IPOS in the R-C1 dataset).

The basin IPOS is well correlated to the urban and agricultural pollution, while the corridor IPOS is not. Thus the correlation between the average ICMi and the corridor IPOS is a response of the fauna to the physical pressures exerted on the river margins.

The relatively low determination coefficients observed in some relationships can be explained by two reasons :

- The land cover pressure index is a general descriptor of the driving forces, mainly urban and agricultural, acting at the basin and riparian corridor scales, and thus cannot represent with a high precision the actual chemical and hydro-morphological pressures exerted on the river site.
- The biological data used here come from the French monitoring network, thus adding an important spatial and temporal variability when compared to smaller datasets especially taken for a scientific project, like AQEM-STAR. (This is evidenced also by the comparison between the AQEM-STAR and ARPA monitoring network datasets in Italy, see annex 2.2.5.2)

Taking into account these limitations, these results demonstrate that the ICM index used in the IC exercise is actually responding to a wide range of combined human pressures, and is relevant to evaluate the invertebrates ecological status in European rivers.

Lyon, october 26th, 2006