Proposal of a new method: the multimetric index based on macroinvertebrate fauna for the ecological assessment of French wadeable rivers (I_2M_2).

1 SUMMARY

The Water Framework Directive 2000/60/EC (WFD; European Council 2000) requires the assessment of the ecological status of European water bodies. This assessment should be performed using information from different biological quality elements including the benthic macroinvertebrate communities. The assessment of macroinvertebrate communities should integrate the information provided by metrics related to (i) taxonomic composition, (ii) abundances, (iii) the ratio of disturbance sensitive to insensitive taxa and (iv) diversity (Annex V of WFD; European Council, 2000). These metrics should react to measurable stressors related to water quality and hydro-morphological pressures.

The French macroinvertebrate index currently used, and successfully intercalibrated (EU commission decision of 20 September 2013) is the French IBGN (AFNOR, 2004). Nevertheless, this index presents several limitations:

- it is not representative of the whole benthic substratum distribution within the river reach,
- it does not fully integrate information on abundances and diversity,
- it mainly focuses on the detection of organic pollution and thus demonstrates weak responses to other pressures, notably hydromorphology.

The new macroinvertebrate-based French index (I_2M_2) was developed to be both fully WFD-compliant (see Table 1) and sensitive to a larger range of anthropogenic pressures (Mondy *et al.*, 2012, Appendix A). After several years of testing on more than 10,000 samples over more than 1,800 sites, the I_2M_2 is now available in a stabilized version, validated at the national level (Usseglio-Polatera *et al.*, 2016). The index has demonstrated a significant gain in sensitivity to pressures compared to the former

IBGN, making it a much more reliable bioassessment method based on benthic macroinvertebrates for French rivers (Mondy *et al.*, 2012; Usseglio-Polatera *et al.*, 2016). It is a multimetric index based on five individual metrics: *1) Shannon index, 2) ASPT score, 3) the relative frequency of polyvoltine organisms in the assemblage, 4) the relative frequency of ovoviviparous organisms in the assemblage and 5) taxonomic richness.* Table 1 summarizes the validation of the WFD requirements by the I_2M_2.

Table 1: I_2M_2 WFD-compliance checking. Codes of metrics: 1) Shannon index, 2) ASPT score, 3) the relative frequency of polyvoltine organisms in the assemblage, 4) the relative frequency of ovoviviparous organisms in the assemblage and 5) taxonomic richness.

Compliance Criteria	Compliance Checking
Ecological status is classified by one of **five classes** (high, good, moderate, poor and bad)	**Yes**
High, good and moderate ecological status are set in line with the WFD's **normative definitions** (**Boundary setting procedure**)	**Yes**
All relevant parameters indicative of the biological quality element are covered. A combination rule to combine parameter assessment into BQE assessment has to be defined. If parameters are missing, Member States need to demonstrate that the method is sufficiently indicative of the status of the QE as a whole	**Yes - abundance** (covered by metrics 1, 3 & 4), **taxonomic composition** (covered by metrics 1, 2, 3, 4 & 5), **diversity** (covered by metrics 1, 2 & 5) and **pollution sensitivity** (covered by metrics 2, 3 & 4)
Assessment is adapted to **intercalibration common types** that are defined in line with the typological requirements of the Annex II WFD and approved by WG ECOSTAT	**Yes** (types R-CB, R-A1, R-A2 & R-M124)
The water body is assessed against type-specific near-natural reference conditions	**Yes**
Assessment results are expressed as **EQRs**	**Yes**
Sampling procedure allows for representative information about water body quality/ecological status **in space and time**	**Yes** (French standard AFNOR, 2016)
All data relevant for assessing the **biological parameters** specified in the WFD's normative definitions are covered by the **sampling procedure**	**Yes** (French standards AFNOR, 2010, 2016)
Selected taxonomic level achieves adequate **confidence and precision** in classification	**Yes** (French standard (AFNOR, 2010) providing the taxonomic level to reach for all potential taxa)

2 I_2M_2 METHODOLOGY OF DEVELOPMENT

2.1 DATA SET

The I_2M_2 was developed using a dataset covering 66 stream types, 1267 streams, 1860 reaches and 10074 sampling events. This dataset contains (i) general information about sampling events (e.g. sampling date, site coordinates, river type, ecoregion), (ii) information on the macroinvertebrate community (list and abundances of taxa) and (iii) information on the quality classes related to the anthropogenic pressures potentially acting on the communities (Table 2).

Table 2. Anthropogenic pressures taken into account for the I_2M_2 development

Water quality	**Hydro-morphological alterations**
Organic matter	Transportation facilities
Nitrates	Riverine vegetation
Nitrogen compounds (except Nitrates)	Urbanization (at the 1 km upper reach scale)
Phosphorous compounds	Clogging risk
Suspended matter	Hydrological instability
Acidification	Catchment anthropization
Mineral micropollutants	Straightening
Pesticides	
PAH	
Organic micropollutants (others)	

Water quality characterization of reaches was performed considering a variable number of parameters among 173 parameters distributed in ten chemical pressure categories (Table 2, cf. Mondy *et al.*, 2012 for the list of individual parameters) for which information on quality class boundaries was available in the French water quality assessment system (i.e. Water Quality Evaluation System or SEQ-Eau; Oudin & Maupas, 2003). The water quality status of a given reach at the macroinvertebrate sampling date was estimated by averaging chemical measures from this reach during the six months before faunal sampling [i.e. 4.16 (± 2.17) measures available, in average, on this period]. Land use and hydromorphology characterization were performed considering ten parameters distributed in seven habitat degradation pressure categories (cf. Table 2 and Mondy *et al.*, 2012 for the list of individual parameters). Individual habitat degradation parameters were measured using ESRI's ArcGis 9.2 software (ESRI, 2006). For each available parameter, pressure level was assessed by comparing the parameter measure with the threshold delimiting 'low' to 'moderate' pressure levels (thresholds are given in Mondy *et al.*, 2012). The pressure level allocated to a given reach for a given pressure category was the worst state (i.e. impaired or not) allocated to this reach by individual parameters [between 1 (nitrates) to 71 (pesticides) parameters according to the pressure type] taken into account by this pressure category.

For each sampling event, a large number of biological metrics (biotic indices, taxonomic metrics, metrics related to bio-ecological traits) was calculated. The data used to build the I_2M_2 thus corresponded to 10074 samples, each described by more than 2500 "metric x sample combination(s)" and quality classes for ten types of water quality pressures and seven types of hydro-morphological alterations (Table 2).

2.2 PRINCIPLES

The I_2M_2 is a multimetric index calculated as the arithmetic average of 17 sub-indices (one per pressure type p); each sub-index being composed of the same N individual metrics :

$$I_2M_2 = \frac{\sum_p^{17} i_2m_2^p}{17} \quad \text{(Eq. 1)}$$

with:

$$i_2m_2^p = \frac{\sum_m^N (DE_m^p \times EQR_m)}{\sum_m^N DE_m^p} \quad \text{(Eq. 2)}$$

$$EQR_m = \frac{obs_m - worst_m}{best_m^t - worst_m} \quad \text{(Eq. 3)}$$

The transformation of raw metric values in ecological quality ratios (EQR; Hering *et al.*, 2006) was done by normalizing the observed values (obs_m) after accounting for the worst value observed at the national scale ($worst_m$) and the stream-type specific reference value observed in the reference sites ($best_m^t$). The 5th and 95th percentiles of the distribution of values for a given metric, were used as 'best' or 'worst' values instead of the highest/lowest values to discard metric values of outliers (Ofenböck *et al.*, 2004). **The pressure-specific discrimination efficiency of a given metric (DE_m^p; Ofenböck *et al.*, 2004) corresponds to the proportion of samples from disturbed sites with EQR values lower than the 25th percentile of the EQR values distribution observed in the reference sites.**

The selection of metrics was performed in order to select those with (i) the ability to respond to a large range of pressures, (ii) a good average DE over the 17 pressure types and (iii) a low variability in type-specific reference sites (to ensure that the I_2M_2 variability is more linked to anthropogenic pressures than to natural variability). Moreover, the selected metrics should bring different taxonomic/functional information about the community, and therefore not be redundant. The I_2M_2 building procedure is described in Figure 1. Following this process, a set of metric combinations was obtained and the combination exhibiting the best trade-off between (i) stability in reference sites, (ii) robustness, (iii) discrimination efficiency over pressure types and (iv) WFD-compliance was selected. The I_2M_2 is thus composed of five metrics: (i) the Shannon index (Shannon, 1948), (ii) the ASPT score (Armitage *et al.*, 1983), the relative frequencies of (iii) polyvoltine and (iv) ovoviviparous organisms (Usseglio-Polatera *et al.*, 2000) in the communities and (v) the taxonomic richness. The REFCOND recommendations, in particular the pressure screening criteria and the definitions of class boundaries were considered carefully during the whole process.

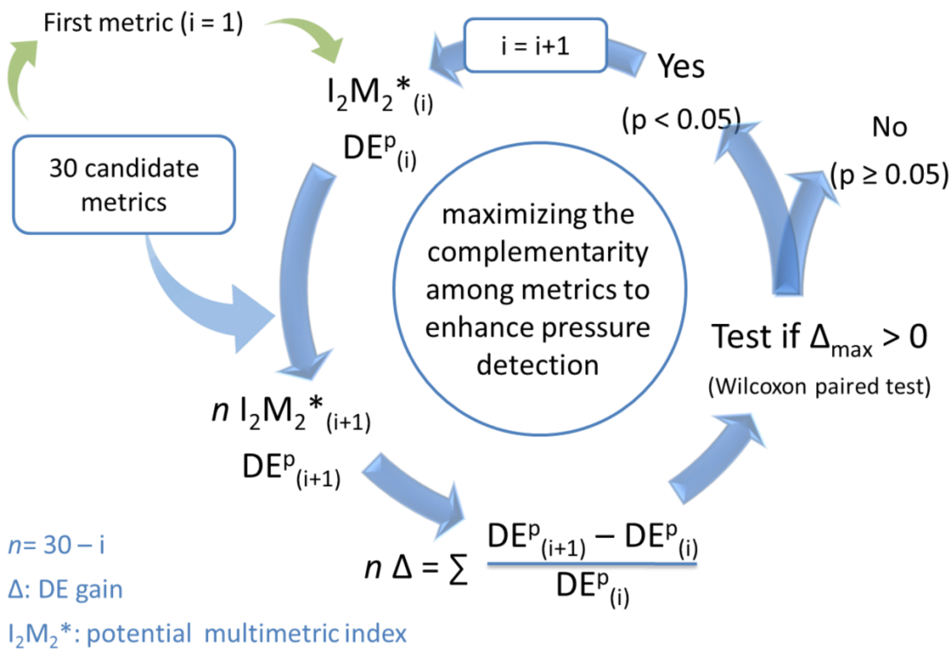

Figure 1. Design of the iterative procedure used to build the I_2M_2 from a selection of candidate metrics. This procedure was repeated using each of the pre-selected metrics as the first metric to include in the I_2M_2. At each step, the procedure tested if the inclusion of an additional metric in the multimetric index resulted in a significant improvement of the index DE (i.e. significantly increased the global ability of the index to detect disturbed situations).

2.3 RESULTS

The new I_2M_2 index has proved to significantly better discriminate reference from disturbed sites than the former IBGN index, with a gain in discrimination efficiency of 26% in average over the 17 pressure type (Figure 2).

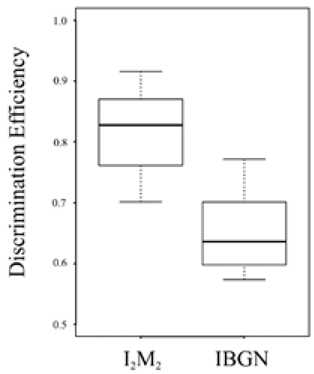

Figure 2. Discrimination Efficiency value distribution for I_2M_2 compared to IBGN over the 17 pressure types considered. Adapted from Mondy et al. (2012). ***The gain in discrimination efficiency is of <u>26% in average</u> over the 17 pressure types.***

This index has also proved to be stable in reference sites, efficient in detecting a large variety of pressures and robust (i.e. keeping its efficiency in new datasets) (Figure 3 & Figure 4).

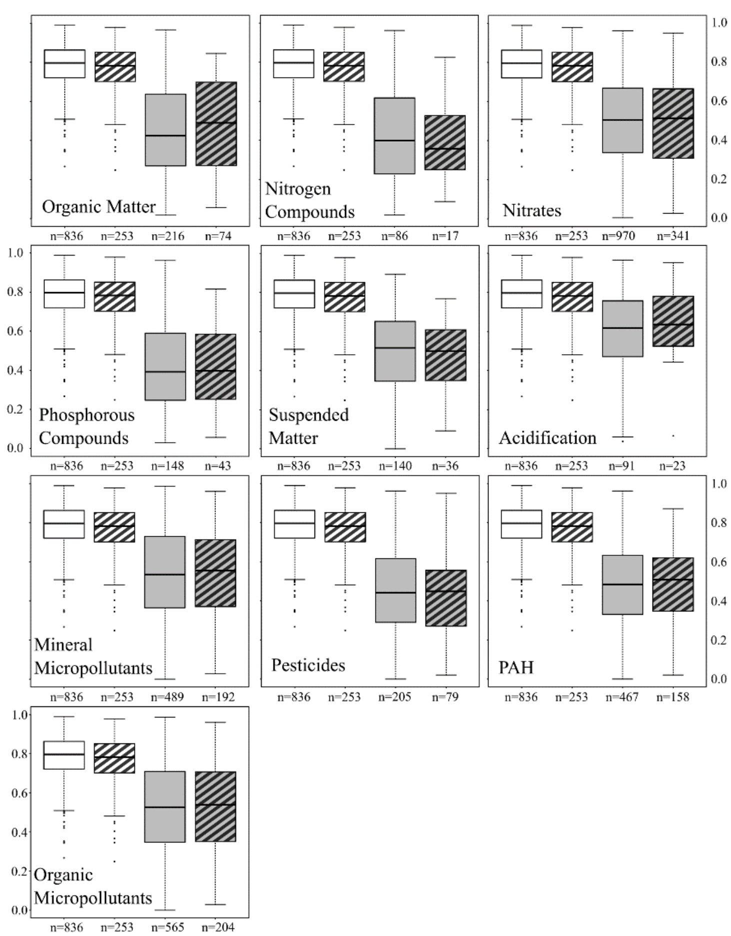

Figure 3. I_2M_2 value distribution in reference sites (white boxes) and disturbed sites (grey boxes) for 10 water quality pressure categories and two datasets (calibration: plain boxes; test: dashed boxes). 'n' represents the number of faunal samples considered in each group. Adapted from Mondy et al. (2012).

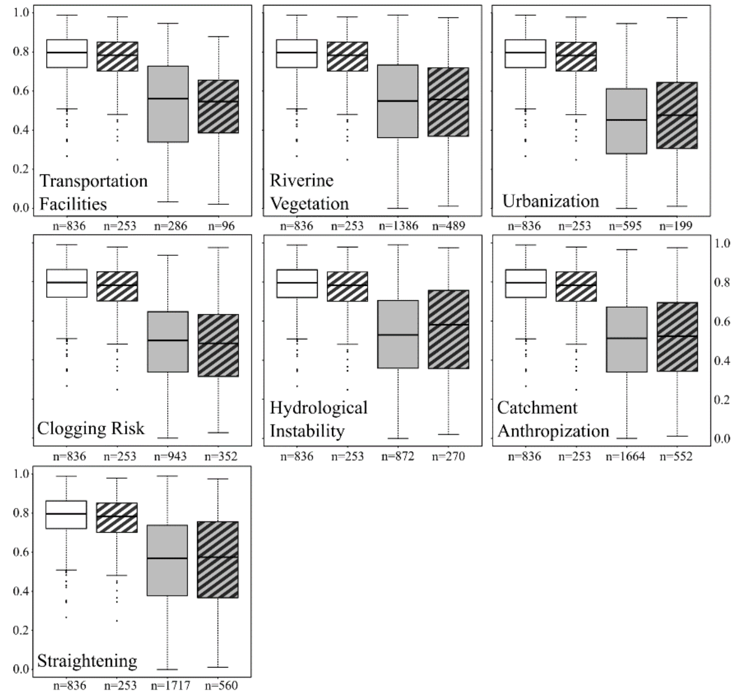

Figure 4. I_2M_2 value distribution in reference sites (white boxes) and disturbed sites (grey boxes) for 7 habitat degradation pressure categories and two datasets (calibration: plain boxes; test: dashed boxes). 'n' represents the number of faunal samples considered in each group. Adapted from Mondy et al. (2012).

Finally, this new I_2M_2 index has proved to respond more efficiently to increasing anthropogenic pressure levels than the former IBGN as illustrated with the level of catchment anthropization (Figure 5).

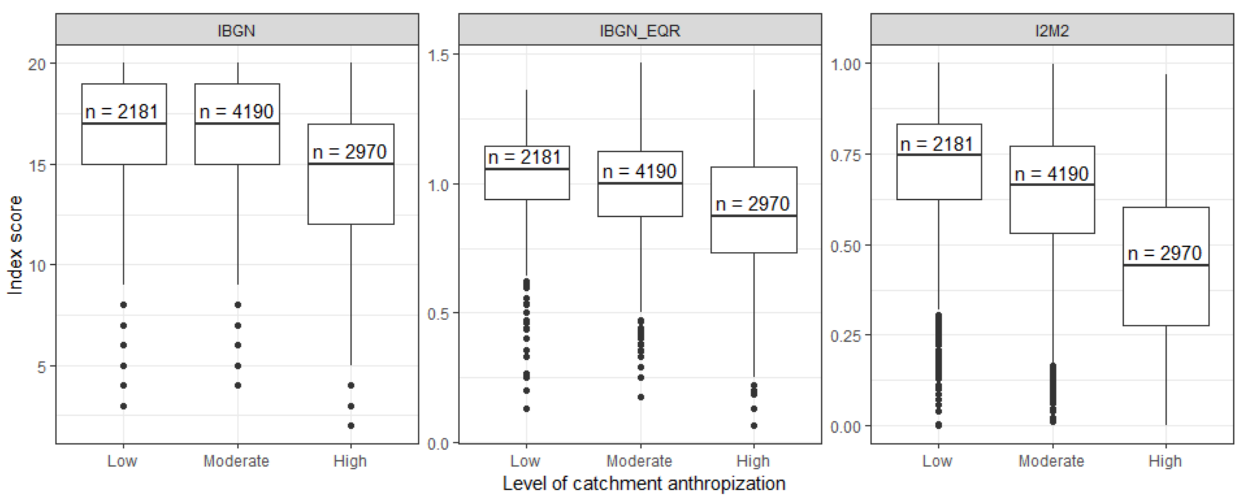

Figure 5. Distribution of the IBGN and I_2M_2 values along a gradient of increasing catchment anthropization level reflecting the overall anthropogenic pressure level of sites acting, at the catchment scale. 'n' represents the number of faunal samples considered in each group

3 I_2M_2 INTERCALIBRATION

3.1 DATA SET

The I_2M_2 was intercalibrated using the same qualified data set as for its development (covering 66 stream types, 1267 streams, 1860 reaches and 10074 sampling events). This data set fulfills the requirements listed in the CIS-Guidance Document on the Intercalibration Process (European Commission, 2011):

- It sufficiently covers the geographical area in which the common types occur within France;
- It encompasses sampling sites covering the entire gradient of the pressure to be intercalibrated, and hence the complete ecological quality gradient ranging from high to poor ecological status;
- It contains non-biological (environmental) and biological data to conduct pressure-impact analyses. The non-biological data are contemporaneous with the accompanying biological data in time and space.

This data set covers the European common river types from the following Geographical Intercalibration Groups (GIGs): Central/Baltic, Alpine and Mediterranean.

3.2 IC METRICS & REFERENCE CONDITIONS

The actually intercalibrated French evaluation method (IBGN) requires data collected with less demanding field and laboratory protocols (in terms of both sampling effort and taxonomic identification level) compared to the new I_2M_2 method. Therefore, the I_2M_2 should be considered as a new evaluation method and not a revision of the IBGN and the intercalibration (IC) option 2 should be applied.

This option requires the use of common IC metrics that enable a GIG-wide comparison of classification results (European Commission, 2011). Regarding the macroinvertebrate biological quality element, these common metrics (ICMs) and the multimetric indices (ICMi) they constitute are different across the several GIGs corresponding to the river types where the I_2M_2 method should be applied (Table 3). The data from the national data set are sufficient to calculate the required ICMs.

Table 3. GIG-specific IC indices and their constitutive IC metrics (see Van de Bund, 2009 for more details). $metric_{EQR}$ refers to EQR-transformed metric values.

Central/Baltic & Mediterranean	Alpine
ICMi = 0.333 * $ASPT_{EQR}$ +	ICMi = 0.25 * $Sfam_{EQR}$ +
0.266 * $log10selEPTD_{EQR}$ +	0.25 * $SfamEPT_{EQR}$ +
0.067 * $oneMinusGold_{EQR}$ +	0.25 * $Ssens_{EQR}$ +
0.167 * $Sfam_{EQR}$ +	0.25 * $ibericASPT_{EQR}$
0.083 * $SfamEPT_{EQR}$ +	
0.083 * $ShannonDiversity_{EQR}$	

For a given ICM, the transformation of raw values to Ecological Quality Ratios (EQR) was performed dividing the raw values by the median of the ICM values observed in reference sites of the corresponding common river type from the qualified national data set. The qualification of reference sites was performed following the REFCOND recommendations (European Commission, 2003). This qualification was performed using two indicators of land use (intensive agriculture and urbanization; at least one 'reference' status and one 'limit' status tolerated) and five indicators of water quality (BOD5, oxygen saturation, concentrations in ammonium, nitrates and orthophosphates; all in 'reference' status) (Table 4).

Table 4. Rules used to qualify a site as 'reference', 'limit' (only for land use criteria) and 'not reference'. x refers to the site-averaged values of the different measured parameters.

		Intensive agriculture (% of the catchment)			Urbanization (% of the catchment)			DBO5	
GIG	River Type	reference	limit	not reference	reference	limit	not reference	reference	not reference
Alpine	1, 2	x ≤ 20	20 < x ≤ 50	x > 50	x ≤ 0.4	0.4 < x ≤ 0.8	x > 0.8	x ≤ 2	x > 2
Central/Baltic	1	x ≤ 20	20 < x ≤ 50	x > 50	x ≤ 0.4	0.4 < x ≤ 0.8	x > 0.8	x ≤ 2.4	x > 2.4
Central/Baltic	2	x ≤ 20	20 < x ≤ 50	x > 50	x ≤ 0.4	0.4 < x ≤ 0.8	x > 0.8	x ≤ 2.4	x > 2.4
Central/Baltic	3	x ≤ 20	20 < x ≤ 50	x > 50	x ≤ 0.4	0.4 < x ≤ 0.8	x > 0.8	x ≤ 2	x > 2
Central/Baltic	4, 5, 6	x ≤ 20	20 < x ≤ 50	x > 50	x ≤ 0.4	0.4 < x ≤ 0.8	x > 0.8	x ≤ 2.4	x > 2.4
Mediterranean	1, 2, 3, 4	x ≤ 20	20 < x ≤ 50	x > 50	x ≤ 0.4	0.4 < x ≤ 0.8	x > 0.8	x ≤ 2.4	x > 2.4

Table 4 (continued)

		Oxygen saturation (%)		ammonium (mg/L)		nitrates (mg/L)		orthophosphates (mg/L)	
GIG	River Type	reference	not reference	reference	not reference	reference	not reference	reference	not reference
Alpine	1, 2	95 ≤ x ≤ 105	x < 95 OR x > 105	x ≤ 0.06	x > 0.06	x ≤ 6	x > 6	x ≤ 0.06	x > 0.06
Central/Baltic	1	95 ≤ x ≤ 105	x < 95 OR x > 105	x ≤ 0.12	x > 0.12	x ≤ 6	x > 6	x ≤ 0.12	x > 0.12
Central/Baltic	2	95 ≤ x ≤ 105	x < 95 OR x > 105	x ≤ 0.06	x > 0.06	x ≤ 6	x > 6	x ≤ 0.09	x > 0.09
Central/Baltic	3	95 ≤ x ≤ 105	x < 95 OR x > 105	x ≤ 0.06	x > 0.06	x ≤ 6	x > 6	x ≤ 0.06	x > 0.06
Central/Baltic	4, 5, 6	95 ≤ x ≤ 105	x < 95 OR x > 105	x ≤ 0.12	x > 0.12	x ≤ 6	x > 6	x ≤ 0.12	x > 0.12
Mediterranean	1, 2, 3, 4	90 ≤ x ≤ 110	x < 90 OR x > 110	x ≤ 0.12	x > 0.12	x ≤ 6	x > 6	x ≤ 0.12	x > 0.12

Applying the rules described in Table 4, and following the river type grouping used during the previously completed IC procedure (Van de Bund, 2009), we qualified 151 reference sites corresponding to 1027 sampling events (Table 5). For each European common river type (or group of river types), we obtained a minimum number of 15 sites as recommended to reliably perform the IC (European Commission, 2011).

Table 5. Distribution of reference sites and sampling events among the four groups of European common river types.

River types	Number of reference sites	Number of reference sampling events
R-A1	15	106
R-A2	49	307
R-CB	19	136
R-M124	67	473

3.3 IC FEASIBILITY CHECK

For each river type, the ICMi scores were regressed against the I_2M_2 scores calculated using the same data. These regressions are summarized in Table 6 and Figure 6. For the IC to be reliably doable, the relationship between the IC metric (here the ICMi) and the national method (I_2M_2) should be strong (with **a significant Pearson's correlation coefficient above 0.5**) and neither too shallow nor too steep (i.e. **a significant slope comprised between 0.5 and 1.5**).

Table 6. IC feasibility analysis for the I_2M_2 in five European common river types.

	Regression			Pearson's correlation	
River type	**R^2**	**Equation**	**p (slope)**	**R**	**p**
R-A1	0.706	y = 0.855x + 0.373	< 0.001	0.840	< 0.001
R-A2	0.614	y = 1.410x - 0.063	< 0.001	0.784	< 0.001
R-CB	0.768	y = 0.572x + 0.528	< 0.001	0.876	< 0.001
R-M124	0.823	y = 0.763x + 0.457	< 0.001	0.907	< 0.001

Following these criteria, the I_2M_2 can be intercalibrated in all the river types considered exhibiting strong linear relationships with the ICMi with slopes comprised between 0.57 (R-CB) and 1.41 (R-A2) (Table 6 and Figure 6).

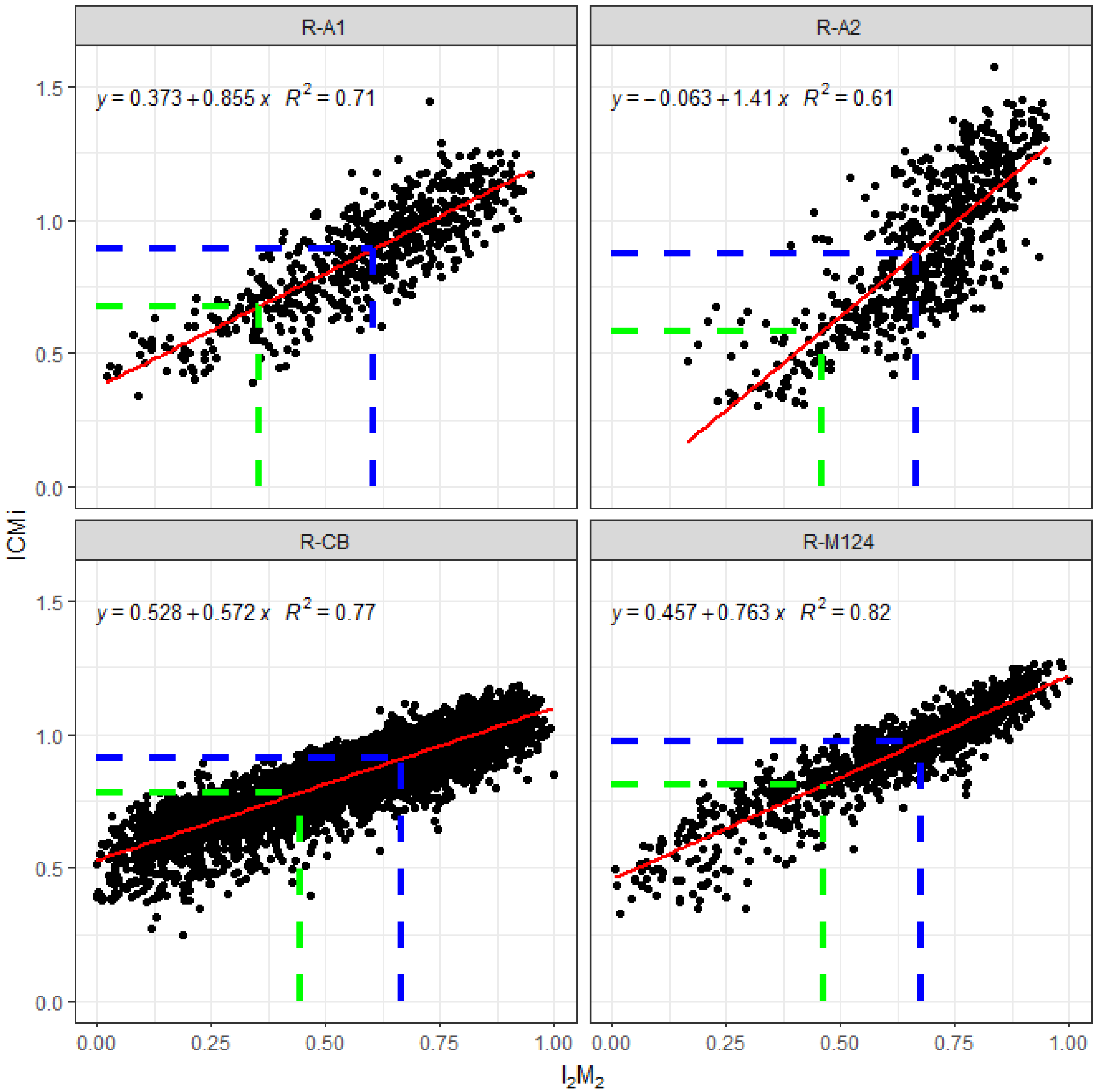

Figure 6. Linear relationship between the I_2M_2 values and the ICMi values for the four groups of European common river types considered (R-A1, R-A2, R-CB and R-M124). The black dots represent the observations, the plain red lines the linear regression models, the dashed blue and green lines the High-Good and Good-Moderate quality class limits, respectively.

3.4 I_2M_2 QUALITY CLASS LIMITS

Since the I_2M_2 was developed to take into account river type-specificities through the use of river-type specific reference conditions during the EQR transformation, a common initial set of quality class limits was defined. The initial High-Good (HG) quality class limit was defined as the 25th percentile of the I_2M_2 distribution in the national set of sites qualified as “reference” following the REFCOND recommendations (Table 4). This HG limit was then divided to obtain four quality classes of equal sizes.

Following a previous work, the HG and Good-Moderate (GM) limits have been adjusted depending on the corresponding IC river types whereas it was kept unchanged for the other national river types that were not considered during the IC process (Table 7).

Table 7. Quality class limits for the I_2M_2 index in the different European common river types. (I_2M_2 units/ICMi units)

River type	HG	GM	MP	PB
R-A1	0.605/0.890	0.354/0.675	0.236/0.574	0.118/0.474
R-A2	0.665/0.872	0.460/0.584	0.306/0.367	0.153/0.152
R-CB	0.665/0.908	0.443/0.781	0.295/0.697	0.148/0.613
R-M124	0.676/0.973	0.464/0.811	0.310/0.694	0.155/0.576
other	0.665/-	0.498/-	0.332/-	0.166/-

The HG and GM limits in I_2M_2 units were then converted in ICMi units (Figure 6 and Table 7) using the regression models built for each river type (Figure 6 and Table 6). The limits in ICMi units were then compared to the global mean view of the HG and GM quality class limits defined in the completed IC exercises. The direction of the deviation between the proposed and the global mean view was then determined as well as the amount of this deviation expressed as the proportion of the corresponding class width.

According to the WFD intercalibration manual (Willby *et al.*, 2014), the following rules are followed to test the comparability criteria:

- The proposed limit falls below the global mean view:
    - The deviation is smaller than ¼ of class width: comparability criteria are met;
    - The deviation is larger than ¼ of class width: the proposed limit has to be raised.
- The proposed limit falls above the global mean view:
    - The deviation is smaller than ¼ of class width: comparability criteria are met;
    - The deviation is larger than ¼ of class width: no obligation to lower the proposed limit.

Table 8. Deviation of the proposed I_2M_2 quality class limits from the European global mean view defined during the completed IC exercises. The deviation is expressed as proportions of class width (negative deviation: proposed limit below global mean view; positive deviation: proposed limit above global mean view)

River Type	HG	GM
R-A1	0.140	-0.163
R-A2	0.111	-0.160
R-CB	-0.156	0.250
R-M124	0.549	0.915

The results of the comparison of the proposed I_2M_2 quality class limits to the global mean views of the intercalibrated methods based on benthic macroinvertebrates are summarized in Table 8 and Figure 7. The proposed limits for the river types belonging to the Alpine GIG fully met the comparability criteria (absolute deviation smaller than ¼ of class width). The HG limit for the Central Baltic river types also met the comparability criteria. The quality class limits proposed for Mediterranean river types are larger (i.e. more strict) than the European global mean view.

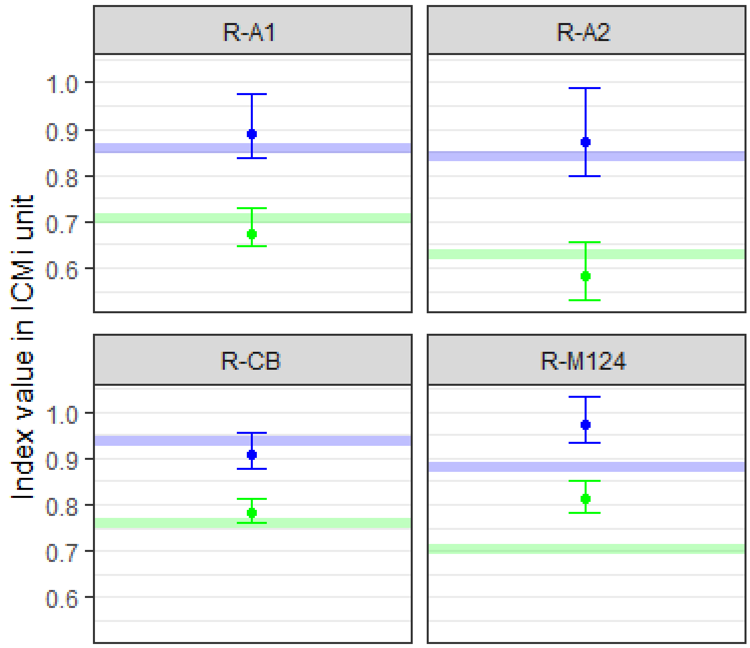

Figure 7. Comparison of the proposed I_2M_2 quality class limits (blue and green dots: High-Good and Good-Moderate limits, respectively) to the global mean views (thick horizontal lines) obtained during the completed IC exercises for the different river types. The whiskers represent +/- ¼ of class width.

4 EVOLUTION OF THE EVALUATION

To assess the consequences of the new evaluation method (I_2M_2 *vs.* intercalibrated IBGN) on the ecological classification, we looked at the distribution of river sites among the five quality classes (i.e. High, Good, Moderate, Poor and Bad) depending on the index used, on two separate time periods (2008-2010 and 2011-2013) (Table 9 & Table 10). For a given site, the evaluation was performed by averaging the index scores over the three year period and allocating the quality class according to the corresponding class limits. We only considered sites belonging to the surveillance monitoring network and for which both the IBGN and the I_2M_2 quality classes can be assigned.

Table 9. Distribution (in %) of French river sites belonging to surveillance monitoring network for the period 2008 to 2010 and allocated to the five different ecological quality classes depending on the index used, at the national level and in each of the six French water districts: AG: Adour-Garonne, AP: Artois-Picardie, LB: Loire-Bretagne, RM: Rhin-Meuse, RMC: Rhône-Méditerranée & Corse and SN: Seine-Normandie.

	France (n = 1287)		AG (n = 282)		AP (n = 55)		LB (n = 348)	
	IBGN	**I_2M_2**	IBGN	I_2M_2	IBGN	I_2M_2	IBGN	I_2M_2
High	**56.3**	**38.2**	41.1	42.6	32.7	1.8	60.1	42.0
Good	**23.5**	**33.1**	32.3	30.1	27.3	9.1	23.9	32.5
Moderate	**16.5**	**16.0**	22.3	15.2	32.7	41.8	13.2	15.8
Poor	**3.4**	**8.8**	3.5	7.8	7.3	29.1	2.9	8.3
Bad	**0.3**	**3.9**	0.7	4.3	0.0	18.2	0.0	1.4

Table 9 (continued).

	RM (N = 80)		RMC (N = 347)		SN (N = 175)	
	IBGN	I_2M_2	IBGN	I_2M_2	IBGN	I_2M_2
High	61.3	25.0	61.1	38.3	69.1	41.1
Good	17.5	46.2	21.0	35.7	14.9	35.4
Moderate	10.0	12.5	14.7	15.9	14.9	11.4
Poor	10.0	6.2	2.9	8.1	1.1	7.4
Bad	1.2	10.0	0.3	2.0	0.0	4.6

Table 10. Distribution (in %) of French river sites belonging to surveillance monitoring network for the period 2011 to 2013 and allocated to the five different ecological quality classes depending on the index used, at the national level and in each of the six French water districts: AG: Adour-Garonne, AP: Artois-Picardie, LB: Loire-Bretagne, RM: Rhin-Meuse, RMC: Rhône-Méditerranée & Corse and SN: Seine-Normandie.

	France (n = 1213)		AG (n = 261)		AP (n = 54)		LB (n = 320)	
	IBGN	**I_2M_2**	IBGN	I_2M_2	IBGN	I_2M_2	IBGN	I_2M_2
High	**58.3**	**35.1**	59.8	46.7	24.1	0.0	59.7	36.6
Good	**24.0**	**35.0**	24.1	28.0	27.8	22.2	25.6	37.5
Moderate	**15.0**	**15.6**	13.0	13.8	42.6	22.2	13.1	13.4
Poor	**2.5**	**10.1**	2.7	8.0	5.6	35.2	1.6	9.7
Bad	**0.2**	**4.2**	0.4	3.4	0.0	20.4	0.0	2.8

Table 10 (continued).

	RM (N = 75)		RMC (N = 339)		SN (N = 164)	
	IBGN	I_2M_2	IBGN	I_2M_2	IBGN	I_2M_2
High	58.7	24.0	53.4	33.6	74.4	33.5
Good	20.0	40.0	27.1	35.7	14.6	41.5
Moderate	12.0	16.0	16.8	19.8	10.4	11.6
Poor	8.0	10.7	2.4	8.8	0.6	8.5
Bad	1.3	9.3	0.3	2.1	0.0	4.9

Due to the better discrimination efficiency of the new I_2M_2 compared to the IBGN regarding to pressures, the use of the I_2M_2 led to changes in the biological quality class distribution of the French river sites for both assessment periods, with less sites corresponding to the good and high quality and more sites considered as significantly impaired (medium, poor and bad quality) (Figure 8). The results were similar for each French hydrographic basin considered independently.

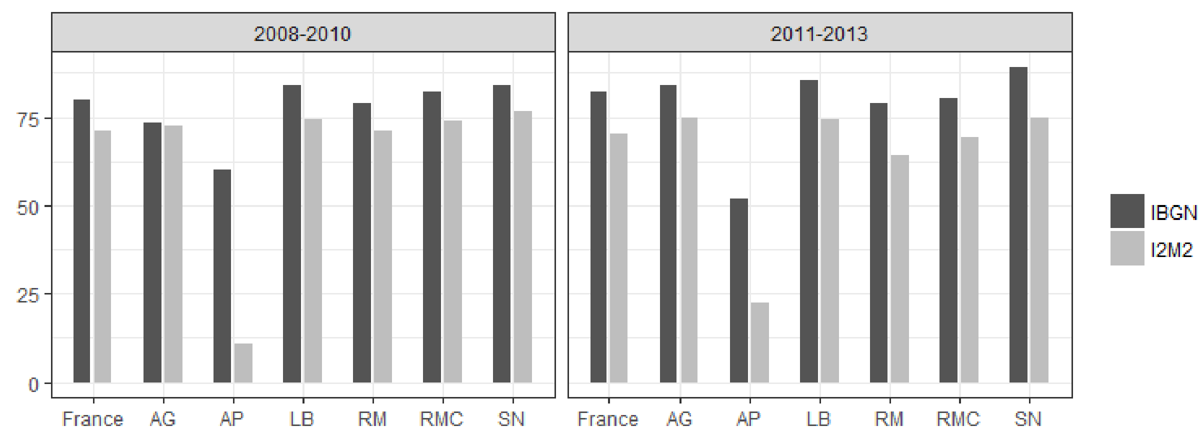

Figure 8. Proportion of the sites classified at least in 'Good' (i.e. 'Good' or 'High') ecological status depending on the index used (dark grey: IBGN index, light grey: I_2M_2 index). The results are given at the national scale (France) and detailed for each water district (AG: Adour-Garonne, AP: Artois-Picardie, LB: Loire-Bretagne, RM: Rhin-Meuse, RMC: Rhône-Méditerranée & Corse and SN: Seine-Normandie) for each of the two investigated period (2008-2010 and 2011-2013).

5 CONCLUSION

The new I_2M_2 is a multimetric index developed to be sensitive to a large panel of anthropogenic pressures related to water quality and hydro-morphological alterations. It includes five metrics and complies with all the WFD requirements: (i) it is expressed in EQR, (ii) it uses a type-specific reference approach, (iii) it includes metrics related to the taxonomic composition, the taxon abundances, the ratio of disturbance sensitive to insensitive taxa and the diversity of benthic invertebrate communities, (iv) it is based on a specifically developed sampling procedure.

The I_2M_2 presents a very good correlation with the common metric ICMi in all the GIGs covered by the French method allowing the intercalibration of the new French index. Moreover, the proposed quality class boundaries complies with the European global mean views of the High-Good and the Good-Moderate limits in the Alpine, Central-Baltic and Mediterranean GIGs.

The better sensitivity of the I_2M_2 compared to the IBGN is especially clear when looking at the discrimination efficiency (Figure 2), at the response to increasing anthropogenic pressures (Figure 5) or at the site distribution among the five ecological quality classes (Table 9 &Table 10) for both indices. This better sensitivity to anthropogenic pressures results in a more severe evaluation of the ecological quality of French streams as illustrated by the proportion of sites no longer with at least a 'Good' status when using the I_2M_2 instead of the former IBGN (Figure 8).

6 LITERATURE

AFNOR (2016) *NF T90-333: Qualité de l'eau - Prélèvement des macro-invertébrés aquatiques en rivières peu profondes*. AFNOR, Paris.

AFNOR (2004) *NF T90-350: Qualité de l'eau - Détermination de l'indice biologique global normalisé (IBGN)*. AFNOR, Paris.

AFNOR (2010) *XP T90-388: Qualité de l'eau - Traitement au laboratoire d'échantillons contenant des macro-invertébrés de cours d'eau*. AFNOR, Paris, France.

Armitage P.D., Moss D., Wright J.F. & Furse M.T. (1983) The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research* **17**, 333–347.

ESRI (2006) *ArcGis 9.2*. ESRI (Environmental Systems Research Institute), Redlands.

European Commission (2003) *Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance document No. 10, rivers and lakes - typology, reference conditions and classification systems*. Office for official publications of the European Communities, Luxembourg.

European Commission (2011) *Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance document No. 14. Guidance on the intercalibration process 2008-2011*. Office for official publications of the European Communities, Luxembourg.

European Council (2000) *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*. Office for official publications of the European Communities, Brussels.

Hering D., Feld C.K., Moog O. & Ofenböck T. (2006) Cook book for the development of a Multimetric Index for biological condition of aquatic ecosystems: Experiences from the European AQEM and STAR projects and related initiatives. *Hydrobiologia* **566**, 311–324.

Mondy C.P., Villeneuve B., Archaimbault V. & Usseglio-Polatera P. (2012) A new macroinvertebrate-based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands: A taxonomical and trait approach. *Ecological Indicators* **18**, 452–467.

Ofenböck T., Moog O., Gerritsen J. & Barbour M. (2004) A stressor specific multimetric approach for monitoring running waters in Austria using benthic macro-invertebrates. *Hydrobiologia* **516**, 251–268.

Oudin L.C. & Maupas D. (2003) *Système d'évaluation de la qualité de l'eau des cours d'eau (Seq-Eau)*. Ministère de l'Ecologie et du Développement Durable - Agences de l'Eau, Paris.

Shannon C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379–423, 623–656.

Usseglio-Polatera P., Bournaud M., Richoux P. & Tachet H. (2000) Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits. *Freshwater Biology* **43**, 175–205.

Usseglio-Polatera P., Larras F., Coulaud R. & Billoir E. (2016) *Finalisation I2M2, chantier Diagnostic multi-EQB et développement d'un indice pour les grands cours d'eau.* Université de Lorraine (LIEC), Metz.

Van de Bund W. (2009) *Water Framework Directive intercalibration technical report - Part 1: Rivers*. Offce for Offcial Publications of the European Communities, Luxembourg.

Willby N., Birk S., Poikane S. & van de Bund W. (2014) *Water Framework Directive Intercalibration Manual: Procedure to fit new or updated classification methods to the results of a completed intercalibration*. Office for official publications of the European Communities, Luxembourg.

7 APPENDIX A

Mondy C.P., Villeneuve B., Archaimbault V. & Usseglio-Polatera P. (2012) A new macroinvertebrate-based multimetric index (I2M2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands: A taxonomical and trait approach. *Ecological Indicators* **18**, 452–467.

Ecological Indicators 18 (2012) 452–467

Contents lists available at SciVerse ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

A new macroinvertebrate-based multimetric index (I_2M_2) to evaluate ecological quality of French wadeable streams fulfilling the WFD demands: A taxonomical and trait approach

Cédric Pascal Mondy[a,*], Bertrand Villeneuve[b], Virginie Archaimbault[c], Philippe Usseglio-Polatera[a]

[a] *Laboratoire des Interactions Ecotoxicologie Biodiversité Ecosystèmes (LIEBE), Université Paul Verlaine – METZ, CNRS UMR 7146, Campus Bridoux, Avenue du Général Delestraint, 57070 Metz, France*
[b] *Cemagref de Lyon, UR MALY, Pôle hydroécologie des cours d'eau ONEMA/Cemagref, Lyon Cedex, France*
[c] *Cemagref de Lyon, UR BEA, Lyon Cedex, France*

A R T I C L E I N F O

Article history:
Received 5 August 2011
Received in revised form
14 December 2011
Accepted 22 December 2011

Keywords:
Macroinvertebrates
WFD
Multimetric index
Pressure–impact relationship
Taxonomic metrics
Bio-ecological traits

A B S T R A C T

Following the Water Framework Directive (WFD) requirements, we designed a new multimetric index (I_2M_2) for the invertebrate-based ecological assessment of French wadeable streams. This index should be able to identify impaired reaches for 17 anthropogenic pressure categories potentially leading to water quality alteration or habitat degradation. Based on a national database, we defined an iterative procedure to select taxonomy- and trait-based metrics exhibiting the best trade-off between (i) high discrimination efficiency, (ii) low specificity and (iii) high stability in least impaired conditions. The I_2M_2, defined as the best combination of such metrics, has been composed by: (i) Shannon diversity index, (ii) original ASPT score, (iii) the relative abundance of polyvoltine taxa, (iv) the relative abundance of ovoviviparous taxa and (v) taxonomic richness. The I_2M_2 was tested against an independent data set. It exhibited good and robust pressure–impact relationship for all the pressure categories, correctly identifying in average 82% of reaches impaired by water quality alterations or habitat degradation. The I_2M_2 significantly improved the detection of impaired reaches by at least 17% for nitrogen compounds and up to 35% for organic micropollutants and clogging risk, when compared to the normalized French biotic index (IBGN). The I_2M_2 has been proposed for future use in the national biomonitoring of wadeable reaches in the context of the WFD implementation.

1. Introduction

The European Water Framework Directive (European Council, 2000) put for the first time the ecological quality in the very heart of the environmental policies of European member states. It requires that countries evaluate the quality of their water bodies using Biological Quality Elements (BQEs): i.e. fish, invertebrates, diatoms, plants and phytoplankton. Among BQEs, invertebrates have a long history as part of biomonitoring tools (Hellawell, 1986; Rosenberg and Resh, 1993; Bonada et al., 2006), being the most widely used biological group in freshwater bioassessment of human impact (Norris and Thorns, 1999; Hering et al., 2006a).

The WFD requires that bioassessment methods implicitly evaluate the ecological status of water bodies, by comparing BQEs between an observed vs. a reference situation. The reference situation should be representative of near natural conditions. Moreover, this comparison has to take into account the typology of water bodies and the metrics selected to evaluate the ecological status of water bodies, and has to regard abundance, diversity and pollution sensitivity of taxa (see annex 5 in European Council, 2000).

In France, the IBGN method (Indice Biologique Global Normalisé) has been used at the national scale and normalized since 1992 (revised in 2004, norm NF T 90-350 in AFNOR, 2004) but is no longer satisfying due to severe inconsistencies with WFD, e.g. the IBGN index is not type specific: the same scoring system and quality class boundaries are used for all types of rivers without considering "reference conditions". Moreover the IBGN sub-metrics [i.e. the faunal indicator group (FIG) and taxonomic richness] did not take into account taxon abundances.

To overcome the technical shortcomings of the French biotic index in the WFD implementation framework, the development of a new biotic index, i.e. the MultiMetric Invertebrate Index (I_2M_2), was decided by the French Ministry of Environment (MEDDTL).

Multimetric indices were first included in biomonitoring approaches with fish communities (Karr, 1981). They have been increasingly used (e.g. Kerans and Karr, 1994; Thorne and Williams, 1997; Buffagni et al., 2004; Böhmer et al., 2004a; Ofenböck et al., 2004; Gabriels et al., 2010) and have become major

* Corresponding author. Tel.: +33 387378431.
E-mail address: mondy@univ-metz.fr (C.P. Mondy).

1470-160X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ecolind.2011.12.013

Fig. 1. Map of the French hydroecoregions and location of the sampling sites. Black squares represent least impaired river reaches and open circles impaired river reaches.

tools in macroinvertebrate-based biomonitoring within the European WFD (e.g. Hering et al., 2004a; Lücke and Johnson, 2009). Indeed a multimetric index has the potential to simultaneously and efficiently evaluate the responses of benthic communities to different categories of pressure because its individual metrics could consider different attributes of communities that specifically respond to different categories of pressure (Karr and Chu, 1997). Several authors have searched for individual metrics that significantly respond to toxic contamination (Archaimbault et al., 2010) or hydromorphological alteration (Lorenz et al., 2004) and have included such metrics in biomonitoring tools. Nevertheless, few works have been done to combine in a single index, metrics able to detect a wide range of anthropogenic pressures at large spatial scale [but see, for example, Ofenböck et al., 2004; Buffagni et al., 2004 (organic contamination + hydromorphological alteration) or Böhmer et al., 2004a (organic contamination + acidification + hydromorphological alteration)].

In this work, we aimed at identifying biological metrics (based on taxonomy or life history traits) that significantly respond to 17 pressure categories potentially leading to water quality or habitat degradation. We selected metrics exhibiting the best trade-off between (i) high mean discrimination efficiency, (ii) low specificity and (iii) high stability in reference conditions. We searched for combinations of those metrics that could be relevant for pressure–impact identification in French wadeable stream-types and selected the best metric combination to build the new multimetric index (I_2M_2). We tested the discrimination efficiency, stability and robustness of this new index on a test data set. In the future intercalibration exercises the new French multimetric index will be compared and intercalibrated to European standards, e.g. with the commonly used European intercalibration multimetric index ICM_{Star} (Buffagni et al., 2006). We already tested the correlation of the I_2M_2 with the ICM_{Star} and compared its discrimination efficiency with those of the ICM_{Star} and the former French biotic index (IBGN).

2. Materials and methods

2.1. Data collection

Fieldwork was performed between 2004 and 2009 by 22 regional environmental agencies on a national network. Selected reaches were representative of 57 stream types of the French hydroecoregion-based typology (Wasson et al., 2002; MEDD, 2005; Chandesris et al., 2006) gathering most of the French wadeable rivers (1305 streams, 1725 reaches and 4132 sampling events; cf. Fig. 1 and Appendix A).

A 'development' data set was formed by randomly selecting 75% of the reaches (1293 reaches, 3112 samples) from the whole data set, while the remainder (i.e. 432 reaches, 1020 samples) was used as a 'test' data set.

Macroinvertebrate communities were sampled in all reaches with a common normalized protocol (Multi-Habitat Sampling, norm XP T 90-333 in AFNOR, 2009). During low flow conditions, twelve sample units per reach were performed on pre-defined mesohabitat types with a normalized Surber net (sampling area 0.05 m^2, mesh size 500 μm). Four sample units from 'marginal habitats' (i.e. with an individual share of less than 5% coverage) were selected according to their hosting capacity ('B1' group) and eight sample units were taken from 'major habitats' (i.e. with an individual share of at least 5% coverage). Four of these samples were selected according to their hosting capacity ('B2' group). The last four sample units were proportionally selected according to the relative coverage of major mesohabitats within the sampling reach ('B3' group), taking into account mesohabitats already sampled in group 'B2' (AFNOR, 2009). Sample units from the same group (B1, B2 or B3) were preserved together with formalin (4% final concentration). In the laboratory, invertebrates were sorted, counted and identified at the normalized taxonomic level [i.e. genus level except for Oligochaeta, some Diptera (mainly family), Trichoptera Limnephilidae, Coleoptera Dytiscidae and Hydrophilidae (sub family); norm XP T 90-388 in AFNOR, 2010; cf. Appendix B].

2.2. Reach characterization

Water quality characterization of reaches was performed considering a variable number of parameters among 173 parameters distributed in ten chemical pressure categories (cf. Table 1 and Supplementary material SI) for which information was available in the French water quality assessment system (i.e. Water Quality Evaluation System or SEQ-Eau; Oudin and Maupas, 2003). The water quality status of a given reach at the macroinvertebrate sampling date was estimated by averaging chemical measures from this reach during the six months before faunal sampling [i.e. 4.16 ($\pm$2.17) measures available, in average, on this period; a

Table 1
Water quality and habitat degradation pressure categories taken into account in this study.

Water quality	Habitat degradation
Organic matter	Transportation facilities
Nitrogen compounds, except nitrates	Riverine vegetation
Nitrates	Urbanization
Phosphorous compounds	Clogging risk
Suspended matter	Hydrological instability
Acidification	Catchment anthropization
Mineral micropollutants	Straightening
Pesticides	
PAH	
Organic micropollutants	

number varying according to the parameter and the reach taken into account].

Land use and hydromorphological characterization was performed considering ten parameters distributed in seven habitat degradation pressure categories (cf. Table 1 and Supplementary material SII). Individual habitat degradation parameters were measured using ESRI's ArcGis 9.2 software (ESRI, 2006). Used geographic data are given in Supplementary material SII.

For each available parameter, pressure level was assessed by comparing the parameter measure with the threshold delimiting 'low' to 'moderate' pressure levels (thresholds are given in Supplementary materials SI and SII). The pressure level allocated to a given reach for a given pressure category was the worst pressure level allocated to this reach by individual parameters from this pressure category.

To define the new WFD-compliant French biomonitoring tool we followed the recommendations of Barbour et al. (1999) for the development of multimetric assessment methods. Barbour decomposed this process in four main steps: (1) stream classification (cf. Wasson et al., 2002), (2) metric identification (cf. Section 2.3), (3) metric normalization (cf. Section 2.3) and (4) index development (cf. Section 2.4). In Fig. 2, the main steps of the applied design were summarized from data collection to the final index and ecological class boundary definition.

2.3. Metric identification and normalization

2.3.1. Metric set

418 biological metrics (see Supplementary material SIII for a detailed list) were calculated considering sample units from (i) all habitats (i.e. 'reach'), (ii) only 'marginal habitats' (i.e. 'B1'), (iii) only major habitats sampled according to hosting capacity (i.e. 'B2') or relative coverage (i.e. 'B3'), (iv) all major habitats (i.e. 'B2 + B3′) and (v) all habitats sampled according to hosting capacity (i.e. 'B1 + B2′). Three supplementary metrics corresponding to the French biotic index (IBGN) and its two sub-indices were only calculated at the 'B1 + B2′ level (because best corresponding to the combination of habitats sampled when applying the IBGN sampling protocol). Fourteen metrics were also specifically calculated at the 'reach' level including the Flemish MMIF and its six sub-indices, three metrics measuring the taxonomic specificity of 'B1', 'B2' or 'B3' within the reach, one between-group (i.e. B1, B2, B3) beta diversity measure and three alien species-related metrics. Then, the 2525 'metric × calculation level' (=metrics hereafter) were allocated to 199 groups (cf. Supplementary material SIII), each group being composed of metrics bringing the same [but calculated at different levels, i.e. (i), (ii), (iii), (iv) or (v)] or very similar biological or ecological information.

In contrast with Barbour et al. (1999), we normalized metrics before selecting the more convenient ones. Indeed, following the WFD requirements, new biomonitoring tools have to be expressed in Ecological Quality Ratio (EQR). This ratio is a number between zero and one, with values from 'reference' reaches close to one and values from reaches with 'bad' ecological status close to zero.

2.3.2. 'Least impaired' and 'impaired' river reaches

To define 'reference' conditions, we selected least impaired river reaches (LIRRs, e.g. Statzner et al., 2005; Dolédec and Statzner, 2008) using first, available data on water quality and habitat degradation, then validating reach status evaluation with a reduced set of biological metrics (including IBGN, ASPT, Shannon diversity, relative richness in Ephemeroptera, Plecoptera and Trichoptera taxa, relative utilization frequency of 'oligotrophic' and 'oligosaprobic' trait categories in reach communities). If not matching the criteria for integrating the LIRRs, river reaches were considered as impaired (IRRs).

2.3.3. Reference and worst metric values

In a previous work (Mondy and Usseglio-Polatera, 2009) we have already demonstrated that, in LIRRs, the inter-annual variability was negligible when compared to the spatial variability in metric values. As a result, data from all the reach sampling dates were simultaneously analyzed, spatial variability being taken into account through the normalization of metrics (cf. Section 2.3.4).

Depending on pressure category, a given metric could exhibit three major response patterns: (i) not simply and/or significantly responding to the pressure (type I), (ii) significantly decreasing in impaired conditions (i.e. pressure level being at least 'moderate'; type II) or (iii) significantly increasing in impaired conditions (type III).

We identified the response pattern of metrics (i.e. the sense of the deviation from values in LIRRs) by transforming metric values into normalized deviations (SES; cf. (1) and Gotelli and McCabe, 2002). SES normalization allowed us to directly compare metric values obtained from different stream types, at large spatial scale.

$$\mathrm{SES} = \frac{\mathrm{Obs}_{\mathrm{type}} - M_{\mathrm{type}}}{\mathrm{sd}_{\mathrm{type}}} \quad (1)$$

with $\mathrm{Obs}_{\mathrm{type}}$ the observed value of the metric in a given reach, M_{type} and $\mathrm{sd}_{\mathrm{type}}$ being respectively the mean and the standard deviation of the metric value distribution in LIRRs from the same stream type.

Then, the discrimination efficiencies (DEs; e.g. Ofenböck et al., 2004) of metrics were calculated. For a given metric and a given pressure category, $\mathrm{DE}_{\mathrm{SES}}$ corresponds to the proportion of samples pre-assigned to IRRs with (i) smaller values than the first quartile of the LIRR value distribution ($\mathrm{DE}_{\mathrm{SES}(25)}$, type II) or (ii) higher values than the third quartile of the LIRR value distribution ($\mathrm{DE}_{\mathrm{SES}(75)}$, type III) (Fig. 3).

Metrics for which neither $\mathrm{DE}_{\mathrm{SES}(25)}$ nor $\mathrm{DE}_{\mathrm{SES}(75)}$ were higher than 0.25 (i.e. the distribution of values from IRR assemblages was not different from the distribution of values from LIRR assemblages) corresponded to type I. Metrics for which $\mathrm{DE}_{\mathrm{SES}(25)}$ was higher than both 0.25 and $\mathrm{DE}_{\mathrm{SES}(75)}$ or for which $\mathrm{DE}_{\mathrm{SES}(75)}$ was higher than both 0.25 and $\mathrm{DE}_{\mathrm{SES}(25)}$ corresponded to type II and III, respectively.

Last, we identified the 'reference' and 'worst' values of each metric. The 'reference' value corresponded to the highest (type I or II) or the lowest (type III) value this metric could take in the LIRRs from a given stream type. The 'worst' metric value corresponded to the lowest (type I or II) or the highest (type III) value a metric could take in the IRRs from the whole data set. The 5th and 95th percentiles of the distribution of values for a given metric, were used as 'reference' or 'true worst' (='worst' hereafter) values instead of the highest/lowest values to discard metric values of outliers (Ofenböck et al., 2004).

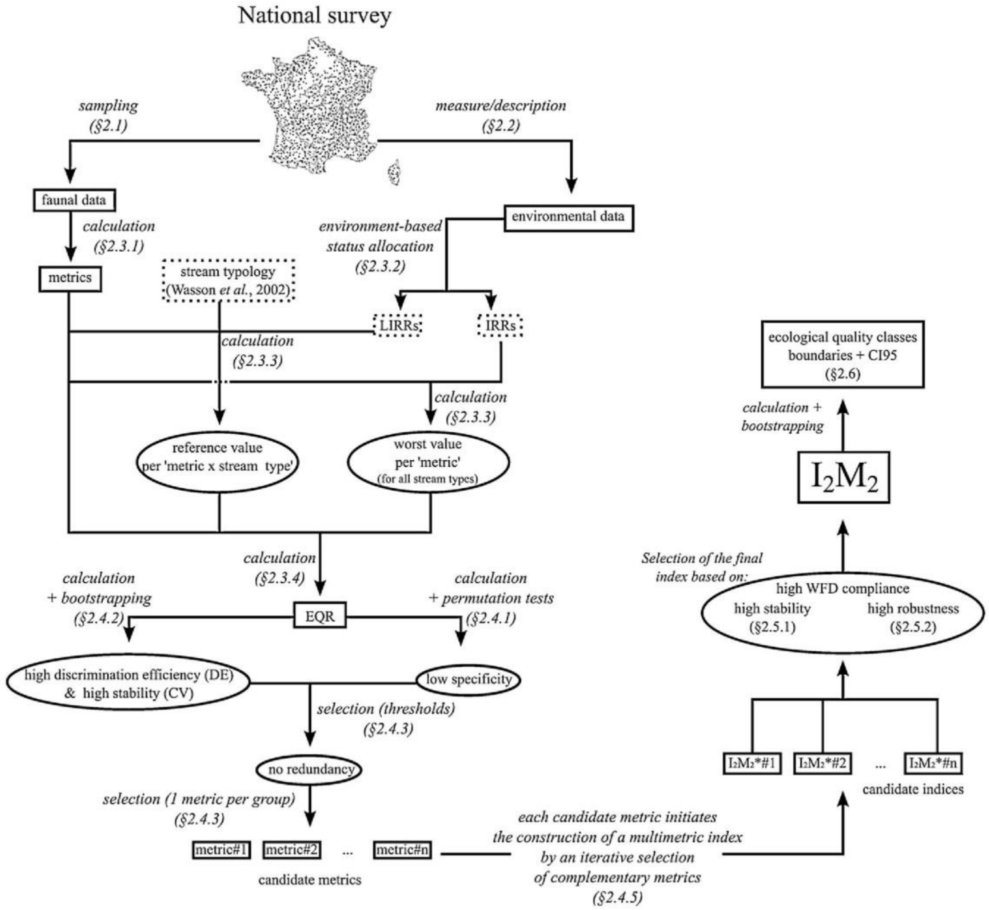

Fig. 2. Flow diagram giving the main steps of the I_2M_2 development strategy. Literature references and sections from this paper precisely describing each step of the index development design were given into brackets.

2.3.4. Metric normalization

Following Hering et al. (2006b), EQR was calculated using Eq. (2) for metrics of types I and II and Eq. (3) for metrics of type III.

$$EQR = \frac{Obs - Lower}{Upper - Lower} \quad (2)$$

$$EQR = 1 - \frac{Obs - Lower}{Upper - Lower} \quad (3)$$

with 'Obs' the metric value for a given sample. In Eq. (2), 'Upper' and 'Lower' correspond to the 'reference' and 'worst' metric values, respectively; whereas in Eq. (3), 'Upper' and 'Lower' correspond to the 'worst' and 'reference' metric values, respectively.

As stipulated in the WFD, EQR values should be bounded between 0 and 1. If observed reaches exhibited metric values out of the 'reference'–'worst' interval for the same stream type, the EQR values were arbitrarily fixed as 1 (if higher quality than the reference value) and 0 (if lower quality than the worst value), respectively. This EQR normalization allowed interpreting metric values from a given reach, regarding their deviation from reference conditions associated to the corresponding stream type.

2.4. Index development

Candidate metrics were selected taking into account four criteria: (i) low specificity, (ii) high discrimination efficiency (DE), (iii) high stability in LIRRs, and (iv) no redundancy. As the estimation of the three first criteria could depend on the development data set composition, we limited this bias using (i) permutation tests for specificity and (ii) bootstrap sub-sampling for robust estimation of DE and stability.

2.4.1. Specificity

A metric was considered as 'specific' if it significantly responded to a low number of pressure categories. We searched for metrics with low specificity, i.e. metrics exhibiting significant difference in the distribution of values in LIRRs vs. IRRs for a high number of

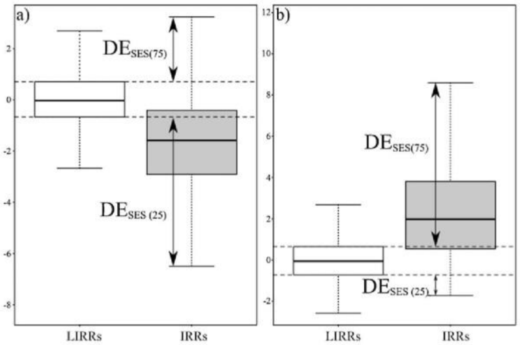

Fig. 3. Discrimination efficiency of normalized metric (DE_{SES}) decreasing (a) or increasing (b) with increasing anthropogenic pressure. Boxplots represent SES value distribution of metrics in least impaired river reaches (LIRRs, white box) and impaired river reaches (IRRs, gray box). The boxes range from the 25th percentile to the 75th percentile. The thick line represents the median and the whiskers extend to extreme values. Black dashed lines represent the 25th and 75th percentiles of SES distributions in LIRRs.

pressure categories. These differences were tested for each pressure category and each metric, with a conditional tree approach (Hothorn et al., 2006), i.e. a dichotomic classification method using Monte Carlo permutation tests ($\alpha = 0.01$, 9999 permutations).

2.4.2. Discrimination efficiency and stability in LIRRs

The DE of a metric for a given pressure category was calculated as the proportion of IRR assemblages with lower EQR values than the first quartile of the LIRR value distribution. The stability of a metric in LIRRs was evaluated using the coefficient of variation (CV) of its EQR value distribution from LIRR assemblages. The calculations of DE and CV were repeated 100 times based on randomly selected sub-samples of 60% of the reach data included in the development data set. A robust estimation of DE and CV for each metric was obtained by averaging the 100 estimations from corresponding sub-sampled data sets.

2.4.3. Selection of candidate metrics

Selected metrics simultaneously exhibited (i) low specificity (significant responses for at least seven from ten 'water quality' and five from seven 'habitat degradation' pressure categories), (ii) high DE (mean robust DE $\geq$0.6) and (iii) high stability in LIRRs (mean robust CV $\leq$1/3).

To avoid redundancy, for each of the groups of metrics giving the same biological or ecological information (cf. Section 2.3.1) only the metric with the highest DE was kept for potential inclusion in the multimetric index.

2.4.4. I_2M_2 calculation rationale

For each pressure category, a sub-index was calculated by averaging the EQR of the selected metrics, each EQR being weighted by its DE for this pressure category, as illustrated in Eq. (4) for PAH contamination.

$$i_2m_2^{PAH} = \frac{\sum (DE_m^{PAH} \times EQR_m^{PAH})}{\sum DE_m^{PAH}} \quad (4)$$

with $i_2m_2^{PAH}$: the sub-index for PAH contamination, EQR_m^{PAH}: the EQR value of the metric 'm' for PAH contamination and DE_m^{PAH}: the robust discrimination efficiency of the metric 'm' for PAH contamination.

The final I_2M_2 score was obtained by averaging the seventeen sub-indices ($i_2m_2^{pressure}$).

2.4.5. Construction of potential index metric combinations

Each of the 'n' candidate metrics (cf. Section 2.4.3) initiates the construction of a 'potential' multimetric index ($I_2M_2^*$) by an iterative selection of complementary metrics performed on the development data set. To reduce the potential bias of the development data set composition in metric selection, the iterative process includes bootstrap sub-sampling of the development data set.

1. One of the candidate metric was selected as the first metric.
2. A smaller data set was obtained by sub-sampling 60% of the reaches from the development data set.
3. In this sub-set, the $I_2M_{2(i)}^*$ and $DE_{(i)}$ values for each pressure category were calculated with the pre-selected metric(s) at step i ($i = 1$, for the first iterative process of metric selection).
4. the $I_2M_{2(i+1)}^*$ and $DE_{(i+1)}$ values were calculated (corresponding to each potential metric combination obtained by adding one of the $(n - i)$ candidate metrics to the metric(s) pre-selected at the beginning of the ith iterative step).
5. The relative increase in DE (Δ) for each pressure category (p) when including an additional metric to the $I_2M_2^*$ was calculated (Eq. (5)).

$$\Delta = \sum \left[\frac{DE_{(i+1)}{}^{p} - DE_{(i)}{}^{p}}{DE_{(i)}{}^{p}} \right] \quad (5)$$

with $DE_{(i+1)}{}^{p}$ and $DE_{(i)}{}^{p}$ the discrimination efficiency of the $I_2M_2^*$ related to pressure category p calculated with the selected metrics after (i) and $(i - 1)$ iterative metric selections.
6. The significance of the increase in DE (considering both water quality and habitat degradation) was statistically tested with unilateral paired Wilcoxon rank sum tests for each additional metric included in the index calculation.
7. The following procedure was applied to select a metric from the set of candidate metrics:
   (a) if only one candidate metric gave a significant increase in DE, this metric was selected;
   (b) if more than one candidate metric gave a significant increase in DE, the metric with the highest Δ (Eq. (5)) was selected.
8. Steps 2–7 were repeated one hundred times on randomly selected sub-sets of the development data set and the candidate metric which was more often selected was included in the $I_2M_2^*$ index.
9. Additional metrics were successively included in the $I_2M_2^*$, following steps 2–8 as long as the increase in DE, calculated on the development data set, was statistically significant ($\alpha < 0.05$).
10. Steps 1–9 were repeated using as first metric, each metric from the candidate metric pool.

2.5. Final selection of the I_2M_2 metric composition

The n $I_2M_2^*$ metric combinations were compared considering their (i) mean DE (the highest is the best), (ii) stability (i.e. no significant differences of index scores in LIRRs between the development and the test data sets), (iii) robustness (i.e. no significant differences in DE between the development and the test data sets) and (iv) compliance with WFD requirements.

2.5.1. Stability of $I_2M_2^*$ values in least impaired conditions

The Kolmogorov–Smirnov test was used to search for significant difference in the distributions of $I_2M_2^*$ values from LIRRs between the test and the development data sets.

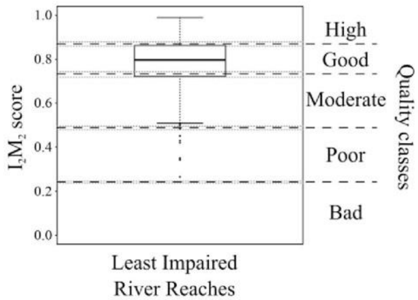

Fig. 4. Ecological quality class boundaries (black dashed lines) and their respective 95% confidence interval (gray dotted lines). The boxplot represents the I_2M_2 score distribution in least impaired river reaches, ranging from the 25th to the 75th percentile of the distribution. The thick line represents the median. The whiskers extend to the extreme data points but no more than 1.5 times the interquartile range from this box. Black dots represent outliers.

2.5.2. Robustness of $I_2M_2{}^*$ discrimination efficiency

The DEs respectively obtained with the development and the test data sets were compared with a bilateral paired Wilcoxon rank sum test.

2.6. Ecological quality class boundaries

As recommended by the WFD, we defined ecological quality class boundaries (i.e. delimiting 'high', 'good', 'moderate', 'poor' and 'bad' classes). Class boundary identification was based on the distribution of the I_2M_2 scores from the LIRRs of the development data set. To limit the influence of the development data set composition on the distribution of I_2M_2 values, a bootstrap sub-sampling approach was used. For each of the one hundred sub-sets (corresponding to 60% of the reaches from the development data set), the 75th and 25th percentiles of the I_2M_2 distribution in LIRRs were calculated. These values were considered as the 'high–good' and the 'good–moderate' boundaries, respectively. For defining the 'moderate–poor' and the 'poor–bad' boundaries, we divided the I_2M_2 scoring range between the minimal value (0) and the 'good–moderate' boundary in three equal classes. The robust estimate of each class boundary was calculated as the median of the one hundred estimates, and the corresponding 95% confidence interval was calculated (Fig. 4).

2.7. Comparison of the I_2M_2 with other indices

2.7.1. Correlation with ICM_{Star}

The ICM_{Star} was calculated following Eq. (6):

$$ICM_{Star} = 0.167 \times S_{fam} + 0.083 \times EPT_{fam} + 0.083 \times H'_{fam} + 0.334 \times ASPT + 0.067 \times (1 - GOLD) + 0.266 \times \log 10(sel\_EPTD + 1) \quad (6)$$

with S_{fam} and EPT_{fam} being respectively the total number of families and the number of families within the EPT orders, H'_{fam}: the Shannon diversity index calculated at the family level, GOLD: the relative abundance of Gastropoda, Oligochaeta and Diptera and log 10(sel_EPTD + 1): the log-transformed sum of Heptageniidae, Ephemeridae, Leptophlebiidae, Brachycentridae, Goeridae, Polycentropodidae, Limnephilidae, Odontoceridae, Dolichopodidae, Stratiomyidae, Dixidae, Empididae, Athericidae and Nemouridae abundances. All these metrics were expressed in EQR (cf. Section 2.3.4) before their aggregation.

We tested if the I_2M_2 was correlated with the ICM_{Star} with a Pearson correlation test.

2.7.2. Comparison of the I_2M_2, ICM_{Star} and IBGN discrimination efficiency

Similarly to I_2M_2, ICM_{Star} and IBGN DEs were calculated for each of the 17 investigated pressure categories as the proportion of IRRs providing scores lower than the first quartile of the score distribution in LIRRs. We tested the null hypothesis asserting there was no difference in DE between methods (i.e. I_2M_2, ICM_{Star} and IBGN) with a Friedman rank sum test for unreplicated blocked data. This test was followed by a multiple comparison test (Siegel and Castellan, 1988) to localize the potential significant differences between the three methods.

All metric calculations and statistical procedures were performed with R software (R Development Core Team, 2009), using packages 'party' (Hothorn et al., 2006) for conditional trees and 'pgirmess' (Giraudoux, 2011) for multiple comparison test after Friedman test.

3. Results

3.1. Index development

3.1.1. Specificity of metrics

Among 2525 tested metrics, 475 metrics significantly responded at least to 7 of the 10 water quality pressure types and 5 of the 7 habitat degradation pressure types.

3.1.2. Discrimination efficiency of metrics

The mean discrimination efficiency of individual metrics ranged from 0.0000 for the relative abundance/richness of several rare groups (taxa, trait categories or bio-ecological groups) to 0.7603 for the richness of the reach EPT taxa identified following the recommendations of Gabriels et al. (2010). Three hundred and ninety three metrics exhibited mean DE greater than 0.6.

3.1.3. Stability of metrics in LIRRs

In LIRRs, the mean CV ranged from 0.099 ("Belgium Biotic Index" calculated from reach faunal assemblages) to 3.132 (relative abundance of Planipennia in marginal habitat assemblages). Eight hundred and seventy two metrics exhibited rather low variability ($CV < 1/3$).

3.1.4. Candidate metrics

One hundred and eighty-six metrics fulfilled all the selection criteria. These metrics belonged to thirty groups gathering metrics giving similar bio-ecological information. As a result, the final set of candidate metrics (Table 2) was composed of thirty metrics, each selected metric exhibiting the highest mean DE of its group.

3.1.5. Final selection of I_2M_2 metrics

The iterative metric selection process provided thirty indices composed of two to eight metrics. Their mean DE on the whole data set ranged between 0.8107 and 0.8442 (cf. Table 3). The stability of indices in LIRRs was estimated with the Kolmogorov–Smirnov *p*-value and ranged from 0.0006 to 0.2479. Robustness was evaluated with the Wilcoxon *p*-value, and ranged from 0.0348 to 1.000.

Five of the 30 metric combinations (i.e. combinations #1, #2, #9, #25 and #30 in Table 3) displayed no significant ($\alpha > 0.10$) stability or robustness differences between the development and the test data sets. Among these five metric combinations, only combination #2 fulfilled all the WFD requirements. As a result, the new multimetric index (I_2M_2) was finally composed of five metrics (Table 4): (i) Shannon diversity index, (ii) original ASPT score

Table 2
Metric candidate to inclusion in the I_2M_2. The (i) full label, (ii) code, (iii and iv) numbers of pressure categories (water quality and habitat degradation) with significant response, (v) mean discrimination efficiency (DE) and (vi) mean coefficient of variation in least impaired river reaches (LIRRs) are given. S = taxonomic richness.

Candidate metrics [calculation level]	Code	No. of pressure categories with significant responses		Mean DE	Mean CV
		Water quality	Habitat degradation		
TAX (number of taxa[a]) [B1 + B2 + B3]	C1	8	7	0.6560	0.2861
Shannon diversity index [B2 + B3]	C2	9	7	0.6395	0.2741
Crustacea (%S) [B2 + B3]	C3	7	6	0.6379	0.2335
EPT (number of taxa[a]) [B1 + B2 + B3]	C4	8	6	0.7603	0.2353
Log 10(sel_EPTD + 1) [B1 + B2 + B3]	C5	8	6	0.6830	0.1992
Revised BMWP [B1 + B2]	C6	8	7	0.7303	0.2582
Original ASPT [B2 + B3]	C7	9	6	0.7479	0.1641
BBI [B1 + B2 + B3]	C8	7	6	0.6173	0.0995
IBGN [B1 + B2]	C9	8	6	0.6440	0.1777
Adult, aquatic stage (%) [B1 + B2 + B3]	C10	7	5	0.6458	0.2457
Aerial, active dispersion (%) [B2]	C11	8	6	0.6882	0.1912
Crawler (%) [B1 + B2 + B3]	C12	8	6	0.6108	0.2495
Microphytes as 'substrate' (%) [B1]	C13	7	6	0.6260	0.2528
Ovoviviparity – trait 'reproduction technique' (%) [B3]	C14	7	5	0.6749	0.2160
Polyvoltinism – trait 'number of cycles per year' (%) [B2 + B3]	C15	7	6	0.7300	0.2261
Oligotrophic – trait 'trophic status' (%) [B1 + B2 + B3]	C16	7	6	0.6368	0.2476
α-mesosaprobic – trait 'saprobity' (%) [B2]	C17	7	6	0.6742	0.2244
Brackish water preferendum (%) [B1 + B2]	C18	7	6	0.7018	0.2139
Temporary water preferendum (%) [B2 + B3]	C19	8	7	0.6326	0.2469
Biological group b (%S) [B3]	C20	7	5	0.6557	0.2152
Biological group f (%S) [B1 + B2]	C21	7	6	0.6517	0.2619
Ecological group B (%S) [B1]	C22	8	6	0.6204	0.2886
Bio-ecological group γ2 (%S) [B1 + B2 + B3]	C23	7	6	0.6292	0.2008
SPEARmetallic [B1 + B2 + B3]	C24	7	7	0.6336	0.1324
SPEARpesticide I (number of taxa) [B1 + B2]	C25	8	6	0.6967	0.2996
Redundancy (dispersal) [B1 + B2]	C26	7	7	0.6723	0.2420
Specialization (maximal potential size) [B3]	C27	8	7	0.6273	0.2208
Specialization (salinity preferendum) [B1 + B2 + B3]	C28	7	6	0.6917	0.1833
Specialization (transversal distribution) [B1 + B2]	C29	7	6	0.6779	0.2870
Specialization (trophic status preferendum) [B1 + B2]	C30	9	6	0.6345	0.3106

%S: relative richness.
[a] Metrics integrated in the Flemish MMIF (Gabriels et al., 2010).

Table 3
Metric composition, discrimination efficiency (DE), stability and robustness of the 30 tested metric combinations. See Table 2 for full labels of metrics.

Combination	Metric								Mean DE	Stability	Robustness
	1	2	3	4	5	6	7	8			
#1	C1	C17	C8						0.8201	0.2057	0.7467
#2	C2	C7	C15	C14	C1				0.8243	0.1617	0.4874
#3	C3	C6	C18						0.8319	0.0855	0.4586
#4	C4	C15	C14	C8					0.8317	0.0246	0.9265
#5	C5	C15	C4						0.8191	0.0542	0.8536
#6	C6	C18							0.8212	0.0846	0.5477
#7	C7	C6	C20	C17					0.8442	0.0213	0.2842
#8	C8	C6	C17	C15					0.8373	0.0069	0.8536
#9	C9	C18	C7						0.8231	0.1145	0.2842
#10	C10	C6	C15	C18	C8				0.8232	0.0300	1.0000
#11	C11	C6	C20	C8	C15				0.8274	0.0374	0.2435
#12	C12	C6	C14	C15	C7	C27	C8		0.8398	0.0179	0.5477
#13	C13	C6	C15						0.8190	0.0359	0.7819
#14	C14	C6	C28						0.8247	0.0120	0.2842
#15	C15	C6	C4	C17	C14	C19	C3	C8	0.8427	0.0176	0.7467
#16	C16	C6	C20						0.8309	0.0478	1.0000
#17	C17	C6	C15	C8					0.8385	0.0069	0.7119
#18	C18	C6							0.8195	0.0846	0.6112
#19	C19	C4	C15	C7	C14				0.8247	0.0259	0.5477
#20	C20	C6	C4	C11	C23				0.8337	0.0189	0.7467
#21	C21	C6	C15						0.8111	0.0362	0.2435
#22	C22	C6	C20	C7	C15				0.8435	0.0006	0.0448
#23	C23	C6	C15	C11	C8				0.8237	0.0304	0.4038
#24	C24	C4	C15	C27	C8	C14	C6		0.8344	0.0190	0.8900
#25	C25	C18	C20						0.8107	0.2051	0.4586
#26	C26	C9	C8	C15					0.8156	0.0073	0.0348
#27	C27	C15	C4						0.8196	0.0388	0.5791
#28	C28	C6	C14	C8					0.8273	0.0095	0.5477
#29	C29	C6	C20	C8					0.8283	0.0509	0.4307
#30	C30	C6	C20	C8					0.8202	0.2479	0.4038

C.P. Mondy et al. / Ecological Indicators 18 (2012) 452–467

Table 4
Response patterns and discrimination efficiency of the I_2M_2 individual metrics, I_2M_2, ICM_{Star} and IBGN. See Section 3.1.5 for a full description of I_2M_2 metrics.

	Shannon (B2 + B3)	Original ASPT (B2 + B3)	Polyvoltinism (B2 + B3)	Ovoviviparity (B3)	TAX (B1 + B2 + B3)	I_2M_2	ICM_{Star}	IBGN
Response pattern	–	–	+	+	–	–	–	–
Organic matter	0.6743	0.7736	0.7658	0.7580	0.6718	0.8345	0.7931	0.6690
Nitrogen compounds (except nitrates)	0.7493	0.8542	0.7983	0.7246	0.7512	0.8641	0.8252	0.7379
Nitrates	0.5994	0.7829	0.7430	0.7562	0.6218	0.8276	0.7658	0.6087
Phosphorous compounds	0.6952	0.8557	0.8261	0.7886	0.7061	0.8900	0.8272	0.7225
Suspended matter	0.6691	0.7432	0.8106	0.6806	0.7681	0.8864	0.8011	0.6989
Acidification	0.5229	0.5372	0.6191	0.4924	0.5645	0.7018	0.5965	0.5789
Mineral micropollutants	0.6105	0.7027	0.6494	0.5928	0.6108	0.7577	0.7357	0.5977
Pesticides	0.6911	0.8838	0.8120	0.7309	0.6903	0.9155	0.8592	0.7711
PAH	0.6757	0.7922	0.7869	0.7020	0.7014	0.8864	0.8288	0.7024
Other organic micropollutants	0.5592	0.7608	0.6918	0.6733	0.5844	0.7867	0.7243	0.5826
Transportation facilities	0.6288	0.6513	0.6910	0.5722	0.6718	0.7853	0.7435	0.6675
Riverine vegetation	0.6004	0.6972	0.6600	0.6233	0.6006	0.7547	0.7120	0.5829
Urbanization	0.7109	0.8047	0.7709	0.7094	0.7065	0.8703	0.8388	0.7015
Clogging risk	0.6380	0.8080	0.7747	0.7662	0.6345	0.8618	0.7985	0.6363
Hydrological instability	0.6233	0.6510	0.6579	0.6230	0.6384	0.7609	0.7040	0.6025
Catchment anthropization	0.6257	0.7644	0.7289	0.6897	0.6288	0.8186	0.7748	0.6277
Straightening	0.5984	0.6515	0.6230	0.5901	0.6006	0.7321	0.6917	0.5736

and (iii) the relative abundance of polyvoltine species in the assemblage, all of them calculated at the major habitat scale (i.e. B2 + B3), (iv) the relative abundance of ovoviviparous species calculated at the B3 level and (v) a measure of taxonomic richness ('TAX') calculated at the reach level following taxonomic identification levels recommended by Gabriels et al. (2010). These five metrics had homogeneous response patterns for all the pressure categories, three were decreasing [i.e. (i), (ii) and (v); type II] and two were increasing [i.e. (iii) and (iv); type III] with increasing pressure gradient. Reference values for each combination of 'metric × stream type' are given in Appendix C.

3.2. Ecological quality class boundaries

The calculated values of the 'high–good', 'good–moderate', 'moderate–poor' and 'poor–bad' boundaries were: 0.8696 (CI_{95} = [0.8603; 0.8796]), 0.7327 (CI_{95} = [0.7195; 0.7411]), 0.4885 (CI_{95} = [0.4797; 0.4941]) and 0.2442 (CI_{95} = [0.2398; 0.2470]), respectively.

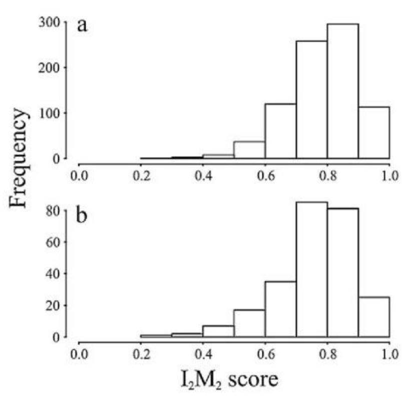

Fig. 5. Histograms of I_2M_2 score frequency distribution in the development (a) and in the test (b) data sets.

3.3. Test of the I_2M_2

3.3.1. I_2M_2 values in least impaired conditions

The distribution of I_2M_2 scores from LIRR assemblages in the development and the test data sets exhibited no significant difference (Kolmogorov–Smirnov test: $D = 0.0805$, $p = 0.1617$; Fig. 5).

3.3.2. Discrimination efficiency

The difference in I_2M_2 discrimination efficiency between the development and the test data sets was not significant (bilateral paired Wilcoxon signed rank test: $W = 61$, $p = 0.4874$; Figs. 6 and 7).

3.4. Correlation of the I_2M_2 with the European intercalibration ICM_{Star} index

The I_2M_2 was strongly and significantly correlated with the ICM_{Star} (Pearson's product moment correlation = 0.9095, p-value < 2.2×10^{-16}).

3.5. Comparison of the I_2M_2, ICM_{Star} and IBGN discrimination efficiency

The three indices exhibited significant differences in discrimination efficiency (Friedman rank sum test: $\chi^2 = 34$, d.f. = 2, p-value = 4.14×10^{-8}; Table 4); the I_2M_2 (mean DE = 0.820 ± 0.064) better performing than the ICM_{Star} (mean DE = 0.766 ± 0.067) and the IBGN (mean DE = 0.651 ± 0.063; cf. Fig. 8; multiple comparison test after Friedman test, $\alpha = 0.05$).

4. Discussion

4.1. Typology specificity

The Water Framework Directive has focused on the need to take into account the specific characteristics of streams from different regions and natural contexts. Many countries using the AQEM approach have defined stream type-specific multimetric indices considering only a low number of stream types: e.g. four in Austria (Ofenböck et al., 2004), three in Portugal (Pinto et al., 2004) and Greece (Skoulikidis et al., 2004) or two in Netherlands (Vlek et al., 2004). Even if 24 stream types have been defined in Germany, type-specific indices have been developed only for five of them (Lorenz et al., 2004).

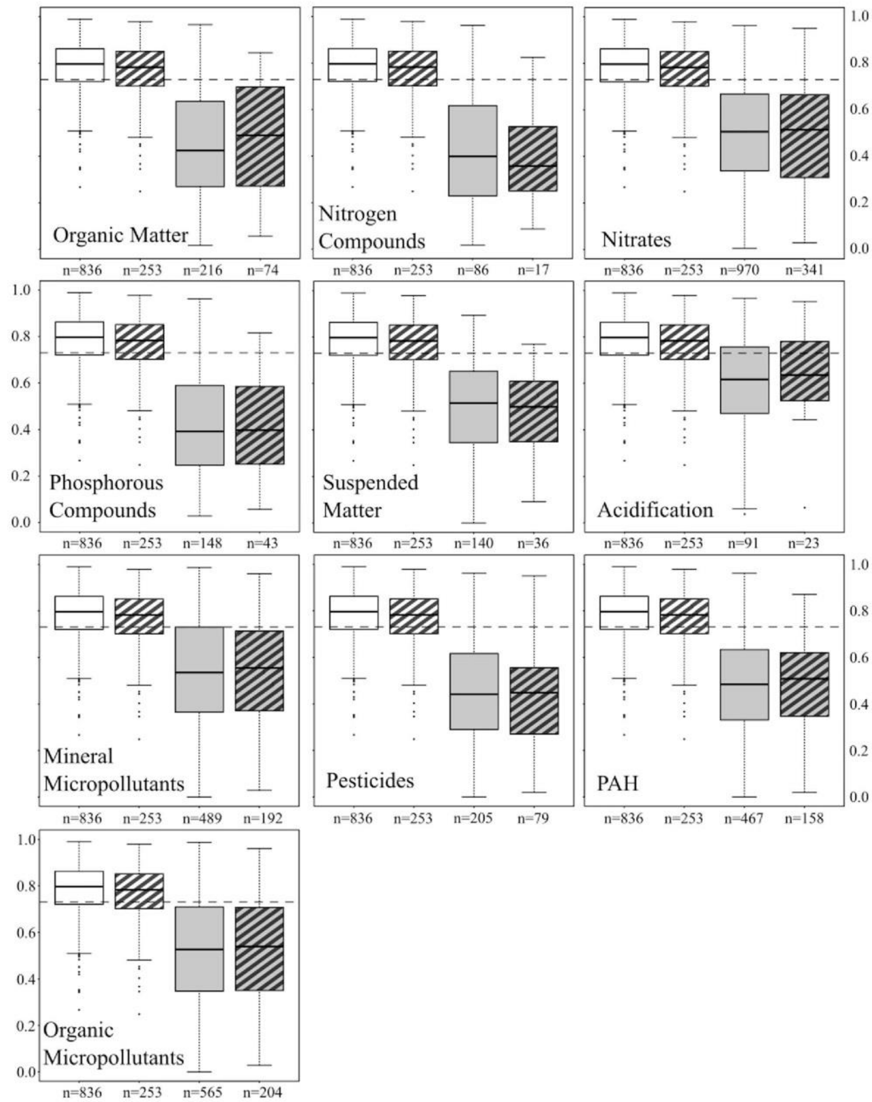

Fig. 6. I_2M_2 score distributions in least impaired river reaches (white boxes) and impaired river reaches (gray boxes) for 10 different water quality pressure categories. Solid boxes represent the I_2M_2 score distribution in the development data set whereas stripped boxes represent the I_2M_2 score distribution in the test data set. Black dashed lines represent the 'low–moderate' pressure level boundary. '*n*' represents the number of faunal samples considered in each group. For further details, see legend of Fig. 4.

This strategy seemed quite unsuitable for French streams because of their high environmental diversity (125 stream types have been defined by Chandesris et al., 2006). A not type-specific approach was necessary, as already developed in Germany (Böhmer et al., 2004b; Hering et al., 2004b) or Flemish Belgium (Gabriels et al., 2010). Even after stream typology simplification (57 stream types), defining one specific index per stream type would be highly difficult due to the low number of available data on reference or least impaired river reaches for several stream types (cf. Appendix A). Moreover, large scale (i.e. between-stream types) comparisons of index values would be difficult due to the potential differences in metric composition of stream type-specific indices.

As a result, we preferred to evaluate the ecological status of rivers using a single common set of metrics for all the stream types, taking into account stream type characteristics when normalizing metrics into Ecological Quality Ratios (EQRs).

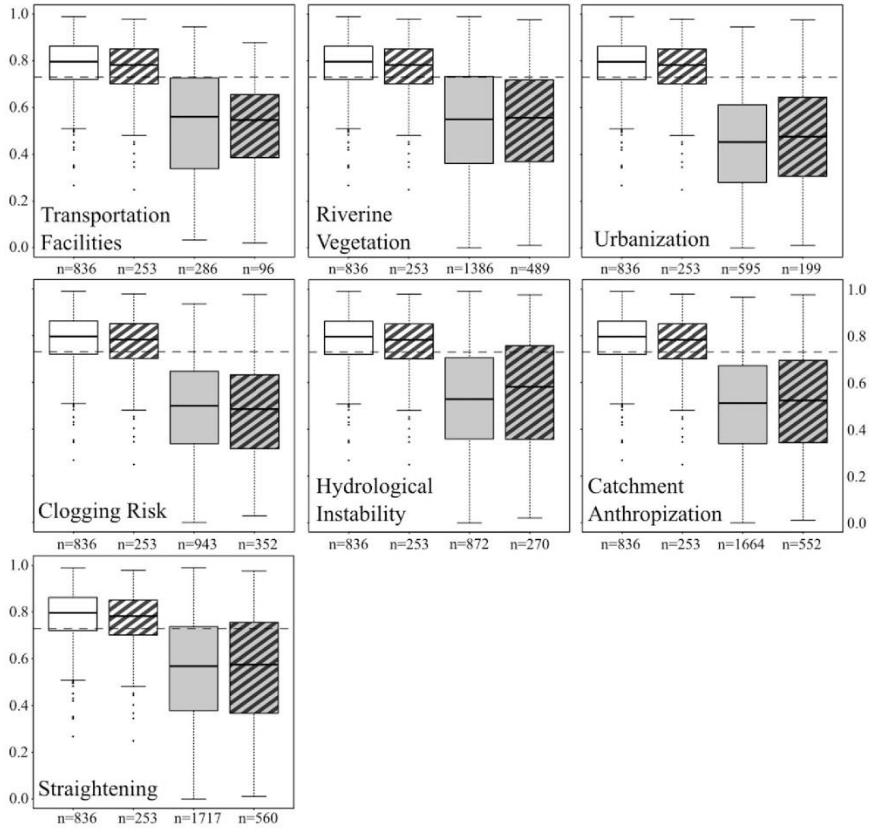

Fig. 7. I_2M_2 score distributions in least impaired river reaches (white boxes) and impaired river reaches (gray boxes) for 7 different habitat degradation pressure categories. For further details, see legends of Figs. 4 and 6.

4.2. Pressure specificity

In their cook book, Hering et al. (2006b) suggested to develop either generalist or pressure-specific multimetric indices. In contrast with several European countries (Böhmer et al., 2004a; Ofenböck et al., 2004; Sandin et al., 2004), we aimed at building a generalist index, usable for a wide spectrum of environmental conditions (not only for a unique combination of 'bioregion × stressor type'; e.g. Ofenböck et al., 2004), by selecting metrics which discriminate anthropogenic pressure from natural variability for a large number of stream types and pressure categories. This choice resulted from several considerations:

1. Pressure-specific index development needs the selection of sampling sites to ensure that: "*environmental stressor gradient[s] is [are] ideally represented by a set of sites of one freshwater ecosystem type covering the whole range [...] of the environmental stressor that is to be targeted by the Multimetric System*" (Hering et al., 2006b). The French survey network was not designed to fulfill this requirement. Indeed, in the two main French National survey networks, reaches were selected to be either least impaired (i.e. Reference Reach National survey) or simply representative of the mean quality of the water body they belonged to (i.e. RCS National survey).
2. The search for pressure category specific metrics, would require that reaches included in the development data set be individually impaired by one or a low number of pressure categories. As a result, complex pressure combinations, which often impair river reaches, would not – or not optimally – be taken into account in the development of pressure-specific indices. As an illustration, the 1725 river reaches included in our database were significantly impaired, in average, by 4.23 (±2.29) of the 17 pre-defined pressure categories (see also Comte et al., 2010). A generalist index seemed to be much more suitable for identifying multiple pressure scenarii.
3. The concentration of many toxicants still remains very difficult (at low level) and expensive to quantify in water and/or sediment, making them scarcely included in routine survey networks (Kolpin et al., 2002). As an example, mineral micropollutants and PAH were respectively measured for only 39.23% and 35.70% of the samples in the used database. We hypothesized that, if selected biological metrics significantly respond to

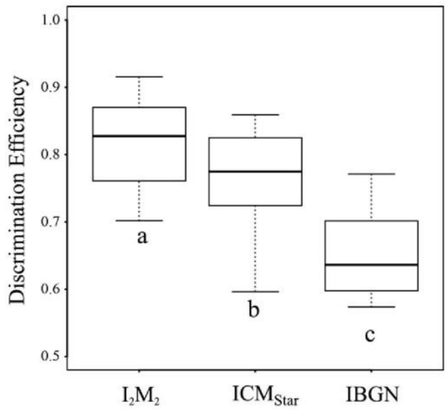

Fig. 8. Distribution of the 17 DEs of I_2M_2, ICM_{Star} and IBGN. The boxes range from the 25th to the 75th percentile. The median is represented by the black thick line. The whiskers extend to the extreme data points. Different letters indicate significant differences in DE (multiple comparison test after Friedman test, Siegel and Castellan, 1988).

a high number of different pressure categories, the multimetric index would have higher chance to identify not targeted (or unexpected) pressure categories.

4. Metrics and indices considered as pressure-specific were, in fact, not so specific. For example, Lorenz et al. (2004) demonstrated that the German Fauna Index, based on taxonomic metrics, designed to specifically identify hydromorphological alterations, significantly responded also to organic contamination. SPEAR indices were developed to identify the biological impact of some specific categories of toxic pollutants: e.g. pesticides (Liess and von der Ohe, 2005), metals (von der Ohe and Liess, 2004), organic micro-pollutants (Beketov and Liess, 2008). Even if they were designed to respond to specific toxic pressure, these indices significantly responded to a more diverse combination of pressure categories, i.e. in average 6.6 ± 1.4 of the 10 predefined water quality pressure categories and 5.3 ± 0.8 of the 7 pre-defined habitat degradation risks, based on our whole database.
5. Last, we found that metric DEs (for the different pressure categories) were strongly correlated (Pearson's coefficient $r = 0.92 \pm 0.04$). In other words, metrics tended to have similar DE for the different pressure categories. As a consequence, it seemed extremely difficult to identify 'truly specific' metrics able to efficiently discriminate only one or a small group of pressure categories.

4.3. Selected metrics

The final I_2M_2 was composed of only five metrics. Three of them are taxonomic metrics that have been widely used in biotic indices: (i) the Shannon's diversity index (Shannon, 1948) included in several European multimetric indices, e.g. in Germany (Böhmer et al., 2004a) or in Belgium (Gabriels et al., 2010); (ii) the "Average Score Per Taxon" (ASPT, Armitage et al., 1983) already involved – sometimes with regional adaptations – in several European multimetric indices developed in the AQEM context, e.g. in Italy (Buffagni et al., 2004), Czech Republic (Brabec et al., 2004), southern Sweden (Dahl and Johnson, 2004) or Portugal (Pinto et al., 2004); and (iii) taxonomic richness, considered as the simplest measure of diversity (Stirling and Wilsey, 2001; Mendes et al., 2008). This metric has been already taken into account in several biotic indices, e.g. the IBGN (France, AFNOR, 2004) and the BBI (Belgium, Pauw and Vanhooren, 1983) and has been included in several multimetric indices (e.g. Royer et al., 2001; Ofenböck et al., 2004; Vlek et al., 2004; Gabriels et al., 2010).

The two last metrics were biological traits, i.e. fuzzy-coded variables (Chevenet et al., 1994) describing various biological attributes of species (Resh et al., 1994; Usseglio-Polatera et al., 2000; Statzner and Bêche, 2010). During the twenty last years, these traits have been increasingly used, first to elucidate the filtering role of habitat on species attributes at various spatial scales (e.g. Townsend and Hildrew, 1994; Townsend et al., 1997; Poff, 1997), then to study the additional filtering role of human activities on biological traits of stream assemblages in a biomonitoring perspective (e.g. Dolédec et al., 1999; Charvet et al., 2000; Usseglio-Polatera and Beisel, 2002; Gayraud et al., 2003; Liess and von der Ohe, 2005; Dolédec and Statzner, 2008; Townsend et al., 2008; Archaimbault et al., 2010; Statzner and Bêche, 2010), but were still rarely included in the composition of multimetric indices. Two trait categories regarding reproduction have been included in the I_2M_2: (iv) 'polyvoltinism' and (v) 'ovoviviparity'. Polyvoltinism – supposed to ensure a higher resilience capacity – is a reproductive strategy expected to occur with higher frequency in unstable conditions compared to a 'reference' situation (Townsend and Hildrew, 1994), whereas 'ovoviviparity' as parental care strategy, would prevent high mortality at egg stage in harsh environmental conditions. An increase in polyvoltine and/or ovoviviparous species frequency in benthic assemblages has been already observed with different pressure categories (e.g. Usseglio-Polatera and Beisel, 2002; Archaimbault, 2003; Piscart et al., 2006; Dolédec and Statzner, 2008).

4.4. Ecological quality class boundaries

Because using LIRRs instead of 'true references', we have not divided the 0–1 range of EQR values in five classes of equal range (e.g. Böhmer et al., 2004a; Gabriels et al., 2010) or used the 25th percentile of the reference value distribution as the 'high–good' boundary before defining the other class boundaries with equal bands (e.g. Munné and Prat, 2009; Poquet et al., 2009). We assigned the 75th and the 25th percentiles of the I_2M_2 scores in LIRRs to the 'high–good' and 'good–moderate' boundaries respectively; equal bands only defining the three other quality classes. Moreover, in the establishment of between-class boundary, we used a bootstrap sub-sampling method that provided (i) a robust estimate of boundary and (ii) the 95% confidence interval associated with each class boundary (Fig. 4), in agreement with the uncertainty measure around class boundaries required by the WFD (European Commission, 2003).

4.5. I_2M_2 efficiency

Testing the I_2M_2 with an independent data set has demonstrated its stability in LIRRs and its robustness regarding discrimination efficiency. The I_2M_2 could also be considered as a highly sensitive index, since there is nearly no overlap between LIRR and IRR inter-quartiles of index score distribution (Royer et al., 2001). In average more than 81% of the reaches pre-classified as 'impaired' on environmental criteria were also considered as 'impaired' by the I_2M_2. This high efficiency of detecting a large panel of pressures (even at moderate level) allows considering the I_2M_2 as a robust and efficient biomonitoring tool (Sandin and Johnson, 2000). Compared to the IBGN, the I_2M_2 significantly improved the detection of impaired reaches by at least 17% for nitrogen compounds and up to 35% for organic micropollutants and clogging risk.

A reasonable proxy for global anthropogenic pressure on reaches could be the anthropization level of their catchment, evaluated with the addition of the relative surfaces respectively

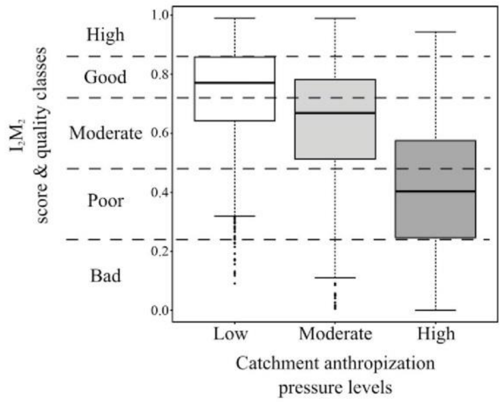

Fig. 9. Boxplot of I_2M_2 scores in relation to the pre-defined reach pressure levels concerning catchment anthropization. Black dashed lines represent ecological quality class boundaries. For further details, see legend of Fig. 6.

used by urbanization, agriculture and industry. The distribution of I_2M_2 reach scores among the different quality classes, closely matches the distribution of corresponding pressure levels of catchment anthropization (Fig. 9), then validating the procedure establishing ecological boundaries.

5. Conclusion

The proposed multimetric index (I_2M_2) (i) completely fulfills the WFD requirements, (ii) significantly improves the detection of impaired reaches when compared to the former French IBGN, (iii) is one of the very first biomonitoring tool designed – from a large national database – to take into account pressure–impact relationships for a high number of pressure categories (including both water quality and habitat degradation of reaches) and considering both taxonomic characteristics and biological traits of benthic macroinvertebrate assemblages. The I_2M_2 has been proposed for future use in the national biomonitoring of wadeable reaches in the WFD implementation framework and for integration in the future French online system 'SEEE' (Système d'Evaluation de l'Etat des Eaux = Water Status Evaluation System) that will provide to managers a simple way to calculate this index (among other metrics describing BQE assemblages) after uploading reach invertebrate assemblage abundance distribution. Moreover, the I_2M_2 is (better performing than and) highly correlated to the European intercalibration multimetric index (ICM_{Star}), which is very promising regarding the future integration of the I_2M_2 in the European pool of WFD-compliant biotic indices.

Acknowledgements

This work was supported by the French Ministry of Environment (MEDDTL) and by the "Office National de l'Eau et des Milieux Aquatiques" (ONEMA). We are grateful to the six water agencies and the 22 regional directions (DREAL) which participated in the abiotic and biotic data collection. We have appreciated also the help of Nicolas Mengin, Martial Ferreol and Nolwenn Bougon (CEMAGREF, Lyon) in data validation and database management and thank Yorick Reyjol (ONEMA) for valuable discussions on this index. We are deeply thankful to the two reviewers for pointing out some improvements of our manuscript.

Appendix A. Distribution of least impaired river reaches (LIRRs) and other reaches (IRRs) among the 57 simplified stream types

HER	Simplified stream type	Least impaired reaches Number of reaches/samples	Other reaches Number of reaches/samples
HER 1: Pyrénées	TP1	5/23	13/21
	P1	7/34	11/29
	GM1	4/8	10/12
HER 2: Alpes internes	TP2	2/8	11/26
	MP2	6/18	25/57
	G2	3/7	7/8
HER 3: Massif Central Sud	TP3	11/42	23/55
	P3	16/54	38/99
	GM3	16/39	32/55
HER 4: Vosges	TP4	4/11	19/76
	P4	3/15	9/17
	M4	1/6	6/16
HER 5: Jura Préalpes du Nord	TP5	4/23	12/42
	P5	7/30	70/149
	GM5	10/39	52/91
HER 6: Méditerranéen	TP6	2/12	51/91
	MP6	5/12	31/60
	GM6	6/17	25/44
HER 7: Préalpes du Sud	TP7	5/19	9/24
	GMP7	10/38	18/42
HER 8: Cévennes	PTP8	7/30	27/64
	GM8	7/20	13/25
HER 9: Tables Calcaires	TP9	7/28	47/136
	P9	13/36	168/357
	M9	4/20	46/108
	G9	1/6	17/22
HER 10: Côtes Calcaires Est	TP10	7/23	49/94
	P10	6/18	31/79
	M10	7/21	30/60
	G10	4/10	18/38
HER 11: Causses Aquitains	PTP11	5/12	14/30
	M11	3/10	2/5
	G11	6/12	2/3
HER 12: Armoricain	TP12	5/15	21/58
	P12	15/59	113/303
	M12	2/8	33/61
	G12	2/7	7/13
HER 13: Landes	PTP13	8/9	15/27
HER 14: Coteaux Aquitains	TP14	2/7	29/43
	P14	6/11	52/67
	GM14	10/10	29/34
HER 15: Plaine de Saône	TP15	1/6	28/43
	MP15	4/10	28/57
HER 16: Corse	M16	9/32	10/21
	G16	2/11	3/6
HER 17: Dépressions Sédimentaires	PTP17	2/6	8/23
HER 18: Alsace	TP18	1/6	5/13
	MP18	4/8	10/23
HER 19: Grands Causses	P19	5/13	4/11
	GM19	2/7	0/0
HER 20: Dépôts argilo-sableux	PTP20	1/6	19/31
HER 21: Massif Central Nord	TP21	9/27	6/11
	P21	20/59	41/110
	M21	12/36	9/16
	G21	4/9	7/11
HER 22: Ardennes	TP22	2/12	1/5
	P22	3/14	7/21

TP – very small; PTP – small and very small; P – small; MP – medium and small; M – medium; GM – large and medium; G – large.

Appendix B. Taxonomic levels required for the Multi-Habitat Sampling protocol

Taxons		Required taxonomic level
Plecoptera		Genus
Ephemeroptera		Genus
Trichoptera	Except *Limnephilidae*	Genus
	Limnephilidae	Sub-family
Coleoptera	Except *Dytiscidae, Hydrophilidae, Curculionidae*	Genus
	Dytiscidae, Hydrophilidae	Sub-family
	Curculionidae	Family
Megaloptera		Genus
Heteroptera	Except *Corixinae*	Family
	Corixinae	Sub-family
Planipennia		Genus
Odonata	Except *Coenagrionidae*	Genus
	Coenagrionidae	Family
Lepidoptera		Family
Hymenoptera		Genus
Diptera		Family
Hydracarina		Sub-class
Crustacea	Except *Asellidae*	Genus
	Asellidae	Family
Bivalvia		Genus
Gastropoda	Except *Planorbidae*	Genus
	Planorbidae	Family
Hirudinea and Branchiobdellida		Family
Oligochaeta		Class
Bryozoa		Phylum
Nematoda		Phylum
Gordiidae		Family
Turbellaria		Family
Hydrozoa		Class
Porifera		Phylum
Nemertea		Phylum

Appendix C. Reference values of the five I_2M_2 individual metrics for the 57 simplified stream types

HER	Simplified stream type	Shannon diversity index (B2 + B3)	Original ASPT (B2 + B3)	Polyvoltinism (B2 + B3)	Ovoviviparity (B3)	TAX (B1 + B2 + B3)
HER 1: Pyrénées	TP1	4.2759	7.2161	0.1684	0.0076	43.0000
	P1	4.6392	7.3273	0.1509	0.0062	51.3500
	GM1	4.6568	7.2233	0.1457	0.0060	44.0000
HER 2: Alpes internes	TP2	4.0179	7.2571	0.1652	0.0070	39.9500
	MP2	3.2448	6.9167	0.1886	0.0126	34.8000
	G2	3.1198	6.5700	0.2786	0.0166	22.0000
HER 3: Massif Central Sud	TP3	4.6387	7.3575	0.1504	0.0095	49.0000
	P3	4.7528	7.3792	0.1656	0.0010	47.3500
	GM3	4.2866	7.2453	0.1983	0.0089	46.4000
HER 4: Vosges	TP4	3.9876	7.2273	0.2301	0.0073	46.5000
	P4	4.3509	7.2538	0.2175	0.0204	45.8000
	M4	4.0218	7.15625	0.2231	0.0102	47.0000
HER 5: Jura Préalpes du Nord	TP5	4.3613	6.9435	0.2139	0.0255	48.4000
	P5	4.3451	6.8548	0.2305	0.0304	47.5500
	GM5	4.1176	7.1490	0.2063	0.0121	42.7000
HER 6: Méditerranéen	TP6	3.6598	6.8708	0.2605	0.0694	50.0000
	MP6	4.0218	7.0150	0.2407	0.0187	51.9000
	GM6	3.9087	6.6207	0.2761	0.0193	53.4000
HER 7: Préalpes du Sud	TP7	4.0933	7.0071	0.2401	0.0074	44.1000
	GMP7	3.9056	7.0453	0.2004	0.0032	47.3000
HER 8: Cévennes	PTP8	4.3600	7.2093	0.2000	0.0116	52.5500
	GM8	3.9826	7.1793	0.2361	0.0160	46.3500
HER 9: Tables Calcaires	TP9	4.3397	6.7665	0.2661	0.0530	52.6500
	P9	4.2396	7.0000	0.2448	0.0611	57.2500
	M9	3.9987	7.0754	0.2459	0.0961	56.0000
	G9	3.9571	6.6429	0.3005	0.1016	47.5000
HER 10: Côtes Calcaires Est	TP10	4.0622	6.9923	0.2529	0.0422	46.9000
	P10	3.8487	6.8046	0.2260	0.0704	47.4500
	M10	4.0517	6.6129	0.2810	0.0565	51.0000
	G10	4.2319	6.8034	0.2998	0.1155	53.1000

HER	Simplified stream type	Shannon diversity index (B2 + B3)	Original ASPT (B2 + B3)	Polyvoltinism (B2 + B3)	Ovoviviparity (B3)	TAX (B1 + B2 + B3)
HER 11: Causses Aquitains	PTP11	4.3977	6.8684	0.2455	0.0873	52.2500
	M11	4.4517	7.2971	0.2199	0.0125	55.5500
	G11	3.6611	7.0188	0.2867	0.0442	48.8000
HER 12: Armoricain	TP12	4.3958	7.0830	0.2001	0.0093	61.3000
	P12	4.5245	7.2179	0.2103	0.0116	55.1000
	M12	4.3451	7.3711	0.2065	0.0099	54.3000
	G12	4.2119	6.5636	0.3006	0.0802	53.7000
HER 13: Landes	PTP13	3.3689	6.7000	0.2308	0.0174	40.0000
HER 14: Coteaux Aquitains	TP14	4.2136	6.5833	0.2310	0.0139	55.7000
	P14	4.4925	7.2004	0.2469	0.0237	45.5000
	GM14	4.1388	6.8106	0.2690	0.0155	42.5500
HER 15: Plaine de Saône	TP15	3.5453	6.5590	0.2359	0.1044	45.7500
	MP15	3.9035	6.9540	0.2286	0.0197	38.6500
HER 16: Corse	M16	3.8343	6.9726	0.2330	0.0095	35.4500
	G16	3.2272	6.5425	0.3106	0.0141	31.0000
HER 17: Dépressions Sédimentaires	PTP17	4.1967	6.6232	0.2512	0.0470	48.5000
HER 18: Alsace	TP18	3.0748	5.3189	0.3676	0.2962	34.0000
	MP18	4.1192	6.6375	0.2713	0.0763	45.9000
HER 19: Grands Causses	P19	4.1831	7.0653	0.2179	0.0635	55.8000
	GM19	3.8044	7.1560	0.2025	0.0150	49.0000
HER 20: Dépôts argilo-sableux	PTP20	3.8253	6.6574	0.2662	0.0616	48.5000
HER 21: Massif Central Nord	TP21	4.4213	7.3904	0.1860	0.0054	44.4000
	P21	4.5879	7.4058	0.1715	0.0121	51.1000
	M21	4.5110	7.4688	0.1730	0.0047	54.0000
	G21	4.6040	7.4207	0.1863	0.0119	46.0000
HER 22: Ardennes	TP22	4.5894	7.2519	0.1664	0.0115	51.4500
	P22	4.4083	7.1629	0.2042	0.0211	55.0000

TP – very small; PTP – small and very small; P – small; MP – medium and small; M – medium; GM – large and medium; G – large.

Appendix D. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolind.2011.12.013.

References

AFNOR, 2004. Qualité de l'eau – Détermination de l'indice biologique global normalisé (IBGN). NF T90-350, 15 pp.
AFNOR, 2009. Qualité de l'eau – Prélèvement des macro-invertébrés aquatiques en rivières peu profondes. XP T90-333, 22 pp.
AFNOR, 2010. Qualité de l'eau – Traitement au laboratoire d'échantillons contenant des macro-invertébrés de cours d'eau. XP T90-388, 21 pp.
Archaimbault, V., 2003. Réponses bio-écologiques des macroinvertébrés benthiques aux perturbations: la base d'un outil diagnostique fonctionnel des écosystèmes d'eaux courantes. Université Paul Verlaine – Metz.
Archaimbault, V., Usseglio-Polatera, P., Garric, J., Wasson, J., Babut, M., 2010. Assessing pollution of toxic sediment in streams using bio-ecological traits of benthic macroinvertebrates. Freshwater Biology 55 (7), 1430–1446.
Armitage, P., Moss, D., Wright, J., Furse, M., 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. Water Research 17 (3), 333–347.
Barbour, M.T., Gerritsen, J., Snyder, B., Stribling, J.B., 1999. Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton Benthic Macroinvertebrates and Fish, 2nd ed. U.S. Environmental Protection Agency, Office of Water, Washington, DC, USA.
Beketov, M.A., Liess, M., 2008. An indicator for effects of organic toxicants on lotic invertebrate communities: independence of confounding environmental factors over an extensive river continuum. Environmental Pollution 156 (3), 980–987.
Böhmer, J., Rawer-Jost, C., Zenker, A., Meier, C., Feld, C.K., Biss, R., Hering, D., 2004a. Assessing streams in Germany with benthic invertebrates: development of a multimetric invertebrate based assessment system. Limnologica – Ecology and Management of Inland Waters 34 (4), 416–432.
Böhmer, J., Rawer-Jost, C., Zenker, A., 2004b. Multimetric assessment of data provided by water managers from Germany: assessment of several different types of stressors with macrozoobenthos communities. Hydrobiologia 516 (1), 215–228.
Bonada, N., Prat, N., Resh, V.H., Statzner, B., 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. Annual Review of Entomology 51 (1), 495–523.
Brabec, K., Zahradkova, S., Nemejcova, D., Paril, P., Kokes, J., Jarkovsky, J., 2004. Assessment of organic pollution effect considering differences between lotic and lentic stream habitats. Hydrobiologia 516 (1), 331–346.
Buffagni, A., Erba, S., Cazzola, M., Kemp, J., 2004. The AQEM multimetric system for the southern Italian Apennines: assessing the impact of water quality and habitat degradation on pool macroinvertebrates in Mediterranean rivers. Hydrobiologia 516 (1), 313–329.
Buffagni, A., Erba, S., Cazzola, M., Murray-Bligh, J., Soszka, H., Genoni, P., 2006. The STAR common metrics approach to the WFD intercalibration process: full application for small, lowland rivers in three European countries. Hydrobiologia 566 (1), 379–399.
Chandesris, A., Wasson, J.G., Pella, H., Sauquet, E., Mengin, N., 2006. Appui scientifique à la mise en œuvre de la Directive Européenne Cadre sur l'Eau. Typologie des cours d'eau de France métropolitaine. Cemagref, 64 pp.
Charvet, S., Statzner, B., Usseglio-Polatera, P., Dumont, B., 2000. Traits of benthic macroinvertebrates in semi-natural French streams: an initial application to biomonitoring in Europe. Freshwater Biology 43 (2), 277–296.
Chevenet, F., Dolédec, S., Chessel, D., 1994. A fuzzy coding approach for the analysis of long-term ecological data. Freshwater Biology 31 (3), 295–309.
Comte, L., Lek, S., de Deckere, E., de Zwart, D., Gevrey, M., 2010. Assessment of stream biological responses under multiple-stress conditions. Environmental Science and Pollution Research 17 (8), 1469–1478.
Dahl, J., Johnson, R.K., 2004. A multimetric macroinvertebrate index for detecting organic pollution of streams in southern Sweden. Archiv für Hydrobiologie 160 (4), 487–513.
Dolédec, S., Statzner, B., 2008. Invertebrate traits for the biomonitoring of large European rivers: an assessment of specific types of human impact. Freshwater Biology 53 (3), 617–634.
Dolédec, S., Statzner, B., Bournard, M., 1999. Species traits for future biomonitoring across ecoregions: patterns along a human-impacted river. Freshwater Biology 42 (4), 737–758.
European Commission, 2003. Common implementation strategy for the Water Framework Directive (2000/60/EC). Guidance document No. 10, rivers and lakes – typology, reference conditions and classification systems. Office for official publications of the European Communities, Luxembourg, 94 pp.
European Council, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy. 72 pp.
ESRI (Environmental Systems Research Institute), 2006. ArcGis 9.2. ESRI, Redlands, CA, USA.

Gabriels, W., Lock, K., Pauw, N.D., Goethals, P.L., 2010. Multimetric Macroinvertebrate Index Flanders (MMIF) for biological assessment of rivers and lakes in Flanders (Belgium). Limnologica – Ecology and Management of Inland Waters 40 (3), 199–207.

Gayraud, S., Statzner, B., Bady, P., Haybach, A., Schöll, F., Usseglio-Polatera, P., Bacchi, M., 2003. Invertebrate traits for the biomonitoring of large European rivers: an initial assessment of alternative metrics. Freshwater Biology 48 (11), 2045–2064.

Giraudoux, P., 2011. pgirmess: data analysis in ecology. R package version 1.5.1.

Gotelli, N.J., McCabe, D.J., 2002. Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. Ecology 83 (8), 2091–2096.

Hellawell, J., 1986. Biological Indicators of Freshwater Pollution and Environmental Management. Elsevier Applied Science Publishers, New York, USA.

Hering, D., Johnson, R.K., Buffagni, A., 2006a. Linking organism groups – major results and conclusions from the STAR project. In: Furse, M.T., Hering, D., Brabec, K., Buffagny, A., Sandin, L., Verdonschot, P.F.M. (Eds.), The Ecological Status of European rivers: Evaluation and Intercalibration of Assessment Methods. Springer, Netherlands, pp. 109–113.

Hering, D., Feld, C., Moog, O., Ofenböck, T., 2006b. Cook book for the development of a multimetric index for biological condition of aquatic ecosystems: experiences from the European AQEM and STAR projects and related initiatives. Hydrobiologia 566 (1), 311–324.

Hering, D., Verdonschot, P.F.M., Moog, O., Sandin, L., 2004a. Integrated assessment of running waters in Europe. Hydrobiologia 516 (1–3), 1–379.

Hering, D., Meier, C., Rawer-Jost, C., Feld, C.K., Biss, R., Zenker, A., Sundermann, A., Lohse, S., Böhmer, J., 2004b. Assessing streams in Germany with benthic invertebrates: selection of candidate metrics. Limnologica – Ecology and Management of Inland Waters 34 (4), 398–415.

Hothorn, T., Hornik, K., Zeileis, A., 2006. Unbiased recursive partitioning: a conditional inference framework. Journal of Computational & Graphical Statistics 15 (3), 651–674.

Karr, J.R., 1981. Assessment of biotic integrity using fish communities. Fisheries 6 (6), 21–27.

Karr, J.R., Chu, E.W., 1997. Biological monitoring: essential foundation for ecological risk assessment. Human and Ecological Risk Assessment: An International Journal 3 (6), 993–1004.

Kerans, B.L., Karr, J.R., 1994. A Benthic Index of Biotic Integrity (B-IBI) for rivers of the Tennessee valley. Ecological Applications 4 (4), 768–785.

Kolpin, D.W., Furlong, E.T., Meyer, M.T., Thurman, E.M., Zaugg, S.D., Barber, L.B., Buxton, H.T., 2002. Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance. Environmental Science & Technology 36 (6), 1202–1211.

Liess, M., von der Ohe, P.C., 2005. Analyzing effects of pesticides on invertebrate communities in streams. Environmental Toxicology and Chemistry 24 (4), 954–965.

Lorenz, A., Hering, D., Feld, C., Rolauffs, P., 2004. A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. Hydrobiologia 516 (1), 107–127.

Lücke, J.D., Johnson, R.K., 2009. Detection of ecological change in stream macroinvertebrate assemblages using single metric, multimetric or multivariate approaches. Ecological Indicators 9 (4), 659–669.

MEDD, 2005. Circulaire DCE 2005/11 relative à la typologie nationale des eaux de surface (cours d'eau, plans d'eau, eau de transition et eaux côtières) en application de la directive 2000/60/DCE du 23 octobre 2000 du Parlement et du Conseil établissant un cadre pour une politique communautaire dans le domaine de l'eau. 18 pp.

Mendes, R.S., Evangelista, L.R., Thomaz, S.M., Agostinho, A.A., Gomes, L.C., 2008. A unified index to measure ecological diversity and species rarity. Ecography 31 (4), 450–456.

Mondy, C.P., Usseglio-Polatera, P., 2009. Invertébrés benthiques: métriques biologiques, conditions de référence, méthodes de bioindication, système-expert. Université Paul Verlaine – Metz, 55 pp.

Munné, A., Prat, N., 2009. Use of macroinvertebrate-based multimetric indices for water quality evaluation in Spanish Mediterranean rivers: an intercalibration approach with the IBMWP index. Hydrobiologia 628 (1), 203–225.

Norris, R.H., Thorns, M.C., 1999. What is river health? Freshwater Biology 41 (2), 197–209.

Ofenböck, T., Moog, O., Gerritsen, J., Barbour, M., 2004. A stressor specific multimetric approach for monitoring running waters in Austria using benthic macro-invertebrates. Hydrobiologia 516 (1–3), 251–268.

Oudin, L.C., Maupas, D., 2003. Système d'évaluation de la qualité de l'eau des cours d'eau (Seq-Eau). Ministère de l'Ecologie et du Développement Durable – Agences de l'Eau, Paris, France.

Pauw, N., Vanhooren, G., 1983. Method for biological quality assessment of watercourses in Belgium. Hydrobiologia 100 (1), 153–168.

Pinto, P., Rosado, J., Morais, M., Antunes, I., 2004. Assessment methodology for southern siliceous basins in Portugal. Hydrobiologia 516 (1), 191–214.

Piscart, C., Usseglio-Polatera, P., Moreteau, J., Beisel, J.-N., 2006. The role of salinity in the selection of biological traits of freshwater invertebrates. Archiv für Hydrobiologie 166 (2), 185–198.

Poff, N., 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. Journal of the North American Benthological Society 16 (2), 391–409.

Poquet, J., Alba-Tercedor, J., Punti, T., del Mar Sanchez-Montoya, M., Robles, S., Alvarez, M., Zamora-Munoz, C., Sainz-Cantero, C., Vidal-Abarca, M., Suarez, M., Toro, M., Pujante, A., Rieradevall, M., Prat, N., 2009. The MEDiterranean Prediction And Classification System (MEDPACS): an implementation of the RIVPACS/AUSRIVAS predictive approach for assessing Mediterranean aquatic macroinvertebrate communities. Hydrobiologia 623 (1), 153–171.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Resh, V.H., Hildrew, A.G., Statzner, B., Townsend, C.R., 1994. Theoretical habitat templets, species traits, and species richness: a synthesis of long-term ecological research on the Upper Rhône River in the context of concurrently developed ecological theory. Freshwater Biology 31 (3), 539–554.

Rosenberg, D.M., Resh, V.H., 1993. Introduction to freshwater biomonitoring and benthic macroinvertebrates. In: Rosenberg, D.M., Resh, V.H. (Eds.), Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York, USA, pp. 1–9.

Royer, T.V., Robinson, C.T., Minshall, G.W., 2001. Development of macroinvertebrate-based index for bioassessment of Idaho rivers. Environmental Management 27 (4), 627–636.

Sandin, L., Dahl, J., Johnson, R.K., 2004. Assessing acid stress in Swedish boreal and alpine streams using benthic macroinvertebrates. Hydrobiologia 516 (1–3), 129–148.

Sandin, R., Johnson, L., 2000. The statistical power of selected indicator metrics using macroinvertebrates for assessing acidification and eutrophication of running waters. Hydrobiologia 422/423, 233–243.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Technical Journal 27 (3–4), 379–423, 623–656.

Siegel, S., Castellan, N.J.J., 1988. Nonparametric Statistics for the Behavioral Sciences, 2nd ed. McGraw-Hill Book Company, New York, NY/England.

Skoulikidis, N., Gritzalis, K., Kouvarda, T., Buffagni, A., 2004. The development of an ecological quality assessment and classification system for Greek running waters based on benthic macroinvertebrates. Hydrobiologia 516 (1), 149–160.

Statzner, B., Bêche, L.A., 2010. Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? Freshwater Biology 55 (Suppl. s1), 80–119.

Statzner, B., Bady, P., Dolédec, S., Schöll, F., 2005. Invertebrate traits for the biomonitoring of large European rivers: an initial assessment of trait patterns in least impacted river reaches. Freshwater Biology 50 (12), 2136–2161.

Stirling, G., Wilsey, B., 2001. Empirical relationships between species richness, evenness, and proportional diversity. The American Naturalist 158 (3), 286–299.

Thorne, R., Williams, P., 1997. The response of benthic macroinvertebrates to pollution in developing countries: a multimetric system of bioassessment. Freshwater Biology 37 (3), 671–686.

Townsend, C.R., Hildrew, A.G., 1994. Species traits in relation to a habitat templet for river systems. Freshwater Biology 31 (3), 265–275.

Townsend, C.R., Uhlmann, S.S., Matthael, C.D., 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45 (6), 1810–1819.

Townsend, C., Dolédec, S., Scarsbrook, M., 1997. Species traits in relation to temporal and spatial heterogeneity in streams: a test of habitat templet theory. Freshwater Biology 37 (2), 367–387.

Usseglio-Polatera, P., Beisel, J.-N., 2002. Longitudinal changes in macroinvertebrate assemblages in the Meuse River: anthropogenic effects versus natural change. River Research and Applications 18 (2), 197–211.

Usseglio-Polatera, P., Bournaud, M., Richoux, P., Tachet, H., 2000. Biomonitoring through biological traits of benthic macroinvertebrates: how to use species trait databases? Hydrobiologia 422–423, 153–162.

Vlek, H.E., Verdonschot, P.F.M., Nijboer, R.C., 2004. Towards a multimetric index for the assessment of Dutch streams using benthic macroinvertebrates. Hydrobiologia 516 (1–3), 173–189.

von der Ohe, P.C., Liess, M., 2004. Relative sensitivity distribution of aquatic invertebrates to organic and metal compounds. Environmental Toxicology and Chemistry 23 (1), 150–156.

Wasson, J.G., Chandesris, A., Pella, H., Blanc, L., 2002. Les hydro-écorégions de France métropolitaine: approche régionale de la typologie des eaux courantes et éléments pour la définition des peuplements de référence d'invertébrés. Cemagref, Lyon, 190 pp.