Ecological Indicators
September 2013 ; Volume 32, Pages 147–156
http://dx.doi.org/10.1016/j.ecolind.2013.03.030

A fish-based index of estuarine ecological quality incorporating information from both scientific fish survey and experts knowledge

A. Tableau[a, 1, *], H. Drouineau[a], C. Delpech[a], M. Pierre[a], J. Lobry[a], O. Le Pape[b], J. Breine[c, d], M. Lepage

[a] Irstea, UR EPBX, 50 av. de Verdun, F-33612 Cestas, France
[b] UMR 985 ESE (INRA-Agrocampus-Ouest), 65 rue de St Brieuc, CS 84215 35042 Rennes Cedex, France
[c] Research Institute for Nature and Forest, Duboislaan 14, B-1560 Groenendaal, Belgium
[d] Laboratory of Aquatic Ecology and Evolutionary Biology, Katholieke Universiteit Leuven, Charles Deberiotstraat 32, B-3000 Leuven, Belgium

[1] Present address: Ifremer – Département EMH, rue de l'Ile d'Yeu, 44311 Nantes Cedex 03, France

*: Corresponding author : Adrien Tableau, tel.: +33 6 59 58 36 00 ; email addresses : adrien.tableau@gmail.com ; adrien.tableau@ifremer.fr

Abstract:

In the Water Framework Directive (European Union) context, a multimetric fish based index is required to assess the ecological status of French estuarine water bodies. A first indicator called ELFI was developed, however similarly to most indicators, the method to combine the core metrics was rather subjective and this indicator does not provide uncertainty assessment. Recently, a Bayesian method to build indicators was developed and appeared relevant to select metrics sensitive to global anthropogenic pressure, to combine them objectively in an index and to provide a measure of uncertainty around the diagnostic. Moreover, the Bayesian framework is especially well adapted to integrate knowledge and information not included in surveys data. In this context, the present study used this Bayesian method to build a multimetric fish based index of ecological quality accounting for experts knowledge. The first step consisted in elaborating a questionnaire to collect assessments from different experts then in building relevant priors to summarize those assessments for each water body. Then, these priors were combined with surveys data in the index to complement the diagnosis of quality. Finally, a comparison between diagnoses using only fish data and using both information sources underlined experts knowledge contribution. Regarding the results, 68% of the diagnosis matched demonstrating that including experts knowledge thanks to the Bayesian framework confirmed or slightly modified the diagnosis provided by survey data but influenced uncertainty around the diagnostic and appeared especially relevant in terms of risk management.

Keywords: Anthropogenic pressure ; Bayesian method ; Expert judgement ; Multimetric fish-based indicator ; Prior information ; Water Framework Directive

1. Introduction

Coastal and estuarine ecosystems are particularly vulnerable to evolution of human activities (Henocque and Denis, 2001; Hoegh-Guldberg and Bruno, 2010) and their degradation is widely observed, e.g. Elliott and Hemingway (2002). In that context, regulation tools such as the European Water Framework Directive (WFD) aims at stopping this degradation process and at restoring aquatic ecosystems to a good ecological status (WFD – Directive 2000/60/EC; European Council, 2000). To fulfil this objective, multimetric indices are widely used to assess the ecological quality of aquatic ecosystems (Hughes and Oberdorff, 1999). A metric is defined as a measurable variable having an ecological meaning, which can be associated to any structural or functional aspect of biological assemblages (Coates et al., 2007). Combining several metrics in a multimetric index assures that the resulting indices are holistic and sensitive (Deegan et al., 1997; Karr and Chu, 1999). A large variety of multimetric indices aims at detecting the ecological impact of stressors, e.g Hering et al. (2006).

As fish integrate a large variety of anthropogenic pressures (Elliott et al., 1988; Karr, 1981), fish assemblages are generally considered as appropriate to develop indicators of ecosystem quality. Consequently, numerous fish-based multimetric indices have been developed in the context of the WFD (Pont et al., 2006), especially in transitional waters, e.g. Borja et al. (2004), Breine et al. (2010), Breine et al. (2007) and Delpech et al. (2010). However, most of those indices suffered from two main weaknesses. First, qualitative estimates of the respective weight of metrics, correlations among them and redundancy in information made combinations of metrics in those indices sensitive to the calculation method. Second, most of those indices did not quantify uncertainty around their assessment (Perez-Dominguez et al., 2012) though it is especially important for managers (Breine et al., 2007; McAllister and Kirkwood, 1998). Indeed, the probability for a water body to reach a score below good status is necessary in terms of risk management (WFD – Directive 2000/60/EC; European Council, 2000). Bootstrap methods were applied with success by Pont et al. (2006) to estimate these probabilities but they required a large amount of data, making them inappropriate in data poor situations, especially in estuarine systems where large standardized sets of surveys data are lacking (Nicolas et al., 2011).

Delpech et al. (2010) proposed an indicator called ELFI to assess the ecological status of French estuaries. This indicator was based on pressure-impact models, as proposed by Courrat et al. (2009), in order to select relevant metrics that are sensitive to anthropogenic pressures. However, this indicator suffered from the two weaknesses previously mentioned. Recently, an original method to build multimetric stressor specific indices was proposed and applied on the French lagoons (Drouineau et al., 2012). This approach based on the Bayesian theory took up both two challenges. First, based on the Bayesian theory, it allows an objective combination of the metrics by incorporating all the information in a probabilistic framework. Secondly, this method provides a measure of uncertainty in its assessment. Consequently, it is hoped that applying this framework to the indicator developed by Delpech et al. (2010) would significantly improve the indicator.

Drouineau et al. (2012) proposed to further incorporate experts knowledge within the French lagoons fish-based index in order to combine both experts knowledge and fish data in the ecological assessment. Integrating experts knowledge in such multimetric index could lead to valuable improvement (Martin et al., 2005; Murray et al., 2009). Experts knowledge may provide a qualitative but global image (Knapp et al., 2011) of the ecological status of an ecosystem not only based on a particular ecological feature (ecological communities or habitats). Furthermore, it does not require expensive scientific surveys. Indeed, most indices are based on surveys with a limited time and space scale, sampling a limited fraction of fish assemblages. Consequently, they are based on restricted image of the ecosystem.

In this context, the present approach developed a fish-based multimetric index for French estuaries, applying the Bayesian method proposed by Drouineau et al. (2012) and incorporating experts knowledge. This index proposed to fulfil weaknesses of the index developed by Delpech et al. (2010) while using its pressure-impact approach and the method developed by Courrat et al. (2009) to select core metrics. The combination of fish data and experts information into the Bayesian framework is described, the impact of the integration of the experts knowledge in the ecological assessment and its interest for monitoring estuarine ecological status are analysed.

2. Materials and method

The proposed multimetric index illustrated in Figure 1 was based on two types of data: fish data from scientific surveys and experts knowledge. Fish data were correlated to an anthropogenic pressure index using pressure-impact statistical models (Courrat et al., 2009; Delpech et al., 2010) (top box in Figure 1). Then models were used to convert fish observations in probabilities of experiencing pressures (Drouineau et al., 2012) (left part of the second box). Experts assessments were aggregated to provide a global assessment per water body (right part of the second box) that was used as a prior in a Bayesian framework that combined both types of data (Drouineau et al., 2012) (third box). This allowed to put forward a pressure level applied on the fish communities of the studied water body. Pressure was decomposed in 5 equal pressure classes. Probability for a water body to be in each class was calculated (fourth box). Last, this level was associated to a quality level (last box).

2.1. Data sets

2.1.1. Fish data and pressure-impact models

A panel of 36 water bodies located along the French coasts of the English Channel and the Bay of Biscay were sampled between 2005 and 2009. A subset of 22 water bodies was sampled in 2010 to complete the WFD schedule (Fig. 2). Indeed, each water body had to be sampled at least 3 years during the 6-year WFD program. A detailed description of the sampling protocol was provided by Delpech et al. (2010) and Lepage and Girardin (2006). Each monitored water body was sampled in spring and autumn with a beam trawl. Hauls were distributed along the salinity gradient and 3 salinity classes were defined (Delpech et al., 2010): oligohaline class ([0–5] $g.L^{-1}$), mesohaline class (]5–18] $g.L^{-1}$) and polyhaline class (>18 $g.L^{-1}$). In each season, at least 6 hauls were carried out in each salinity zone of each water body. A minimum of 12 hauls was carried out in the water bodies having only one salinity zone (e.g. Downstream Seine). In each trawl haul (sample), each fish was identified to the species level and each species was assigned to functional ecological guilds related to its diet and its use of the estuarine ecosystems along its life cycle (Elliott and Dewailly, 1995).

The lack of pristine estuaries to define reference conditions involved the use of statistical modelling (Delpech et al., 2010; Pont et al., 2006). A solution was to develop pressure-impact models (Borja et al., 2006). However, a proxy of anthropogenic pressure is generally required to link fish data to human disturbances. In this study, an index developed by Courrat et al. (2009) was used. This index was based on a principal component analysis carried out on heavy metals concentrations measured by the French national network of quality of the French marine environment since 1979 in suspension feeder molluscs from various French estuaries. This index was normalized on the range [0;1].

Based on those data, Courrat et al. (2009) proposed to build pressure-impact models (generalized linear models) to select metrics that are sensitive to anthropogenic pressure. Delpech et al. (2010) selected 4 metrics among the 12 metrics tested (density of (i) benthic fish, (ii) diadromous species, (iii) marine juveniles migrants and (iv) total density of fish respectively denoted DB, DDIA, DMJ and TD) because they significantly responded to a variation of anthropogenic pressures, and consistently with expert judgments (Delpech et al., 2010). The pressure-impact models were then used to predict the expected value of each metric at 3 distinct levels of anthropogenic pressures, providing thresholds for each metric. A scoring method was then applied to combine the 4 metrics in the indicator.

Pressure-impact models as proposed by Courrat et al.(2009) and Delpech et al. (2010) appeared appropriate to select metrics negatively correlated to pressure, and potentially relevant to be included in the index. Consequently, the metrics selected by Delpech et al. (2010) were used, except the total density metric (TD) given its redundancy with the combination of the 3 others (DB, DDIA and DMJ). Pressure-impact models were fitted and included in the framework proposed by Drouineau et al. (2012). Models options depended on data distribution of the different metrics. The use of linear models, consistent for DB metric, was inappropriate for the other metrics composed of 0 inflated. Thus, a delta type model that consisted in a combination of two models was used: one for presence/absence modelling, another one for positive values modelling. Those models were similar to models developed by Delpech et al. (2010), except for the use of generalized linear mixed models (GLMMs) rather than generalized linear models (GLMs). Indeed, an “estuary” random effect was incorporated in the model to avoid correlation between hauls carried out within an estuary. To assess estuaries quality in 2010, the models were fitted on previous fish data collected between 2005 and 2009, using R software (R Development Core Team, 2011). This dataset (2005-2009) was considered for calibration of pressure-impact models. Akaike Information Criterion was used to select the most relevant and parsimonious models (Akaike, 1973).

GLMM of a metric $M_{(i)}$ can be written on a matrix form (Drouineau et al., 2012):

$$g_{(i)}(E(M_{(i)})) = \alpha_{(i)}X + \beta_{(i)}\,Pr + Reffect$$

with $g_{(i)}(E(M_{(i)}))$ the link transformed (function $g_{(i)}$) expected value of the i-th metric, $\alpha_{(i)}$ the regression parameters for covariables, X the model matrix for the covariables, $\beta_{(i)}$ the regression parameter for the pressure, Pr the vector of pressure values and *Reffect* the random effect linked to estuaries. Two kinds of covariables detailed in Delpech et al. (2010) were tested: (i) 'season' and 'salinity class' were effects of protocol and (ii) 'size' and 'ecoregion' were estuarine features. Fitted GLMMs enabled to establish the following likelihood function with pressure as a parameter:

$$L(\textit{fish data};\textit{pressure}) = f(\textit{fish data}|\textit{pressure})$$

With $f(\textit{fish data}|\textit{pressure})$, the density of probability for a fish observation given a level of pressure.

This was used to convert 2010 fish data into probability densities; so that metrics were combined on a common scale (Drouineau et al., 2012).

2.1.2. Experts knowledge

A panel of 23 experts was interviewed to collect knowledge on the different water bodies. Experts having information on fish data collected in 2010 were not included in the panel to avoid hindsight bias (redundant information between fish data and experts opinions) (Morgan and Henrion, 1990). Experts were selected for their abilities to assess a global pressure level for one part of the sampled water bodies. Finally, the 23 experts provided 100 assessments for the 22 studied water bodies (between 3 and 8 assessments per water body except Risle with only 2 assessments).

A user-friendly graphical interface was developed to collect experts assessment per water body. A cursors system was used to assess both mean level of pressure and to provide a reliability measure of the assessment (Fig. 3). This reliability box is a self-evaluation of the level of confidence the expert have on his assessment.

In order to guide his reflection, three sub-pressures (pollution, morphology and hydrology) had to be considered by the expert before assessing the global pressure (Fig.3). A manual provided explanations about the definition of each sub-pressure, the scale corresponding to the WFD reference conditions and the use of probability distributions. Moreover, a table at the end of the questionnaire presented a summary of all water bodies assessments realized by the expert, so that he could compare and adjust his assessments.

Experts assessments were aggregated in a unique informative prior per water body. To represent the experts diversity and get a stable prior, a minimum number of opinions is required (Kuhnert et al., 2005), but a limited number of 3 to 5 experts is suggested to provide a compromise between redundant information and a good representation of experts intervariability (Clemen and Winkler, 1985; Ferrell, 1985; Makridakis and Winkler, 1983). To get a consistent framework between priors given the limited number of available assessments per water body, 3 experts distributions were used to build each prior. Given that experts could assess a water body even with a low knowledge level, the assessments with the highest precisions were used to assimilate a good quality of information in priors.

A beta distribution was fitted on the mixture of the 3 assessments to build a consensus prior (O'Hagan, 1998) for each water body (wb), ranging from [0;1] and denoted $prior_{wb}$.

2.2. Computing the index

Experts priors and fish data collected in 2010 were used to assess the quality of the different water bodies using the following Bayesian equation (Drouineau et al., 2012):

The probability that the pressure level applied on a water body denoted *wb* was in a restricted pressure class, given *J* fish observations, is (1):

$$(1)\ P(pressure_{wb} \epsilon\ class_i\ |fish\ data_{wb}) = \frac{\int_{c_{i_{min}}}^{c_{i_{max}}} \prod_j f\left(fish\ data_{wb_j} \middle| pressure_{wb}\right) * g(pressure_{wb}).d\ pressure}{\int_{p_{min}}^{p_{max}} \prod_j f\left(fish\ data_{wb_j} \middle| pressure_{wb}\right).d\ pressure}$$

With $[c_{i\ min};c_{i\ max}]$ and $[p_{min},p_{max}]$ the domains of definition respectively of the pressure class *i* and the whole pressure, $f(fish\ data_{wb\ j}\ |pressure_{wb})$ the density of probability of fish observation *j (j in [1:J])* given a pressure level directly calculated from the outputs of GLMMs, $g(pressure_{wb})$ a density of probability coming from a prior distribution of pressure.

This equation was summarised as the following relationship in Figure 1 (2):

$$(2)\ P(pressure_{wb}|fish\ data_{wb}) \propto P(fish\ data_{wb}|pressure_{wb}) * P(pressure_{wb})$$

To analyse the influence of experts knowledge, two indices were computed for each water body. The first index included the prior based on experts knowledge (the resulting distribution is denoted $index_{wb,exp}$) corresponding to the final indicator. The second index did not use experts prior but an uninformative prior which was a uniform distribution between [0;1] (the resulting distribution is denoted $index_{wb,non}$). WinBUGS (Lunn et al., 2000) was used to compute the a posteriori distribution, with a

Gelman Rubin test (Brooks and Gelman, 1998) to check chains convergence. Probability densities corresponding to each index were arbitrarily decomposed into 5 equal classes corresponding to 5 quality classes.

For each water body, both the posterior probability in percentage in each quality level and the related mean class were computed for the 3 variables ($prior_{wb}$, $index_{wb,exp}$ and $index_{wb,non}$). The comparison of these variables would inform on (i) the agreement level between experts knowledge and information from field data and on (ii) the effect of the combination of both information sources on quality statuses assessment. Results were synthesised in a Principal Components Analysis (PCA) where the 22 water bodies were individuals and the variables the 3 mean quality values.

3. Results

3.1. Generalised linear models

The selected models were detailed in Table 1. Regression parameters of pressure were negative demonstrating that pressure and fish data were negatively correlated as expected. Regarding log normal models (positive fish data for metrics DDIA and DMJ), the standard deviations for the among-estuaries random effect, *estuary sd*, of the 3 metrics DB, DDIA and DMJ were respectively 0.70, 0.45 and 0.88, while residuals variations among all fish data, *model sd*, were respectively 1.35, 1.22 and 1.26. The relative high values of standard deviations of the random effect clearly justified the interest to include it in the models to provide accurate uncertainty assessments.

3.2. Quality diagnosis

3.2.1. Overview of the diagnosis provided by the 3 indices

The results of the 3 indices are summarized in Table 2. Mean classes for $index_{wb,non}$ (i.e. diagnosis from fish data) and $prior_{wb}$ (i.e. diagnosis from experts knowledge) were similar for only 32% of the water bodies. However, only 18% differed from more than one class, demonstrating a rather good consistency between fish data and experts knowledge (correlation rate of 0.63), with nonetheless an important discrepancy in the case of the Vilaine estuary. The final indicator, $index_{wb,exp}$ (i.e. diagnosis from fish data and experts knowledge), was distributed between all quality classes, from high quality for Vilaine and Baie du Mont Saint Michel to bad quality for Bidassoa, Downstream and Central

Gironde, Upstream, Central and Downstream Seine. The associated credibility intervals at 95% were heterogeneous covering from 1 (e.g. Central Seine) to 3 (e.g. Seudre) quality classes. Consequently, the index precision was variable from one water body to another; it was lower for upstream water bodies of estuaries and for small estuaries. Comparing $index_{wb,exp}$ to the two other indices showed that mean class was either between $index_{wb,non}$ and $prior_{wb}$ (14% of water bodies), or that it matched with both of them (32%), or only with one of them (36% with $index_{wb,non}$ and 18% with $prior_{wb}$).

3.2.2. Quality diagnosis and Water Framework Directive objectives

According to the normalized scale of quality retained in the present approach, 5 out of 22 water bodies fulfilled the objective to be at least in the good quality status and only 2 (Baie du Mont St Michel and Vilaine) without significant risk, i.e. with insignificant probability to be in the quality statuses lower than the good status. The Blavet had a high probability to be in a lower quality range and the risk was also significant respectively at 18% and 14% for Baie des Veys and Laita (Table 2).

3.2.3. Global comparison between the 3 indices

The three variables (mean values of quality for $index_{non,wb}$, $prior_{wb}$ and $index_{exp,\ wb}$) were correlated to the first principal component of the PCA (Fig. 4), representing 86 % of the inertia and separating deteriorated from low-impacted water bodies. $Index_{exp,\ wb}$ was in intermediate position between $index_{non,wb}$ and $prior_{wb}$, providing a consensus between fish data and experts knowledge but giving more weight to fish data. Indeed, the correlation rate of $index_{exp,\ wb}$ and $index_{non,wb}$ is 0.96 while the $index_{exp,\ wb}$ and $prior_{wb}$ one is 0.75 and the $index_{non,wb}$ and $prior_{wb}$ one is 0.63.

3.2.4. Diagnosis general trends

The position of the water bodies in the first PCA plane highlighted geographic contrasts (Fig. 4). The negative correlation between the first principal component and the Adour Garonne district water bodies revealed their bad quality contrasting with the Loire Bretagne district status. The Seine Normandie district water bodies were distributed in all quality scale. Moreover, the water bodies of the largest estuaries, Seine, Loire and Garonne/Gironde, were the most deteriorated of their respective districts.

Three main types of experts knowledge effects on the assessments are illustrated in Figure 5. In the Charente, the mean classes obtained with $index_{wb,non}$ and $prior_{wb}$ were very different (respectively low pressure and high pressure) though the precision was similar. As a consequence, $index_{wb,exp}$ provided a consensus, i.e. medium class with an important uncertainty. A second situation occurred when mean classes obtained with $index_{wb,non}$ and $prior_{wb}$ were adjacent such as for Baie du Mont Saint Michel. For this water body, $index_{wb,exp}$ provided an assessment consistent with $index_{wb,non}$ but with a lower precision. Finally, for some water bodies, $index_{wb,non}$ and $prior_{wb}$ were consistent, such as for the Risle. In this case, $index_{wb,exp}$ was consistent with both $index_{wb,non}$ and $prior_{wb}$ with a greater precision.

4. Discussion

Many fish indicators were recently developed in the context of the Water Framework Directive. However, many of those indicators suffered from two main drawbacks: (i) assessments of uncertainties around quality diagnostics are missing (Perez-Dominguez et al., 2012) and (ii) the choice of core metrics combination is rather subjective. Regarding French estuaries, an indicator called ELFI was developed by Delpech et al. (Delpech et al., 2010), based on pressure-impact models proposed by Courrat et al. (Courrat et al., 2009). Pressure-impact approach proved to be a relevant method to choose appropriate metrics, however ELFI still suffered from the two weaknesses. Consequently, this paper developed an indicator based on ELFI pressure-impact models but applying a Bayesian framework developed by Drouineau et al. (2012). In addition, a method was proposed to collect and summarise expert judgments in order to combine them in the Bayesian framework with fish data. This approach found several interests detailed further.

4.1. The pressure-impact Bayesian models

The index was based on generalised linear mixed models, and not on previously used generalised linear models (Delpech et al., 2010). Indeed, fish data were not perfect replicates but pseudo-replicates: several trawl hauls were collected in each estuary, generating dependency. Given that the estuary effect could not be taken into account, the independence assumption of data was corrupted. A rigorous solution was to use random effect in mixed models, with, as a consequence, an increase of the standard error and uncertainty around the parameters. A random effect of year was also tested but was not significant. Other random effects such the interaction between salinity class and estuary were also considered but the present GLMMs were finally considered as suitable. The index was finally assessed on a Bayesian framework (Drouineau et al., 2012) and presented the advantages provided by this framework: (i) an objective combination of the core metrics and (ii) a quantification of the uncertainty around the diagnostic.

GLMMs were fitted with a pressure index based on water contamination in heavy metals (Courrat et al., 2009). As mentioned by Courrat et al. (2009), this index is highly correlated to many anthropogenic activities (industry, urbanisation, agriculture), consequently resulting pressure-impact models should be more considered as the global effect anthropogenic pressures than a direct effect of heavy metal contaminations. However, comparing results with models fitted on others global pressure indices could be relevant to assess the robustness of this hypothesis. It would also be interesting to apply the method to each different type of pressure separately (for example pollution, hydrology and morphology) to identify the more likely type of pressure impacting the ecological status. However, the water bodies are generally affected by various pressures simultaneously, consequently pressures are highly correlated and it would probably not be possible to conclude on the most likely pressure impacting a given water body. Moreover, to fulfil the WFD requirements, a method to combine the results on the different pressures in a unique indicator would be required. Bayesian Network may be an interesting method in the future to fulfil those challenges.

Regarding the indicator calibration and validation, it was chosen to calibrate the indicator on data collected from 2005 to 2009 and then to apply it on data collected in 2010. Usually, to cross-validate a model, a random subset of the entire dataset is used to calibrate and the other part of the data-set is used to validate the model. However, as proposed by Wenger and Olden (2012), it was chosen to apply a non-random cross-validation strategy which is appropriate to assess time transferability of the model. This strategy is consistent with the Water Framework Directive context: to assess the ecological status in a given year, all data from previous years will be used to calibrate the indicator since time series are often limited.

4.2. Taking experts knowledge into account

The assessment realized in the Bayesian framework was based on both experts opinions and scientific surveys. Only another WFD-like index integrated directly experts knowledge in its architecture (Cabral et al., 2012). Indeed, experts knowledge was rarely considered in ecological indicators (Carpenter, 2002).

4.2.1. A method to collect and standardize knowledge

Before assessing the global pressure, it was asked the experts to assess three sub-pressures for every water body. But assessing sub-pressures before the global one may introduce bias since it may influence experts judgments. However, it is required to enhance the reproducibility of those experts assessments. To limit the potential bias, no weighting rates were proposed to establish global pressure, so each expert provided them on his own.

In addition, experts assessments may contain biases linked to the lack of neutrality (Choy et al., 2009; Dennis, 1996). Moreover, experts are usually overconfident (Hora et al., 1992; Kadane and Wolfson, 1998; Winkler, 1967a, b). Consequently, too precise priors may drive parameters assessment rather than only provide a direction (Dennis, 1996). To get around both issues, knowledge of several experts was used to build priors to counterbalance both the bias linked to each expert background (Martin et al., 2005) and overconfidence. Indeed, the inter-variability between experts is generally higher than the variability of any expert opinion (Kuhnert et al., 2005; Uusitalo et al., 2005). Consequently, prior precision was mainly influenced by the agreement level between experts rather than the confidence level of each assessment.

Alternative methods exist to collect experts opinions. For instance, a simple choice in a range of different values is efficient if the number of experts is sufficient (Kuhnert et al., 2005; Martin et al., 2005). In other cases, and especially in the present approach, the limited number of experts involved the use of probability distributions. Assessing successive percentiles (Garthwaite and Dickey, 1996; Kadane and Wolfson, 1998) is a precise method but time consuming. A faster but less precise method (O'Hagan, 1998) aims at assessing a mean and a credibility interval of 50%. Both these methods can suffer of underestimation of distribution tails (Hora et al., 1992; Winkler, 1967a, b). Accordingly, assessing a graphical distribution directly was here preferred for its simplicity and user-friendliness, in order to use the instinct qualities of experts. A pressure scale with a colour gradient was developed instead of any common numerical scale to prevent the halo effect (Nisbett and Wilson, 1977), i.e. bias linked to the distorted perception of numerical scales of a repeated question.

Different solutions may be investigated to improve the precision of priors. The DELPHI method (Jacobs, 1995; Linstone et al., 1975) allows a prior to be elicited from several experts. A consensus is constructed indirectly after several assessment rounds. Between each round, each expert consults the assessments and associated rationales provided by the others before updating his opinion. This kind of method generally provides a more precise prior though a direct consensus between experts could also be supported (O'Hagan, 1998). However, such method is time and costs consuming for the experts. In addition, they may suffer from a potential strong influence of few dominating experts, leading to a precision overestimation (Kuhnert et al., 2005). Another method (Coolen and Newby, 1994) consists in defining a range of possible values, i.e. possible pressure levels here, selected by experts. A uniform prior is defined on the range. Such a prior is very informative because some pressure levels are unconsidered in the final index, leading to an unsuitable minor influence of fish data, and often to some problems in the convergence of Monte-Carlo Markov Chains.

4.2.2. Insights provided by experts knowledge

The questionnaire used in the present approach proved to be efficient, with an answer rate greater than 75%. As a positive side-effect, it also proved to be an interesting communication tool: experts were interested by sharing their opinions and felt involved in the index development. On the whole, experts knowledge was rather consistent with fish data, demonstrating that the indicator provides an objective consensus between both data sources. Nevertheless, some discrepancies allowed to point out that experts knowledge may provide information not included in the data. Especially, the Vilaine estuary appeared rather specific, with a large disagreement between fish data and experts knowledge. In this estuary, a dam was built in 1970 and greatly impacted the water body (i.e. meso-haline and oligo-haline have entirely disappeared and this water body is considered as heavily modified in the WFD (Borja and Elliott, 2007)) explaining the negative assessment provided by the experts. However, the impact was not detected from surveys in the remaining poly-haline zone and the assessment based on fish data only was positive. Indeed, the impacts of removing parts of estuarine system were difficult to quantify from fish data given samplings were performed in remaining surfaces (Courrat et al., 2009). Though the small prior precision did not allow the assessment to be influenced significantly, this example legitimises the interest of experts knowledge in the indicator.

On one hand, when experts disagreed with the conclusion provided by fish data (e.g Vilaine but also Charente and Baie du Mont Saint-Michel), the impact of the prior on the indicator depended on two factors. The more precise and in contradiction with fish data was the experts consensus, the more the final main quality class was modified. So, the weighting between fish data and experts knowledge in the final index is directly linked to their relative precisions. On the other hand, when experts knowledge and fish data matched, experts knowledge increased the final index precision (e.g. Risle). This last point illustrates another yet main advantage when using experts knowledge as the final assessment gains precision in a legitimate way. Applying experts knowledge appeared particularly essential in those situations where the most reliable assessment as possible should be obtained.

4.3. Ecological status assessment of estuarine water bodies from fish data and experts knowledge

Using the present approach, the probability of the fish based index to be in any range of stressor values can be easily calculated. This allocation was here based on an arbitrary decomposition into 5 equal classes of quality. A calibrating phase of the thresholds with the other European indices would be necessary to provide a relevant index with the WFD. Indeed, this approach focused on the interest of incorporating experts knowledge in fish index to improve quality assessment. It appeared especially appropriate to visualize sensitivity in allocation on a range of classes with well distributed data. In a WFD context, these classes do not necessary fit with five classes of environmental status. A calibrating phase of the thresholds with the other European indices will be necessary to provide a relevant index with the WFD. Nevertheless, this approach allowed to analyse patterns in quality assessment and associated variability among both estimation methods and water bodies.

The current index appeared particularly appropriate in the context of risk management, a notion developed by McAllister and Kirkwood (1998), and precautionary approach. In the WFD context, the risk may be defined by the probability of not being in a good ecological status which is provided by the Bayesian method (Drouineau et al., 2012). This might lead to different management measures for two water bodies such as Baie des Veys and Blavet. Both were qualified as having a good ecological status, but with different levels of risks (respectively 18% and 45%). In that context, though it had a moderate effect on the final mean class, including experts knowledge was a significant improvement because it often had a significant effect on the assessment precision.

According to the indicator, large estuaries (Seine, Gironde and Loire) were shown to be the most deteriorated estuaries of their district. As precision was high for these estuaries, this statement could be considered as reliable. As a consequence, specific restoration effort should be dedicated to large estuaries. Assessments from fish data tended to be less precise for upstream water bodies than for corresponding downstream water bodies. This indicated more restricted knowledge, perhaps linked to the choice of metrics or the fishing gear (beam trawl) used during the surveys, less appropriate for upstream water body. Similar remarks could be made when comparing small estuaries, which quality estimates were uncertain, to large ones. Two converging facts explained that point. First, the size of small estuaries limited sometimes the number of trawl hauls and consequently the assessment precision. Secondly, given small estuaries were not as studied as the large ones in the past, prior precision was globally smaller and did not match with fish data as much as for the large water bodies. In that context, new data acquisition should be obtained from upstream and small water bodies in priority, either from other surveys or by consulting specific experts.

The index proposed here appears especially relevant in data poor situation. For example in France, only 22 water bodies have presently been monitored (for 14 more, a previous estimate was realized between 2005 and 2009) though 54 estuarine transitional water bodies are listed in the WFD context. For the remaining estuaries, given the cost of scientific surveys, the index may first provide an assessment only based on experts knowledge. However, in this situation, these very preliminary assessments would be based on experts opinion including all the implied subjectivity. Moreover, the lack of quantitative assessment will prevent from measuring effects of restoration if water bodies are not considered in good ecological status.

In conclusion, the Bayesian approach is a generic method fulfilling WFD index objectives. It can be used for any index based on a pressure-impact approach. Its main advantages are (i) its flexibility in the modelling phase, (ii) its estimates of uncertainty and (iii) its possibility to integrate easily and rigorously experts knowledge. The presented method to collect and combine experts opinions in a prior should not be ignored given its cost-efficiency and its adequacy with the common availability of experts.

Acknowledgements

Authors would like to thank the French Water Agencies and the Onema for their collaboration and all the partners involved in the collect of survey data. They also would like to thank the experts who have participated to the study. This work is part of the WISER European project (Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery) funded by the European Union under the 7th Framework Programme, Theme 6 (Environment including Climate Change) (contract No. 226273). http://www.wiser.eu

Bibliography

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in: Petrov, B.N., Csaki, F. (Eds.), 2nd International Symposium on Information Theory. Akademiai Kiado, Tsahkadsor, pp. 267-281.

Borja, Á., Elliott, M., 2007. What does 'good ecological potential' mean, within the European Water Framework Directive? Mar. Pollut. Bull. 54, 1559-1564.

Borja, Á., Franco, J., Valencia, V., Bald, J., Muxika, I., Jesús Belzunce, M.a., Solaun, O., 2004. Implementation of the European water framework directive from the Basque country (northern Spain): a methodological approach. Mar. Pollut. Bull. 48, 209-218.

Borja, Á., Galparsoro, I., Solaun, O., Muxika, I., Tello, E.M., Uriarte, A., Valencia, V., 2006. The European Water Framework Directive and the DPSIR, a methodological approach to assess the risk of failing to achieve good ecological status. Estuar. Coast. Shelf Sci. 66, 84-96.

Breine, J., Maes, J., Quataert, P., Bergh, E., Simoens, I., Thuyne, G., Belpaire, C., 2007. A fish-based assessment tool for the ecological quality of the brackish Schelde estuary in Flanders (Belgium). Hydrobiologia 575, 141-159.

Breine, J., Quataert, P., Stevens, M., Ollevier, F., Volckaert, F.A.M., Van den Bergh, E., Maes, J., 2010. A zone-specific fish-based biotic index as a management tool for the Zeeschelde estuary (Belgium). Mar. Pollut. Bull. 60, 1099-1112.

Brooks, S.P., Gelman, A., 1998. General Methods for Monitoring Convergence of Iterative Simulations. J. Comp. Graph. Stat. 7, 434-455.

Cabral, H.N., Fonseca, V.F., Gamito, R., Gonçalves, C.I., Costa, J.L., Erzini, K., Gonçalves, J., Martins, J., Leite, L., Andrade, J.P., Ramos, S., Bordalo, A., Amorim, E., Neto, J.M., Marques, J.C., Rebelo, J.E., Silva, C., Castro, N., Almeida, P.R., Domingos, I., Gordo, L.S., Costa, M.J., 2012. Ecological quality assessment of transitional waters based on fish assemblages in Portuguese estuaries: The Estuarine Fish Assessment Index (EFAI). Ecol. Indic. 19, 144-153.

Carpenter, S.R., 2002. ECOLOGICAL FUTURES: BUILDING AN ECOLOGY OF THE LONG NOW[1]. Ecology 83, 2069-2083.

Choy, S.L., O'Leary, R., Mengersen, K., 2009. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology 90, 265-277.

Clemen, R.T., Winkler, R.L., 1985. Limits for the Precision and Value of Information from Dependent Sources. Oper. Res. 33, 427-442.

Coates, S., Waugh, A., Anwar, A., Robson, M., 2007. Efficacy of a multi-metric fish index as an analysis tool for the transitional fish component of the Water Framework Directive. Mar. Pollut. Bull. 55, 225-240.

Coolen, F.P.A., Newby, M.J., 1994. Bayesian reliability analysis with imprecise prior probabilities. Reliab. Eng. Syst. Safe. 43, 75-85.

Courrat, A., Lobry, J., Nicolas, D., Laffargue, P., Amara, R., Lepage, M., Girardin, M., Le Pape, O., 2009. Anthropogenic disturbance on nursery function of estuarine areas for marine species. Estuar. Coast. Shelf Sci. 81, 179-190.

Deegan, L., Finn, J., Ayvazian, S., Ryder-Kieffer, C., Buonaccorsi, J., 1997. Development and validation of an estuarine biotic integrity index. Estuaries 20, 601-617.

Delpech, C., Courrat, A., Pasquaud, S., Lobry, J., Le Pape, O., Nicolas, D., Boët, P., Girardin, M., Lepage, M., 2010. Development of a fish-based index to assess the ecological quality of transitional waters: the case of French estuaries. Mar. Pollut. Bull. 60, 908-918.

Dennis, B., 1996. Discussion: Should Ecologists Become Bayesians? Ecol. Appl. 6, 1095-1103.

Drouineau, H., Lobry, J., Delpech, C., Bouchoucha, M., Mahévas, S., Courrat, A., Pasquaud, S., Lepage, M., 2012. A Bayesian framework to objectively combine metrics when developing stressor specific multimetric indicator. Ecol. Indic. 13, 314-321.

Elliott, M., Dewailly, F., 1995. The structure and components of European estuarine fish assemblages. Neth. J. Aquat. Ecol. 29, 397-417.

Elliott, M., Griffiths, A.H., Taylor, C.J.L., 1988. The role of fish studies in estuarine pollution assessment. J. Fish Biol. 33, 51-61.

Elliott, M., Hemingway, K., 2002. Fishes in Estuaries. Blackwell Science Ltd, Oxford.

Ferrell, W., 1985. Combining Individual Judgments, in: Wright, G. (Ed.), Behavioral Decision Making. Plenum Press, New York, pp. 111-145.

Garthwaite, P.H., Dickey, J.M., 1996. Quantifying and using expert opinion for variable-selection problems in regression. Chemometr. Intell. Lab. 35, 1-26.

Henocque, Y., Denis, J., 2001. A Methodological Guide: Steps and Tools Towards Integrated Coastal Area Management, Manuals and guides. UNESCO, Paris.

Hering, D., Feld, C., 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.

Hoegh-Guldberg, O., Bruno, J.F., 2010. The Impact of Climate Change on the World's Marine Ecosystems. Science 328, 1523-1528.

Hora, S.C., Hora, J.A., Dodd, N.G., 1992. Assessment of probability distributions for continuous random variables: A comparison of the bisection and fixed value methods. Organ. Behav. Hum. Dec. 51, 133-155.

Hughes, R.M., Oberdorff, T., 1999. Applications of IBI concepts and metrics to waters outside the United States, in: Simon, T.P. (Ed.), Assessing the Sustainability and Biological Integrity of Water Resource Quality Using Fish Communities. CRC Press, Boca Raton, pp. 79–96.

Jacobs, R., 1995. Methods for combining experts' probability assessments. Neural Comput. 7, 867-888.

Kadane, J., Wolfson, L.J., 1998. Experiences in elicitation. J. R. Stat. Soc. 47, 3-19.

Karr, J.R., 1981. Assessment of Biotic Integrity Using Fish Communities. Fisheries 6, 21-27.

Karr, J.R., Chu, E.W., 1999. Restoring Life in Running Waters: Better Biological Monitoring. Island Press, Washington.

Knapp, C.N., Fernandez-Gimenez, M., Kachergis, E., Rudeen, A., 2011. Using Participatory Workshops to Integrate State-and-Transition Models Created With Local Knowledge and Ecological Data. Rangeland Ecol. Manag. 64, 158-170.

Kuhnert, P.M., Martin, T.G., Mengersen, K., Possingham, H.P., 2005. Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. Environmetrics 16, 717-747.

Lepage, M., Girardin, M., 2006. Inventaire Poisson dans les eaux de transition. Protocole d'échantillonnage de la façade Atlantique et Manche. Cemagref - groupement de Bordeaux, Cestas, p. 32.

Linstone, H.A., Turoff, M., Helmer, O., 1975. The Delphi Method: Techniques and Applications. Addison-Wesley, Boston.

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat. Comput. 10, 325-337.

Makridakis, S., Winkler, R.L., 1983. Averages of Forecasts: Some Empirical Results. Manage. Sci. 29, 987-996.

Martin, T.G., Kuhnert, P.M., Mengersen, K., Possingham, H.P., 2005. THE POWER OF EXPERT OPINION IN ECOLOGICAL MODELS USING BAYESIAN METHODS: IMPACT OF GRAZING ON BIRDS. Ecol. Appl. 15, 266-280.

McAllister, M.K., Kirkwood, G.P., 1998. Bayesian stock assessment: a review and example application using the logistic model. ICES J. Mar. Sci. 55, 1031-1060.

Morgan, M.G., Henrion, M., 1990. Uncertainty: a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York.

Murray, J.V., Goldizen, A.W., O'Leary, R.A., McAlpine, C.A., Possingham, H.P., Choy, S.L., 2009. How useful is expert opinion for predicting the distribution of a species within and beyond the region of expertise? A case study using brush-tailed rock-wallabies *Petrogale penicillata*. J. Appl. Ecol. 46, 842-851.

Nicolas, D., Chaalali, A., Drouineau, H., Lobry, J., Uriarte, A., Borja, A., Boët, P., 2011. Impact of global warming on European tidal estuaries: some evidence of northward migration of estuarine fish species. Reg. Environ. Change. 11, 639-649.

Nisbett, R.E., Wilson, T.D., 1977. The halo effect: Evidence for unconscious alteration of judgments. J. Pers. Soc. Psychol. 35, 250-256.

O'Hagan, A., 1998. Eliciting expert beliefs in substantial practical applications. J. R. Stat. Soc. 47, 21-35.

Perez-Dominguez, R., Maci, S., Courrat, A., Lepage, M., Borja, A., Uriarte, A., Neto, J.M., Cabral, H., Raykov, V.S., Franco, A., Alvarez, M.C., Elliott, M., 2012. Current developments on fish-based indices to assess ecological-quality status of estuaries and lagoons. Ecol. Indic. 23, 34-45.

Pont, D., Hugueny, B., Beier, U., Goffaux, D., Melcher, A., Noble, R., Rogers, C., Roset, N., Schmutz, S., 2006. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. J. Appl. Ecol. 43, 70-80.

R Development Core Team, 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Uusitalo, L., Kuikka, S., Romakkaniemi, A., 2005. Estimation of Atlantic salmon smolt carrying capacity of rivers using expert knowledge. ICES J. Mar. Sci. 62, 708-722.

Wenger, S.J., Olden, J.D., 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol. Evol. 3, 260-267.

WFD – Directive 2000/60/EC; 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. Off. J. Eur. Commun. L327, 1-72.

Winkler, R.L., 1967a. The Assessment of Prior Distributions in Bayesian Analysis. J. Am. Statist. Assoc. 62, 776-800.

Winkler, R.L., 1967b. The Quantification of Judgment: Some Methodological Suggestions. J. Am. Statist. Assoc. 62, 1105-1120.

Figure captions:

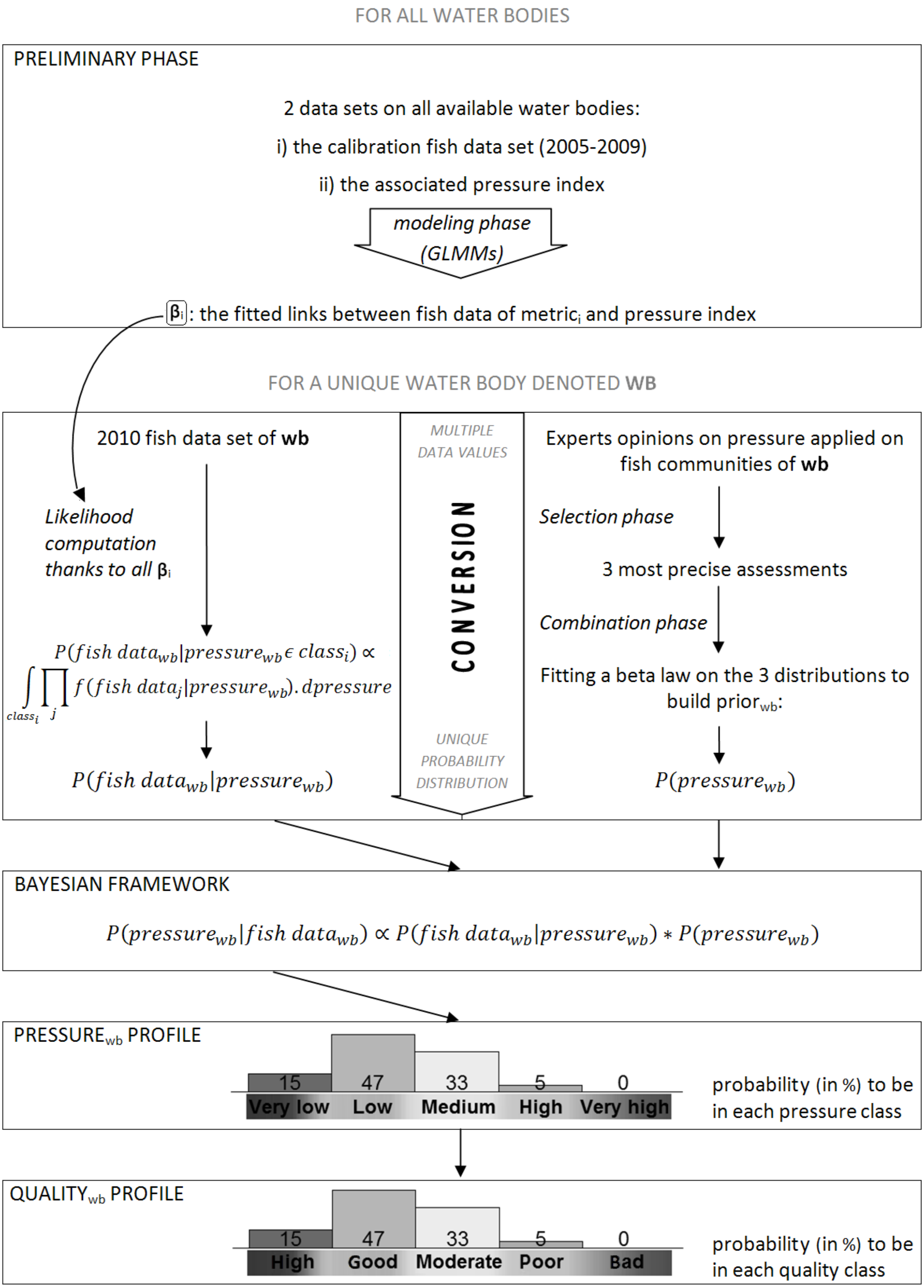
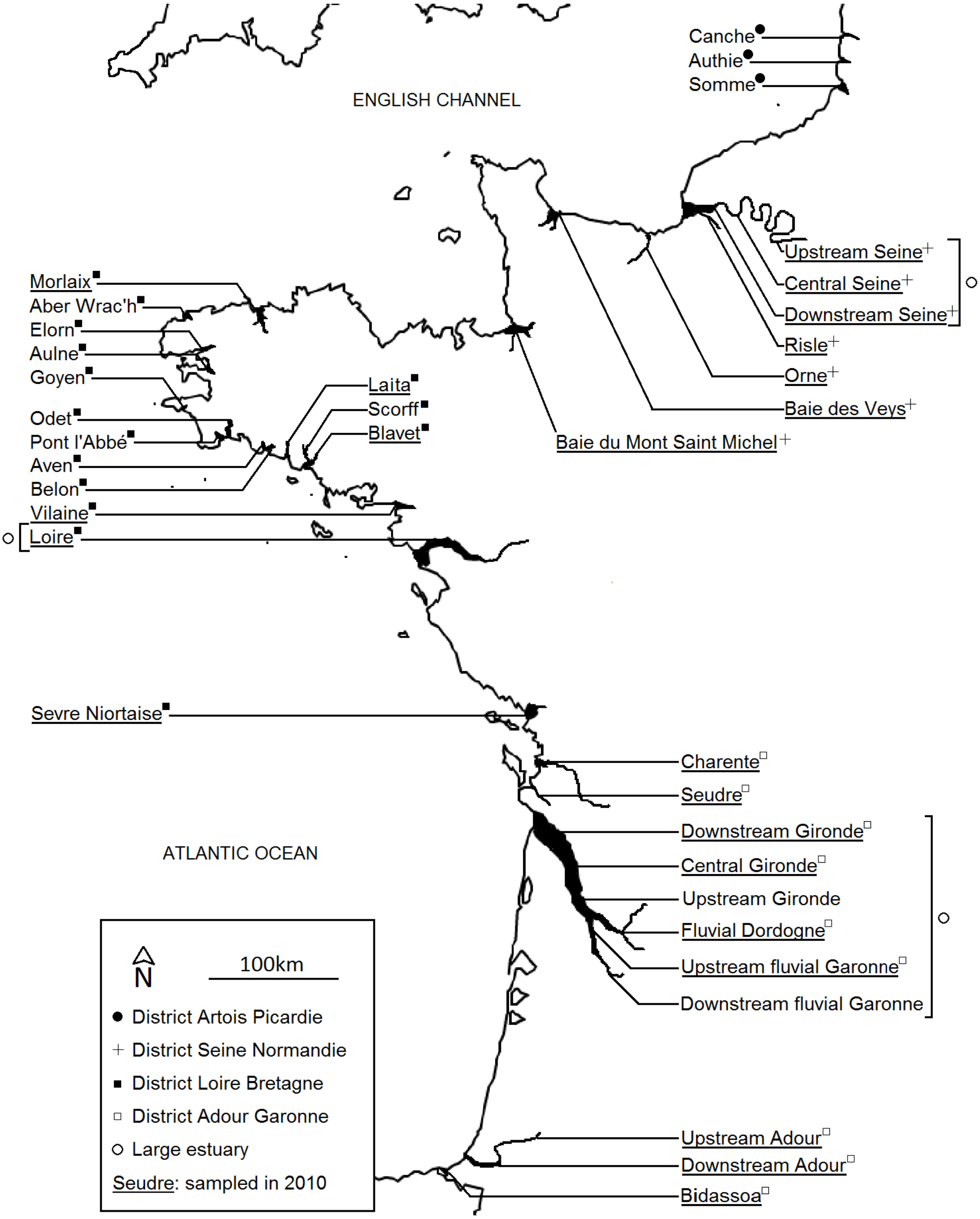


Fig. 1: General methodology to develop the index



577

578 Fig. 2: Study area and investigated water bodies in 2010 gathered in 3 geographical groups

579 corresponding to 3 districts; an estuary can be made up of several water bodies

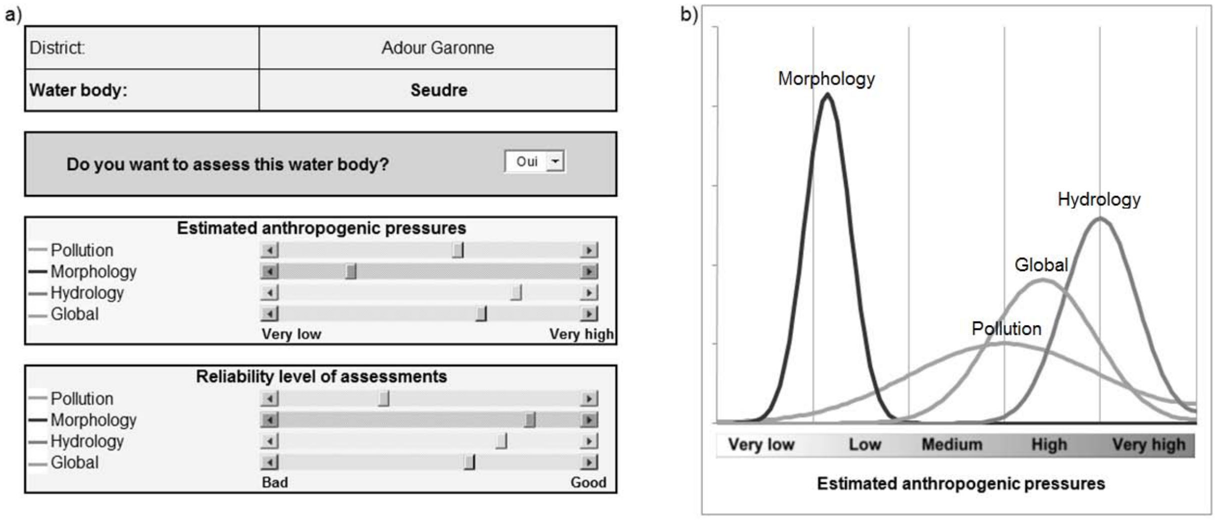


Fig. 3: Interface provided for experts to assess the pressure level of each water body: a) Table to fill by each expert; b) related representation

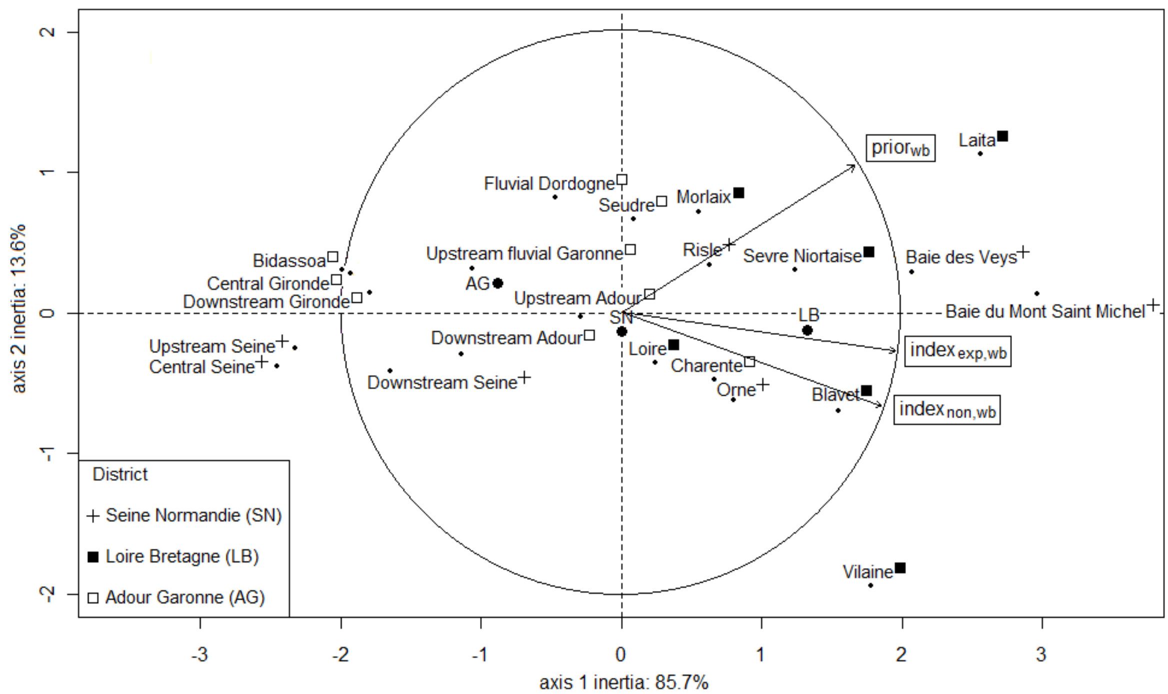


Fig. 4: Correlation circle of the PCA on the 3 indices ($index_{wb,non}$, $prior_{wb}$ and $index_{wb,exp}$), the scale is twice the original one; individuals (water bodies) are gathered according to their respective district in the first factorial plan of the PCA

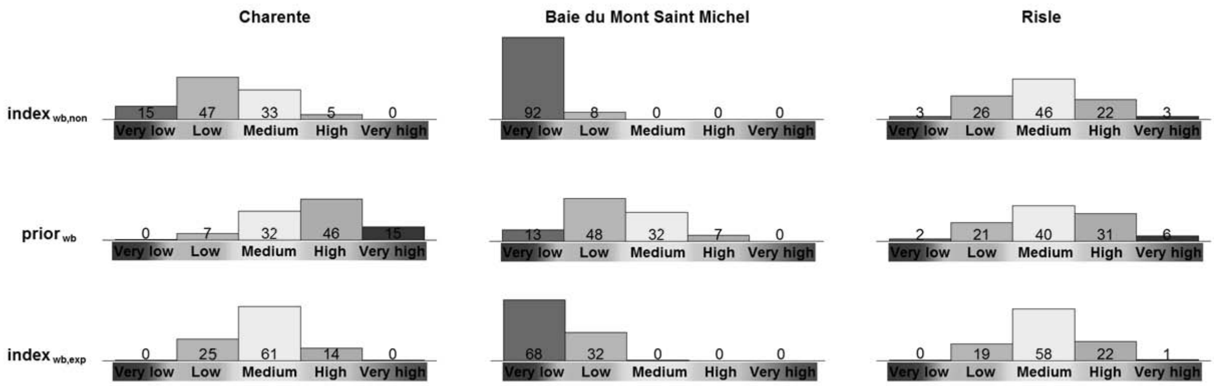


Fig. 5: Distributions of each index ($index_{wb,non}$, $prior_{wb}$ and $index_{wb,exp}$) for 3 water bodies (Charente, Baie du Mont Saint Michel, Risle), the number in each stick represents the rounded probability in percentage to be in the associated pressure class

Metric (density of fish)	Model type		Model
Benthic (DB)	Log normal		Log(DB+1) ~ season + salinity + size + pressure + estuary random effect pressure rp: -1.62, estuary sd: 0.70, model sd: 1.35
Diadromous (DDIA)	Delta	Bernoulli	Logit($DDIA_{0/1}$) ~ season + salinity + size + ecoregion + pressure + estuary random effect pressure rp: -1.36, estuary sd: 1.19
		Log normal	Log($DDIA_{+}$) ~ season + salinity + size + ecoregion + pressure + estuary random effect pressure rp: -2.57, estuary sd: 0.45, model sd: 1.22
Marine juveniles (DMJ)	Delta	Bernoulli	Logit($DMJ_{0/1}$) ~ season + size + ecoregion + pressure + estuary random effect pressure rp: -4.78, estuary sd: 1.16
		Log normal	Log(DMJ_{+}) ~ season + salinity + size + pressure + estuary random effect pressure rp: -2.75, estuary sd: 0.88, model sd: 1.26

Table 1: Models structure of the 3 metrics used; an arbitrary value of 1 is added to each value of DB metric; delta type models composed of one sub model for presence/absence ($_{0/1}$) and another one for positive values (+); pressure regression parameter is denoted *pressure rp*; standard deviation of the estuary random effect is denoted *estuary sd*; model residuals standard deviation is denoted *model sd*

Water body	High	Good	Moderate	Poor	Bad
DISTRICT: SEINE NORMANDIE					
Upstream Seine*	0	0	0	7	**93**
Central Seine*	0	0	0	4	**96**
Downstream Seine*	0	0	0	40	**60**
Risle	0	19	**58**	22	1
Baie des Veys	13	**69**	17	1	0
Orne	1	37	**53**	9	0
Baie du Mont Saint Michel	**68**	32	0	0	0
DISTRICT: LOIRE BRETAGNE					
Morlaix	0	12	**62**	25	1
Laita	31	**55**	14	0	0
Blavet	0	**55**	44	1	0
Vilaine	**71**	29	0	0	0
Loire*	0	4	**57**	38	1
Sevre Niortaise	0	31	**66**	3	0
DISTRICT: ADOUR GARONNE					
Charente	0	25	**61**	14	0
Seudre	0	5	**50**	43	2
Upstream fluvial Garonne*	0	0	11	**63**	26
Fluvial Dordogne*	0	5	20	**44**	31
Central Gironde*	0	0	0	23	**77**
Downstream Gironde*	0	0	0	35	**65**
Upstream Adour	0	2	34	**57**	7
Downstream Adour	0	0	4	**72**	24
Bidassoa	0	0	0	15	**85**

Table 2: For each water body ranked by district, posterior probabilities (in %) from the index with experts prior, $index_{wb,exp}$, to be in each quality class; the mean classes are indicated for $index_{wb,exp}$ (bold character), $index_{wb,non}$ (box) and $prior_{wb}$ (shaded); the water bodies of large estuaries are indicated with the symbol: *