A Pragmatic and Integrated Approach for the Evaluation of Environmental Impact of Oil and Chemical Spills at sea: Input to European Guidelines (PRAGMA)

Contract nº 07.030900/2005/429172/SUB/A5

FINAL TECHNICAL REPORT

Main authors: Th. Baussant[1], M. Cajaraville[2], M. Theron[3] & S. Le Floch[4]

[1] IRIS (beneficiary); [2] UPV (partner 3); [3] UBO (partner 2); [4] Cedre (partner 1)

http:\\www.iris.no\pragma

Table of contents

1 General reminder of project objectives

Despite of the extensive transport of fuel oil and other chemicals at sea, there is a lack of knowledge of the effects of many of them on the marine biota. Several accidental spills in European coastal areas have resulted in the release of different toxic compounds into the marine environment. In recent oil spills such as those of the tankers “Erika” (1999) and “Prestige” (2002) heavy fuel oil type nº 2 was released, heavily contaminating coastal areas of Brittany and Galicia and the Bay of Biscay. In 2000 the loss of the tanker “Levoli Sun” in the English Channel resulted in the release of several tonnes of the highly toxic and corrosive hydrocarbon styrene. There is a clear need to develop tools that might allow assessing the impact of these accidental spills on aquatic organisms. In response to these problems, European policy-makers and intergovernmental agencies have developed regulatory guidelines to control and supervise the activities of maritime users and to support member states in the field of response to accidental and deliberate marine pollution. A necessary step for regulators and advisors in decision-making is the evaluation of the **risk** posed by spilled substances at sea. Two types of entities can conduct risk assessment:

- Scientists who communicate to governments and public opinion about the state of the environment and who pull the alarm when problems need to be signalled.
- Environmental managers who make sure that the company activities comply with the legislation framework and are respectful of ecosystems.

Both have responsibilities to decide upon the tolerability of risks and the standards against which measurements should be compared. This inevitably leads to the type of tools and techniques used to perform the measurements.

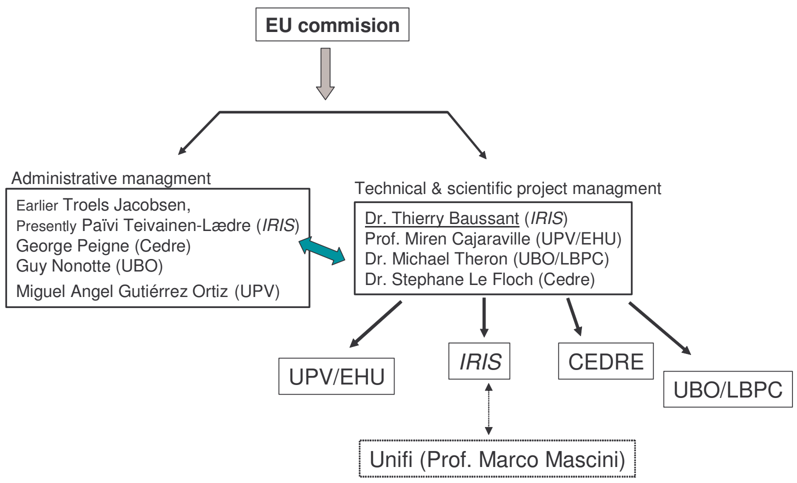


Figure 1 – General project managment in Pragma

The ultimate goal is to use them for field situation either to establish background knowledge before a spill eventually occurs or post-spill as a way to follow up recovery and assess the quality of the water. Monitoring surveys are conducted in the field to assess the condition of the environment. Before this is even possible, monitoring technologies and tools undergoes tests and validation in laboratories. This constitutes the main technical framework of PRAGMA funded under the 2005 marine pollution call by the DG-Environment of the European Union (*Community framework for cooperation in the field of accidental or deliberate marine pollution*). The project aimed at:

- *implementing well-established methodologies based on biological marker measurements and other cost-effective analytical tools* in a tier approach for the assessment of environmental impact of oil and chemical spill at sea
- *integrating them in the current regulations and in future pollution monitoring programme* of EU member states for decision-making criteria

The expected deliverables are:

1. An assessment of the suitability of several analytical methodologies and tools for their implementation in pollution monitoring programmes
2. An integrated approach whereby the selected methodologies are combined to provide a manageable environmental index summarizing the quality of the water and used by decision-makers
3. A recommendation towards a harmonisation of environmental criteria for pollution impact assessment of hazardous substances in marine water at the European level.

IRIS (Norway) is the leader of the project and three other partners are involved, *Le Cedre* (France), University of Brest (UBO, France) and University of the Basque Country (UPV/EHU, Spain). In addition, work related to biosensors has been contracted with the University of Florence (Italy) (see fig.1).

2. General summary of project implementation

2.1 General overview

IRIS has received the leadership and administrative responsibility towards the DG-Environment. Practically, the core experimental studies of this pilot project were carried out in two different locations: the marine facilities of Cedre (Brest, France) were used to perform the work on fish. Cedre got strong support from UBO for the setting up and analytical work, and UPV for sampling and analytical work. Part of the analytical work on fish was also sub-contracted to Unifi. For the experimental work with mussels, the Akvamiljø marine centre facilities were used by IRIS to set up the experimental system. UPV participated in some sampling events (start and end) and in the analyses all the samples from of that work.

The responsibilities and share were assigned as following:

IRIS – Project management, responsible for mussel experiment/setup, responsible for mussel analytical work (chemistry, blood/physiological parameters: immunology, genotoxicity, lipid, general condition, bioassay, valve gape)

UPV – responsible for fish and mussel analytical work (genotoxicity, histopathology, histochemistry, endocrine markers, novel “omics” markers)

CEDRE – responsible for fish experiment/setup, responsible for fish analytical work (chemistry)

UBO – responsible for fish analytical work (blood/physiological parameters: immunology, gas, ion balance, enzymatic responses, metabolites).

Work sub-contracted to UNIFI (Italy) – Analytical work on biosensors

2.2 Initial and actual time schedule

The original agreement between IRIS and EC was established for an original eligible period of 14 months but was extended to 19 months due to several technical problems or difficulties. At Cedre, the first fish pilot experiment had to be end up earlier than planned due to unexpected high mortality observed from the 11th to the 25th September 2006. A new experimental setup started at the beginning of October 2006. Concerning the mussel pilot study, the original timing of implementation was not well suited due to the reproduction period of mussels. Reproduction has a seasonal pattern accompanied by large physiological changes. In order to better control the type of methodological assessment and tools proposed for that study, we requested to postpone the implementation of the mussel experiment to the autumn 2006 which delayed also the original implementation period of the project.

2.3 Planned and used resources

The total eligible costs of this project for which the Commission grant was awarded were estimated originally at EURO 280.492,00. Following several technical challenges in the project, the final total eligible cost are estimated at EURO 303.605,00. The contribution of the partners increased by EURO 23.113 (see the final "Standard statement of expenditure and income" for details)

2.4 Expected and actual results

No actual expectation of the results existed at the onset of this project. However, the partners are already well aware of the conceptual ideas and approach developed in this project and all of them are currently involved in developing, testing and applying biomonitoring tools for environmental matters related to acute or chronic pollution. Hence the results achieved in this project support other observations. The novelty was however in the type of exposure simulation realised and the integration of measurements in a simple environmental index assessment for decision-making criteria. Although we feel more validation work need to be achieved before the proposed approach can actually be implemented in a regulatory framework at the European level, the observations reported here constitute a good platform towards the integration in European guidelines.

3. Project management/implementation process

The management of the project was demanding due to the complex coordination of activities and role in separate facilities located in different countries. However, the roles of partners were well defined so that each of them was assigned a well-defined task in the project implementation. Following up the project implementationwas instructive and good management process was achieved by regular communication between partners in order to coordinate the different actions as best possible.

Several technical challenges have raised to obtain comparable exposure systems at IRIS and Cedre. Unexpected mortality with fish due to acute effects of both fuel oil and styrene obliged Cedre/UBO to interrupt experiments, re-adjust/optimize the experimental system and start again. Usually, working with living animals in the laboratory require good quality of water, careful maintenance and a controlled planning of sampling events in relation to the natural seasonal cycle of the species.

Clearly, this project has raised key issues which are currently debated both by scientists involved in environmental issues and legislators needing comprehensible assessment that can be used for decision-making. The partnership of PRAGMA is already largely implicated in their respective national environmental programmes but also very much internationally involved. Key issues related to biological effect monitoring, pollution monitoring, impact assessment, technical assistance to combat pollution and communication to industrial or governmental authorities, are part of their regular work. Many of these issues are regularly debated, updated and revised within international working groups like ICES, OSPAR, IMO/GESAMP. By raising these questions through the project, a European dimension has been given and we do hope that the outcomes of PRAGMA may also help in the future revision/recommendation of methods and assessment approach concerning environmental monitoring of spill in the European waters.

4. Activities

All the activities initially planned in the original project proposal have actually been implemented during the duration period of PRAGMA.
A web page has been created for PRAGMA (http://www.iris.no/pragma) where general information, events, progress reports are been made opened for the public. The link has also an internal restricted page for the exchange of data by the partners.
Dissemination of the project framework and results included participation to international seminar (see also copy of poster presented at SETAC conference in Porto; May 2007) and workshop (http://www.ecotox.uquebec.ca/archives/ecobim_2006/rapport_final.htm and more recently http://www.cedre.fr/uk/publication/workshop/pollution-assessment.html).

5. Detailed presentation of the technical results and deliverables

5.1 Suitability of a set of analytical methodologies and tools in pollution monitoring programmes – Pilot studies in PRAGMA

5.1.1. General consideration

Monitoring programmes are used to assess the condition of the environment, that is:

- *to provide an early warning signal of changes,*
- *or to diagnose the cause of an environmental problem.*

The aim of monitoring programmes is the recognition of areas at risk and needing intervention. In that sense, these programmes are clearly tools for ecological risk assessment.

Ecological risk assessment is an analytical process applied to environmental issues and used as a policy tool to assist regulators in decision-making. Following risk assessment, a decision may be made to proceed with a management action (Cf. Fig. 2). Monitoring can help determine if the desired result of a management action is achieved.

Ideally ecological risk assessment should be mainly undertaken by industries and scientific organisations to make sure that human activities comply with the legislation framework and are respectful of ecosystems. Policy-makers (governments / international institutions) advised by scientific groups, are supposedly involved in deciding upon the tolerability of risks and the standards against which measurements should be compared. In fact, the reality is usually far more complicated than what is summarised in the following diagram (Cf. Fig. 3).

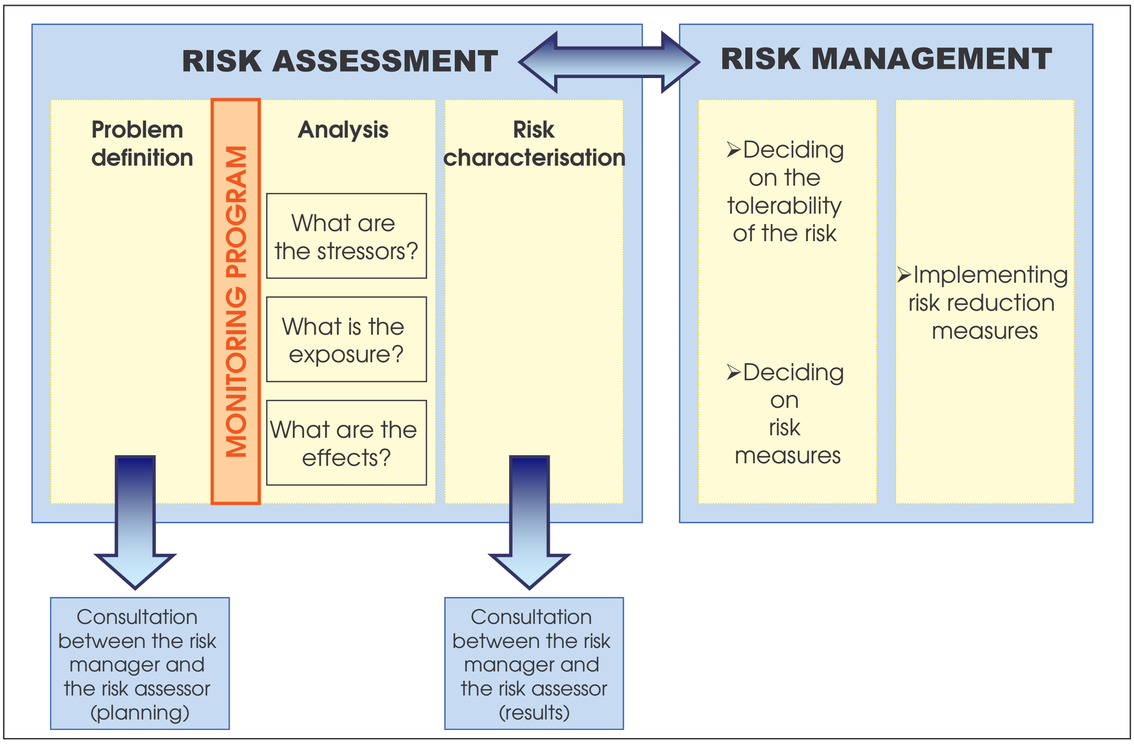


Figure 2: Steps in risk management process (redrawn from the Health Council of the Netherlands, 1995).

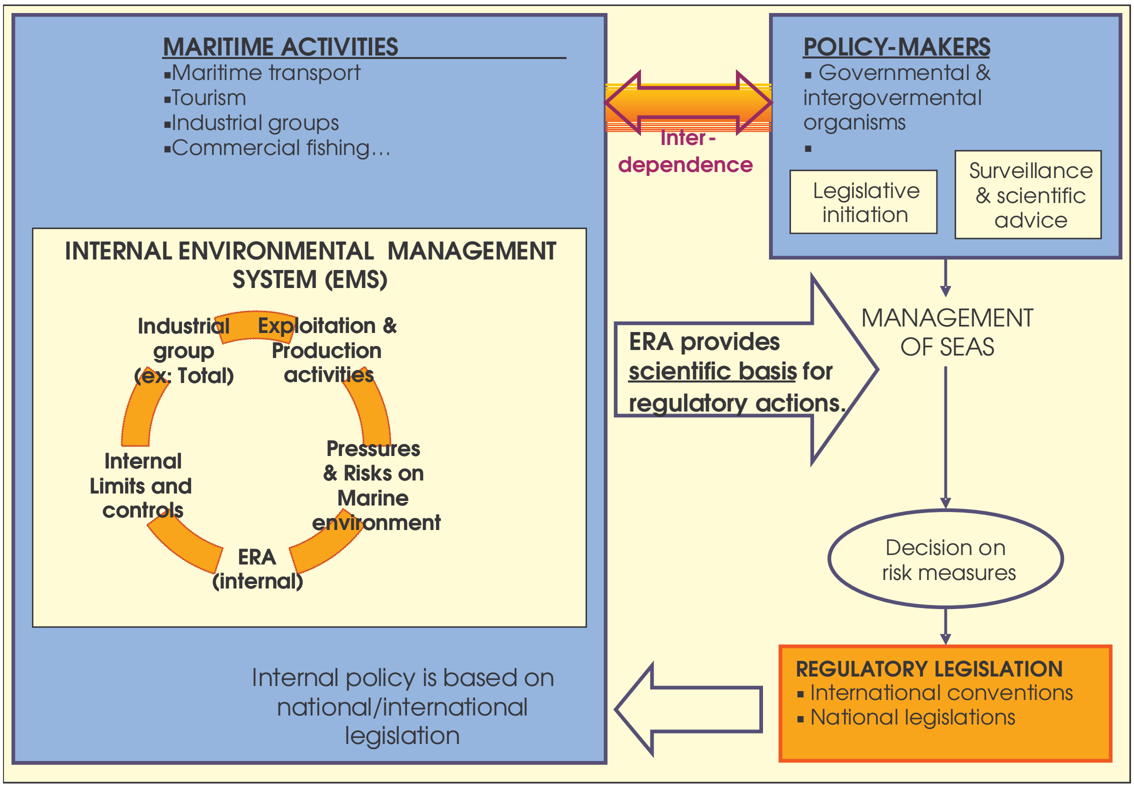
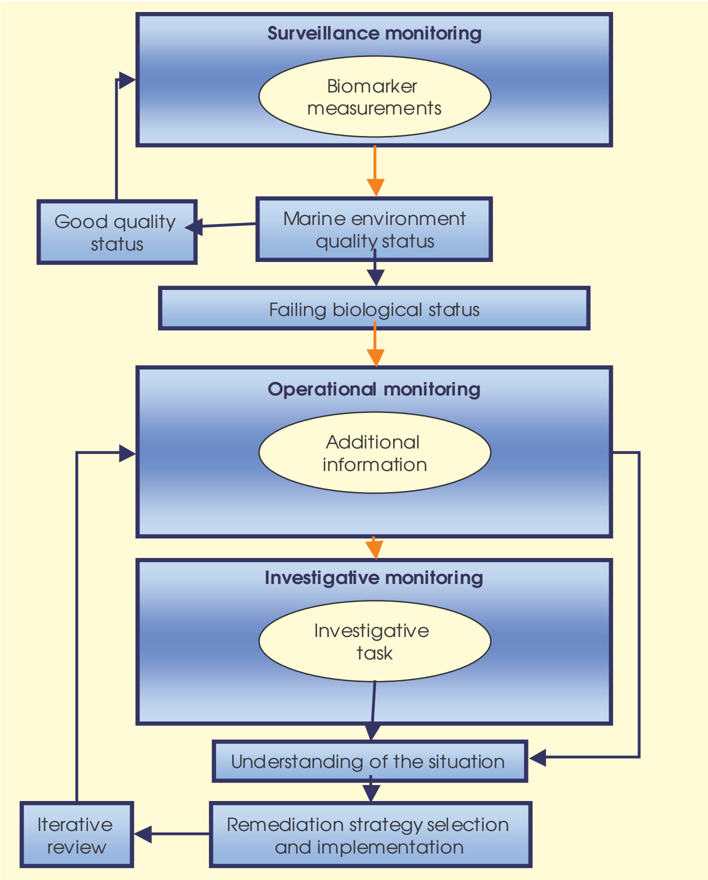


Figure 3: Importance of ecological risk assessment (ERA) as a policy tool to assist regulators in decision-making.

5.1.2. Different monitoring strategies

Three modes of monitoring regime are specified in the European Union's Water Framework Directive (WFD) and are supposed to be introduced in management plans since December 2006 (Allan et al. 2006). These include (Cf. Fig.4):

- *surveillance monitoring* aimed at assessing long-term water quality changes and providing baseline data on river basins allowing the design and implementation of other types of monitoring,
- *operational monitoring* aimed at providing additional and essential data on water bodies at risk or failing environmental objectives of the WFD,
- *investigative monitoring* aimed at assessing causes of such failure.



If "good quality" status is achieved only surveillance monitoring is required to ensure this is maintained. "Good quality" status means that the levels of chemicals in the water body comply with environmental quality standards (EQSs). These EQSs have yet to be stated. However, for water bodies, which are determined to be at risk, or of moderate or poor quality, further information will be needed so that adequate remediation strategies can be implemented and subsequently monitored.

Figure 4: Simplified scheme for the three types of monitoring embedded in the Water Framework Directive (redrawn from Allan et al., 2006)

In the concept of JAMP (Joint Assessment and Monitoring Programme, OSPAR 2004), this process is similarly applied and recommended for the maritime area under regulation of the Oslo Paris Convention (Lehtonen, 2005). The monitoring of biological effects of contaminants is typified in two categories:

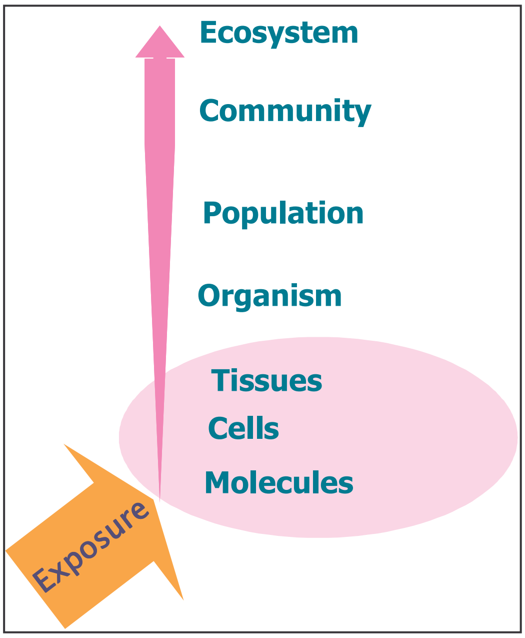
- *General biological effects monitoring*: "general quality status" of the marine environment (indicators of long-term change; broad screening for biological effects of contaminants by using non-specific "early warning" methods). Identification of known or suspected areas of contamination (bioassays; biomarkers; population/community responses).
- *Contaminant-specific biological effects monitoring*: application of methods specific to a certain contaminant/classes of contaminants.

Monitoring spreads out over a spatio-temporal dimension. A distinction can be made between:

- *Routine regional monitoring programs*, which are surveys established on large spatial scale: set of different stations, sampled several times at different sites.
- And *site-specific assessment* or local surveys: samples are taken along a pollution gradient during a given period.

5.1.3. Biomarkers: tools for biological effects monitoring

Definition



Several authors have tried to define the concept of biomarker. One definition is "*biochemical, cellular, physiological, or behavioural variations that can be measured in tissue or body fluid samples, or at the level of whole organisms, to provide evidence of exposure and / or effects from one or more contaminants*" (Depledge and Galloway, 2005). Figure 5 illustrates that changes in biomarkers at sub-individual or individual level can give rise later to changes in populations, communities and ecosystems.

Figure 5 – Monitoring with biomarkers. Biomarkers are measurements made to assess changes at molecular, cellular or tissue level (coloured bubble). They can used to estimate the impact for individuals and further the ecological risk for a population

Challenges in the development and use of biomarkers as ecological indicators

The use of biomarkers has prevailed for more than 35 years in pollution research. While the focus has traditionally been and continues to be on the comparison between contaminant levels measured in the environment to thresholds levels established by international / national agencies as standard practices, biomarkers are considered as complementary tools in addition to measures of chemical contaminants. Some scientists believe that the use of biomarkers in environmental monitoring confers significant advantages over traditional chemical measurements:

- Biomarkers may indicate the presence of *biologically available* contaminant.
- Using a suite of biomarkers may reveal the presence of contaminants that were not suspected initially.
- Biomarker responses often persist long after a transient exposure to a contaminant that can no longer be detectable. Thus these responses may detect intermittent pollution events that chemical monitoring may miss.
- Biomarker analyses are in many cases, easier to perform and considerably less expensive than a wide range of chemical analyses.

Biomarkers currently show that organisms have been exposed to contaminants and/or that exposure is associated with deterioration in health. However, the extent to which biomarkers are able to provide unambiguous and ecologically relevant indication of exposure to or effects of toxicants remains highly controversial.

Biomarkers development results partly from a need for early warning measures that respond before measurable effects on individual performance and population/community dynamics occur, and partly as an aid to identifying the causes of the effects observed at population and community levels (higher ecological levels). The aims of the process are to employ biomarkers to indicate that organisms have been or are being exposed to certain pollutants or that they are suffering or likely will suffer future impairments of ecological relevance.

For some, biomarkers assume that the relationship between pollutant concentration in the environment and the response of the target cells, tissues or organisms are known, or at least correlated. Regrettably, in the fluctuating environmental conditions which occur *in situ*, it is hard to attribute effect to a cause and it is even less certain how exposure-response relationships potentially demonstrated in the laboratory are modified. Unless these exposure/adverse effects-time relationships are understood, the biomarker approach may have limited utility for predicting the ecologically relevant effects (Depledge and Galloway, 2005).

Also, the successful application of the biomarker approach is still hampered by the lack of a simple and reliable integration system able to overcome the obvious difficulty of predicting harmful biological effects and their subsequent consequences for environmental health.

5.1.4. Considerations to be addressed before incorporating biomarkers in monitoring programs.

Different types of biomarkers

The field of biomarkers is large; these responses can be measured at several sub-individual or individual levels. Biomarkers are usually divided into the following:

– Biomarkers of exposure

This type of biomarkers indicates that an organism has experienced exposure to toxicant. *Biomarkers of exposure* correspond to responses such as induction or inhibition of specific enzymes involved in biotransformation and detoxification as a consequence of chemical exposure. In most cases, these responses are early biomarkers for specific toxicants at a low level of biological organisation (molecular or cellular).

– Biomarkers of toxic effects

Biomarkers of toxic effects reflect pathological endpoints and are determined at each level of biological organisation. In contrast to the *biomarkers of exposur*e these effects mostly cannot be attributed to the impact of a single contaminant and therefore serve as integrative markers of complex toxicities.

The appropriate battery of biomarkers

No single biomarker can indisputably measure environmental degradation. Only suites of biomarkers are likely to provide a comprehensive indicator of ecosystem health as it will minimize the influence of natural variation and allow discrimination of clean healthy and polluted/unhealthy sites (Forbes et al., 2006; Hagger et al., 2006; Handy et al., 2003; Moore et al., 2004).

Moreover the battery of biomarkers must be seriously considered in relation to the aim of the monitoring program (strategy definition). Various kinds of parameters may be used in biomonitoring campaign but selectors must be fully aware of what parameters represent. Biomarkers are selected according to their sensitivity/specificity, the biological level or function they represent, the time needed for the response, their regulation properties…, depending on which attributes reflect the most the pollution-induced effects.

Usually, *biomarkers of exposure* are selected for their early response and their specificity of reaction. While complex contamination are not reflected, biomarkers of exposure are mostly used for the monitoring of hot spots of pollution or clearly defined point source inputs as well as for the characterisation of chronic unknown chemical inputs. *Biomarkers of toxic effects*, on the other hand, are selected for their high ecological relevance. They give a general picture of the status of environmental deterioration, but in most cases the quality of contamination remains speculative.

Integrated biological effects monitoring should comprise the implementation of the widest possible range of different biomarkers which reflect the most pollution-induced effects at several levels of the biological organisation. Therefore only a combination of both exposure and effect biomarkers provides sufficient information for the assessment of responses reflecting the quality as well as the level of environmental deterioration.

5.1.5. Relevance of the biomarkers chosen in PRAGMA

The PRAGMA project could be related to a particular type of chronic pollution characterised by peaks and hollows in pollutant concentration, so that the chronic exposure is a long series of shorter pollution events. This refers to one of the main problems particularly dreaded when spills of heavy oil occur: substances trapped within sediments in low-energy coastal environment can re-contaminate the environment when remobilised by storms or tides. In the project the focus has been made on the analysis and the detection of long-term adverse effects which can be encountered in after either type of events.

A suite of both exposure and effect biomarkers were selected. A set of biomarkers representing two main physiological control systems (immune and endocrine) were chosen for their ability to detect history of injury to biological function associated with a low quality environment. Histological biomarkers offer the advantage of detecting long-term injury. Indeed to reflect long-term adverse effects, biomarkers need to relate to a body-system or cell-type that accumulates injury over long-time scales, or at least have some form of biochemical memory to store the response that is associated with some critical biological function. In both projects, the stress was laid on selecting biomarkers of high ecological relevance to reflect the general state of organisms. This deliberated choice allows scientists to have a better idea of the general state of the entire ecosystem.

5.1.6. Other novel analytical tools used in PRAGMA

Biosensors and other sensing analytical tools

Biosensors are simple and cost-effective devices which have a potential for use in biomonitoring studies (Bagni et al, 2005; Kröger and Law, 2005; Rogers, 2006). A DNA biosensor was tested as a mean to detect substances binding to DNA at the electrode surface. The DNA biosensor is based on the immobilisation of dsDNA on the surface of a graphite screen-printed electrode and on the use of voltammetry to investigate the changes in the DNA redox properties (i.e. the oxidation of the guanine base) after interaction with an analyte. In this system, the changes in guanine oxidation peak after interaction with the analysed sample can then be interpreted as the presence of toxicants with a possible carcinogenic or mutagenic effect. The strength of the changes are expressed by an index S% which is simply the ratio between the guanine oxidation peak in the analysed sample to that measured in the reference sample. **A sample with low affinity for DNA will then give a high S% while a sample with high affinity will give a low S%** (figure 6)

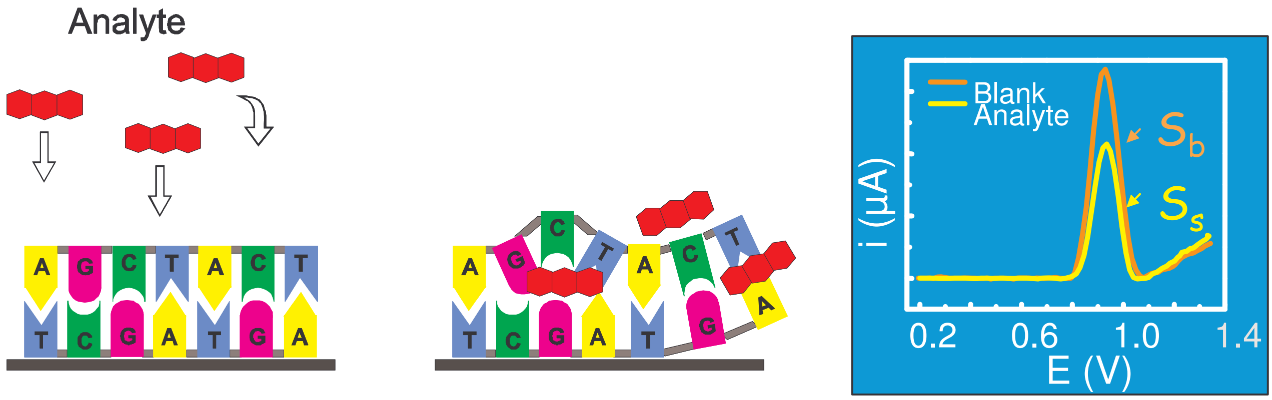


Figure 6 – A) Schematic representation of the interaction between analyte and the dsDNA B) Example of a square wave voltametric scan obtained after signal baseline correction. The interaction of the DNA with pollutant molecules modifies the redox status of the DNA guanine base that is displayed at +1.0 V (vs. Ag-SPE). This is measured by a decrease of the guanine peak oxidation and expressed as S % = (**Ss / Sb**) * 100. **Sb:** guanine peak after the interaction with the blank solution; **Ss:** guanine peak after the interaction with the sample. Note that the adenine peak is also detected at +1.25 V (vs. Ag-SPE) but is not shown on this figure.

Usually, this device is applied in water bodies to detect and quantify contaminant but in PRAGMA, it was applied as a simple screening tool to detect the presence of contaminant derived mutagenic metabolites in the bile of fish.

Figure 7 – Valve gape system measuring closing/opening of the mussel

Another technique used in the project was biomonitoring using an on-line biological early-warning system (BEWS) with mussels. Compared to the biosensor, the BEWS with mussels can provide a direct continuous measurement of the biological stress level under exposure. The system used in PRAGMA was based on valve closure or valve movement frequency as a behavioural defence mechanism against contaminated water. The prototype was developed in another on-going project at IRIS and was tested in PRAGMA to as part of the validation process required for this type of novel instrumentation (figure 7).

Omics biomarkers

Apart from the so-called "conventional" biomarkers, the use of "omic" approaches, as tools for the development of molecular level biomarkers, looks promising for the assessment of environmental pollution. Omics techniques have in the recent years been highly focusing on regulation processes of hundreds to thousands of proteins or genes at once. The new field of ecotoxicogenomics studies profiles of gene transcription in cells, tissues or organisms after exposure to pollutants in an attempt, to fingerprint the exposure to a given chemical compound or class of compounds in terms of their biological activity. These techniques can hence be directly indicative of the molecular mechanisms behind the toxic effects. In PRAGMA, we used these techniques in fish samples but only with a research purpose. Owing to the novelty of this technique and the development still needed in this field of research at this stage, this type of analysis was not integrated in the final assessment of this project.

Table 1 below summarizes the type of methodologies selected in this project for the different species.

5.1.7. Relevance of the species chosen in PRAGMA

We used two species representative of two different taxonomic groups, invertebrate and vertebrate. The blue mussel *Mytilus edulis* is often used as a sentinel invertebrate species due to its worldwide distribution in marine ecosystem. This sedentary species is well suited to the monitoring of biological effects. The turbot fish *Scophthalmus maximus* is often used in ecotoxicology studies and is one of the so-called OSPAR species. Standardized acute tests are often made using this fish species. Both species are closely related to the bottom of the sea where chemicals may accumulate after sedimentation from the water column. In addition, both species are commercially important as target species in aquaculture.

BIOTEST	SPECIES CHOICE MUSSEL	TURBOT
Exposure biomarker		
	peroxisome proliferation	peroxisome proliferation EROD, CYP P450 activity
	body burden	bile metabolites
Effect biomarker		
genotoxicity	comet assay, micronucleus	micronucleus
endocrine	alkali labile phosphate assay	vitellogenin/Zona radiata protein assay
physiological	lysosomal membrane stability, haemolymph (NRRT assay)	haematology
	clearance rate	blood gases (Po_2, Pco_2)
	survival in air	acid base status
	condition index	hydromineral balance
histopathological	lysosomal membrane stability, digestive gland (LP assay)	lysosomal membrane stability, liver (LP assay)
	gonad somatic index	gonad somatic index
immunological	phagocytosis, cell count	haematology, cell count
Bioassay		
embryo-larval bioassay	fecundity success developmental abnormalities, growth	
Novel biotools		
'-omics' biomarkers	SSH studies (styrene)	DNA array (crude oil)
biosensors	DNA affinity biosensor	DNA affinity biosensors
	amperometric electrode	
BEWS:biological early warning system	valve gape	

Table 1 - biotests methodologies studied under **PRAGMA**

5.1.8. Type of exposure simulated in PRAGMA

In Pragma, we put emphasis on two type of chemical – heavy oil (“fuel number 2”) and styrene – simply because these have caused large spills and environmental concern following recent accidents in European waters.

The first technical challenge of Pragma was to select and optimise exposure systems adequately to the type of situation envisaged in case of spill by these chemicals. Factors related to weathering, time and actual persistence of the compounds were taken into account. Hence, the spill discharges were made using exposure systems able to simulate the fate of these chemicals in a manner that represented as realistically as possible their natural behaviour and hazard for the marine biota. As reminded earlier, the focus of this study was not on acute effects but rather the longer term effects of these substances.

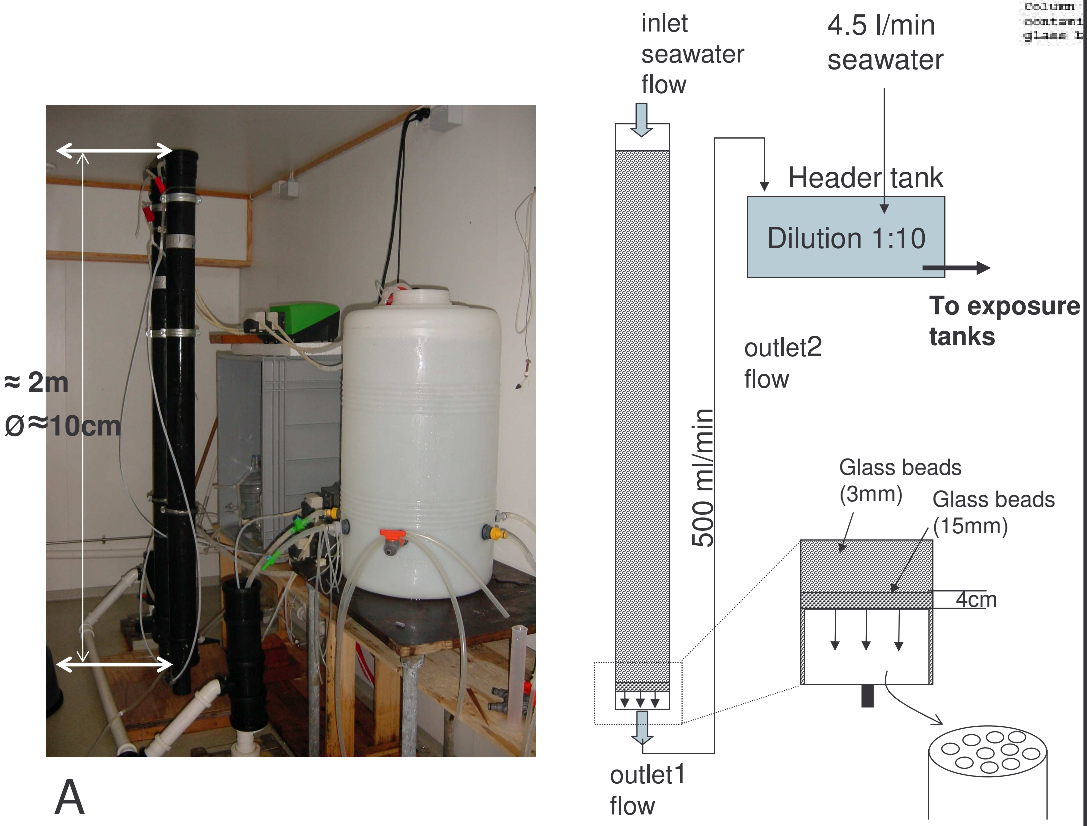
For the fuel oil exposure, we used a system based on that described by Short & Heintz, 1997; 2003. Briefly, a column filled with glass beads (3mm) coated with oil (≈40g oil per kg beads) was used throughout the pilot study. A flow-through of seawater was maintained at a rate of 500 ml/min in the exposure column. By this means, the most soluble components are first washed out of the column into the exposure tanks. With time, their relative concentration and contribution in the water will decrease. Hence, this dosing technique simulates the weathering of oil which naturally follows a spill event. In the mussel experiment, the column was reloaded two times with fuel oil to simulate the effect of oil resuspension in the water column following for example a storm. However, the first reload did not work well due a technical problem and did not give as high as expected increase of organic compounds in the water. The second reload worked successfully.

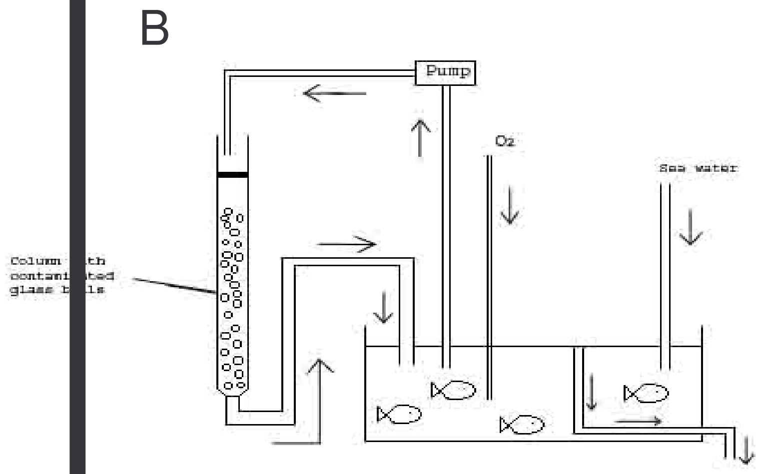
For styrene, the chemical was pumped to a large tank together with seawater in order to create a saturated solution. The seawater loaded with styrene was then used to supply the exposure tank with the animals. A flow-through of seawater was maintained at an adequate rate in the exposure tank. (fig. 8).

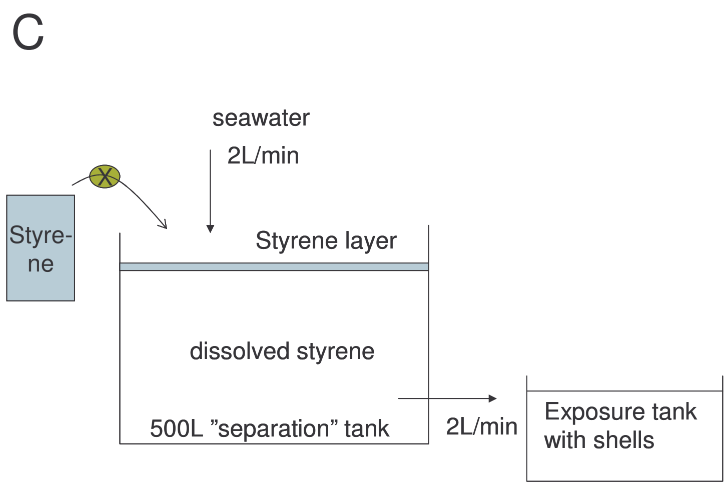
The same principle and exposure techniques were used both at IRIS and CEDRE although slight adaptations were made by each institute depending on their existing facilities and supply of seawater.

Figure 8 – Exposure systems in PRAGMA
A) Column system used to simulate fuel oil exposure and generate water-soluble fraction (WSF) of oil at IRIS
B) System used in the fish experiment at Cedre
C) System used for the styrene experiment at IRIS

.







5.1.9. Exposure conditions and sampling events used in PRAGMA

The fuel oil exposure was maintained longer in the pilot studies as many chemicals in oil degrade at a much slower rate, hence are more persistent than styrene. Also, heavy fuel oil in shallow water of coastal areas can be resuspended in the water following for example storm or high/low tide events. Mussels were exposed longer than fish in the pilot studies as these organisms may experience longer period of exposure in nature due to their sedentarity type of life.

Styrene was analysed either by GCMS or a fixed fluorescence method using λ_{ex} = 265 nm and λ_{em} = 305 nm. PAHs from WSF generated by the column system were measured by GCMS analysis (table 2)

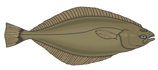
SUPPLIER FARM	Tredarzec (France) 	Ryfylke fjord (Norway) 
AGE/SIZE	≈ 350 g (juvenile < 1 year), ≈ 28 cm	7 ± 0.8 cm
FEEDING	Commercial dry pellets	Continuously with InstantAlgae “Shell fish diet” microalgae concentrate (≈10 to 30x10^3 cells/ml) (see http://www.reedmariculture.com/microalgae/shelldiet.asp)
STYRENE	*02-05/2007 - 22-05/2007*	*13-10/2006 - 15-11/2006*
Seawater	16.5±0.2°C 38±0.1ppt	12±1°C 35±1ppt
Flow	≈1.2L/min	≈2L/min
Sampling events	T*0* 3 days 7 days recovery (1week)	T*0* 3 days 7 days 19 days recovery (2weeks)
FUEL OIL	*02-10/2006 - 30-11/2006*	*29-09/2006 - 12-04/2007*
Seawater	17.8±0.2°C 39±0.2ppt	8.5±2°C 35±1ppt
Flow	≈5L/min	≈2L/min
Sampling events	T*0* 7 days 21 days recovery (1week)	T*0* 3 days 7 days 31 days → *1st oil refilling* 40 days → *2nd oil refilling* 76 days 159 days recovery (4weeks)

Table 2 – Exposure conditions used in pilot experiment with fish and mussel

5.1.10. Chemical uptake in organisms

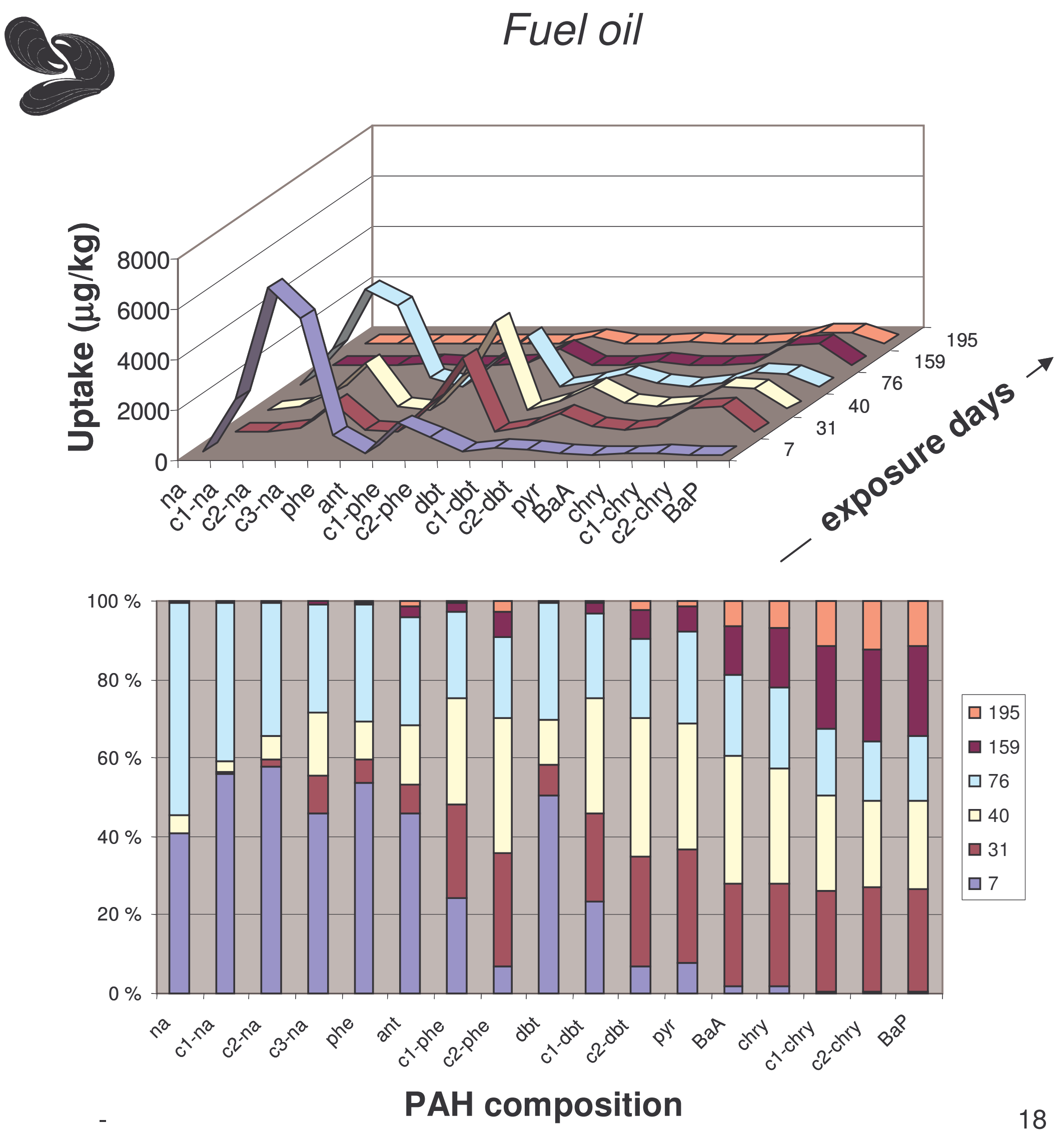


Figure 9 – PAH uptake during the exposure with fuel oil and mussel

Figure 9 shows the PAH uptake in the mussel exposure with fuel oil in the course of the experiment. The PAH pattern reflects the exposure in the seawater and the weathering processes generated by the exposure system as described above. Generally PAH levels were high in body tissues, increasing with the size of the PAH. At the onset of the exposure and after refilling the column system with oil (day 76), there is a clear increase in the uptake of the low molecular weight PAHs which dominate the burden in the animals. In between these events and as the oil gets older (washed out by the flow-through), the body burden is dominated by the larger molecular weight PAHs which represent a larger proportion in the seawater. At the end of the exposure (day 159) and after recovery (day 195), the PAH composition in mussel is still skewed towards an increasing proportion of the larger PAHs although these compounds were under detection limit in the seawater.

Styrene in the exposure system was maintained at < 1 and ≈ 3.5 mg/L respectively in the fish and in the mussel experiment. Compared to PAHs, the bioconcentration of styrene in mussel and in fish at the end of the exposure was small (BCF≈6.5 and ≈ 20, respectively). Even though styrene was detectable, this results confirms the general reported observation that this compound does not accumulate to a large extent in body tissue of aquatic organisms. Styrene was not detectable in body tissue at the end of the recovery period.

5.1.11. Summary of suitability of the selected parameters for biomonitoring in fish

Fuel oil experiment

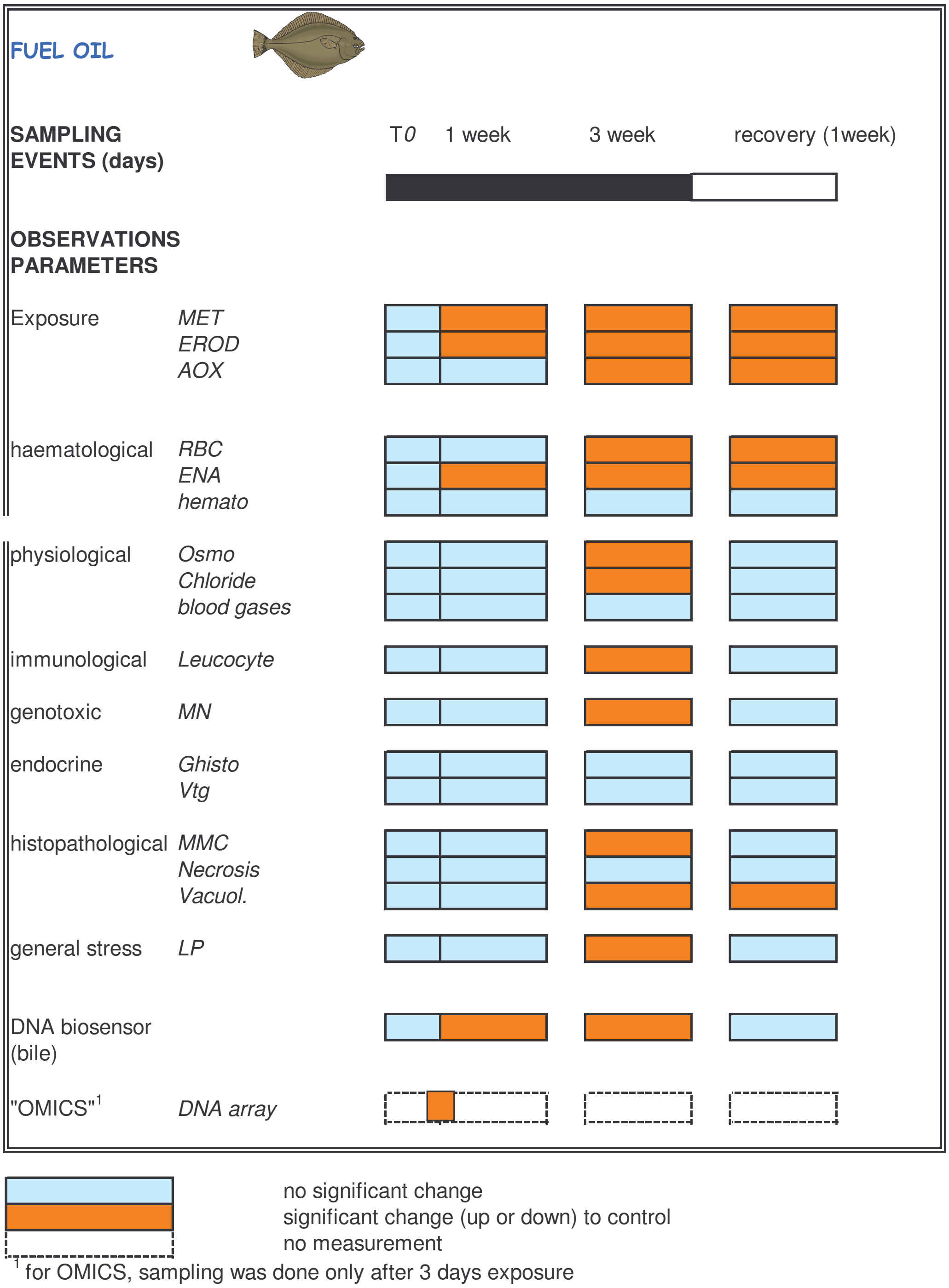


Figure 10 –Statistical significant changes in parameters measured in fuel oil exposed fish compared to control fish – Code as: *MET*-metabolites (fluorescence), *EROD*- EROD activity, *AOX*-Acyl-CoA oxidase activity (peroxisome), *RBC*-Red blood count, *ENA*-erythrocytic nuclear abnormalities, *hemato*-hematocrit, *Osmo*-Osmolarity; *Chloride*-Plasmatic chloride concentration; *MN*-micronuclei, *Ghisto*-Gonad histology, *Vtg*-Vitellogenin level (blood), *MMC*-Melanomacrophage centres (liver), *Vacuol.*-cell vacuolization (liver), *LP*-lysosomal membrane stability (liver), *DNA array*-gene up/down regulation. All statistical significance tested at p<5% using parametric or non-parametric tests.

Styrene

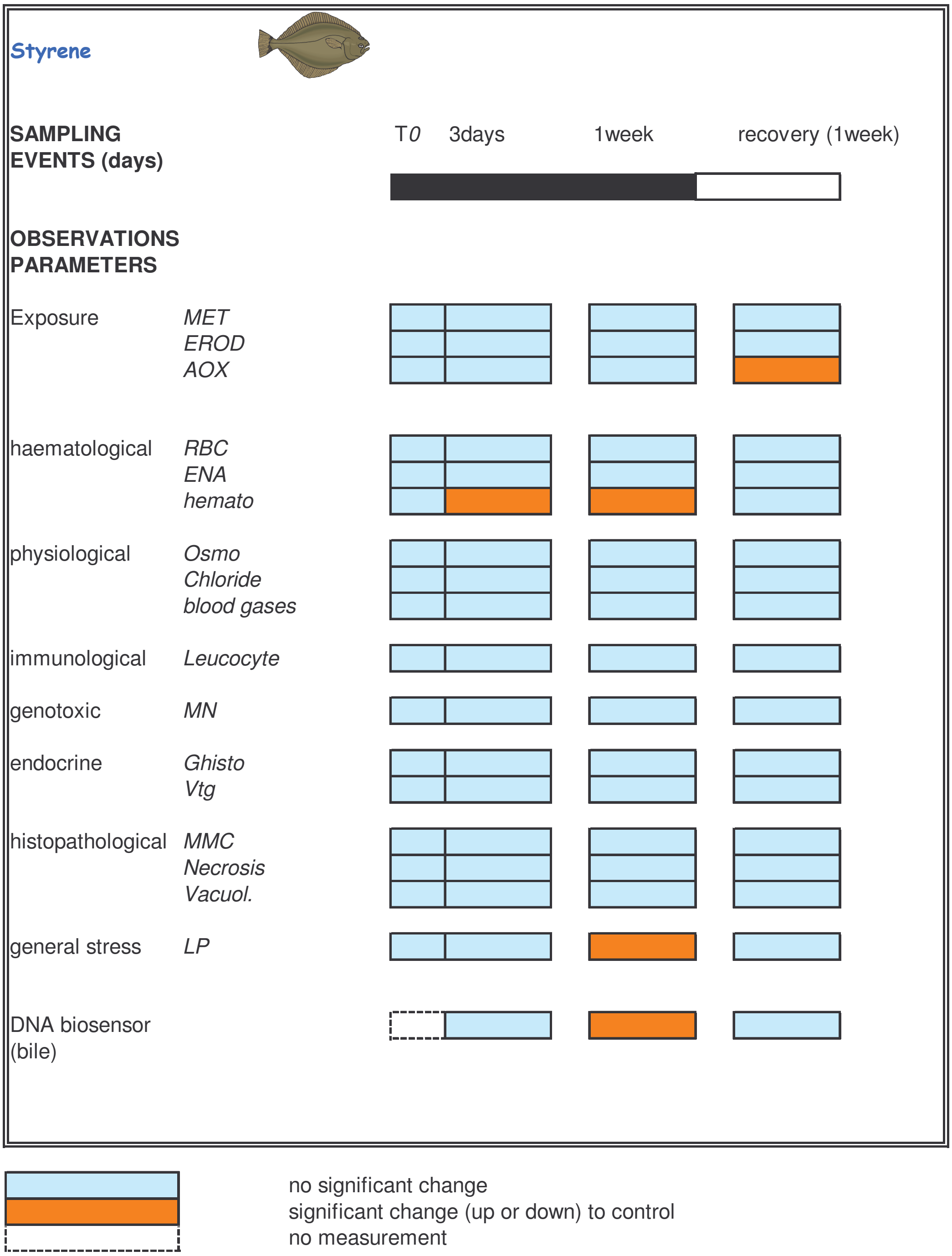


Figure 11 –Statistical significant changes in parameters measured in styrene exposed fish compared to control fish – Code as: *MET*-metabolites (fluorescence), *EROD*- EROD activity, *AOX*-Acyl-CoA oxidase activity (peroxisome), *RBC*-Red blood count, *ENA*-erythrocytic nuclear abnormalities, *hemato*-hematocrit, *Osmo*-Osmolarity; *Chloride*-Plasmatic chloride concentration; *MN*-micronuclei, *Ghisto*-Gonad histology, *Vtg*-Vitellogenin level (blood), *MMC*-Melanomacrophage centres (liver), *Vacuol.*-cell vacuolization (liver), *LP*-lysosomal membrane stability (liver). All statistical significance tested at p<5% using parametric or non-parametric tests.

5.1.12. Summary of suitability of the selected parameters for biomonitoring in mussel

Fuel oil experiment

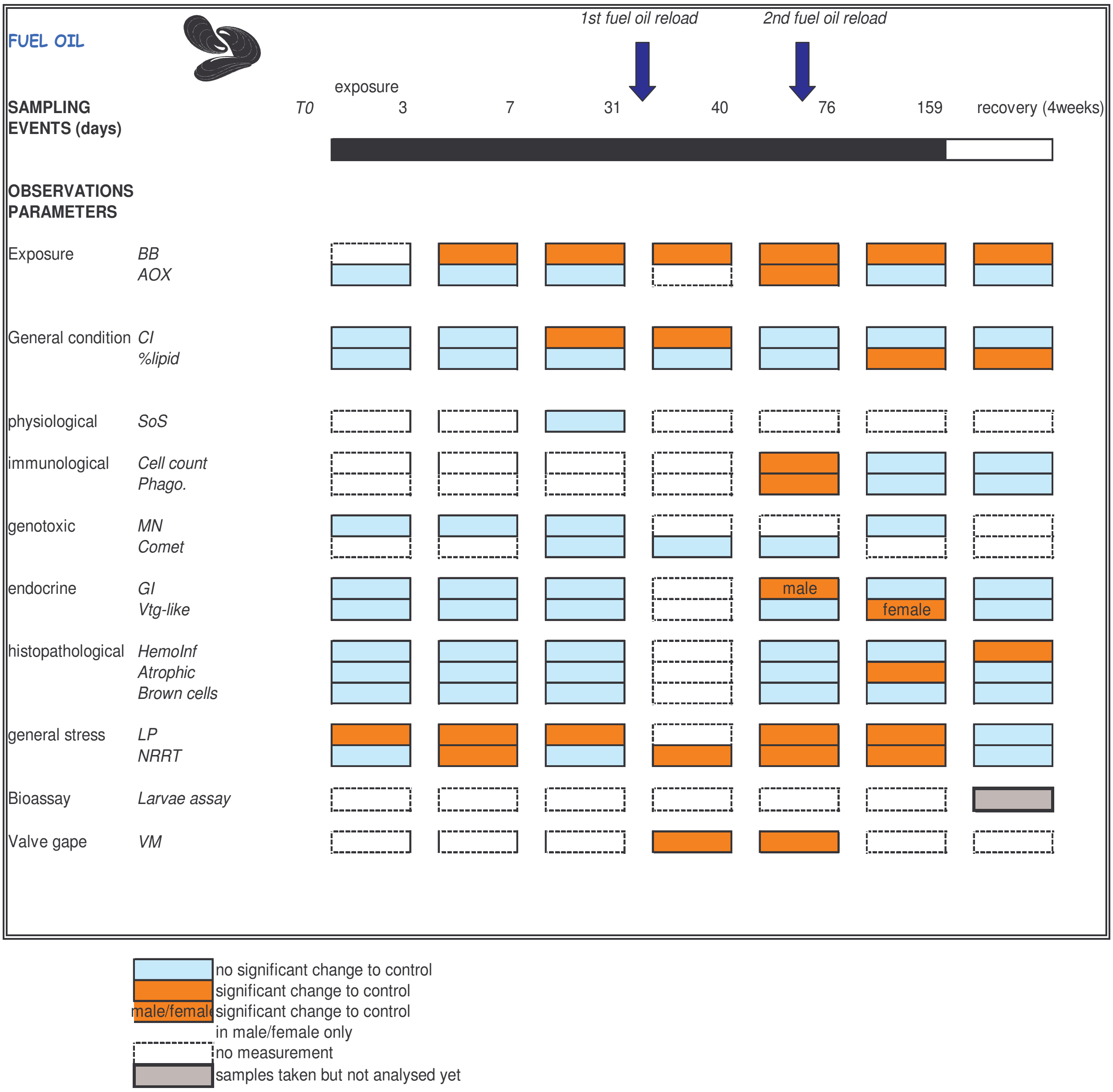


Figure 12 –Statistical significant changes in parameters measured in fuel oil exposed mussels compared to control mussel – Code as: *BB*-PAH body burden (GC/MS), *AOX*-Acyl-CoA oxidase activity (peroxisome), *CI*-Whole organism condition index , *%lipid*- percentage total lipid, *SoS*- Survival in air test, *Phago.*- phagocytosis activity (hemolymph) *MN*-micronuclei, *Comet-DNA strand break detection* (comet test, blood)), *GI*-Gonad index, *Vtg-like*-Vitellogenin-like protein level (gonad), *HemoInf.*-Hemocytic infiltration (digestige gland, DG), A*trophic*- Atrophic tubules (DG), *LP*-lysosomal membrane stability (DG), *NRRT*-Neutral red retention time in lysosome (hemolymph), *VM*-valve movement (whole organism). All statistical significance tested at p<5% using parametric or non-parametric tests.

Styrene

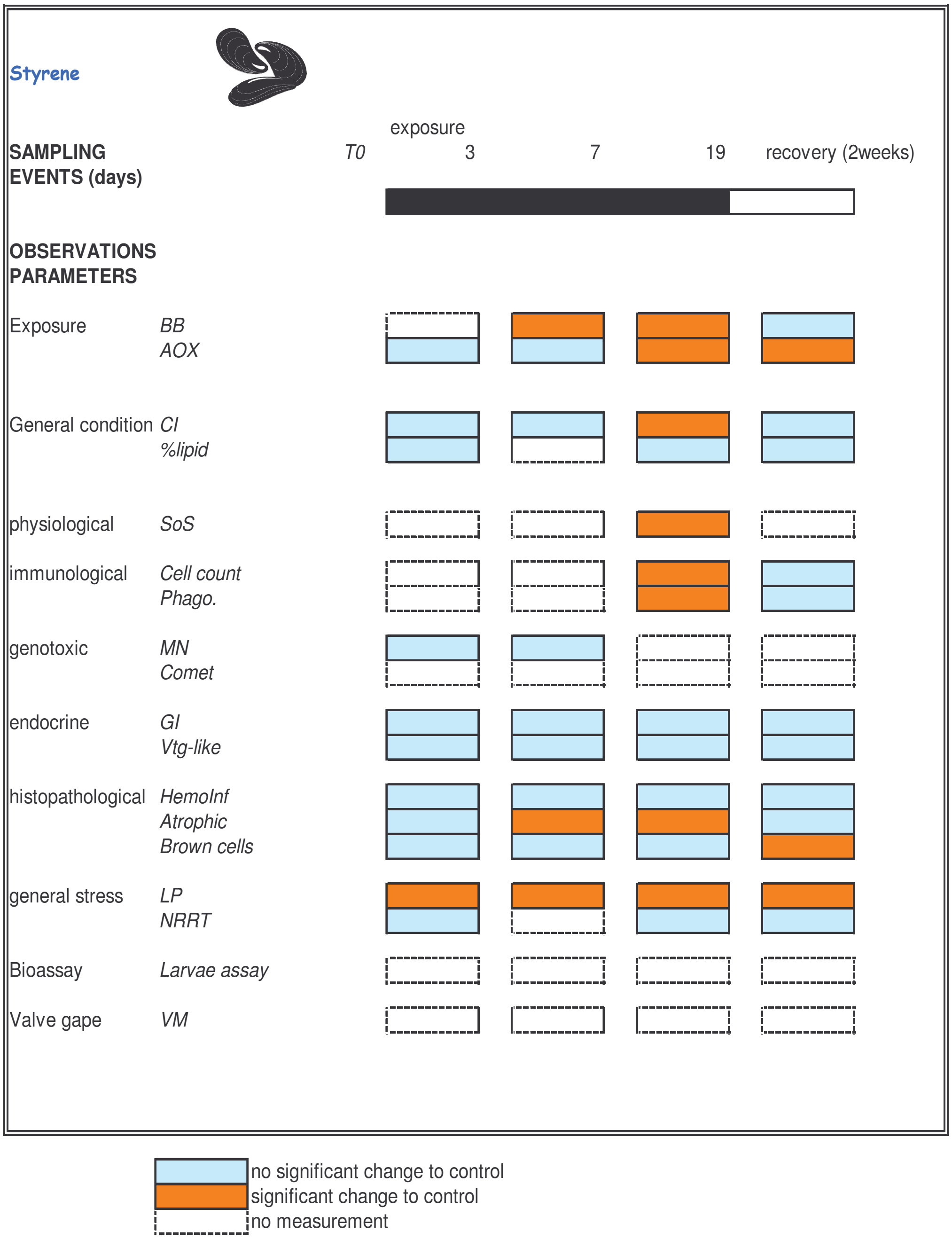


Figure 13 –Statistical significant changes in parameters measured in styrene exposed mussels compared to control mussel – Code as: *BB*-PAH body burden (GC/MS), *AOX*-Acyl-CoA oxidase activity (peroxisome), *CI*-Whole organism condition index , *%lipid*- percentage total lipid, *SoS*- Survival in air test, *Phago.*- phagocytosis activity (hemolymph) *MN*-micronuclei, *Comet-DNA strand break detection* (comet test, blood)), *GI*-Gonad index, *Vtg-like*-Vitellogenin-like protein level (gonad), *HemoInf.*-Hemocytic infiltration (digestige gland, DG), A*trophic*- Atrophic tubules (DG), *LP*-lysosomal membrane stability (DG), *NRRT*-Neutral red retention time in lysosome (hemolymph), *SSH*-gene up/down regulation. All statistical significance tested at p<5% using parametric or non-parametric tests.

5.2 An integrated approach

At first, the goal of the present section was to review some biomarker-based indices that have been developed so far and to assess their theoretical applicability for data derived from the PRAGMA project Some biomarker-based indices were selected from published papers:

- Narbonne's Biomarker Index (*Narbonne et al., 1999*)
- Integrated Biomarker Response (IBR) (*Beliaeff & Burgeot, 2002*)
- Chèvre's Biomarker Index (*Chèvre et al., 2003*)
- Bioeffect Assessment Index (BAI) (*Broeg et al., 2005*)
- Expert System (*Dagnino et al., 2007*)

Mulitmetric EPA index (http://www.epa.gov/owow/monitoring/tech/appdixe.html)

The selection of multi-biomarkers indices presented is not exhaustive but was made to illustrate and to appraise the variety of approaches.

5.2.1. The need for an integrative tool able to estimate risk

The relative lack of effective and simple tools for aiding interpretation of pollutant effect biomarker data, in the context of ecological risk assessment, has been a major limiting factor in the adoption of biomarkers in environmental management. Indeed the interest in biomarkers would be particularly highlighted if the entire biological measures from biological effects monitoring could be drawn together as a coherent whole. Environmental managers are seeking a simple and reliable integration system able to overcome the obvious difficulties in:

i. relating changes in biomarker data to environmental quality
ii. classifying the sites on a scale according to pollutant-induced changes in the health status of the organisms

Indices summarize information on the health status of marine organisms. While the original information on health status is expressed by several biomarker values, an index is expected to represent the most relevant information by standardized numeric scores. Instead of complex interpretation of individual variables, biological information is then transformed to one global value providing a reasonable apprehension of ecological risk. Such an index should reduce the complexity and uncertainty of multidimensional data for purposes of assessment, remediation and communication of results to the public and decision-makers.

5.2.2. Some principals concerning indices

General requirements

The dictionary definition of an "index" is: "a numerical scale used to compare variables with one another or with some reference number". To be relevant an index should meet the following criteria:

- *Provide a proper summary of the situation*: the index should be integrative and sensitive to stresses i.e. no loss of important information is to be feared and should exhibit a plausible biological situation.
- *Easily measured*: the index should be calculable from accessible data and easy to implement.
- *Objective*: the index should not depend on the researcher's personal feeling on how to deal with the measurements obtained.

There is a wealth of possibilities to calculate a summarizing index and this is reflected in the various biomarker-based indices developed so far and reviewed in the scientific literature.

General construction principals

- Step 1: *General mathematical design*

Different combinations, or in other words, mathematical algorithms combining the set of biological measures exist in index construction. In simple (and most) cases an index is formally a weighted sum of measurement results. In all situations the task of the index is to highlight important aspects of the data.

- Step 2: *Standardisation procedures*

Often a set of different quantities on different measurement scales constitutes the inputs of index calculation. Mathematical transformation of the original data set is therefore required to overcome this problem i.e. to generate a common scale for all components. Standardisation procedures are thus frequently used prior to index calculation in order to prevent the dominance of a component over others as a simple consequence of the units in which they are measured. Standardisation also accounts for the orientation of the quantity (increasing or decreasing). Negative "weights" are usually applied when a parameter is decreasing to ensure that all parameters are oriented the same way.

- Step 3: *Weighting*

The assignment of weights is a very crucial step because it may affect the sensitivity of an index. The rationale that led to these weights needs to be well-argued and to rely on strong scientific knowledge. Usual principles to derive weights are:

- Parameters expressing the presence of a worse health state are assigned a higher weight compared to those expressing less severe state.
- Weights are chosen such that the constructed index preserves as much as possible of the variation in all measurements.

The first principle requires more biological knowledge than is often available, while the second one can be followed relatively easy but may fail to generate relevant index.

Multi-Biomarkers indices of environmental quality: case-study

All of the selected biomarker-based indices are sum of scores related to the original biomarkers values (Cf. Fig 14). In additive index, the number of biomarkers plays an important role affecting the "relative weight" of each biomarker in the final index value. Indeed when the set is relatively large the weight of one factor is markedly reduced. The weighting step (Cf. step 3, Fig. 14) might be one way of correcting this bias.

From this set of published indices, two types of standardisation procedures were encountered:

- Transformation of original biomarkers values into graphic coordinates (graphical visualisation) like in the case of the IBR: this technique allows the visualisation of data into graphs as well as the calculation of quantitative values. Of course, computation of graphic coordinates requires pre-processing steps of the original data (ex: z-scores calculation, negative weighting in case the biological effect corresponds to inhibition and zero values set to zero).
- Establishment of classes. There are two fundamental approaches to classification: *"a priori"*, where "known rules" are applied to classifying

biomarker values and *"a posteriori"* where rules for classifying biomarker values are derived from the biomarker values themselves. In *"a priori"* classification (e.g. BAI), classes are established on expert judgement (biological knowledge and experience) based on beforehand background data. They are thus considered absolute, allowing the generalization of the index. Naturally the absolute character of these classes can be argued considering the reliance on empirical judgement. On the contrary, *"a posteriori"* classes (US EPA index) are relative as they only rely on the acquired biomarkers values. They apply to limited conditions i.e. the survey they are derived from. In case of spill however, this limitation may not be considered a problem as one focalised only on the regional impact. Yet, comparison of such index values across surveys more challenging or even impossible.

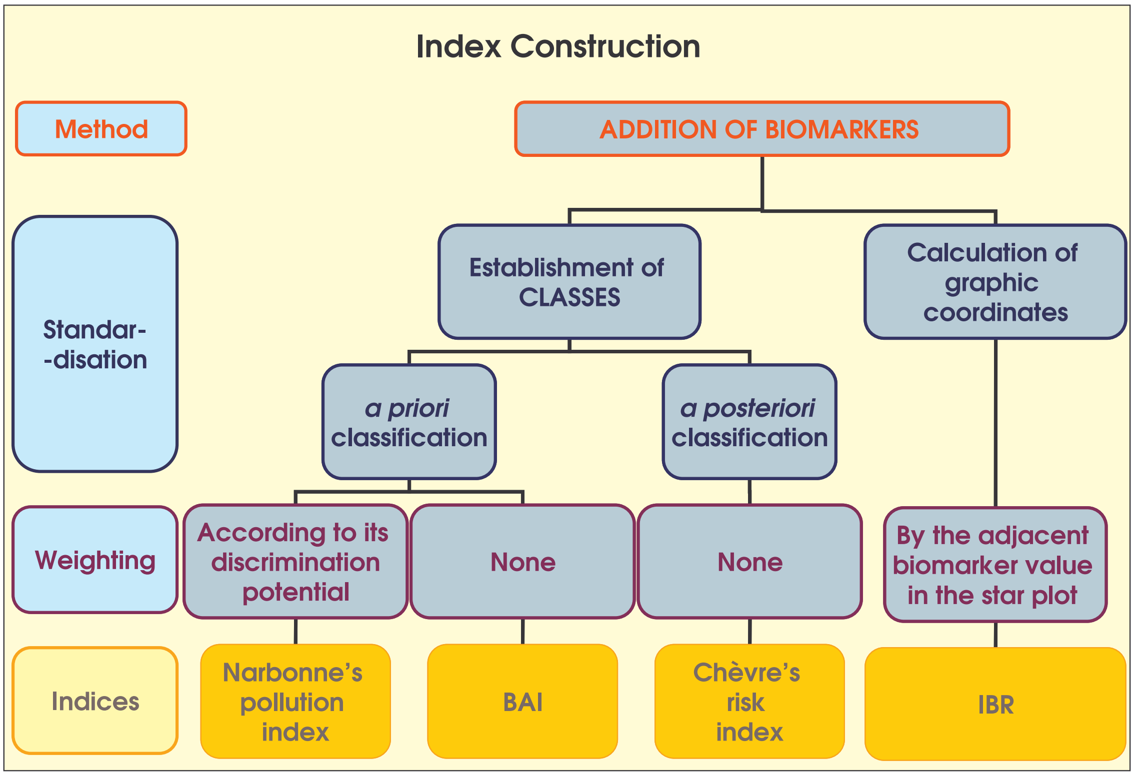


Figure 14: Construction proposal of some published biomarker-based indices.

Weighting is the final step in index construction but it is not obligatory. In some indices, biomarker values were not weighted prior to summation. Chèvre mentioned in her publication: "*although we might suppose that some biomarkers are more important than others, it is difficult to justify the selection of weight values. It would be preferable to over- or underestimate certain effects than to introduce unsuitable weights leading to incoherent results*". In the Integrated Biomarker Response index (IBR), one biomarker is weighted by the biomarker value, which is directly adjacent in the star plot. This approach probably emphasise/undervalue the impact of some pairs depending on the given order of biomarkers. In Narbonne's biomarker index, weights were chosen such that they accentuate variation in measurements. This process emphasizes the discrimination between sites to allow better visualization. Unfortunately no justification is given for the weight values by the author.

General characteristics of the reviewed indices **(excluded US EPA)**

Indices	Article	Authors	Date of release	Summary
IBR	"Integrated biomarker response: a useful tool for ecological risk assessment. Environmental Toxicology and chemistry	Benoit Beliaeff & Thierry Burgeot	2001	The IBR index is a simple multivariate graphic method which allows visual integration of a set of early warning responses measured with biomarkers (easy-to-use graphic means of summarizing available data).
Chèvre's biomarker index	"Development of a biomarker-based index for assessing the ecotoxic potential of aquatic sites." Biomarkers	N. Chèvre et al.	2002	In this index, the numerical scale has been generated by a mathematical process called local dicretization based on an algorithm developed by Fayyad & Irani (1993). The local discretization is the first step of the analysis by rough sets performed to generate rules for site discrimination.
Narbonne's biomarker index	"Scale of classification based on biochemical markers in mussels: application to pollution monitoring in European coast." Biomarkers	J.F. Narbonne , M. Daubeze, et al.	1999	Narbonne's Pollution Index Scale is based on the assignment of weighted score according to biomarkers discrimination properties. Sites are then classified on a 5 levels pollution scale according to the global response index.
BAI	"The "bioeffect assessment index" (BAI) a concept for the quantification of effects of marine pollution by an integrated biomarker approach." Marine Pollution Bulletin	K. Broeg et al.	2005	The BAI index allows integration of several pathological endpoints to test the general state of toxically-induced alterations in fish. The numerical scale is established on a knowledge-based approach where one parameter (lysosomal membrane stability) serves as a guide parameter and allows calibration of the other parameter.
Expert system	"Development of an expert system for the integration of biomarker responses in mussels into an animal health index." Biomarkers	A. Dagnino et al.	2007	The expert system is quite different from the other index as it is not a sum of scores. It is based on a set of rules in the "if…then" form relying on extended biomarker database. The final diagnostic is expressed in a five-level status index. Expert systems are typically used in clinical decision support system.

Establishment of comparative criteria

This evaluation inquiry is based on the review of scientific publications in the field of integrated multi-biomarkers indices of environmental quality. The basic reasoning behind the appraisal is the comparison and classification of statements and facts collected through this review. For analysis purposes, the assessment has been implemented using similar criteria which:

i. facilitate, on the one hand, the comparison of different indices based on a common scale and
ii. allow, on the other hand, the establishment of a general picture of the reviewed indices with better understanding of their signification and their construction principles.

These criteria (Cf. Fig. 15) were based on the general requirements and construction principals mentioned previously. The whole set were organised around five fundamental questions:

1. *Conceptual relevance*: What is the index significance? Is it based on well-established principles? Is it scientifically renowned?
   - General importance
   - Conceptual basis & prior knowledge
   - Reliability

2. *Field of application*: Is it widely applicable? Is it easy to implement? What type of input measurements does it rely on? What type of monitoring does it apply to?
   - Type of data
   - Type of parameters
   - Type of monitoring program

3. *Construction relevance*: How does it fill the construction principles?
   - Standardisation procedures
   - Weighting
   - General mathematical design

4. *Field of interpretation*: Is it a sensitive index? What type of conclusions can be drawn from this index?
   - Calibration
   - Dependant from biotic/abiotic factors
   - Clear diagnosis

5. *Linkage to management action*: What type of tool is it? Is it relevant for ecological risk assessment (management concerns)?
   - Easy to set-up
   - Rapid / Effective
   - Presentation of the data

Conceptual relevance	General importance	Definition
		Significance
	Conceptual basis & Prior knowledge	Biological concepts
		Prior knowledge
	Reliability	Published in
		Used in
		Mentioned in at least
Field of application	Type of data	Amount
		Data quality
		Data distribution
	Type of parameters	Numbers
		Typology
		Species
	Type of surveys	Spatio-temporal monitoring
		Type of contamination
		Regional/local campaigns
Construction relevance	Standardisation procedures	Input data
		Parameter's orientation
		Dimensionless quantities
		Objectivity
	Weighting	Assignment of weights
		Justification
	General mathematical design	Additive index
		Global response index
Field of interpretation	Calibration	Index construction
		Choice of parameters
	Dependant from biotic/abiotic factors	Yes/No
	Clear diagnosis	Justification of choices
		Awareness and carefulness
		Expertise
Linkage to management action	Easy to set-up	Practical convenience
		Special needs
	Rapid/effective	Setting up
		Interpretation
	Presentation of the data	Graphic tool

Figure 15: Overview of the comparative criterions used for the indices appraisal.

Recapitulative table

The establishment of comparison tables allowed a more comprehensive overview of the integrated multi-biomarkers indices . The establishment of a recapitulative table summarizing all these comparison tables was achieved using synthetic criteria. These criteria correspond to the indices characteristics (orange cells in figure 16) and are deduced from the previous criteria (yellow cells in figure 16):

Conceptual relevance			Field of application		Construction Relevance		Field of interpretation	Linkage to management action		
✓ **General importance** ✓ **Conceptual basis & Prior Knowledge** ✓ **Reliability**			✓ **Type of data** ✓ **Type of parameters** ✓ **Type of surveys**		✓ **Standardisation procedures** ✓ **Weighting** ✓ **General mathematical design**		✓ **Calibration** ✓ **Dependant biotic/abiotic factors** ✓ **Clear diagnosis**	✓ **Easy to set-up** ✓ **Effective** ✓ **Presentation of the data**		
Method's Insights	Based on extended biomarker knowledge	Supplemented by chemical analysis	Input constraints	Applications in monitoring	Statistical analysis prior to index computation	Rationale	Absolute (can sense a wide range of stresses in any biosystem) / Conditional index	Easily practicable	Type of "Decision-support" system (DSS)	Operational tool

Figure 16: Summary table layout.

- Method's insights: Overview of the method used.
- Based on extended biomarker knowledge: Integration of previous validated knowledge gives an indication on the index's credibility.
- Supplemented by chemical analysis: Stand-alone or needs further analysis for confirmation (prior knowledge about site polluted situation).
- Input constraints: Input data and parameters constraints due to the index computation.
- Applications in monitoring: Range of applications in biological effects monitoring.
- Statistical analysis prior to index computation: Direct or indirect index.
- Rationale: The index construction's reasoning is disputable or not (ex: subjectivity).
- Absolute (can sense a wide range of stresses in any biosystem)/ Conditional index: Range of applicability in monitoring. Index is conditional when it can sense some stresses in some particular species otherwise it is absolute.
- Easily practicable: Easy to set-up and to compute.
- Type of "decision-support" system (DSS): Indices are considered as "decision-support" system in environmental management. DSS is a tool (usually computerized) for helping make decisions. Using the mode of assistance as a criteria, DSS can be differentiated as:

- *Communication-led approach*: an approach driven by the need to communicate the data and to be used by the broadest range of people possible.
- *Model-led approach*: an approach driven by the need to analyse the situation through a model (the model being entirely dependant from the data)
- *Knowledge-led approach*: It provides specialized problem solved by expertise i.e. by the general knowledge stored as facts, rules...

° <u>Operational tool:</u> It strictly fits in with the general requirements set earlier.

Indices	Conceptual relevance			Field of application		Construction Relevance		Field of interpretation	Linkage to management action		
	✓ **General importance** ✓ **Conceptual basis & Prior Knowledge** ✓ **Reliability**			✓ **Type of data** ✓ **Type of parameters** ✓ **Type of surveys**		✓ **Standardisation procedures** ✓ **Weighting** ✓ **General mathematical design**		✓ **Calibration** ✓ **Dependant biotic/abiotic factors** ✓ **Clear diagnosis**	✓ **Easy to set-up** ✓ **Effective** ✓ **Presentation of the data**		
	Method's Insights	Based on extended biomarker knowledge	Supplemented by chemical analysis	Input constraints	Applications in monitoring	Statistical analysis prior to index computation	Rationale	Absolute (can sense a wide range of stresses in any biosystem) / Conditional index	Easily practicable	Type of "Decision-support" system (DSS)	Operational tool
IBR	Simple mathematical transformation of the data set allowing multiple visual comparisons.	No	Yes	Low	Vast	No	Disputable	Absolute	Yes	• Exploratory approach. • Communication-led DSS.	Yes
Chèvre's biomarker index	Diagnostic is based on the discrimination abilities of biomarkers.	No	Yes	Medium	Limited	Yes	Undisputable	Absolute	No	• Conclusive approach. • Model-led DSS.	Yes
Narbonne's biomarker index	Diagnostic is based on the discrimination abilities of biomarkers.	?	Yes	Medium	Vast	Yes	Disputable	Conditional	Yes	• Conclusive approach. • ?	No
BAI	Diagnostic is based on the screening of responses for stress syndrome. Similar approach in medicine.	Yes	No	High	Limited	No	Disputable	Conditional	Yes	• Exploratory approach. • Knowledge-led DSS.	No
Expert system	Diagnostic is based on the screening of responses for stress syndrome. Similar approach in medicine.	Yes	No	Medium	Vast	Yes	Disputable	Conditional	Yes	• Conclusive approach. • Knowledge-led DSS.	Yes

Figure 17: Recapitulative table.

5.2.3. Application of a selection of indices from the project – case study applied on data from the mussel experiment

The ultimate objective of this study is to propose an appropriate and integrative method that could apply to the management of environmental data following spill event and as decision-making criteria. Considering the outcome of this assessment, several data treatment methods have been tested using the available data from the PRAGMA project:

- an existing multi-biomarkers index of environmental quality based on the IBR star plot representation method
- a scoring method derived from US Environmental Protection Agency technical guidance documents for biomonitoring and assessing water quality.

If one of these approaches proved to be effective in integrating and communicating all the results, the interest in biomarkers could be enhanced and the selected methodologies could be proposed to be included in EU guidelines.

Integrated biomarker approach with the IBR

The IBR is an index related to the exploratory approach. This method can be employed as a mean of detecting changes on a routine basis and to follow up recovery of the impacted zone The IBR can be supplemented by chemical analysis (required for comparing results). Prior to the application of IBR, identification of the source of pollution, or at least some information on the contaminants mixture present in the studied environment is regarded as an advantage. The IBR is thus appropriate for *specific investigation* where the stressor is identified prior to the implementation of monitoring. The IBR can hence be considered as an easily measurable tool, usually providing a proper summary of the situation. At an exploratory level i.e. as a first step in environmental quality assessment, the meeting of these conditions is sufficient to be regarded as operational.

The general computation of the IBR can be found in Beliaeff and Burgeot (2002). Here some adaptations have been adopted based on the measurements done in PRAGMA and the work based on Broeg et al. (2006).

The IBR has been developed and used originally only on a series of so-called biochemical "early-warning" biomarkers but in the present assessment, both histochemical and general health parameters have been used since little attention was paid on the early, often transient responses following spill or pollution situation.

This index requires a minimum number of parameters to be reliable and as a rule, the power of the IBR increases with the number of parameters used in the investigation. Here we used a minimum of four and a maximum of ten parameters depending on the sampling event time. Also, since the principal calculation of the IBR is obtained by summing up scores derived from the biological data values, it is thus directly dependent on the number of parameters used or available for each sampling interval. During a monitoring survey, data are rarely obtained regularly from all parameters for several reasons e.g. logistic or economical. Hence, the final IBR values have here been weighed by the number of parameters measured by sampling time. These values have been referred as IBR/n (Fig.18 & 19).

Figure 18 – Application of the weighted IBR on fuel oil data from the mussel experiment – Leftt panel: bar charts showing IBR/n computation in control/exposed animals and most contributing parameters. Right panel: Top - star plot of PAH body burden (PAHbb) and IBR/n; Bottom – Relative IBR/n changes in exposed animals plotted against PAHbb.

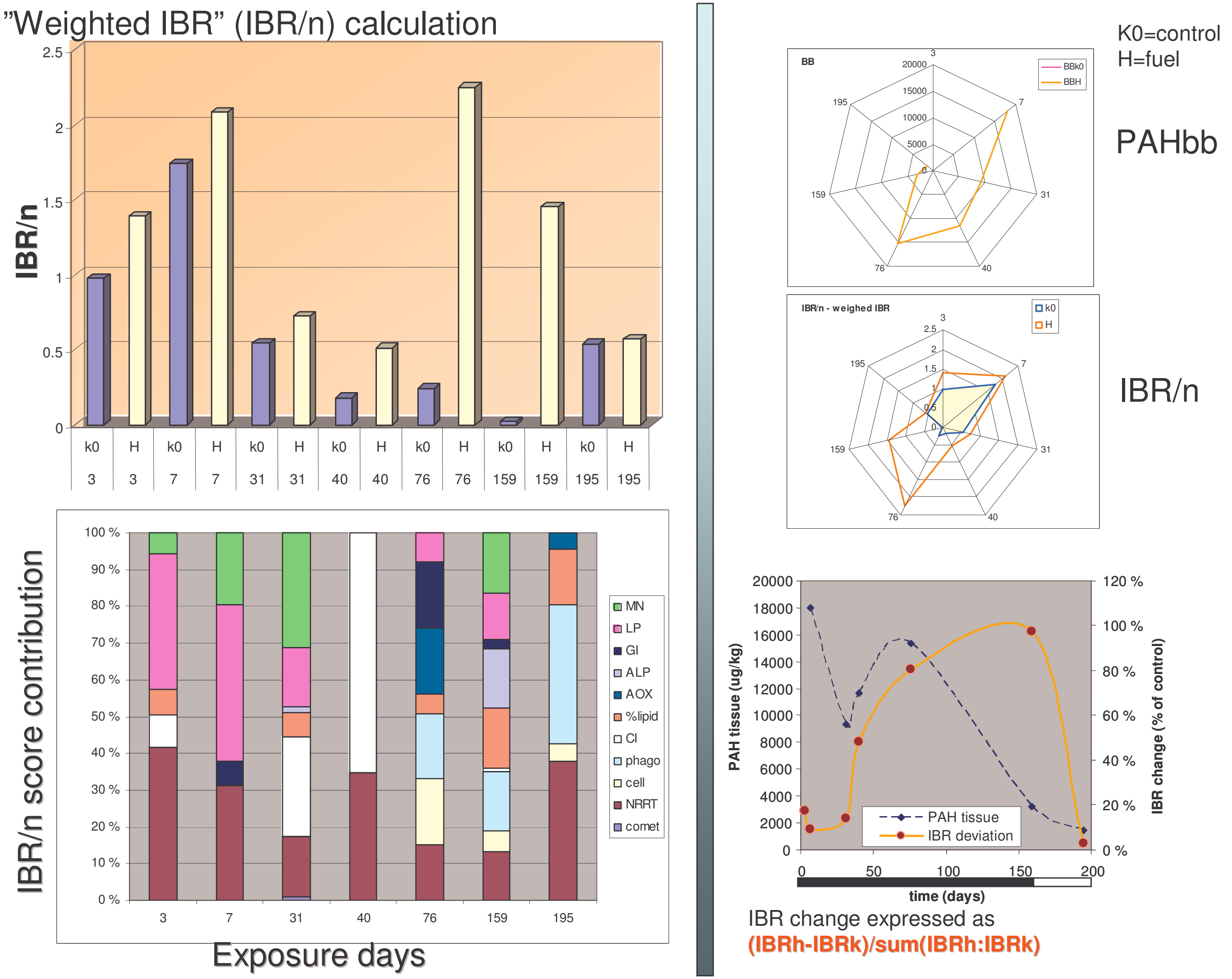
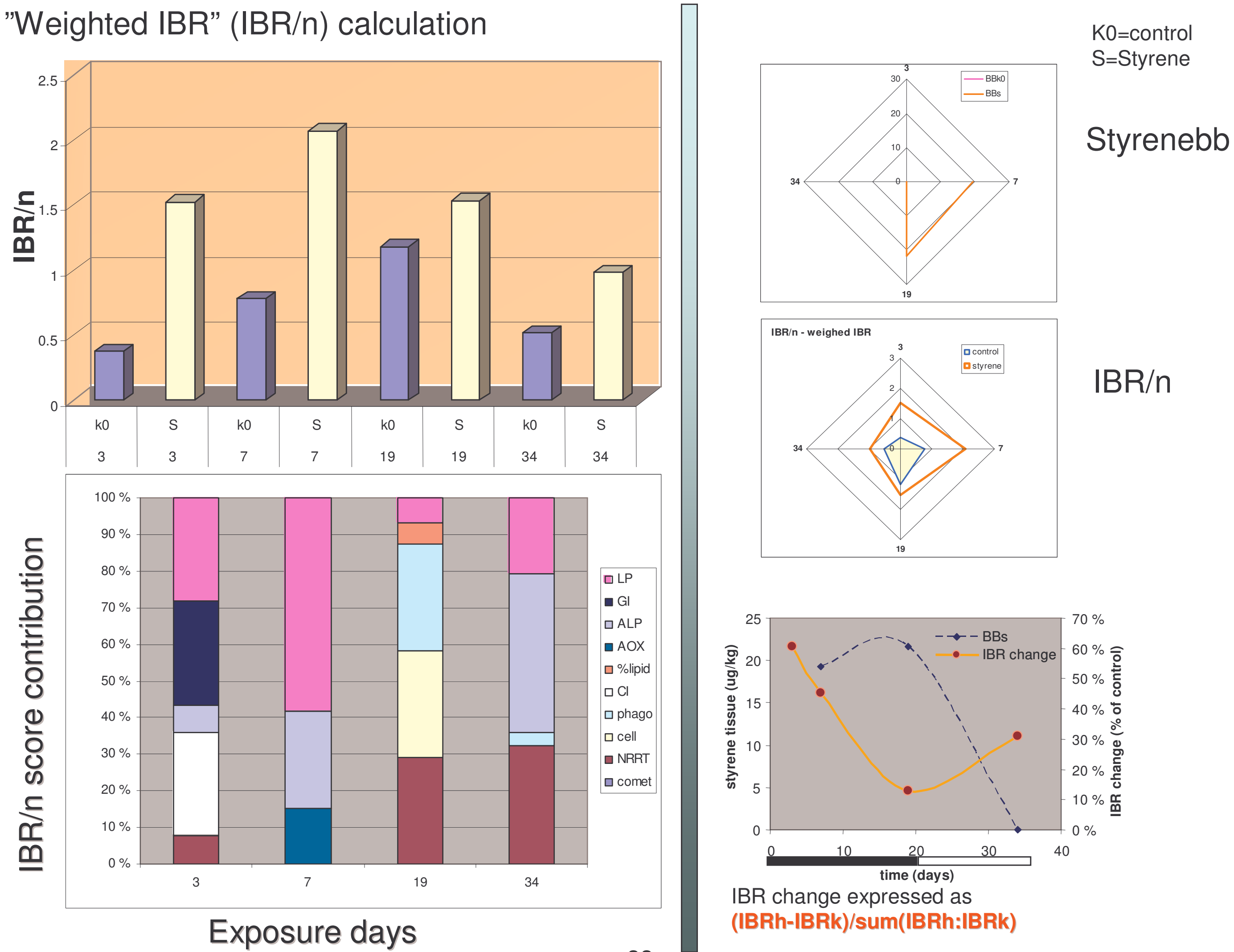


Figure 19 – Application of the weighted IBR on styrene data from the mussel experiment. Leftt panel: bar charts showing IBR/n computation in control/exposed animals and most contributing parameters. Right panel: Top - star plot of styrene body burden (styrenebb) and IBR/n; Bottom – Relative IBR/n changes in exposed animals plotted against styrenebb.



US EPA multimetric system

Usually, metrics are judged responsive if there are statistical significant differences in central tendency or in variance between reference and impaired sites (Fig. 20). The hypothesis is that, if the test sites are affected by anthropogenic pollution or disturbance, then mean or median values of responsive metrics should be substantially different between reference and test sites. Several situations may be found: but generally, if the test sites simply do not meet reference criteria then the variance in the test sites should be larger than that in the reference sites. In many cases however, there may be a change of only some values out of the total number of observations done (mixed situation where some individuals are impaired, some are not).

Biologists are using statistical significance tests to select metrics and report their observation with some strength, but slavish reliance on significance tests does not contribute to biological understanding and actually result in false interpretation. If sample size is small (say, $n \approx 5$ in both reference and impact sites), then significance tests (at $p< 0.05$) will have low power and responsive metrics may be rejected. On the other hand, if sample size is large (say, $n = 30$ in both site categories), then it would be possible to detect a statistically significant difference that is possibly biologically meaningless or is not *actually* existing. In this case, metrics that do not contribute to meaningful assessment could be selected, simply because statistical significance was detected. An alternative measure is the expected frequency with which a metric will fall below a threshold to register impairment. Frequency can be estimated with a box and whisker plot, but not with a significance test. For example, if the median of impaired sites is below the quartile of reference sites (Fig. 20), then we estimate that impaired test sites will be below the reference quartile in at least 50% of all observations.

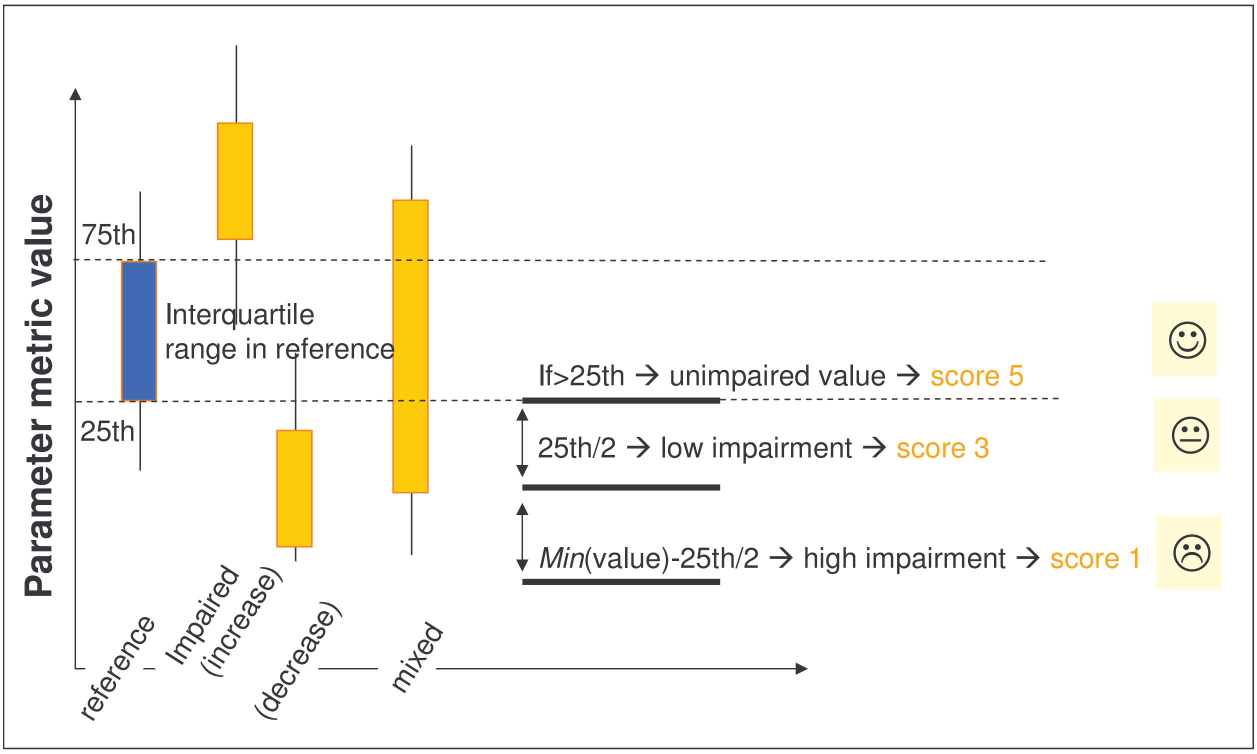


Figure 20 – Illustration of scoring method using the whisker box plot of the referente distribution.
Note – If impairment results in decrease then the 25th quartile of the reference distribution was used; in case of increase, the 75th quartile of the reference distribution was used.

All of these require comparison to some measure of the reference value distribution: an upper percentile, a lower percentile, or a central tendency

Here we have chosen a scoring criteria based on the quartile value of the reference (control) distribution In this method, values above the 25th percentile are considered unimpaired (similar to reference conditions) and values below the 25th percentile are considered impaired to some degree. The range from 0 to the 25th percentile is bisected, with values in the top half receiving a score of 3 and those in the bottom half receiving a score of 1 (Fig. 20).
A multimetric index has been calculated by simply summing up the score obtained by this method for each parameter.
We have applied this technique to the data collected during the Pragma project and used here 7 parameters from the pilot experiment with mussel exposed to fuel oil (fig 21).

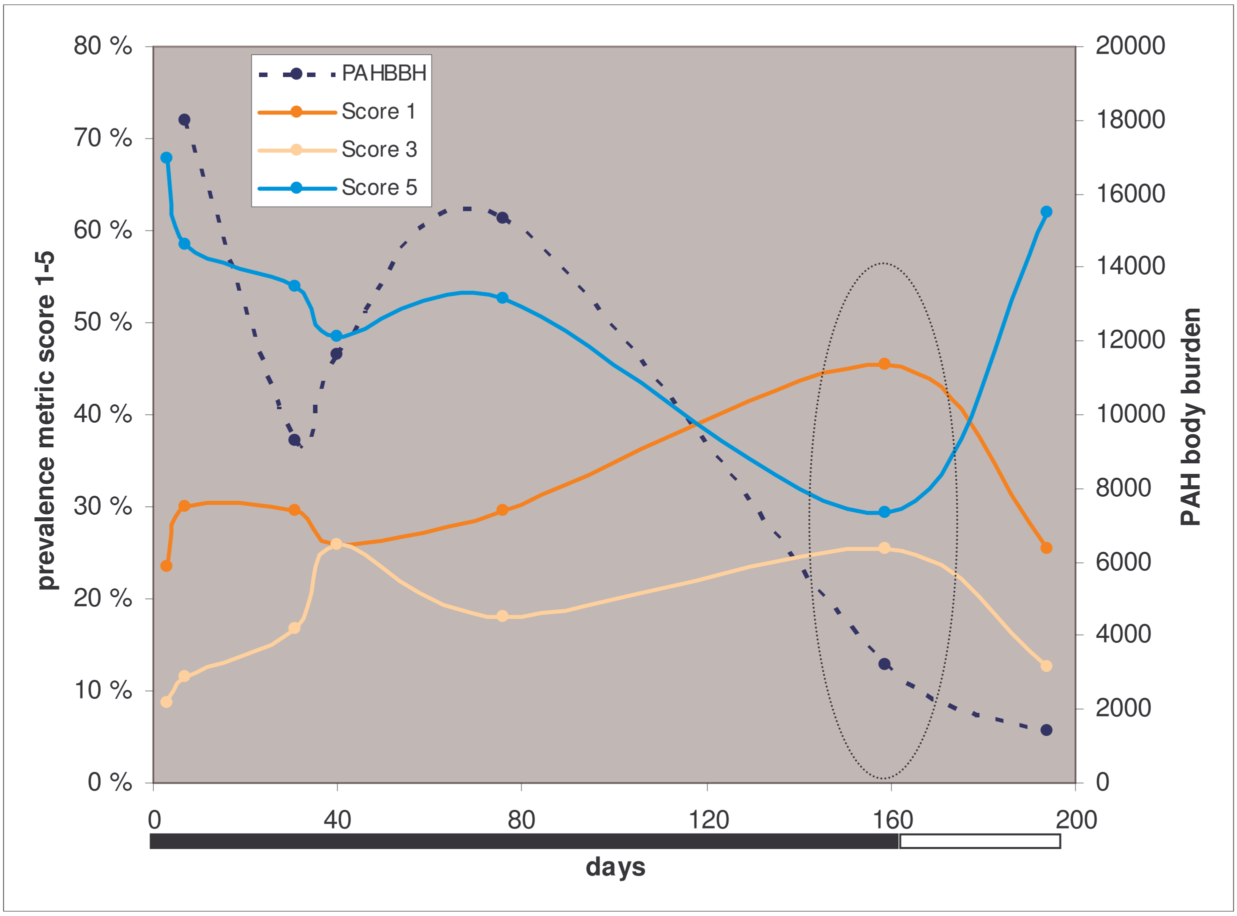


Figure 21 – Prevalence (%) of the score 1,3 and 5 based on 7 measured parameters during the mussel pilot study with fuel oil. Score 1 – *high level of impairment*; score 2 – *low level of impairment*, score 5 – *no impairment relative to control group*. The parameters used are CI, NRRT, %lipid, AOX, ALP, LP and GI (see legend of **fig.11** for coding). The circle shows the time (end of exposure, 159 days) where the number of data with high impairment is above critical level.

Evaluation of both IBR and US EPA metric system

Both metrics are easy to compute and do not require complicated calculations. They can easily be made on an excel worksheet. While the IBR values represent the total integrated biomarker response of the individuals at each sampling events, the US EPA metric divides intentionally data at each sampling event in 3 categories (=scores) where category 5 have no deviation from reference data and category 1 shows high deviation from reference. The decision criteria is based on a simple *a posteriori* assessment based on the reference data. Hence, it is important to obtain reliable and good reference data. Comparison of figures 18 and 21 obtained with the two respective methodologies shows the same trends. During the first 30 days of exposure to fuel oil, changes in biological measures are observed but the total integrated deviation in the exposed mussel population is moderate based on the selected parameters. Thereafter, when the exposure simulates a re-suspension event boosting the total PAH content in the water and consequently in mussels, there is a gradual increase of the deviation to the reference population until the end of the exposure. It is remarkable to observe that this trend is opposite to the general pattern of chemical burden during the exposure period. Indeed, while the exposed population are gradually getting more impaired (deterioration visualized as an increase in IBR/n or increase of score 1 prevalence), their total chemical burden is reduced because of the aging of oil. Hence there is a time lag between the total chemical burden and the biological responses. This is of critical importance for decision-making criteria. Based on chemistry alone, the mussels would appear relatively healthy at the end of the exposure and on their way towards recovery. Based on the selected parameters of PRAGMA, the conclusions would differ completely as their actual health quality indexes are poorer (=high IBR/n or high score 1).

Hence, these results validate the approach used in this project and confirm the general observation that biological measurements should be added to chemical measurements for a better environmental assessment of sites following spill events.

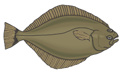
Both the IBR and the US EPA metrics show that the impacted mussels recovered post-contamination and after one month in clean seawater.

The computation and visualisation of these metrics are easy to read and understand. Based on our evaluation, they could serve in the assessment of the environmental quality status and for decision-making purposes.

5.3 Recommendation of environmental criteria and harmonisation at EU level

Based on the pilot work with fish

FUEL OIL
Styrene



suitability for field monitoring

CRITERIA CODE

- ☺ *recommended parameter*
- 😐 *recommended parameter although little response in present study*
- ☹ *not recommended*

Time and cost

- 👍 *low*
- 👎 *high*
- ≈ *in between*

OBSERVATIONS PARAMETERS				time	cost
Exposure	*MET*	☺	bile	👍	👍
	EROD	☺	liver	👍	👍
	AOX	☺	liver	👍	👍
haematological	*RBC*	☺	blood	👍	👍
	ENA	☺	blood	👍	👍
	hemato	☺	blood	👍	👍
physiological	*Osmo*	☺	blood	👍	👍
	Chloride	☺	blood	👍	👍
	blood gases	☹	blood	👎	👎
immunological	*Leucocyte*	☺	blood	👍	👍
genotoxic	*MN*	😐	blood	👎	👎
endocrine	*Ghisto*	😐	gonad	👎	≈
	Vtg	😐	liver	👍	👍
histopathological	*MMC*	☺	liver	👎	≈
	Necrosis	😐	liver	👎	≈
	Vacuol.	☺	liver	👎	≈
general stress	*LP*	☺	liver	👍	👍
DNA biosensor		☺	bile	👍	👍
"OMICS"	*DNA array*	☹	liver	👎	👎

Based on the pilot work with mussel

FUEL OIL
Styrene

suitability for field monitoring

CRITERIA CODE

- ☺ *recommended parameter*
- 😐 *recommended parameter although little response in present study*
- ☹ *not recommended*

Time and cost

👍 *low*   👎 *high*   ≈ *in between*

OBSERVATIONS	PARAMETERS			time	cost
Exposure	*BB*	☺	whole	👎	👎
	AOX	☺	DG	👍	👍
General condition	*CI*	☺	whole	≈	👍
	%lipid	☺	whole	👎	≈
physiological	*SoS*	😐	whole	👎	👍
immunological	*Cell count*	☺	*hemo.*	👍	👍
	Phago.	☺	*hemo.*	👍	👍
genotoxic	*MN*	😐	*hemo.*	👎	👎
	Comet	😐	*hemo.*	👎	👎
endocrine	*GI*	😐	*gonad*	≈	👎
	Vtg-like	😐	*DG*	👍	👍
histopathological	*HemoInf*	😐	*DG*	👎	≈
	Atrophic	☺	*DG*	👎	≈
	Brown cells	😐	*DG*	👎	≈
general stress	*LP*	☺	*DG*	👍	👍
	NRRT	☺	*hemo.*	👍	👍
Bioassay	*Larvae assay*	😐	*larvae*	👎	👎
Valve gape	*VM*	☺	whole	👍	👍
"OMICS"		☹		👎	👎

6. General evaluation of the technical results and deliverables

In general, this pilot project gives support to the use of a more biological-oriented monitoring approach to follow up environmental impact of a toxic spill situation. The methodologies tested herein provided good indication of changes at several levels of biological organization in biota placed in realistic exposure situations. Hence, a major outcome of this project is to confirm the relevance of such approach in monitoring programmes related to spill of heavy oil or other noxious chemicals.

Many of the parameters used in this study are already well-implemented techniques often routinely performed by several laboratories in Europe. They comply also with the recommended list proposed by international European group like ICES-WGME. For many of them a standardized procedure exists.

For field monitoring, factors related to sensitivity, cost-effectivity, repeatability and practicability and relevancy of methods and tools are important to assess in the final tool box selection the scientists and the governmental authorities can integrate into a contingency framework. A method can be informative of a change but practically too challenging to use for economical reason or practicability of deployment. Hence a sound balance between all these factors has to be made.

In this project, we have deliberately chosen an approach where methods measuring transient responses were not favoured. Even though these methods can be interesting to understand the mechanisms behind the observed effects, they may not be used in practice because their amplitude is changing too rapidly in time and there seems to be a fast recovery. Yet, absence of these responses does not necessarily mean absence of longer-time effects. For that reason, we selected parameters which are not specifically related to a type of pollution but indicate either general stress condition or conditions that can be related to serious impairments in the individuals.

Clearly, the combination of several parameters that measure changes at various levels of biological organisation (cell, tissue, individual) seems to be a sound approach. The causality of these changes may express at different interval of time and for different exposure conditions. However, for mussels, the chemical burden (BB) and the general stress parameters that measure lysosomal membrane stability (LP in the digestive gland of mussel/liver of fish or NRRT in the hemolymph) were clearly different between reference and exposed situation at all sampling events. Compared to LP, NRRT does not necessitate killing the individual which if carefully done – can be re-used on several occasions. For fish, the measurement of metabolites by fluorescence technique, the measurement of EROD activity (a measurement of the uptake and biotransforamtion of the compounds) and simple blood parameters were very indicative of impairment at almost all sampling events.

A large effort needs to be done at the European level to understand better how biological parameters used or recommended for biomonitoring programmes behave even in the absence of pollution event. Indeed, there is a strong need to address the reference levels issue at the European level

- in order to describe better the “normal” pre-spill situation (natural seasonal variation etc) and establish clear sampling protocols (i.e. species selection, number of individuals, sensitivity to biotic and abiotic conditions etc)
- in order to be able to decide on recovery post-spill

Novel devices related to sensing techniques proved to be very informative of a change at a screening level both in mussel and in fish (bile). With these devices, the quest is to obtain relatively rapid hence cost-effective measurements that can potentially be done at

a much higher frequency both in time and in space. In addition, these tools are prone to be integrated in other platforms combining several measurements that can be performed on-line and controlled remotely. Hence, the interest for such development for environmental field monitoring has increased tremendously during the last decade although there is still few operational systems in place for the marine environment.
Owing to this system, there exist current barriers to the realization of their potential for routine use in monitoring like:

- stability
- price of development and production
- instrumentation requirements
- acceptance of these tools into regulatory context
- field validation of proof-of-concept and relevancy to other chemical and biological responses
- standardisation

In one hand, the selection of methodologies recommended in EU contingency and environmental guidelines is an important step but to be really operational and accredited in decision-making process, there is a request to integrate them in a simple assessment tool that summarizes the status of the water and simplifies the communication of the results. This was a major deliverable of Pragma. Here, we used to techniques derived from the IBR and the US EPA metric system
Based on the results obtained in the present study, it can be noticed that together the indices show coherent results compared to the conclusions based on statistical analysis alone.

As mentioned before the IBR is an operational graphical tool that allows a simple visual integration of a set of biological responses. For that reason this index was selected to analyse the results from this project. This index is easy to set-up. Moreover, this trial proved that the IBR goes beyond a simple exploratory approach, as it distinguishes exposed from non-exposed animals and can help to reveal biomarkers contribution. It can be used in both approaches:

- the exploratory approach or the routine inquiry where indices considered as preliminary methods can be employed as a mean of detecting problems on a routine basis.
- the explanatory approach or the comprehensive inquiry where indices can be applied as means of understanding the situation (situation analysis): the contamination processes and localisation of the source of contamination. This inquiry is deemed as a verification of the previous one.

The inconvenient is that it requires re-calculation anytime modifications or new information are added. Modification may concern changes in the positioning order, the amount as well as the content of biomarkers. However, here we used a weighed IBR index that theoretically was independent of the number of biomarker used. To avoid useless calculations, the order in which biomarkers are plotted on the star needs to be well thought through beforehand. Also the implementation of this index is quite simple results interpretation can be sometimes confusing. It has been noticed that the IBR's standardisation process is involuntarily making zero values emerge. In fact this has a great impact on the multiplication process where the appearing of zero values is amplified. This might be the reason why IBR index scores can be different according to the positioning order of biomarkers on the star plot.The IBR users must always remember that each biomarker contribution to the IBR depends on the values of two measurements: this very same biomarker and its neighbour in the star plot. Nevertheless, star plots allow simple visual comparison between sampling sites.

The US EPA multimetric approach provided results comparable to the IBR method. This approach is very simple to use. The basic idea is to establish the metric scoring based on the metric distribution in the reference sites. Metrics are given ordinal scores corresponding to impaired, intermediate or unimpaired data. Hence, that approach requires that the background level of the reference site is well described since the metric scoring is established on it.

The use of integrated indices describing contaminant-induced stress can be considered as useful management and exploratory tools as they have proved to be consistent and coherent with the polluting situation simulated in the PRAGMA project. Yet this data-treatment method depends on other data analysis for interpretation. Site's background knowledge acts as a determinant support of the indices outcome. Indices results can hardly be interpreted without this background. To achieve and assess good status of marine water bodies, experts advocate the integration of various disciplines and analysis such as, in this scenario, biomarker measurements and novel/emerging on-line/on-site tools combined with chemical analysis.

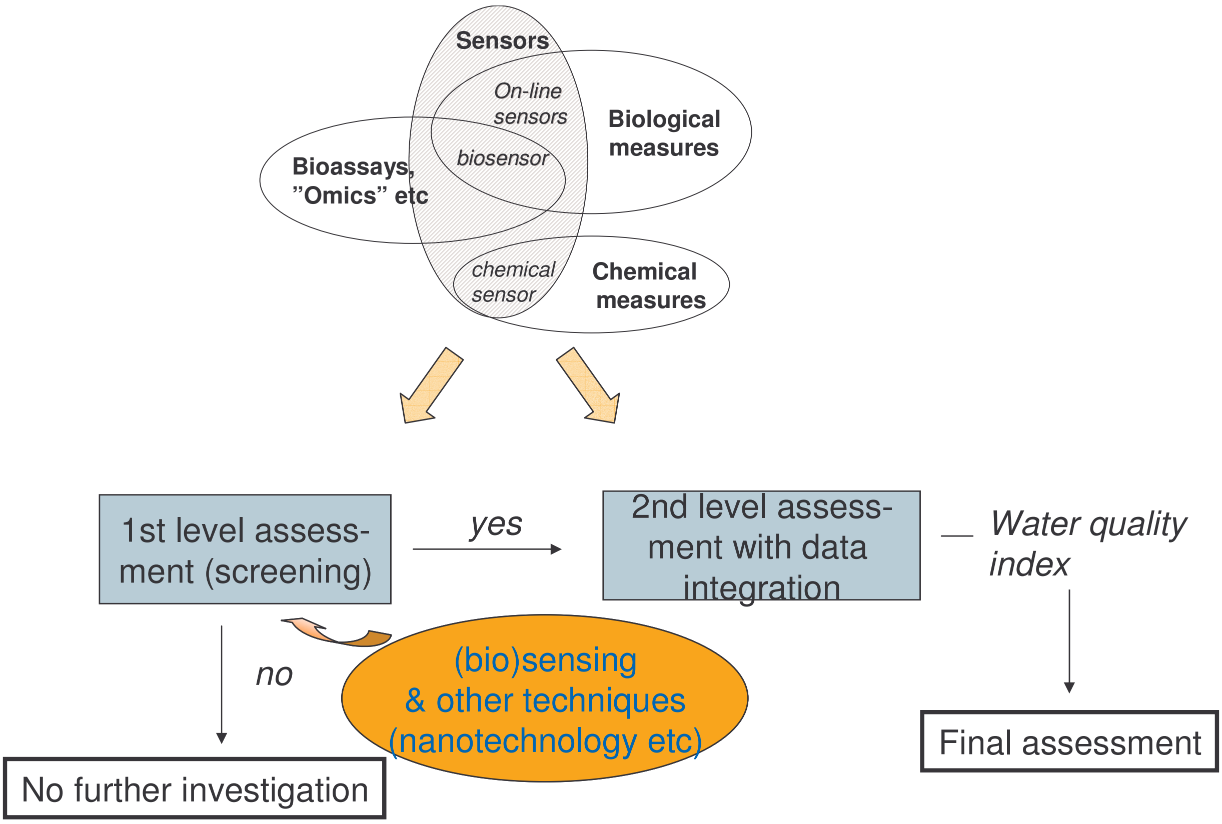


Fig. 22 – Recommended cost-effective approach for environmental monitoring and decision-making process in impacted marine areas.

In summary of this project, it appears that a good practice for monitoring environmental damage of oil and HNS chemical spills would be a cost-effective combination of several methods and technologies developed in a tier approach (Fig. 22). With the on-going development and potential use of sensing techniques for environmental monitoring in pre- or post-spill operations, a sound incorporation of these techniques at least at the screening level appears advisable. New directions in environmental monitoring indicate that novel tools like chemical-, bio- or nanosensors have been showing considerable progresses. With the improvement of sensitivity, selectivity, durability as well as miniaturization, it is likely that these devices will become operational to the point that they can be incorporated at least in the initial screening phase. This may demand further funds from European and international organizations. Undoubtedly, chemical assessment is basically needed both in water and in biota, but is not sufficient to obtain an actual health assessment of an impacted region. Other methods based on a rationale and suitable selection of biological measurements like those suggested in this project is recommended. Yet, the combination of these and possibly other emerging parameters ("omics") into an environmental decision-making tool for managers and national authorities is essential in order to interpret, summarize and communicate the water quality status during the different stage post-spill. This process can also be used in the development of a technical-juridical-economical methodology to assess the damage impaired to the marine environment from an economic point of view (economic estimate of the injuries to all marine resources i.e. water, sediment, biota). This would help decision makers in their assessment toward the public stakeholders.

This project showed that this approach can give a sound and simple expression of the health assessment that can be useful for scientists to have methodologies at hand for the evaluation of environmental consequences, stakeholders and authorities that need to translate these consequences into regulation, and for the public that requires a better communication.

7. Works derived from PRAGMA

7.1 master thesis

- Ruiz, P. Exposure and effect biomarkers in mussels and fish exposed to styrene and fuel oil. Master Environmental Contamination and Toxicology. September 2007, pp 78.
- Vingen S. Biological responses in mussel (*Mytilus edulis*) exposed to heavy oil and styrene in the laboratory: Linking physiological condition to contaminant accumulation – Method evaluation. Master Thesis at UiS (University of Stavanger, Norway), June 2007, pp. 62.
- Sinet E. Integrated multi-biomarkers indices of environmental quality: means of data-treatment in marine biological effects monitoring. Master Degree at ENSAIA (National Engineering School of Agronomy and Food Sciences, Nancy, France), September 2007, pp. 50.

7.2 presentations in congresses

- 17th ANNUAL MEETING OF THE SOCIETY OF ENVIRONMENTAL TOXICOLOGY AND CHEMISTRY-EUROPE, Porto (Portugal), 2007:
    - Baussant, T, Le Floch, S, Theron, M and Cajaraville, MP. A pragmatic and integrated approach for the evaluation of environmental impact of oil and chemical spills at sea.
- 18th ANNUAL MEETING OF THE SOCIETY OF ENVIRONMENTAL TOXICOLOGY AND CHEMISTRY-EUROPE, Warsaw (Poland), 2008:

- Cancio, I, Diaz de Cerio, O, Martínez, P and Cajaraville, MP. Gene expression analysis of n° 2 fuel oil exposure in the liver of juvenile turbot (*Scophthalmus maximus*) using a custom oligonucleotide microarray.

- 7th IBERIAN AND 4th IBEROAMERICAN CONGRESS OF ENVIRONMENTAL CONTAMINATION AND TOXICOLOGY, Lisbon (Portugal) 2008:
  - Ruiz, P, Ortiz-Zarragoitia, M, Orbea, A, Baussant, T and Cajaraville, MP. Short- and long-term responses of mussels *Mytilus edulis* exposed to styrene and heavy fuel oil nº2.
  - Ruiz, P, Ortiz-Zarragoitia, M, Orbea, A, Cancio, I, Theron, M, Le Floch, S and Cajaraville, MP. Exposure and effect biomarkers in turbot (*Scophthalmus maximus*) exposed to styrene and heavy fuel oil nº 2.

8. Follow-up

New funding (Respil - Grant Agreement N ° 07.030900/2006/448357/SUB/A3) by the DG-Environment have been allocated to IRIS and partners Cedre and UBO/Lemar for a pilot study within the 2006 marine framework call. The resources will help implementing further the developed approach in Pragma to other type of chemicals transported in large quantities along the European waters and for which fewer monitoring methodologies currently exist. We hope this new grant will also contribute to a better implementation at the european level of the proposed approach and recommendations made in Pragma. Hence the experience and network built in Pragma will be transferred and followed up in Respil.
It is clear that a larger effort should be made in establishing better the natural variability (from biotic and abiotic factors), the sensitivity of the biological measurements to other species and other baseline information related to climate change, chronic pollution and other long term anthropogenic impacts. Hence, new incentive in future European marine call may seek to obtain this necessary information to be better prepared.
Moreover, there is a large gap to fill to understand the actual ecological meaning of these measurements. Clearly, a way forward is to strengthen the link between biological measurements made in individuals and ecological consequences at the community/population level. This was also one major conclusion from the EU-funded workshop organised by Cedre (see also http://www.cedre.fr/uk/publication/workshop/pollution-assessment.html).

Concerning the acceptance of the proposed methodologies into a regulatory context, there seems to be a need for a better comprehension and communication towards legislators so that a full recognition and accreditation of the validity of this approach is made. This may be achieved by a better harmonisation across the European countries to ensure quality and comparability of the data. This process could be validated during European workshop/ring test in the laboratory and in field/real situation.

9. References

- Allan, I. J., B. Vrana, et al. 2006. A “toolbox” for biological and chemical monitoring requirements for the European Union’s Water Framework Directive. Talanta 69: 302-322.

-Bagni G., Ravera M. Osella D. Mascini M. 2005. Electrochemical biosensors as a screening tool of in vitro dna-drug interaction. Current pharmaceutical analysis 1 (3): 217-224.

-Beliaeff B, Burgeot T. 2002. Integrated biomarker response: a useful tool for ecological risk assessment. Environmental Toxicology and Chemistry 21(6):1316-1322.

-Broeg K, Lehtonen KK. 2006. Indices for the assessment of environmental pollution of the Baltic Sea coasts: Integrated assessment of a multi-biomarker approach. Marine Pollution Bulletin in press.

-Chevre N., F. Gagne, C. Blaise. 2003. Development of a biomarker-based index fro assessing the ecotoxic potential of aquatic sites. Biomarkers. 8 (3-4): 287-298.

-Dagnino A, Allen IJ, Moore MN, Broeg K, Canesi L, Viarengo A. 2007. Development of an expert system for the integration of biomarker responses in mussels into an animal health index. Biomarkers 12(2):155-172.

-Depledge M.H. ; Galloway T.S. 2005. Healthy animals, healthy ecosystems. Front. Ecol. Environ. 3(5): 251-258.

-Forbes, V. E., A. Palmqvist, et al. (2006). "The use and misuse of biomarkers in ecotoxicology." Environmental Toxicology and Chemistry 25(1): 272-280.

-Hagger J. A., M. B. Jones, et al. (2006) ”Biomarkers and Integrated Environmental Risk Assessment: Are there more questions than answers?”. Integrated Environmental Assessment and Management 2(4): 312-329.

-Handy RD, Galloway TS, Depledge MH. 2003. A Proposal for the Use of Biomarkers for the Assessment of Chronic Pollution and in Regulatory Toxicology. Ecotoxicology 12:331-343.

-Kröger S and Law R.J. 2005. Sensing the sea. Trends in biotechnology 23 (5): 250-256.

-Lehtonen KK 2005 Monitoring of biological effects of contaminant in the Baltic Sea: Current status and outline for a regional monitoring strategy. Stockholm, Sweden: Project to review the HELCOM Monitoring and Assessment Programmes. HELCOM MON-PRO 2/2005.

-Moore MN, Depledge MH, Readman JW, Leonard PDR. 2004. An integrated biomarker-based strategy for ecotoxicological evaluation of risk in environmental management. Mutation Research 552:247-268.

-Narbonne, J., M. Daubeze, et al. (1999). "Scale of classification based on biochemical markers in mussels: application to pollution monitoring in European coast." Biomarkers 4(6): 415-424.

-Rogers K.R. 2007. Recent advances in biosensor techniques for environmental monitoring. Analytica Chimica Acta 568 (1-2): 222-231.

-Short, J.W. and R. Heintz. 1997. Identification of Exxon Valdez oil in sediments and tissues from Prince William Sound and the Northwestern Gulf of Alaska based on a PAH weathering model. Environ. Sci. Technol. 31: 2375-2384.

-Short, J.W. and R.A. Heintz. 2003. Normal Alkanes and the Unresolved Complex Mixture as Diagnostic Indicators of Hydrocarbon Source Contributions to Marine

Sediments of the Northern Gulf of Alaska. In Proceedings of the Twenty-sixth Arctic and Marine Oilspill Program (AMOP) Technical Seminar, pp.155-168 (June 10-12, 2003, Victoria, British Columbia; Environment Canada, Ottawa, Ontario, Canada) AMOP conference.

10. Acknowledgements

IRIS, Cedre, the University of Western Britany (UBO) and the University of Basque Country (UPV) are very grateful to the European Community funding under the 2005 Community framework for cooperation in the field of accidental or deliberate marine pollution. IRIS would like to thank Total Norge as for their industrial support in the implementation of their shared activities. UPV is grateful to the Basque Government (ETORTEK-IMPRES) and UPV/EHU (consolidated research group). The partners acknowledge the excellent sub-contracted work carried out by the University of Firenze (Unifi) on biosensors.

A special thank to Pamela Ruiz, Sjur Vingen and Emilie Sinet whom dedicated a large and significantly important part of their Master degree work to this project. We cannot list all the names of people having contributed to the completion of this work. Hence, we would like also to thank all personnel and staff from each institute for their cooperation, help and support at different levels for the scientific and administrative implementation of this project.