



**HAL**  
open science

## Mining the sequential patterns of water quality preceding the biological status of waterbodies

Corinne Grac, Xavier Dolques, Agnes Braud, Michele Tremolieres, Jean-Nicolas Beisel, Florence Le Ber

### ► To cite this version:

Corinne Grac, Xavier Dolques, Agnes Braud, Michele Tremolieres, Jean-Nicolas Beisel, et al.. Mining the sequential patterns of water quality preceding the biological status of waterbodies. *Ecological Indicators*, 2021, 130, 10.1016/j.ecolind.2021.108070 . hal-03325504

**HAL Id: hal-03325504**

**<https://hal.science/hal-03325504>**

Submitted on 22 Aug 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

1           **TITLE: Mining the sequential patterns of water quality preceding**  
2           **the biological status of waterbodies**

3

4 Authors:

5 Corinne Grac<sup>a,b</sup>, Xavier Dolques , Agnès Braud<sup>c</sup>, Michèle Trémolières<sup>b</sup>, Jean-  
6 Nicolas

7 Beisel<sup>a,b</sup>, Florence Le Ber<sup>c,d</sup>

8

9           <sup>a</sup> ENGEES, LIVE UMR 7362, F-67000 Strasbourg, France

10           [corinne.grac@engees.unistra.fr](mailto:corinne.grac@engees.unistra.fr)

11           <sup>b</sup> Université de Strasbourg, CNRS, LIVE UMR 7362, F-67000 Strasbourg, France

12           <sup>c</sup> Université de Strasbourg, ICube UMR 7357, F-67400 Illkirch-Graffenstaden,  
France

13           <sup>d</sup> ENGEES, ICube, UMR 7357, F-67000 Strasbourg, France

1           **TITLE: Mining the sequential patterns of water quality preceding the**  
2           **biological status of waterbodies**

3  
4    **Abstract**

5  
6    We have implemented a specific data mining process to explore the relationship  
7    between biological indices and physico-chemical pressures in rivers. Data were  
8    collected in the framework of the French National monitoring network set up to  
9    assess the ecological status of rivers under the European Water Framework Directive  
10   (WFD). Chemical parameters and biological indices were collected regularly from  
11   1.781 locations in metropolitan France from 2007 to 2013. The sequential pattern  
12   mining process generates closed partially ordered patterns representing a  
13   succession of physico-chemical events that precede a given biological index in a  
14   given status, validated using a subset of data. This paper focuses on the patterns  
15   and their occurrence. We showed that biological statuses depend on these temporal  
16   successions of alterations and not only on the last alterations. The physico-chemical  
17   statuses of water bodies usually appeared to be higher than their biological statuses,  
18   suggesting synergism between toxicants and/or an additive impact of other stressors  
19   related to hydromorphology or hydrology. Patterns found in the highest biological  
20   status for the biological indices based on macroinvertebrates, diatoms, macrophytes  
21   or fish, were characterised by the constancy of a high physico-chemical status over  
22   time. By contrast, before indices based on macroinvertebrates and macrophytes, two  
23   types of patterns were observed for bad biological status: (1) a chronic multi-pressure  
24   pattern, in which pressure categories such as nitrates, pesticides and other organic  
25   hydrocarbons, in moderate, poor or bad status, repeated themselves several times  
26   over time, or (2) a single occurrence of a degraded pressure category, such as one

27 moderate nitrogen, excluding nitrate, or one poor oxidizable organic matter, among  
28 other pressure categories in good status. Extracting such patterns is a promising  
29 solution both to disentangle the effects of the different stressors on water quality, and  
30 to identify the key temporal sequences among them in a context of multi-stress  
31 conditions, which is a challenge currently facing the WFD.

## 32 **Keywords**

33

34 Rivers, water quality, biological status, data mining, temporal patterns, physico-  
35 chemical status.

36

37

38

## 39 **Introduction**

40

41 In rivers and streams, several human activities produce a combination of pressures.  
42 These pressures alter the abiotic components of the ecosystem, affect the biological  
43 communities and hence their ecological status. The European Water Framework  
44 Directive (WFD) (European Council, 2000) requires the achievement of a good  
45 ecological status for the conservation or restoration of aquatic ecosystems, in the  
46 short (2021) and medium term (2027). Currently there is the urge to have access to  
47 biological tools that not only able to assess this status but also to disentangle and  
48 identify the different pressures in order to propose the appropriate restoration actions  
49 to achieve good ecological status (Feld et al., 2016a; Reyjol et al., 2014).

50 The assessment of aquatic ecosystems relies on monitoring, which generates large  
51 volumes of heterogeneous data from multiple sources (Hering et al., 2010) at  
52 different temporal scales. Data mining methods are able to analyze large datasets

53 and may be a good alternative to traditional statistical methods (Giraudel and Lek,  
54 2001). These methods can produce readable results, thereby facilitating interactions  
55 between data miners and experts (Džeroski et al., 1997). Two categories of data  
56 mining methods exist: supervised and unsupervised methods. Supervised methods  
57 require a learning dataset to build a specific model adapted to a given issue,  
58 including variables and expected results. When the model is built, it has to be tested  
59 on a second dataset. After this validation stage, it can be applied to a third dataset to  
60 predict results expected in similar conditions. Unsupervised methods do not require a  
61 learning dataset and make it possible to explore and identify unexpected rules in a  
62 dataset.

63 Several authors (Larras et al., 2017; Mondy and Usseglio-Polatera, 2013; Villeneuve  
64 et al., 2015) have already used water quality data to test the potential ability of  
65 supervised data mining methods to identify key anthropic pressures using different  
66 biological organisms. A large set of supervised learning techniques includes some  
67 artificial neural networks (D'heygere et al., 2006; Dakou et al., 2006; Everaert et al.,  
68 2016; Tsai et al., 2016) and decision trees (Džeroski, 2001; Feld et al., 2016b; Larras  
69 et al., 2017; Mondy and Usseglio-Polatera, 2013; Villeneuve et al., 2015). Methods  
70 such as Markov random fields or Bayesian networks could be used to take into  
71 account spatial and temporal dependencies (Adriaenssens et al., 2004; Forio et al.,  
72 2016; Fytilis and Rizzo, 2013; Landuyt et al., 2016; Van Looy et al., 2015). However,  
73 these off-the-shelf tools still need to be adapted to the nature of the available data as  
74 well as to the specific problem at hand. To check biological responses with respect to  
75 physico-chemical variables is necessary in order to discover relevant knowledge  
76 (Marzin et al., 2012; Oberdorff and Hughes, 1992) but remains still difficult to  
77 interpret. Temporal aspects are insufficiently taking into account. Most studies used

78 multiple physico-chemical variables measured on the same date or just before the  
79 biological measurement while responses of biological compartment to alterations are  
80 dynamic. An acute toxic pollution can alter durably biodiversity as observed in the  
81 Rhine after the Sandoz accident (Van Urk et al., 1993). Then the restoration of the  
82 former aquatic populations can require several months (Schulz and Liess, 1999).  
83 Biodiversity could also be altered progressively and durably by non-toxic alteration,  
84 such as nutrients (Pringle, 1990). The aim of this study is to take into account the  
85 successive values of physico-chemical parameters of water and to analyze their  
86 impact on biological populations, and the ecological status of waterbodies at a given  
87 date.

88

89 (Agrawal and Srikant, 1995)(Agrawal and Srikant, 1994)(Ren et al., 2009)(Sallaberry  
90 et al., 2011)(George and Binu, 2012)(Fabrègue et al., 2013)

91

92 In the present study, we used a tool generating temporal patterns with the aim of  
93 exploring the relationship between biological indices in different statuses and the  
94 succession of physico-chemical events which precede them. Indices based on four  
95 biological groups: diatoms, macrophytes, macroinvertebrates and fish were taken into  
96 account. Our questions were: 1) do patterns allow us to disentangle and identify the  
97 pressure categories implicated in a degraded status? 2) are pressure categories  
98 found in patterns specific to a biological group? 3) is there a correspondence  
99 between physico-chemical and biological statuses?

## 100 **1. Materials and methods**

### 101 **1.1 The dataset used**

102

103 Available data were collected in 1,781 sampling sites (Fig. 1) in the framework of the  
104 French network created to assess the ecological status of waterbodies, management  
105 unit, according to the WFD.



106

107 **Fig. 1: Location of the 1,781 French sampling sites for the national ecological**  
108 **assessment of rivers**

109

110 Fieldwork was undertaken by regional environmental agencies on the French  
111 national river network between 2007 and 2013. Table 1 lists the volume of the data we  
112 used. The 23,071,909 physico-chemical results concerned 1,201 parameters. Among  
113 them, 2% were major parameters (parameters with concentrations in milligram per  
114 liter, e.g. pH, nitrogen) and 98% were minor parameters, including micro-pollutants  
115 (parameters with concentrations in microgram per liter, e.g. copper, atrazine).  
116 analyzes of the major elements were conducted 12 times a year and analyzes of  
117 minor elements four or six times a year. The completeness of the physico-chemical  
118 data varied considerably over the study period, ranging from 100% to 1%. Above, we

119 used the physico-chemical results which were the most complete: they represented  
 120 1,146,544 results and concerned 189 parameters (completeness ranged between  
 121 98% and 33% for polychloro-biphenyls, PCB). The 24,593 biological results  
 122 concerned four biological groups and five French standardised biological indices  
 123 (identified below by their French acronym): macroinvertebrates (33% of results) with  
 124 IBGN (AFNOR, 2004) and I2M2 (AFNOR, 2016a, 2010), diatoms (35%) with IBD  
 125 (AFNOR, 2016b), fishes (23%) with IPR (AFNOR, 2011), macrophytes (10%) with  
 126 IBMR, (AFNOR, 2003). Calculation of I2M2 is detailed in Mondy et al. (2012). IBGN,  
 127 the former French macroinvertebrate index is expected be replaced by I2M2, but at  
 128 present, the both are still calculated. The frequency of biological sampling was once  
 129 a year for macroinvertebrates and diatoms, once every two years for fishes and  
 130 macrophytes. The overall biological results were 69% complete.

131

132

Table 1: Volume of data used

Type of data	Number
Sampling site	1,781
Water sampling for physical chemical analyzes	122,765
Biological sampling	26,072
Physical chemical parameters	1,201
Physical chemical results	23,071,909
Biological groups	5
Biological index results	24,593

133

134 We created a database specifically for these data.

135

## 136 1.2 Data preprocessing

137

138 We discretized chemical and biological data using the five levels represented by  
139 colors, which symbolize the different statuses of the WFD; i.e. blue: high, green:  
140 good, yellow: moderate, orange: poor, and red: bad.

141 For the physico-chemical data, we used two sets of thresholds: the first was created  
142 in France before the application of the WFD, hereafter referred to as SEQ (MEDD  
143 and AE, 2003) and a second one upon application of the WFD, hereafter called the  
144 WFD guide (MEEM, 2012). We performed physico-chemical discretization following  
145 a three-step process for a given date: (1) by grouping parameters in the pressure  
146 categories listed in Table 2 for SEQ and in Table 3 for the WFD guide; (2) by  
147 discretizing of each parameter in a pressure category according to the corresponding  
148 thresholds; (3) by attributing the final level of each category keeping the worst level of  
149 parameters considered, based on the one out, all out principle. With SEQ, we used  
150 192 physico-chemical parameters and with the WFD guide, 58 parameters. The  
151 thresholds of the two sets are listed in appendix 1 (for SEQ) and appendix 2 (for the  
152 WFD guide).

153 Table 2: List of physico-chemical pressure categories, their acronyms based on SEQ  
 154 (MEDD and AE, 2003) and the number of associated parameters

N°	Acronym	Pressure categories	N° of parameters
1	MOOX	Oxidizable organic matter (e.g. O <sub>2</sub> , DBO)	7
2	AZOT	Nitrogen excluding nitrate	3
3	NITR	Nitrate	1
4	PHOS	Phosphorous	2
5	EPRV	Effect of eutrophication	2
6	PAES	Suspended matter	2
7	TEMP	Temperature	1
8	ACID	Acidification parameters	2
9	MINE	Mineralisation	8
10	MPMI	Heavy metals	10
11	PEST	Pesticides	74
12	HAP	Polycyclic aromatic hydrocarbons	15
13	PCB	Polychloro-biphenyls	8
14	MPOR	Other organic hydrocarbons	57

155  
 156 Table 3 : List of physico-chemical pressure categories, their acronyms based on the  
 157 WFD guide (MEEM, 2012) and the number of associated parameters

N°	Acronym	Pressure categories	N° of parameters
1	TEMP	Temperature	1
2	ACID	Acidification parameters	1
3	BILO2	Oxygen balance	4
4	NUTRI	Nutrients	5
5	POSPE	Specific pollutants	9
6	SDP	Priority substances and priority hazardous substances	38

158  
 159 For each biological index, the thresholds from MEEM (2012) were applied. In most  
 160 cases, about two-thirds of the results correspond to the status required by the WFD  
 161 guide in class 1, high status, or in class 2, good status and one third of degraded  
 162 quality (Class 3, moderate status; class 4: poor status; class 5: bad status) except for  
 163 IPR (52% in classes 1 and 2 and 48% in classes 3, 4 and 5).

164  
 165

### 166 1.3 A pattern extracting tool named “PRESTOR”

167

168 To explore data resulting from monitoring the quality status of rivers, sequential  
169 pattern mining, an unsupervised data mining tool able to treat large volumes of  
170 data and to account for the temporal aspect of events has been used. This  
171 method first introduced by Agrawal and Srikant (1995) is a temporal extension of  
172 association rules (Agrawal and Srikant, 1994). It was then developed for and used  
173 in engineering software (Ren et al., 2009), medicine (Sallaberry et al., 2011), or  
174 marketing (George and Binu, 2012). Sequential pattern mining allows to reveal  
175 frequent temporal regularities (called sequential patterns) in a sequence  
176 database, each pattern being associated to the set of sequences that contain it.  
177 Unfortunately, the number of extracted patterns can be high and they provide  
178 redundant information, making their interpretation difficult. In order to reduce this  
179 formation redundancy and to limit the number of patterns and hence the volume  
180 of the result, Fabrègue et al. (2013) proposed to use closed partially ordered  
181 patterns (CPO-patterns). Actually, a CPO-pattern is a summary of a set of  
182 sequential patterns shared by a set of sequences (Casas-Garriga, 2005). CPO-  
183 patterns can be used in all kinds of sequential databases and have three main  
184 advantages: (1) they provide more detailed information on order among elements;  
185 (2) they are depicted by a directed acyclic graph, which is easy to understand; (3)  
186 they summarize sequential pattern sets.

187

188

189 **PRESTOR (temporal PRESSure categories patterns extracTOR)** is a tool  
190 designed for the analysis of a big temporal database. It was implemented in C++

191 allowing the best computing performances (along with intel TBB for parallel  
192 programming and BOOST and QT to handle different kinds of interfaces). PRESTOR  
193 is a command line software, most of the configuration is done in a "config.ini" file. It  
194 can fetch data directly from the database or from well-formatted files. Several  
195 configuration options are provided (described below) to select specific data subsets.  
196 The result consists in pictures and a few files providing different kinds of statistical  
197 values to help understanding the results.

198

199 **Frequent Closed Partially Ordered Patterns** The main algorithm at the core of  
200 PRESTOR generates frequent closed partially ordered patterns (CPO-patterns) as  
201 proposed by Fabrègue et al. (2013).

202 A sequential pattern is a succession of physico-chemical events that can be checked  
203 by the samples preceding a biological status assessment at a sampling site on one or  
204 more occasions. Formally, it is a sequence of itemsets. An item is a value to be  
205 recognised and an itemset is a set of values assessed at the same point in time  
206 (values from the same sample).

207 For example, having at the same time a *high* value for the alteration *PHOS* (noted  
208 *High PHOS*) and a bad value for the alteration *MPOR* (noted *Bad MPOR*) is an  
209 itemset we denote  $(High\ PHOS, Bad\ MPOR)$ .  $\langle (High\ TEMP), (High\ PHOS, Bad$   
210  $MPOR) \rangle$  is then a sequential pattern checked at all stations where the itemset  $(High$   
211  $TEMP)$  is found before the itemset  $(High\ PHOS, Bad\ MPOR)$ .

Table 4: Toy database presenting the history of 3 sampling sites

Station	Date	Value
1	2/6/2015	<i>Bad PHOS</i>
1	31/5/2015	<i>High IBGN</i>
1	29/5/2015	<i>High PHOS</i>
1	29/5/2015	<i>Bad MPOR</i>
1	28/05/2015	<i>Moderate AZOT</i>
1	27/5/2015	<i>Poor MINE</i>
1	27/5/2015	<i>High TEMP</i>
2	28/6/2015	<i>High IBGN</i>
2	27/6/2015	<i>High TEMP</i>
2	27/6/2015	<i>High PHOS</i>
2	27/6/2015	<i>Bad MPOR</i>
2	26/6/2015	<i>Moderate AZOT</i>
3	30/4/2015	<i>High IBGN</i>
3	28/4/2015	<i>Bad MPOR</i>
3	27/4/2015	<i>High PHOS</i>
3	27/4/2015	<i>High TEMP</i>
3	26/4/2015	<i>Moderate AZOT</i>
3	18/3/2014	<i>Good TEMP</i>

213

214 Let us consider the toy database shown in Table 4, presenting the history of 3 stations.

215 In this paper, we focus on the biological indices (e.g. IBGN) we want to characterize

216 using the preceding samples. To this end, we build sequences of itemsets from this

217 database following two simple rules: the sequence must end on a biological sample,

218 called the generating context, and the duration in months between the first and the

219 last samples is bounded by a value (here 6 months) defined by the operator. Three

220 sequences were extracted:

221 -S1: <(High TEMP, Poor MINE),(Moderate AZOT),(Bad MPOR, High PHOS),(High

222 IBGN)>

223 -S2: <(Moderate AZOT),(Bad MPOR, High PHOS, High TEMP),(High IBGN)>

224 -S3: <(Moderate AZOT),(High TEMP, High PHOS),(Bad MPOR),(High IBGN)>

225 The sample collected at station 1 on 2/6/2015 was not included in a sequence as it

226 did not precede a biological sample. The sample collected at station 3 on 18/3/2014

227 was not included in a sequence even though it did precede a biological sample

228 because it was too old with respect to the time-length threshold chosen in this toy  
229 example..

230 In the sequence database, only sequence S1 confirmed the sequential pattern  
231  $\langle(\textit{High TEMP}),(\textit{High PHOS}, \textit{Bad MPOR})\rangle$  (it was said that {S1} supported the  
232 pattern), even though other itemsets exist between  $(\textit{High TEMP})$  and  $(\textit{High PHOS},$   
233  $\textit{Bad MPOR})$ .

234 Here we consider the order of events, this is why sequences S2 and S3 did not  
235 confirm the sequential pattern. On S2  $\textit{High TEMP}$  occurred at the same time as  $\textit{High}$   
236  $\textit{PHOS}$  and  $\textit{Bad MPOR}$  whereas to confirm the pattern, it would need to occur before.  
237 On S3, the problem was that  $\textit{High PHOS}$  and  $\textit{Bad MPOR}$  did not occur at the same  
238 time.

239 For a set of sequences, several sequential patterns may be valid e.g.  $\langle(\textit{Moderate}$   
240  $\textit{AZOT})(\textit{High PHOS}),(\textit{High IBGN})\rangle$ ;  $\langle(\textit{Moderate AZOT})(\textit{Bad MPOR}),(\textit{High IBGN})\rangle$ ;  
241  $\langle(\textit{High TEMP}),(\textit{High IBGN})\rangle$ ;  $\langle(\textit{High PHOS}),(\textit{High IBGN})\rangle$ ;  $\langle(\textit{Moderate AZOT}),(\textit{High}$   
242  $\textit{IBGN})\rangle$ ;  $\langle(\textit{Bad MPOR}),(\textit{High IBGN})\rangle$ ;  $\langle(\textit{High IBGN})\rangle$ ;  $\langle(\textit{Moderate AZOT})(\textit{High}$   
243  $\textit{PHOS})\rangle$ ;  $\langle(\textit{Moderate AZOT})(\textit{Bad MPOR})\rangle$ ;  $\langle(\textit{High TEMP})\rangle$ ;  $\langle(\textit{High PHOS})\rangle$ ;  
244  $\langle(\textit{Moderate AZOT})\rangle$ ;  $\langle(\textit{Bad MPOR})\rangle$ ;  $\langle\rangle$  were all valid sequential patterns confirmed  
245 by sequences S1, S2 and S3.

246 However, some patterns may be redundant as the information they convey is already  
247 contained in other patterns. For example, the information in  $\langle(\textit{High PHOS}),(\textit{High}$   
248  $\textit{IBGN})\rangle$  is already conveyed by the pattern  $\langle(\textit{Moderate AZOT})(\textit{High PHOS}),(\textit{High}$   
249  $\textit{IBGN})\rangle$ . Hence, there is always a smallest set of sequential patterns conveying all  
250 the information regarding a set of sequences. A sequential pattern from this set is  
251 called a closed sequential pattern. A sequential pattern is closed with respect to a set  
252 of sequences if no other pattern conveys the same information, e.g.  $\{\langle(\textit{High}$

253 *TEMP*),(High *IBGN*)>,<(Moderate *AZOT*)(High *PHOS*),(High *IBGN*)>,<(Moderate

254 *AZOT*)(Bad *MPOR*),(High *IBGN*)>} is the set of closed sequential patterns for

255 sequences S1, S2 and S3; which means that it contains the same information as the

256 set of all sequential patterns presented earlier.

257 Closed sequential pattern sets are relevant because they represent the smallest and

258 most complete description possible of a set of sequences with sequential patterns.

259 In order to facilitate the reading of closed sequential pattern sets, Fabrègue et al.

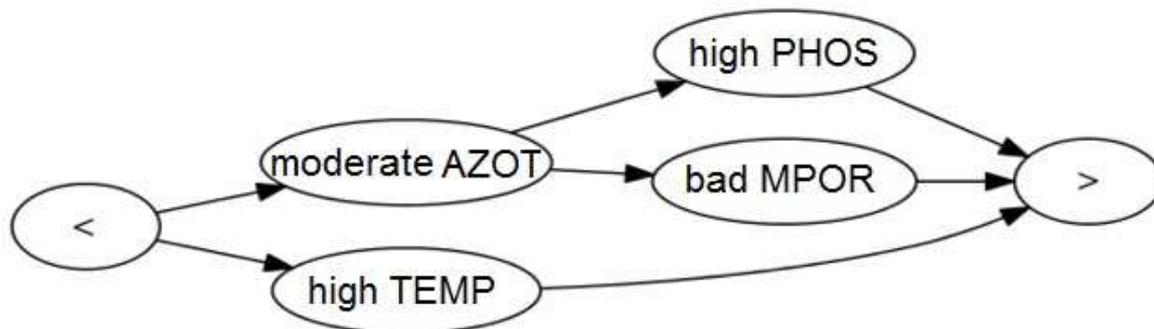
260 (2013) summarized them in a single structure named closed partially ordered

261 patterns (CPO-patterns). A CPO-pattern is a directed acyclic graph in which each

262 path is a sequential pattern. It is built in such a way that the number of nodes is

263 minimized.

264



265

266 **Fig. 2: CPO-pattern for sequences S1, S2 and S3**

267

268 The algorithm extracts the CPO-patterns (Fig. 2) only for a specific biological index in

269 a specific status, here *High IBGN*, which is the generating context. Hence we did not

270 display the generating context of the pattern but this is important information for the

271 evaluation. The CPO-pattern obtained is interpreted as follows: for each sequence

272 which confirms it, *High IBGN* is preceded by *High TEMP* and by *Moderate AZOT*

273 followed by *High PHOS* and *Bad MPOR* in any order. All the paths in the graph are  
274 checked by referring to all the supporting sequences.

275 Implementing the method proposed by Fabrègue et al. (2013), PRESTOR extracts all  
276 the frequent CPO-patterns for a given sequence database. The frequency of a  
277 pattern is the ratio of the size of its support to the size of the sequence database. In  
278 some extracts, CPO-patterns extracted in several contexts may be identical. Each  
279 has a generating context, e.g. *High IBGN*, *High IBMR*. For this set of identical CPO-  
280 patterns, the context of the CPO-pattern with the highest frequency is called the  
281 dominant context. The frequency of a pattern in its dominant context is denoted by  
282  $f_{\max}$ .

283 Hereafter, we refer to CPO-patterns simply as patterns.

284

### 285 **Choice of parameters (region, time, reference thresholds, etc.)**

286 The operator can modify several parameters to obtain different kinds of results. A  
287 filter can be applied to limit the analysis to a subset of the database restricted to a  
288 given region or a period of time. It is also possible to choose between the two quality  
289 reference norms for discretizing the physico-chemical values (SEQ or the WFD  
290 guide). The operator has to define a frequency threshold below which the patterns  
291 are not extracted. The algorithm itself can also be parameterized to consider different  
292 time lengths (in months) for the input sequences. One can also limit the values  
293 considered to remove irrelevant values. One example of an irrelevant value is *High*  
294 *TEMP* because it appears in more than 90% of the sequences in the database. This  
295 generates a significant number of parasite nodes which provide no information. Note  
296 that only the *High TEMP* is filtered but *Bad TEMP* is kept as such because this lowest

297 general value may provide some information. The choices of these parameters for an  
298 extraction specify its configuration.

299

#### 300 4. Metrics for the evaluation of patterns

301

302 In addition to the frequency (f), we propose three metrics of interest on patterns to  
303 describe the results: 1) their complexity (C) (Equation 1), 2) their scarcity (S) (Equation  
304 2), 3) their emergence (E) (Equation 3).

305

306 The complexity (C) of a pattern P indicates its relative size compared to the biggest  
307 pattern found in a given extraction. Its value range is  $[0-1]$ : it is zero when the pattern  
308 P is empty and 1 when pattern P is the biggest one.

309

$$310 \text{Equation 1} \quad C = n / N$$

311 where n is the number of items in pattern P; N is number of items found in the biggest  
312 pattern in the chosen extraction.

313

314

315 Scarcity (S) conveys the level of specificity of pattern P in its generating context. Its  
316 value range is  $[0-1]$ : it is zero when pattern P is found in all 25 contexts (5 biological  
317 indices in five statuses) and 1 when pattern P is found only in one context.

318

$$319 \text{Equation 2} \quad S = (1 - n') / 24$$

320 where n' is number of repetitions of the pattern in other extraction contexts.

321

322 The emergence (E) of pattern P, in a given extraction, is calculated only if the generating  
323 context is dominant: then its frequency is  $f_{\max}$  as defined before; we denote by  $f_{(\max-1)}$   
324 the next highest frequency of the pattern in the others contexts. The bigger E is, the  
325 more specific P is for its targeted biological value. Its value range is  $[1-\infty[$ .

326

327 **Equation 3**  $E = f_{\max} / f_{(\max-1)}$

328

## 329 **2. Results**

330

331 In our conditions (Computer INTEL COR I7-4790 3.6 Go and 16 Go RAM), extractions  
332 take from few seconds to few minutes, depending on the size of the sequence dataset  
333 and the configuration parameters. The number of patterns increases with an increase in  
334 time-length and with a decrease in the minimum frequency or in the number of  
335 sampling sites, but without constant correlations. The number of patterns obtained  
336 using the WFD thresholds guide is always higher than with SEQ thresholds in the same  
337 configuration.

338

## 339 2.1 Batch of patterns in a specific configuration

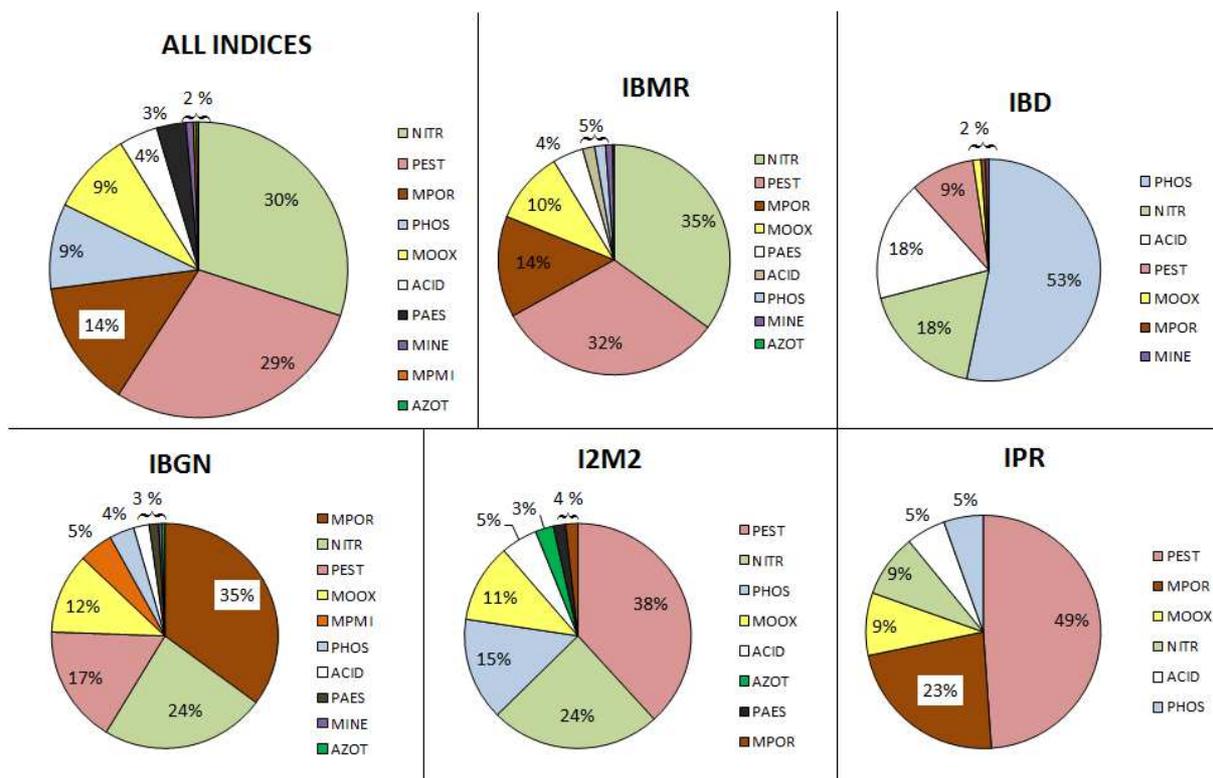
340

341 The results were obtained with the following configuration: [area: France, period: 2007-  
342 2013, table of thresholds: SEQ, time-length: 24 months, minimum frequency: 0.6]. We  
343 first eliminated 4 pressure categories for which a single status dominates over 80%:  
344 TEMP, ACID and EPRV in high status and HAP in moderate status, to avoid patterns with  
345 repetitive and no discriminating items.

346 A first batch of patterns was generated for a minimal frequency of 0.96. As 90% of the  
347 patterns obtained contained only five pressure categories in a single status, high or good  
348 (good PAES, high MINE, good MOOX, good AZOT and good PHOS), hence masking the  
349 impact of other parameters, we also eliminated these pressure categories in these  
350 statuses from the dataset. A second batch of patterns was generated without these  
351 pressure categories, for a minimum frequency of 0.6.

352

353 The distribution of the alteration categories in the different statuses found in the  
354 patterns differed from that in the dataset. The distribution of pressure categories for  
355 major parameters, was comparable: (8-9%: from 90,681 results for NITR to 102,006  
356 results for TEMP compared to the total number of physico-chemical results: 1,146,544),  
357 while nitrogen was dominant in the patterns (30% = 953/3,124). The micro-pollutants  
358 were less abundant in the dataset (from 3% -count: 36,158- for PCB to 6% -count:  
359 66,523- for PEST). Conversely PEST was abundant in the patterns (29%). The  
360 distribution of sequences per index varied from 2,416 (8%) for IBMR to 8,404 (26%) for  
361 IBD, while 67 (8% of) patterns had I2M2 as context and 433 (i.e. 53%) had IBMR as  
362 context.



364

365 Fig. 3: Overall distribution of the pressure categories in the 809 patterns obtained for  
 366 the configuration [area: France, period: 2007-2013, table of thresholds: SEQ, time-  
 367 length: 24 months, minimum frequency: 0.6] for each biological index and ranked  
 368 according to the top down percentage of appearance in patterns (Pressure categories:  
 369 NITR nitrates-; PEST – pesticides -; MPOR – other organic hydrocarbons-; PHOS –  
 370 phosphorous - ; MOOX – oxidizable organic matter -; ACID – acidification parameters –;  
 371 PAES – suspended matter; MINE– mineralistion; MPMI – heavy metals; AZOT – nitrogen  
 372 excluding nitrate).

373

374 Five main pressure categories were found in the 809 patterns (Fig. 3–All indices), in  
 375 decreasing order: nitrate (NITR) in good and moderate statuses ((present in 30% of  
 376 total patterns); pesticides (PEST) only in poor and bad statuses (present in 29% of  
 377 patterns), phosphorous (PHOS) and oxidizable organic matter (MOOX), with 9.0 % in all

378 patterns) mainly in the moderate status (present in 9% of patterns). It should be  
379 recalled that MOOX with good status was removed before the extraction because of its  
380 dominance. We observed the same distribution of four pressure categories for IBMR, but  
381 phosphorous (PHOS) disappearing and suspended matter (PAES) coming. Two main  
382 pressure categories were associated with the IBD index: phosphorous (PHOS) and  
383 acidification parameters (ACID) (. By studying patterns on the IBD index in detail, we  
384 observed that PHOS was mainly in high status and ACID only in good status – PHOS in  
385 good status and ACID in high status were removed before the extraction because of their  
386 dominance. The other organic hydrocarbons (MPOR was found in moderate status  
387 mainly in the IBGN index and in poor status in the IBMR and IPR indices. Heavy metals  
388 (MPMI) was found only in the IBGN index in poor status while PEST was the main  
389 pressure which impacts I2M2, the second index based on macroinvertebrates as IBGN.  
390 Nitrogen excluding nitrate (AZOT) () was found only in the I2M2 index in moderate  
391 status. The first pressure categories NITR, PEST, MPOR and MOOX representing at least  
392 45% of the total distribution were the same in the IBMR, IPR and IBGN index but not in  
393 the same order of importance. PHOS, NITR and ACID were the main pressures in IBD,  
394 PEST, NITR and PHOS in I2M2.

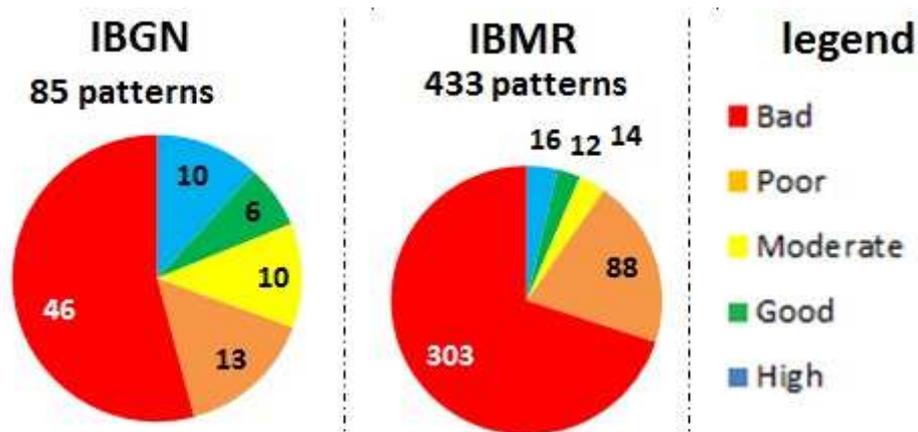
395

## 396 **2.2 Characteristic patterns of IBGN and IBMR indices**

397

398 A total of 85 characteristic patterns were obtained for IBGN (Fig. 4), (respectively 10, 6,  
399 10, 18 and 46 in biological statuses high, good, moderate, poor and bad) and 433  
400 patterns for IBMR (respectively 16, 12, 14, 88 and 303 in biological statuses high, good,  
401 moderate, poor and bad).

402



403

404 **Fig. 4** : Repartition by status classes of the patterns extracted for the two French  
 405 biological indices IBGN and IBMR obtained with the configuration: [area: France, period:  
 406 2007-2013, table of thresholds: SEQ, time-length: 24 months, minimum frequency: 0.6]

407

408 The largest number of patterns was obtained in the bad status context, which had the  
 409 fewest sequences: 54% of patterns in the context “bad IBGN” versus less than 1% of  
 410 sequences, and 70% of patterns in the context “bad IBMR” versus 1% of sequences. This  
 411 was not the case for all the indices. For instance, we obtained 27% of patterns in the  
 412 context “high IBD” versus 63% of sequences. That is why we chose to study the two  
 413 indices IBGN and IBMR in more detail.

414 In patterns, the nitrate pressure (NITR) appeared in only two statuses: good and  
 415 moderate, even when the two biological indices were in poor or bad statuses. The  
 416 pesticide pressure (PEST) appeared only in two statuses: poor and bad, even when the  
 417 two biological indices were in high, good or moderate statuses. The other organic  
 418 hydrocarbon pressure (MPOR) appeared only in two statuses: moderate and poor  
 419 respectively when IBGN was in poor or bad biological status and when IBMR was in bad  
 420 biological status. The phosphorous pressure (PHOS) appeared most frequently in high  
 421 status which corresponds to high and good statuses in the two biological indices, the

422 oxidizable organic matter pressure (MOOX) in moderate status when the two biological  
 423 indices were in moderate, poor and more particularly in bad statuses.

424 Patterns were selected and analyzed per context according to the following criteria: 1)  
 425 highest frequency (f), 2) highest emergence (E), 3) highest complexity (C), 4) highest  
 426 scarcity and 5) item diversity. Table 5 lists the first five patterns for each context except  
 427 for bad IBMR, for which we included six patterns in order to have the pressure category  
 428 AZOT at least once.

429  
 430 **Table 5: Selection of five first patterns generated for all the contexts of the French**  
 431 **biological indices IBGN and IBMR according to their frequency (f), emergence (E),**  
 432 **complexity (C) and scarcity (S) and the result (f x C x S +E); The dominant context is the**  
 433 **context in which the pattern is the most frequent; the last row indicates the items in**  
 434 **each pattern; configuration: [area: France, period: 2007-2013, table of thresholds: SEQ,**  
 435 **time-length: 24 months, minimum frequency: 0.6]; in bold: first pattern with no micro-**  
 436 **pollutants.**

Pattern	Generating Context	Dominant Context	f	E	C	S	f X C X S + E	Items
<b>229</b>	high IBGN	high IBD	<b>0.609</b>		<b>0.150</b>	<b>0.800</b>	<b>0.073</b>	<b>3 high PHOS</b>
228	high IBGN	high IBMR	0.604		0.150	0.640	0.058	3 poor PEST
221	high IBGN	high IPR	0.660		0.050	0.600	0.020	1 high MOOX
223	high IBGN	bad IBMR	0.619		0.050	0.160	0.005	1 bad PEST
220	high IBGN	bad IBMR	0.651		0.050	0.080	0.003	1 good NITR
<b>307</b>	<b>good IBGN</b>	<b>bad IBMR</b>	<b>0.637</b>		<b>0.100</b>	<b>0.360</b>	<b>0.023</b>	<b>2 good NITR</b>
303	good IBGN	high IBD	0.636		0.050	0.560	0.018	1 high PHOS
308	good IBGN	bad IBMR	0.624		0.100	0.240	0.015	2 bad PEST
305	good IBGN	bad IBMR	0.719		0.050	0.080	0.003	1 good NITR
304	good IBGN	bad IPR	0.738		0.050	0.000	0.000	1 good ACID

240	moderate IBGN	bad IBMR	0.629		0.200	0.640	0.080	4 poor PEST
239	moderate IBGN	bad IBMR	0.671		0.150	0.480	0.048	3 bad PEST
<b>236</b>	<b>moderate IBGN</b>	<b>poor IBMR</b>	<b>0.633</b>		<b>0.100</b>	<b>0.560</b>	<b>0.035</b>	<b>2 moderate NITR</b>
237	moderate IBGN	bad IBMR	0.605		0.100	0.360	0.023	2 good NITR
232	moderate IBGN	bad IBGN	0.675		0.050	0.440	0.015	1 moderate MOOX
254	poor IBGN	high IBMR	0.617		0.150	0.640	0.059	3 poor PEST
251	poor IBGN	poor IBMR	0.613		0.100	0.960	0.059	2 moderate MPOR
<b>253</b>	<b>poor IBGN</b>	<b>bad I2M2</b>	<b>0.613</b>		<b>0.100</b>	<b>0.560</b>	<b>0.034</b>	<b>2 moderate NITR</b>
242	poor IBGN	bad IBMR	0.613		0.050	0.920	0.028	1 moderate AZOT
244	poor IBGN	high IBMR	0.637		0.050	0.760	0.024	1 moderate PHOS
300	bad IBGN	bad IBGN	0.600	2.100	0.400	0.960	2.330	5 moderate MPOR, 1 moderate NITR, 1 moderate MOOX
266	bad IBGN	bad IBGN	0.600	1.719	0.250	0.960	1.863	4 poor MPMI & 1 moderate NITR
270	bad IBGN	bad IBGN	0.600	1.260	0.200	0.960	1.375	3 poor PEST & 1 moderate MOOX
289	<b>bad IBGN</b>	<b>bad IBGN</b>	0.600	1.273	0.100	0.960	1.331	<b>1 poor MOOX &amp; 1 moderate NITR</b>
259	bad IBGN	bad IBGN	0.667	1.215	0.150	0.960	1.311	3 bad PEST
324	high IBMR	high IBMR	0.637	1.058	0.200	0.880	1.170	4 poor PEST
<b>325</b>	<b>high IBMR</b>	<b>high IBD</b>	<b>0.612</b>		<b>0.200</b>	<b>0.880</b>	<b>0.108</b>	<b>4 high PHOS</b>
323	high IBMR	bad IBMR	0.602		0.150	0.760	0.069	3 good NITR
316	high IBMR	high IPR	0.619		0.100	0.880	0.054	2 high MOOX
320	high IBMR	bad IBMR	0.604		0.100	0.240	0.015	4 bad PEST
745	good IBMR	high IBMR	0.625		0.150	0.640	0.060	3 poor PEST
746	good IBMR	bad IBMR	0.621		0.150	0.480	0.045	3 bad PEST

<b>738</b>	<b>good IBMR</b>	<b>high IPR</b>	<b>0.667</b>		<b>0.050</b>	<b>0.600</b>	<b>0.020</b>	<b>1 high MOOX</b>
739	good IBMR	high IBD	0.643		0.050	0.560	0.018	1 high PHOS
735	good IBMR	bad IBMR	0.613		0.050	0.440	0.014	1 moderate MOOX
340	moderate IBMR	bad IBMR	0.6478		0.200	0.640	0.083	4 bad PEST
<b>338</b>	<b>moderate IBMR</b>	<b>poor IBMR</b>	<b>0.608</b>		<b>0.150</b>	<b>0.800</b>	<b>0.073</b>	<b>3 moderate NITR</b>
334	moderate IBMR	bad IBMR	0.637		0.100	0.480	0.031	3 poor PEST
335	moderate IBMR	high IBMR	0.659		0.100	0.360	0.024	2 good NITR
328	moderate IBMR	bad IBMR	0.704		0.050	0.440	0.016	1 moderate MOOX
410	poor IBMR	poor IBMR	0.604	1.228	0.300	0.960	1.402	4 bad PEST, 2 moderate NITR
<b>363</b>	<b>poor IBMR</b>	<b>poor IBMR</b>	<b>0.6923</b>	<b>1.119</b>	<b>0.150</b>	<b>0.960</b>	<b>1.219</b>	<b>2 moderate NITR, 1 moderate MOOX</b>
360	poor IBMR	poor IBMR	0.625		0.150	0.760	0.071	3 good NITR
421	poor IBMR	poor IBMR	0.604		0.100	0.960	0.058	1 moderate NITR, 1 moderate PAES
419	poor IBMR	poor IBMR	0.604		0.100	0.920	0.056	1 moderate NITR, 1 good ACID
623	bad IBMR	bad IBMR	0.619	4.647	0.500	0.960	4.944	2 bad PEST, 4 good NITR, 2 moderate MOOX, 2 moderate PAES
689	bad IBMR	bad IBMR	0.619	3.714	0.950	0.960	4.279	9 bad PEST, 4 good NITR, 3 moderate MOOX, 3 poor MPOR
<b>564</b>	<b>bad IBMR</b>	<b>bad IBMR</b>	<b>0.619</b>	<b>3.714</b>	<b>0.350</b>	<b>0.960</b>	<b>3.922</b>	<b>3 good NITR, 1 moderate MOOX, 3 moderate PAES</b>
674	bad IBMR	bad IBMR	0.619	3.200	0.600	0.960	3.557	3 bad PEST, 4 good NITR, 2 moderate NITR, 3 poor MPOR
494	bad IBMR	bad IBMR	0.619	2.758	0.500	0.960	3.055	5 good NITR, 4 moderate MOOX, 1 good MINE

442	bad IBMR	bad IBMR	0.61905	2.207	0.250	0.960	2.356	1 moderate AZOT, 1 good MINE, 3 good NITR
-----	-------------	-------------	---------	-------	-------	-------	-------	---

437

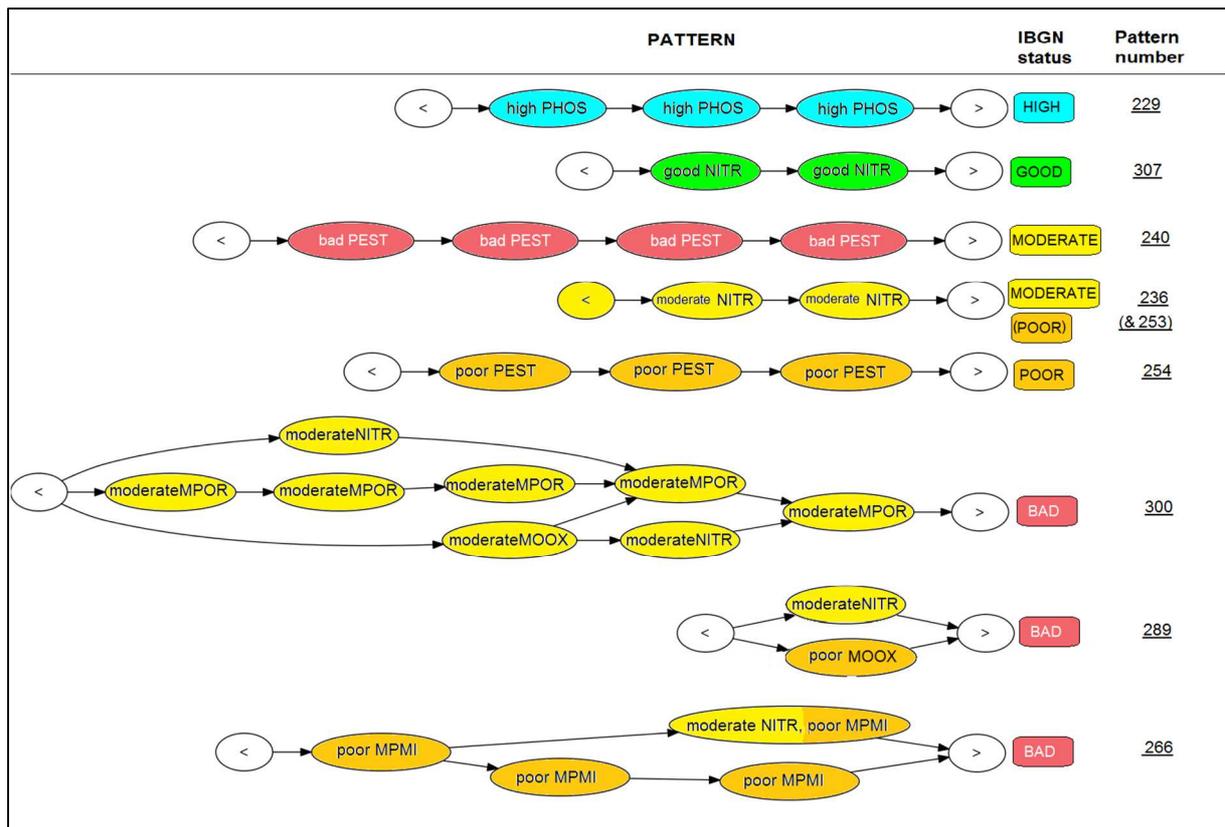
438 Among the 50 selected patterns, 14 were emerging patterns (they have a E value, the  
439 highest values were for bad IBMR) and they belong to only four contexts: IBMR in high,  
440 poor and bad statuses and IBGN in bad status. There were eight situations where the  
441 generating context is the dominant context, the four previously cited and four more  
442 contexts based on other indices than IBGN and IBMR (high IBD, high IPR, bad I2M2 and  
443 bad IPR, not shown in the table). The other generating contexts, based on IBGN and  
444 IBMR indices, were never dominant for any given pattern. Only the patterns of four  
445 generating contexts had a complexity higher than 0.25 (in this configuration, the biggest  
446 pattern had 20 items and  $C=0.25$  corresponded to a pattern with five items): the  
447 patterns for IBMR in high, poor and bad statuses and IBGN in bad status. Patterns for  
448 only four contexts had a scarcity less than 0.76 (in this configuration,  $S=0.76$  was  
449 obtained for a pattern found in more than six contexts): the patterns for IBMR and for  
450 IBGN in good and moderate statuses.

451 In each context, we performed detailed analysis of the first pattern according to the  
452 selected classification, and the first one with macro-pollutants.

453 The first pattern for high IBGN (pattern 229, Fig. 5) and the second for high IBMR  
454 (pattern 325, Fig. 6) showed that the pressure PHOS remained stable in high status,  
455 during respectively three and four successive measurements. Pattern 738 (Fig. 6),  
456 obtained for the generating context high IBMR, was the only one with another pressure  
457 category in high status, i.e. one item with high MOOX. The stability of pressure PEST in  
458 poor or bad statuses, measured successively three times (for example patterns 228 for  
459 high IBGN and 254 for poor IBGN in Table 5 and Fig. 5) or four times (for example pattern

460 340 for moderate IBMR in Fig. 6), was not discriminating with respect to a specific  
461 context. It appeared in high to bas biological context. The pressure category “good  
462 NITR” did not appear to be discriminating; as whatever the context, it appeared: once in  
463 high and good IBGN (patterns 220 and 305, Table 5), twice in good, moderate IBGN  
464 (patterns 307, Fig. 5 and pattern 237, Table 5) and moderate IBMR (pattern 334, Table 5),  
465 three times in moderate IBGN and poor IBMR (patterns 323 and 360, Table 5). Moderate  
466 NITR appeared more specifically in poor and bad contexts of both indices (patterns 236,  
467 253, Table 5, and 338, Table 5 and Fig. 6). Pattern 289 (Fig. 5) was specific to bad IBGN  
468 (E= 1.273 and S=0.960): meant that only one poor MOOX was measured before or after  
469 one moderate NITR. Pattern 300 (Fig. 5) was also specific to bad IBGN (E= 2.100 and  
470 S=0.960): meant that pollution with moderate MPOR measured five times remained  
471 stable, and two moderate NITR and one moderate MOOX were measured after or before  
472 one these measurements. In pattern 442 (Fig. 6), only one measurement of moderate  
473 AZOT between one good MINE and three good NITR were observed before a bad IBMR.  
474 Pressure MPMI (heavy metals) appeared only in bad IBGN contexts, e.g. in pattern 266  
475 (Fig. 5) composed of four poor MPMI and one moderate NITR. In the bad IBMR context,  
476 patterns 623 and 564 (Fig. 6) represented multi-stress conditions: moderate MOOX, and  
477 moderate PAES for the first, with added bad PEST for the second.

478



479

480 Fig. 5: Nine major patterns extracted from the five contexts of IBGN in the configuration:

481 [area: France, period: 2007-2013, table of thresholds: SEQ, time-length: 24 months,

482 minimum frequency: 0.6]

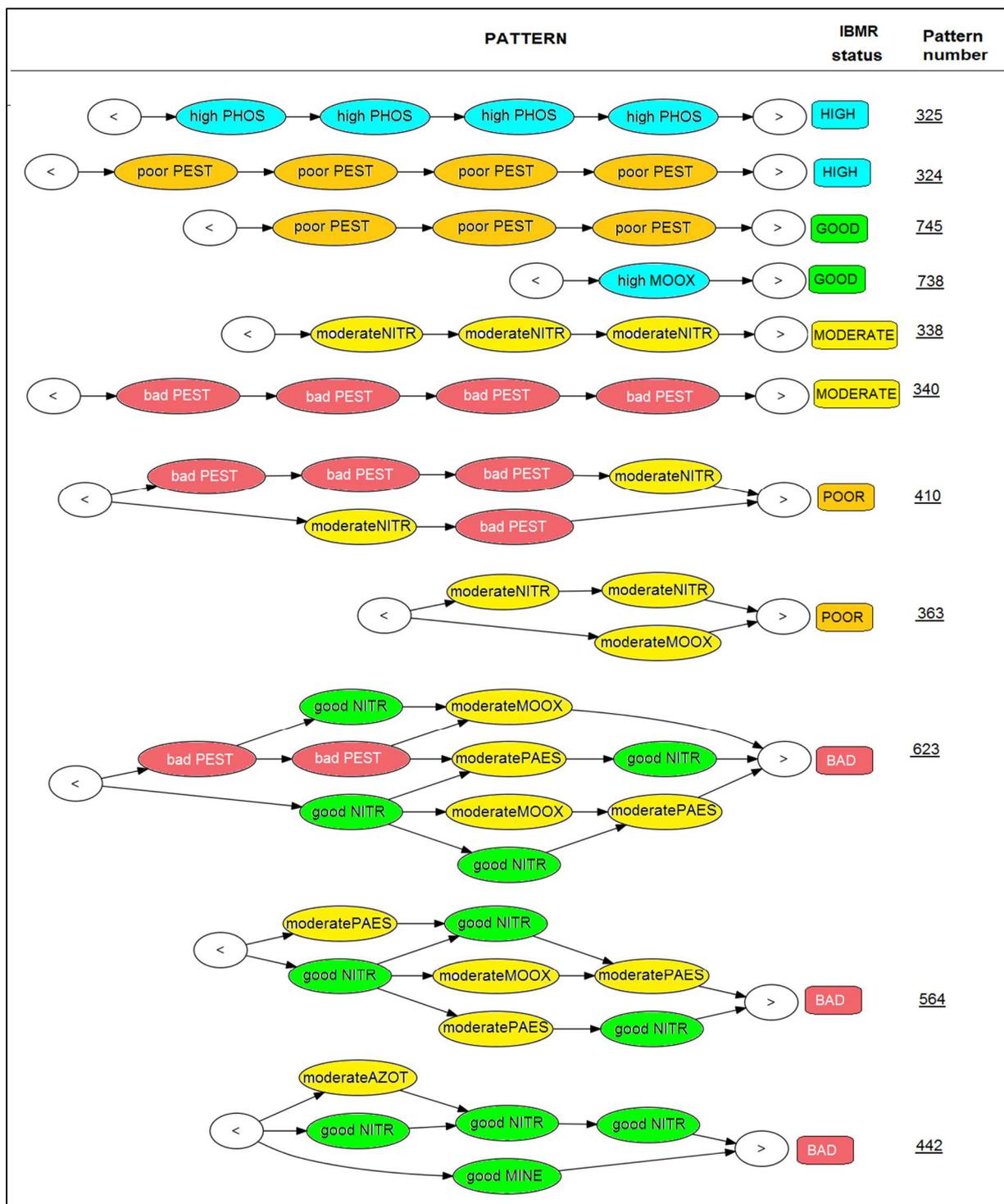
483 (Pressure categories: PHOS – phosphorous - ; NITR nitrates-; PEST – pesticides -; MPOR

484 – other organic hydrocarbons-; MOOX – oxidizable organic matter -; MPMI – heavy

485 metals).

486

487



488

489

490

491

492

Fig. 6: Eleven major patterns extracted from the five contexts of IBMR in the configuration: [area: France, period: 2007-2013, table of thresholds: SEQ, time-length: 24 months, minimum frequency: 0.6] (Pressure categories: PHOS – phosphorous - ; PEST – pesticides -; MOOX – oxidizable organic matter -; MPOR – other organic hydrocarbons-

493 ; NITR nitrates-; PAES – suspended matter; ACID – acidification parameters –;MINE–  
494 mineralistion; AZOT – nitrogen excluding nitrate).

495

496

### 3. Discussion

497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521

#### **A promising tool to reveal temporal sequences preceding a biological status**

Fabrègue et al. (2014) were the first to show the promising potential of patterns to describe sequential data concerning quality in river ecosystems. We continued this work using a bigger dataset (here 1,146,544 physico-chemical data and 24,593 biological data) and by multiplying the possible configurations. The PRESTOR tool we created was specifically applied to extract patterns based on different criteria (time-length, minimum frequency, etc.) directly from hydro-ecosystem databases. Whereas extraction is stable and rapid, many tests have to be conducted to tune the parameters, and, like in multivariate analysis, the dominant pressure categories had to be removed in order to analyze all the data in detail using a top-down approach. As shown by Serrano Balderas et al. (2017), selecting features for data reduction is an essential data preprocessing technique to ensure the final results are valid.

To proceed from a potentially useful to an operational method using PRESTOR, we proposed a double process before and after extraction to control the number of patterns extracted and to identify the most relevant ones. Analysing of the 809 patterns we extracted is irrelevant for an expert. Some are nested in others, for example, a pattern of three successive “poor PESTs” included in another one of four successive “poor PESTs” (Fig. 6). After extraction, the three metrics (complexity (C), scarcity (S) and emergence (E)) help selecting the most relevant patterns among hundreds or even thousand ones.

The process generated more patterns for the smaller sub-datasets. This bias is shared by other methods as well as by data mining and statistical methods, which is why the pre

522 and post-processing are recommended particularly for environmental data mining  
523 (Gibert et al., 2008). Here, there can be many patterns in a given context for different  
524 reasons: 1) few sequences were available: in our case, IBMR represented 8% of  
525 sequences and extracted 53% of patterns, because the threshold of minimum frequency  
526 is easily reached; 2) data covered all the range of values: dataset for IBMR were more  
527 heterogeneous than those for IBD, which represented the maximum number of  
528 sequences (27%) and extracted only 18% of patterns; 3) data were numerous but  
529 homogeneous, as were data for high IBD in the dataset. This problem was amplified by  
530 unbalanced data distribution. The advantage of this dataset is that it is the complete set  
531 of French data from a survey network covering six years, but we nevertheless had to  
532 deal with the real distribution of these data: more major parameters than minor  
533 pollutants, a different number of sequences available for each index and fewer  
534 sequences for bad ecological status.

535 Selecting the same number of the most relevant patterns for all biological indices and  
536 statuses with a given status, thanks to the three metrics, limited this bias. Before  
537 focusing on the most relevant patterns, to be sure the method is efficient, we studied the  
538 complete batch of patterns extracted for a time-length of two years.

539

#### 540 **Feedback on the pressure categories analyzed and the quality of data**

541 Six types of pressure categories have been identified: the first one resulted from the  
542 analysis of the original data and the five others from the analysis of the extracted  
543 patterns.

544 The first type concerned several pressure categories considered to remain stable in a  
545 single status throughout the French territory, i.e. temperature (TEMP), acidification  
546 parameters (ACID) and effect of eutrophication (EPRV) in high status and polycyclic

547 aromatic hydrocarbons (HAP) in moderate status. For temperature, in the SEQ (op. cit)  
548 or in the WFD guide (op.cit.), the quality class is defined by higher temperatures and the  
549 temperature difference between high and good class thresholds is 1.5 °C. This difference  
550 may correspond to the 1.4 °C to 1.7 °C increase observed by Durance and Ormerod  
551 (2007) in streams in Wales between 1981 and 2005 or to the forecasts of 1- 9 °C  
552 announced by Webb (1996) for 2050. But the mechanisms which control freshwater  
553 temperature behavior in the context of global climate change are complex, in particular  
554 because temperature increases have been observed in winter while summer  
555 temperatures have remained relatively stable (Webb et al., 2008). This could explain  
556 why, in the French dataset, temperatures remained stable in high status, and underlines  
557 the need to change the way we define this pressure category. In Europe and North  
558 America, acidification of freshwater decreased throughout the 1980s and 1990s thanks  
559 to the reduction of atmospheric emissions of sulphur dioxide (Stoddard et al., 1999).  
560 Therefore, pressure due to acidification is now localized rather than widespread, which  
561 is probably why it is difficult to observe this pressure in a study conducted at national  
562 scale. The quality class for the effect of eutrophication (EPRV), in the SEQ (op. cit), is  
563 generally defined for water combining basic pH and oversaturated in oxygen. In our  
564 dataset, this pressure category was mainly in high quality based on this definition, which  
565 is why it was not included in the analysis of patterns. Again, HAP are ubiquitous  
566 contaminants to a moderate degree in our study area, as also shown by Motelay-Massei  
567 et al. (2004) in the soils of the Seine River Basin.

568 The five following type of pressure categories are selected according to the patterns  
569 extracted from the complete dataset with a time length of two years.

570 The second type pressure categories was measured in different statuses but considered  
571 as dominant in only one status. There were macro-pollutants in high or good quality

572 statuses: mineralisation (MINE) in high status, suspended matter (MES), oxidisable  
573 organic matter (MOOX), nitrogen excluding nitrate (AZOT) and phosphorus matter  
574 (PHOS) in good status. They were removed from the pattern extractions, but they can be  
575 considered as regularly measured in each site.

576 The third type includes pressure categories that were absent or rare in the patterns.  
577 Polychloro-biphenyls (PCB) did not appear in any patterns. This pressure category was  
578 the one with the lowest number of measurements in the dataset (completeness rate:  
579 33% versus at least or more than 50% for the other micro-pollutants). Teil et al. (2004)  
580 estimated that the pressure exerted by PCB in the Seine River Basin to be beyond the  
581 global national estimation of industrial inputs, but we did not observe it at the national  
582 scale. Pressure categories ACID and MINE in other classes than the high one accounted  
583 for less than 4% of all appearances of pressures.

584 The fourth type represents pressure categories that were dominant in patterns but  
585 poorly discriminant, regardless of the biological status: pesticides (PEST) was only  
586 found in poor or bad statuses in the patterns we extracted. This pressure category was  
587 the one in which the number of measurements of micro-pollutants was the highest in  
588 the dataset (completeness rate: 61%). According to the Millenium Ecosystem  
589 Assessment Programme (2005), pesticides are one of the major stressors in freshwater  
590 ecosystems. The patterns extracted revealed the spatial and temporal dominance of this  
591 pressure category in the French dataset. After nitrates, it was the second most dominant  
592 category in all the patterns extracted, but was found in all biological statuses from high  
593 to bad. There was no graduation either in status or in the number of appearances in the  
594 patterns extracted for the IBGN index. Conversely, there was a small gradation in status  
595 and in the number of appearance in patterns extracted for the IBMR index: poor PEST  
596 appeared only in patterns extracted for high to moderate IBMR, whereas patterns

597 extracted for poor and bad IBMR contained only bad PEST. The proportion of herbicides  
598 (45%) versus insecticides (31%) in the parameters included in PEST in SEQ (op.cit.)  
599 could partly explain the small graduation observed in patterns extracted for IBMR. Even  
600 if macroinvertebrates are known to be the most sensitive groups to pesticides in  
601 freshwater communities (Schäfer et al., 2012), we observed different responses of the  
602 two indices we tested; IBGN sensitivity was higher than that of I2M2 (Mondy et al.,  
603 2012).

604 The fifth type grouped distinguishing pressure categories for a given biological status in  
605 all biological indices: nitrates (NITR) in good and moderate status, phosphorus (PHOS)  
606 in high and moderate status, oxidizable organic matter (MOOX) in high, moderate and  
607 poor status, and other organic hydrocarbons (MPOR) in moderate and poor status. Like  
608 in recent studies dealing with the effects of multiple stressors on freshwater  
609 biodiversity, we found nutrients among the parameters that have the most impact  
610 (Lemm and Feld, 2017; Stendera et al., 2012) along with organic matter (Comte et al.,  
611 2010; Villeneuve et al., 2015). We are the first to underline the importance of other  
612 organic hydrocarbons (MPOR) in patterns especially for IBGN, IPR and IBMR contexts.

613 The sixth type is heavy metals (MPMI) in poor status. This pressure category is specific  
614 to patterns extracted for IBGN index in bad status. It was missing in patterns associated  
615 to I2M2 the other French index based on macroinvertebrates whereas Mondy et al.  
616 (2012) found a better correlation between this stressor and I2M2 than IBGN.

617

#### 618 **A specific biological response to a specific pressure category?**

619 Except for heavy metals (MPMI), none of the pressure categories was specific to one  
620 biological index in the patterns extracted from the French dataset we used. Many  
621 authors who used large national indices and several biological indices reported that all

622 biological groups responded firstly to macro-pollutants in general (Dahm et al., 2013;  
623 Marzin et al., 2012), particularly to organic matter and nutrients (Villeneuve et al.,  
624 2015). Haury et al. (2006) showed that the IBMR was sensitive to trophic disruption but  
625 also to heavy organic pollution. Nutrients did not appear to be specific to patterns  
626 extracted for IBMR or IBD indices. However high and good IBMRs (generating contexts)  
627 were associated with High PHOS and MOOX or good NITR. PHOS was also the dominant  
628 pressure category for patterns extracted for IBD index (dominant context).  
629 Nevertheless, excess nutrients are known to have indirect effects on aquatic organisms,  
630 especially on macroinvertebrates (Dolédec et al., 2006; Lemm and Feld, 2017). The  
631 pressure acidification (ACID) was well represented (18%) in the patterns extracted for  
632 the IBD index, whereas Larras et al. (2017) found a weak correlation between this  
633 pressure and diatoms, but their approach using life history traits differs substantially  
634 from our approach. Fishes are known to have lower responses to physico-chemical  
635 pressures (Dahm et al., 2013; Marzin et al., 2012; Villeneuve et al., 2015). Indeed, the  
636 patterns extracted for IPR contained the lowest number of pressures (6/10). Few  
637 authors who used large datasets tested biological responses to all micro-pollutant  
638 categories, except Larras et al. (2017) and Mondy and Usseglio-Polatera (2013)  
639 respectively for diatoms and macroinvertebrates using the same French dataset. These  
640 authors found good biological responses for pesticides but poor responses for polycyclic  
641 aromatic hydrocarbons (PAH), heavy metals (MPMI) and other organic hydrocarbons  
642 (MPOR), whereas the last pressure category was widely represented in the patterns we  
643 extracted in our study.

644

645 **Gap between physico-chemical statuses and biological statuses**

646 In the differentiating pressure categories, physico-chemical statuses were often better  
647 distinguished than the biological statuses in the patterns. Oberdorff and Hughes (1992)  
648 already observed a gap between the precursor of the IPR index and SEQ results in the  
649 Seine catchment. Except for suspended matter (PAES) and pesticides (PEST) found in  
650 bad status in some patterns, the worst status found for other pressure categories was  
651 poor status. Defining class boundaries is a critical step in the design of methods of  
652 assessment (Birk et al., 2012). Using new French thresholds (MEEM, 2012) to discretize  
653 physico-chemical data, we obtained less significant results. This can be partly explained  
654 by the fact that these pressure categories include more parameters, but also by the  
655 change in the boundary between the good and moderate threshold for nitrates in the  
656 nutrient pressure category, which increased from 10 mg/L in SEQ (MEDD and AE, 2003)  
657 to 50 mg/L in the WFD guide (MEEM, 2012). This is at least the second time that this  
658 threshold has been revised upwards: 50 years ago, Nisbet and Verneaux (1970) set the  
659 boundary for excess nitrates in French freshwater at 3 mg/L. Although good-bad  
660 thresholds for micro-pollutants were discussed at the European level for the WFD guide,  
661 there was no harmonization of the thresholds for macro-pollutants, and the literature  
662 comparing them in Europe is really poor. Current French thresholds are among the  
663 highest in Europe probably because the majority of countries take the mean or the  
664 median as the annual value while France takes the 90<sup>th</sup> percentile. The French good-  
665 moderate thresholds compared to the European ranges are 1) for NH<sub>4</sub>: 0.5 mg/L in the  
666 range [0.05-1.6 mg/L] (Claussen et al., 2012), 2) for O<sub>2</sub>: 6 mg/l in the range [6-10 mg/L] -  
667 but the higher values are used by countries specifying different thresholds for different  
668 stream typologies (based on size, climate, geology, geographical location), which is not  
669 the case in France - (Claussen et al., 2012), 3) for orthophosphates: 0.5 mg/L in the  
670 range [0.05-1 mg/L] (Arle et al., 2016), 4) for nitrogen adding the French good-moderate

671 thresholds for Kjeldahl nitrogen, nitrites and nitrates, the boundary is 4.5 mg/L for SEQ  
672 (MEDD and AE, 2003) and 14 mg/L for WFD guide (MEEM, 2012) versus the European  
673 range [0.7-10 mg/L] (Claussen et al., 2012).

674

675 The stronger reactions of living organisms evidenced by worse states than their physico-  
676 chemical status, can also be explained by possible synergism between stressors and/or  
677 an additive impact of other stressors related to hydromorphology or hydrology.  
678 According to Lemm and Feld (2017), combined nutrient and hydromorphological stress  
679 can strengthen an individual's reactions to each single stressor for different biological  
680 traits.

681 Extreme ecological status classes are the easiest to define (Birk et al., 2012). This is  
682 surely why in the configuration we chose, or in other configurations, intermediate  
683 biological statuses i.e. good and moderate, never displayed specific patterns. Conversely,  
684 extreme statuses, high and bad, often displayed specific patterns. For the poor status,  
685 there were some specific patterns for IBMR, but not for IBGN.

686

### 687 **Characteristic physico-chemical successions for high and bad biological statuses**

688

689 Patterns of high biological status were characterized by consistently high physico-  
690 chemical status especially for phosphorus (PHOS) measured three or four times and  
691 probably in the consistency of the pressure categories we had to remove before the  
692 extraction: mineralisation (MINE) in high status, suspended matter (MES), oxidizable  
693 organic matter (MOOX), nitrogen excluding nitrate (AZOT).

694 We found two types of patterns in bad biological status: (1) a chronic multi-pressure  
695 one, in which pressure categories such as nitrates, pesticides and other organic

696 hydrocarbons, in moderate, poor or bad status, repeated themselves several times over  
697 time, or (2) a single occurrence of a degraded pressure category, such as one moderate  
698 nitrogen excluding nitrate, or one poor oxidizable organic matter, among other pressure  
699 categories in good status (respectively pattern 289, Fig. 5 and pattern 442, Fig. 6).

700

701 Studies of the change effect in pressure categories over time are scarce or limited to  
702 general trends in the case of degradation such as the hydraulic management of large  
703 rivers (e.g. Fruget et al., 2001; Trémolières, 1994) or more recently in the case of  
704 restoration (e.g. Meyer et al., 2013; Staentzel et al., 2017). However databases storing  
705 temporal monitoring data on rivers do exist. Most studies used data on each site on each  
706 sampling occasion as not accounting for time information to increase the gradient of  
707 measurements before timeless treatments, because the methods used were not able to  
708 incorporate the temporal dimension (D'heygere et al., 2006; Dahm et al., 2013; Larras et  
709 al., 2017; Marzin et al., 2012; Mondy and Usseglio-Polatera, 2013; Villeneuve et al.,  
710 2018). This is why the temporal patterns extracted by PRESTOR are innovative and  
711 offer new opportunities to explore data on rivers. Here the life span of the different  
712 biological compartment studied were not taken into account. To go further, we will  
713 search the specific time length of sequences to consider with respect to biological  
714 indices to obtain the most relevant patterns. Moreover hydrological and  
715 hydromorphological pressures will be added in the dataset to produce richer patterns  
716 that will deepen the analysis.

717

718

## 719 **Conclusion**

720 Our questions were: 1) do patterns allow us to disentangle and identify the pressure  
721 categories implicated in a degraded status? 2) are pressure categories found in patterns  
722 specific to a biological group? 3) is there a correspondence between physico-chemical  
723 and biological statutes?

724 Applying a new data mining method for river ecological assessment required iterative  
725 collaboration between computer scientists and experts in the domain to (1) adapt the  
726 method to a particular question and to the specific format of the data, (2) analyze the  
727 complete results and not only those that “matched well” to ensure the methods are  
728 efficient, (3) to propose selection criteria for the evaluation of patterns, (4) to select the  
729 most important results particularly for data mining methods which have led to  
730 exponential increase in results. **PRESTOR (temporal PRESSure categories patterns  
731 extractOR)** was implemented specifically to extract patterns from a hydro-ecosystem  
732 database. The operator chooses several criteria such as the time-length of patterns or  
733 their minimum frequency. To check the efficiency of the method, the 809 patterns  
734 extracted from data collected all over France with a time-length of 24 months, were  
735 analyzed. Three metrics, complexity, scarcity and emergence, were used to select  
736 significant patterns according to their ecological status.

737 Water managers need simple and transparent methods. Selected patterns extracted by  
738 PRESTOR are easily readable and match managers’ needs. They could be used at large  
739 scale and as such, considered as a holistic approach, the kind of approach that is  
740 urgently needed (Demars et al., 2012; Stendera et al., 2012). In a multi-pressure  
741 environment (Reyjol et al., 2014), hydromorphological and physico-chemical pressures  
742 should be analyzed together: patterns including various time scales have to build

743 therefore. In this study, patterns were extracted over a large territory, here France, but  
744 the method can be applied in the same way to a small territory, climatically and  
745 geologically homogeneous, in order to extract more specific patterns. Finally, patterns,  
746 highlighting a sequence of alteration events before an observed biological response, are  
747 a possible solution to disentangle the effects of different pressures.

748

## 749 **Acknowledgments**

750

751 This work was funded by the French National Agency of Biodiversity (research support  
752 agreement n°460, December 23rd, 2014).

753

## 754 **References**

755

756 Adriaenssens, V., Goethals, P.L.M., Charles, J., De Pauw, N., 2004. Application of Bayesian  
757 Belief Networks for the prediction of macroinvertebrate taxa in rivers. *Ann. Limnol.*  
758 40, 181–191. <https://doi.org/10.1051/limn/2004016>

759 AFNOR, 2016a. Qualité de l'eau: prélèvements des macro-invertébrés aquatiques en  
760 rivières peu profondes. Norme Française NF T 90-333.

761 AFNOR, 2016b. Qualité de l'eau: détermination de l'Indice Biologique Diatomées (IBD).  
762 Norme Française NF T90-354.

763 AFNOR, 2011. Qualité de l'eau: détermination de l'Indice poissons rivière (IPR). Norme  
764 Française NF T90-344.

765 AFNOR, 2010. Qualité de l'eau: traitement au laboratoire d'échantillons contenant des  
766 macro-invertébrés de cours d'eau. Norme Française Expérimentale XP T 90-388.

767 AFNOR, 2004. Qualité de l'eau: détermination de l'Indice Biologique Global Normalisé  
768 (IBGN); Norme Française NF T90-350.

769 AFNOR, 2003. Qualité de l'eau: détermination de l'Indice Biologique Macrophytique en  
770 Rivière (IBMR). Norme Française NF T90-395.

771 Agrawal, R., Srikant, R., 1995. Mining sequential patterns, in: Proceedings of the Eleventh  
772 International Conference on Data Engineering. IEEE Comput. Soc. Press, pp. 3–14.  
773 <https://doi.org/10.1109/ICDE.1995.380415>

774 Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules, in: Proc. 20th  
775 IVLDB Conference.

776 Arle, J., Mohaupt, V., Kirst, I., 2016. Monitoring of Surface Waters in Germany under the  
777 Water Framework Directive — A Review of approaches, methods and results.  
778 Water 8, 217. <https://doi.org/10.3390/w8060217>

779 Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund,  
780 W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface  
781 waters: An almost complete overview of biological methods to implement the  
782 Water Framework Directive. Ecol. Indic. 18, 31–41.  
783 <https://doi.org/10.1016/j.ecolind.2011.10.009>

784 Casas-Garriga, G., 2005. Summarizing Sequential Data with Closed Partial Orders, in:  
785 Proceedings of the 2005 SIAM International Conference on Data Mining. pp. 380–  
786 391. <https://doi.org/10.1137/1.9781611972757.34>

787 Claussen, U., Müller, P., Arle, J., 2012. WFD CIS ECOSTAT WG A Report “ Comparison of  
788 envi- ronmental quality Objectives , threshold values or water quality targets Set for  
789 the demands of the European Water Framework Directive ”. Version 1 . Internal  
790 report , 2012 .

791 Comte, L., Lek, S., de Deckere, E., de Zwart, D., Gevrey, M., 2010. Assessment of stream  
792 biological responses under multiple-stress conditions. *Environ. Sci. Pollut. Res.* 17,  
793 1469–1478. <https://doi.org/10.1007/s11356-010-0333-z>

794 D’heygere, T., Goethals, P.L.M., De Pauw, N., 2006. Genetic algorithms for optimisation of  
795 predictive ecosystems models based on decision trees and neural networks. *Ecol.*  
796 *Modell.* 195, 20–29. <https://doi.org/10.1016/j.ecolmodel.2005.11.005>

797 Dahm, V., Hering, D., Nemitz, D., Graf, W., Schmidt-Kloiber, A., Leitner, P., Melcher, A.,  
798 Feld, C.K., 2013. Effects of physico-chemistry, land use and hydromorphology on  
799 three riverine organism groups: A comparative analysis with monitoring data from  
800 Germany and Austria. *Hydrobiologia* 704, 389–415.  
801 <https://doi.org/10.1007/s10750-012-1431-3>

802 Dakou, E., Goethals, P.L.M., D’heygere, T., Dedeker, A.P., Gabriels, W., De Pauw, N.,  
803 Lazaridou-Dimitriadou, M., 2006. Development of artificial neural network models  
804 predicting macroinvertebrate taxa in the river Axios (Northern Greece). *Ann.*  
805 *Limnol. J. Limnol.* 42, 241–250. <https://doi.org/Doi 10.1051/Limn/2006025>

806 Demars, B.O.L., Potts, J.M., Trémolières, M., Thiébaud, G., Goucelin, N., Nordmann, V.,  
807 2012. River macrophyte indices: Not the Holy Grail! *Freshw. Biol.* 57, 1745–1759.  
808 <https://doi.org/10.1111/j.1365-2427.2012.02834.x>

809 Dolédec, S., Phillips, N., Scarsbrook, M., Riley, R.H., Townsend, C.R., 2006. Comparison of

810 structural and functional approaches to determining landuse effects on grassland  
811 stream invertebrate communities. *J. North Am. Benthol. Soc.* 25, 44–60.  
812 [https://doi.org/10.1899/0887-3593\(2006\)25\[44:COAFA\]2.0.CO;2](https://doi.org/10.1899/0887-3593(2006)25[44:COAFA]2.0.CO;2)

813 Durance, I., Ormerod, S.J., 2007. Climate change effects on upland stream  
814 macroinvertebrates over a 25-year period. *Glob. Chang. Biol.* 13, 942–957.  
815 <https://doi.org/10.1111/j.1365-2486.2007.01340.x>

816 Džeroski, S., 2001. Applications of symbolic machine learning to ecological modelling.  
817 *Ecol. Modell.* 146, 263–273. [https://doi.org/10.1016/S0304-3800\(01\)00312-X](https://doi.org/10.1016/S0304-3800(01)00312-X)

818 Džeroski, S., Grbović, J., Walley, W.J., Kompare, B., 1997. Using machine learning  
819 techniques in the construction of models. II. Data analysis with rule induction. *Ecol.*  
820 *Modell.* 95, 95–111. [https://doi.org/10.1016/S0304-3800\(96\)00029-4](https://doi.org/10.1016/S0304-3800(96)00029-4)

821 European Council, 2000. Directive 2000/60/EC of the European Parliament and of the  
822 Council of 23 October 2000 Establishing a Framework for Community Action in the  
823 Field of Water Policy. Office for official publications of the European Communities,  
824 Brussels., Office for official publications of the European Communities, Brussels.

825 Everaert, G., Bennetsen E., Goethals, P.L.M., 2016. An applicability index for reliable  
826 and applicable decision trees in water quality modelling. *Ecol. Inform.* 32, 1–6.  
827 <https://doi.org/10.1016/j.ecoinf.2015.12.004>

828 Fabrègue, M., Braud, A., Bringay, S., 2013. OrderSpan: Mining Closed Partially Ordered  
829 Patterns, in: Tucker A., Höppner F., Siebes A., Swift S. (Eds) *Advances in Intelligent*  
830 *Data Analysis XII. IDA 2013. Lecture Notes in Computer Science, Vol 8207.* Springer,  
831 Berlin, Heidelberg. Springer, Berlin, Heidelberg, pp. 186–197.  
832 [https://doi.org/https://doi.org/10.1007/978-3-642-41398-8\\_17](https://doi.org/https://doi.org/10.1007/978-3-642-41398-8_17)

833 Fabrègue, M., Braud, A., Bringay, S., Grac, C., Le Ber, F., Levet, D., Teisseire, M., 2014.  
834 Discriminant temporal patterns for linking physico-chemistry and biology in hydro-  
835 ecosystem assessment. *Ecol. Inform.* 24, 210–221.  
836 <https://doi.org/10.1016/j.ecoinf.2014.09.003>

837 Feld, C.K., Birk, S., Eme, D., Gerisch, M., Hering, D., Kernan, M., Maileht, K., Mischke, U., Ott,  
838 I., Pletterbauer, F., Poikane, S., Salgado, J., Sayer, C.D., Van Wichelen, J., Malard, F.,  
839 2016a. Disentangling the effects of land use and geo-climatic factors on diversity in  
840 European freshwater ecosystems. *Ecol. Indic.* 60, 71–83.  
841 <https://doi.org/10.1016/j.ecolind.2015.06.024>

842 Feld, C.K., Segurado, P., Gutiérrez-Cánovas, C., 2016b. Analysing the impact of multiple  
843 stressors in aquatic biomonitoring data: A ‘cookbook’ with applications in R. *Sci.*  
844 *Total Environ.* 573, 1320–1339. <https://doi.org/10.1016/j.scitotenv.2016.06.243>

845 Forio, M.A.E., Van Echelpoel, W., Dominguez-Granda, L., Mereta, S.T., Ambelu, A., Hoang,  
846 T.H., Boets, P., Goethals, P.L.M., 2016. Analysing the effects of water quality on the  
847 occurrence of freshwater macroinvertebrate taxa among tropical river basins from  
848 different continents. *AI Commun.* 29, 665–685. [https://doi.org/10.3233/AIC-](https://doi.org/10.3233/AIC-160712)  
849 [160712](https://doi.org/10.3233/AIC-160712)

850 Fruget, J.-F., Centofanti, M., Dessaix, J., Olivier, J.-M., Druart, J.-C., Martinez, P.-J., 2001.  
851 Temporal and spatial dynamics in large rivers : example of a long-term monitoring  
852 of the middle Rhone River. *Ann. Limnol. - Int. J. Limnol.* 37, 237–251.  
853 <https://doi.org/10.1051/limn/2001021>

854 Fytilis, N., Rizzo, D.M., 2013. Coupling self-organizing maps with a Naïve Bayesian  
855 classifier: Stream classification studies using multiple assessment data. *Water*

856 Resour. Res. 49, 7747–7762. <https://doi.org/10.1002/2012WR013422>

857 George, A., Binu, D., 2012. DRL-Prefixspan: A novel pattern growth algorithm for  
858 discovering downturn, revision and launch (DRL) sequential patterns. *Open*  
859 *Comput. Sci.* 2, 426–439. <https://doi.org/10.2478/s13537-012-0030-8>

860 Gibert, K., Izquierdo, J., Holmes, G., Athanasiadis, I., Comas, J., Sanchez-Marre, M., 2008.  
861 On the role of pre and post-processing in environmental data mining Volume 3,  
862 1937–1958.

863 Giraudel, J.L., Lek, S., 2001. A comparison of self-organizing map algorithm and some  
864 conventional statistical methods for ecological community ordination. *Ecol. Modell.*  
865 146, 329–339. [https://doi.org/10.1016/S0304-3800\(01\)00324-6](https://doi.org/10.1016/S0304-3800(01)00324-6)

866 Haury, J., Peltre, M.-C., Trémolières, M., Barbe, J., Thiébaud, G., Bernez, I., Daniel, H.,  
867 Chatenet, P., Haan-Archipof, G., Muller, S., Dutartre, A., Laplace-Treyture, C.,  
868 Cazaubon, A., Lambert-Servien, E., 2006. A new method to assess water trophy and  
869 organic pollution - The Macrophyte Biological Index for Rivers (IBMR): Its  
870 application to different types of river and pollution. *Hydrobiologia* 570, 153–158.  
871 <https://doi.org/10.1007/s10750-006-0175-3>

872 Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.S.,  
873 Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., de Bund, W. van, 2010. The European  
874 Water Framework Directive at the age of 10: A critical review of the achievements  
875 with recommendations for the future. *Sci. Total Environ.* 408, 4007–4019.  
876 <https://doi.org/10.1016/j.scitotenv.2010.05.031>

877 Kristensen, P., 2004. The DPSIR Framework. UNEP Headquarters, on a  
878 comprehensive/detailed assessment of the vulnerability of water resources to

879 environmental change in Africa using basin approach. Nairobi, Kenya.

880 Lalande, N., 2013. Impacts multi-échelles de l'occupation du sol sur l'état écologique des  
881 cours d'eau - Elaboration et test d'un cadre d'analyse et de modélisation. Thèse.  
882 AgroParisTech Montpellier.

883 Landuyt, D., Broekx, S., Engelen, G., Uljee, I., Van der Meulen, M., Goethals, P.L.M., 2016.  
884 The importance of uncertainties in scenario analyzes - A study on future ecosystem  
885 service delivery in Flanders. *Sci. Total Environ.* 553, 504–518.  
886 <https://doi.org/10.1016/j.scitotenv.2016.02.098>

887 Larras, F., Coulaud, R., Gautreau, E., Billoir, E., Rosebery, J., Usseglio-Polatera, P., 2017.  
888 Assessing anthropogenic pressures on streams: A random forest approach based on  
889 benthic diatom communities. *Sci. Total Environ.* 586, 1101–1112.  
890 <https://doi.org/10.1016/j.scitotenv.2017.02.096>

891 Lemm, J.U., Feld, C.K., 2017. Identification and interaction of multiple stressors in central  
892 European lowland rivers. *Sci. Total Environ.* 603–604, 148–154.  
893 <https://doi.org/10.1016/j.scitotenv.2017.06.092>

894 Marzin, A., Archaimbault, V., Belliard, J., Chauvin, C., Delmas, F., Pont, D., 2012. Ecological  
895 assessment of running waters: Do macrophytes, macroinvertebrates, diatoms and  
896 fish show similar responses to human pressures? *Ecol. Indic.* 23, 56–65.  
897 <https://doi.org/10.1016/j.ecolind.2012.03.010>

898 MEDD, AE, 2003. Système d'évaluation de la qualité de l'eau des cours d'eau (SEQ-Eau),  
899 version 2 - étude Inter-Agences N°52 réalisée par le Ministère de l'Ecologie et du  
900 Développement Durable (MEDD) et les Agences de l'Eau (AE), France.

901 MEEM, 2012. Guide technique Evaluation de l'état des eaux de surfaces continentales  
902 (Cours d'eau, Canaux, Plans d'eau) édité par le Ministère de l'Environnement de  
903 l'Energie et de la Mer, France.

904 Meyer, A., Combroux, I., Schmitt, L., Trémolières, M., 2013. Vegetation dynamics in side-  
905 channels reconnected to the Rhine River: what are the main factors controlling  
906 communities trajectories after restoration? *Hydrobiologia* 714, 35–47.  
907 <https://doi.org/10.1007/s10750-013-1512-y>

908 Millenium Ecosystem Assessment Programm, 2005. Ecosystems and human well-being :  
909 synthesis, Island Pre. ed.

910 Mondy, C.P., Usseglio-Polatera, P., 2013. Using conditional tree forests and life history  
911 traits to assess specific risks of stream degradation under multiple pressure  
912 scenario. *Sci. Total Environ.* 461–462, 750–760.  
913 <https://doi.org/10.1016/j.scitotenv.2013.05.072>

914 Mondy, C.P., Villeneuve, B., Archaimbault, V., Usseglio-polatera, P., 2012. A new  
915 macroinvertebrate-based multimetric index ( I 2 M 2 ) to evaluate ecological quality  
916 of French wadeable streams fulfilling the WFD demands : A taxonomical and trait  
917 approach. *Ecol. Indic.* 18, 452–467. <https://doi.org/10.1016/j.ecolind.2011.12.013>

918 Motelay-Massei, A., Ollivon, D., Garban, B., Teil, M.J., Blanchard, M., Chevreuil, M., 2004.  
919 Distribution and spatial trends of PAHs and PCBs in soils in the Seine River basin,  
920 France. *Chemosphere* 55, 555–565.  
921 <https://doi.org/10.1016/J.CHEMOSPHERE.2003.11.054>

922 Nisbet, M., Verneaux, J., 1970. Composantes chimiques des eaux courantes: discussion et  
923 proposition de classes en tant que bases d'interprétation des analyses chimiques.

924 Ann. Limnologie 6, 161–190. <https://doi.org/10.1051/limn/1970015>

925 Oberdorff, T., Hughes, R.M., 1992. Modification of an Index of Biotic Integrity Based on  
926 Fish Assemblages to Characterize Rivers of the Seine Basin ., *Hydrobiologia* 228,  
927 117–130. <https://doi.org/10.1007/BF00006200>

928 Pringle, C.M., 1990. Nutrient spatial heterogeneity: effects on community structure,  
929 physiognomy, and diversity of stream algae. *Ecology* 71, 905–920.  
930 <https://doi.org/10.2307/1937362>

931 Ren, J., Wang, L., Dong, J., Hu, C., Wang, K., 2009. A Novel Sequential Pattern Mining  
932 Algorithm for the Feature Discovery of Software Fault, in: 2009 International  
933 Conference on Computational Intelligence and Software Engineering. IEEE, pp. 1–4.  
934 <https://doi.org/10.1109/CISE.2009.5367106>

935 Reyjol, Y., Argillier, C., Bonne, W., Borja, A., Buijse, A.D., Cardoso, A.C., Daufresne, M.,  
936 Kernan, M., Ferreira, M.T., Poikane, S., Prat, N., Solheim, A.L., Stroffek, S., Usseglio-  
937 Polatera, P., Villeneuve, B., van de Bund, W., 2014. Assessing the ecological status in  
938 the context of the European Water Framework Directive: Where do we go now? *Sci.*  
939 *Total Environ.* <https://doi.org/10.1016/j.scitotenv.2014.07.119>

940 Sallaberry, A., Pecheur, N., Bringay, S., Roche, M., Teisseire, M., 2011. Sequential patterns  
941 mining and gene sequence visualization to discover novelty from microarray data. *J.*  
942 *Biomed. Inform.* 44, 760–774. <https://doi.org/10.1016/J.JBI.2011.04.002>

943 Schäfer, R.B., Von Der Ohe, P.C., Rasmussen, J., Kefford, B.J., Beketov, M.A., Schulz, R.,  
944 Liess, M., 2012. Thresholds for the effects of pesticides on invertebrate communities  
945 and leaf breakdown in stream ecosystems. *Environ. Sci. Technol.*  
946 <https://doi.org/10.1021/es2039882>

947 Schulz, R., Liess, M., 1999. A field study of the effects of agriculturally derived insecticide  
948 input on stream macroinvertebrate dynamics. *Aquat. Toxicol.* 46, 155–176.  
949 [https://doi.org/10.1016/S0166-445X\(99\)00002-8](https://doi.org/10.1016/S0166-445X(99)00002-8)

950 Serrano Balderas, E.C., Berti-Equille, L., Hernández, M.A.A., Grac, C., 2017. Principled data  
951 preprocessing: Application to biological aquatic indicators of water pollution. *Proc.*  
952 *- Int. Work. Database Expert Syst. Appl. DEXA 2017-Augus*, 52–56.  
953 <https://doi.org/10.1109/DEXA.2017.27>

954 Staentzel, C., Arnaud, F., Combroux, I., Schmitt, L., Trémolières, M., Grac, C., Piégay, H.,  
955 Barillier, A., Chardon, V., Beisel, J.-N., 2017. How do instream flow increase and  
956 gravel augmentation impact biological communities in large rivers: A case study on  
957 the Upper Rhine River. *River Res. Appl.* <https://doi.org/10.1002/rra.3237>

958 Stendera, S., Adrian, R., Bonada, N., Cañedo-Argüelles, M., Hugueny, B., Januschke, K.,  
959 Pletterbauer, F., Hering, D., 2012. Drivers and stressors of freshwater biodiversity  
960 patterns across different ecosystems and scales: a review. *Hydrobiologia* 696, 1–28.  
961 <https://doi.org/10.1007/s10750-012-1183-0>

962 Stoddard, J.L., Jeffries, D.S., Lükewille, A., Clair, T.A., Dillon, P.J., Driscoll, C.T., Forsius, M.,  
963 Johannessen, M., Kahl, J.S., Kellogg, J.H., Kemp, A., Mannlo, J., Monteith, D.T.,  
964 Murdoch, P.S., Patrick, S., Rebsdorl, A., Skjelkvale, B.L., Stainton, M.P., Traaen, T., Van  
965 Dam, H., Webster, K.E., Wleting, J., Willander, A., 1999. Regional trends in aquatic  
966 recovery from acidification in North America and Europe. *Nature* 401, 575–578.  
967 <https://doi.org/10.1038/44114>

968 Teil, M.-J., Blanchard, M., Chevreuil, M., 2004. Atmospheric deposition of organochlorines  
969 (PCBs and pesticides) in northern France. *Chemosphere* 55, 501–514.

970 <https://doi.org/10.1016/J.CHEMOSPHERE.2003.11.064>

971 Trémolières, M., 1994. Changes in aquatic vegetation in Rhine floodplain streams in  
972 Alsace in relation to disturbance. *J. Veg. Sci.* 5, 169–178. [https://doi.org/doi:](https://doi.org/doi:10.2307/3236149)  
973 [10.2307/3236149](https://doi.org/doi:10.2307/3236149)

974 Tsai, W.-P., Huang, S.-P., Cheng, S.-T., Shao, K.-T., Chang, F.-J., 2016. A data-mining  
975 framework for exploring the multi-relation between fish species and water quality  
976 through self-organizing map. *Sci. Total Environ.* 579, 474–483.  
977 <https://doi.org/10.1016/j.scitotenv.2016.11.071>

978 Van Looy, K., Piffady, J., Tormos, T., Villeneuve, B., Valette, L., Chandesris, A., Souchon, Y.,  
979 2015. Unravelling River System Impairments in Stream Networks with an  
980 Integrated Risk Approach. *Environ. Manage.* 55, 1343–1353.  
981 <https://doi.org/10.1007/s00267-015-0477-1>

982 Van Urk, G., Kerkum, F., Van Leeuwen, C.J., 1993. Insects and insecticides in the Lower  
983 Rhine. *Water Res.* 27, 205–213. [https://doi.org/10.1016/0043-1354\(93\)90077-U](https://doi.org/10.1016/0043-1354(93)90077-U)

984 Villeneuve, B., Piffady, J., Valette, L., Souchon, Y., Usseglio-Polatera, P., 2018. Direct and  
985 indirect effects of multiple stressors on stream invertebrates across watershed,  
986 reach and site scales: A structural equation modelling better informing on  
987 hydromorphological impacts. *Sci. Total Environ.* 612, 660–671.  
988 <https://doi.org/10.1016/J.SCITOTENV.2017.08.197>

989 Villeneuve, B., Souchon, Y., Usseglio-Polatera, P., Ferréol, M., Valette, L., 2015. Can we  
990 predict biological condition of stream ecosystems? A multi-stressors approach  
991 linking three biological indices to physico-chemistry, hydromorphology and land  
992 use. *Ecol. Indic.* <https://doi.org/10.1016/j.ecolind.2014.07.016>

993 Webb, B.W., 1996. Trends in stream and river temperature. *Hydrol. Process.* 10, 205–  
994 226. [https://doi.org/10.1002/\(SICI\)1099-1085\(199602\)10:2<205::AID-](https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<205::AID-)  
995 [HYP358>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<205::AID-HYP358>3.0.CO;2-1)

996 Webb, B.W., Hannah, D.M., Moore, R.D., Brown, L.E., Nobilis, F., 2008. Recent advances in  
997 stream and river temperature research. *Hydrol. Process.*  
998 <https://doi.org/10.1002/hyp.6994>

999

1000 AFNOR, 2016a. Qualité de l'eau: prélèvements des macro-invertébrés aquatiques en  
1001 rivières peu profondes. Norme Française NF T 90-333.

1002 AFNOR, 2016b. Qualité de l'eau: détermination de l'Indice Biologique Diatomées (IBD).  
1003 Norme Française NF T90-354.

1004 AFNOR, 2011. Qualité de l'eau: détermination de l'Indice poissons rivière (IPR). Norme  
1005 Française NF T90-344.

1006 AFNOR, 2010. Qualité de l'eau: traitement au laboratoire d'échantillons contenant des  
1007 macro-invertébrés de cours d'eau. Norme Française Expérimentale XP T 90-388.

1008 AFNOR, 2004. Qualité de l'eau: détermination de l'Indice Biologique Global Normalisé  
1009 (IBGN); Norme Française NF T90-350.

1010 AFNOR, 2003. Qualité de l'eau: détermination de l'Indice Biologique Macrophytique en  
1011 Rivière (IBMR). Norme Française NF T90-395.

1012 Agrawal, R., Srikant, R., 1995. Mining sequential patterns, in: *Proceedings of the Eleventh*  
1013 *International Conference on Data Engineering.* IEEE Comput. Soc. Press, pp. 3–14.  
1014 <https://doi.org/10.1109/ICDE.1995.380415>

1015 Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules, in: Proc. 20th  
1016 IVLDB Conference.

1017 Arle, J., Mohaupt, V., Kirst, I., 2016. Monitoring of Surface Waters in Germany under the  
1018 Water Framework Directive — A Review of approaches, methods and results.  
1019 Water 8, 217. <https://doi.org/10.3390/w8060217>

1020 Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund,  
1021 W., Zampoukas, N., Hering, D., 2012. Three hundred ways to assess Europe's surface  
1022 waters: An almost complete overview of biological methods to implement the  
1023 Water Framework Directive. Ecol. Indic. 18, 31–41.  
1024 <https://doi.org/10.1016/j.ecolind.2011.10.009>

1025 Casas-Garriga, G. Summarizing sequential data with closed partial orders. In SIAM  
1026 International Conference on Data Mining, SDM, 2005.

1027 Claussen, U., Müller, P., Arle, J., 2012. WFD CIS ECOSTAT WG A Report “ Comparison of  
1028 envi- ronmental quality Objectives , threshold values or water quality targets Set for  
1029 the demands of the European Water Framework Directive ”. Version 1 . Internal  
1030 report , 2012 .

1031 Comte, L., Lek, S., de Deckere, E., de Zwart, D., Gevrey, M., 2010. Assessment of stream  
1032 biological responses under multiple-stress conditions. Environ. Sci. Pollut. Res. 17,  
1033 1469–1478. <https://doi.org/10.1007/s11356-010-0333-z>

1034 D'heygere, T., Goethals, P.L.M., De Pauw, N., 2006. Genetic algorithms for optimisation of  
1035 predictive ecosystems models based on decision trees and neural networks. Ecol.  
1036 Modell. 195, 20–29. <https://doi.org/10.1016/j.ecolmodel.2005.11.005>

- 1037 Dahm, V., Hering, D., Nemitz, D., Graf, W., Schmidt-Kloiber, A., Leitner, P., Melcher, A.,  
1038 Feld, C.K., 2013. Effects of physico-chemistry, land use and hydromorphology on  
1039 three riverine organism groups: A comparative analysis with monitoring data from  
1040 Germany and Austria. *Hydrobiologia* 704, 389–415.  
1041 <https://doi.org/10.1007/s10750-012-1431-3>
- 1042 Dakou, E., Goethals, P.L.M., D’heygere, T., Dedeker, A.P., Gabriels, W., De Pauw, N.,  
1043 Lazaridou-Dimitriadou, M., 2006. Development of artificial neural network models  
1044 predicting macroinvertebrate taxa in the river Axios (Northern Greece). *Ann.*  
1045 *Limnol. J. Limnol.* 42, 241–250. [https://doi.org/Doi 10.1051/Limn/2006025](https://doi.org/Doi%2010.1051/Limn/2006025)
- 1046 Demars, B.O.L., Potts, J.M., Trémolières, M., Thiébaud, G., Goucelin, N., Nordmann, V.,  
1047 2012. River macrophyte indices: Not the Holy Grail! *Freshw. Biol.* 57, 1745–1759.  
1048 <https://doi.org/10.1111/j.1365-2427.2012.02834.x>
- 1049 Dolédec, S., Phillips, N., Scarsbrook, M., Riley, R.H., Townsend, C.R., 2006. Comparison of  
1050 structural and functional approaches to determining landuse effects on grassland  
1051 stream invertebrate communities. *J. North Am. Benthol. Soc.* 25, 44–60.  
1052 [https://doi.org/10.1899/0887-3593\(2006\)25\[44:COAFA\]2.0.CO;2](https://doi.org/10.1899/0887-3593(2006)25[44:COAFA]2.0.CO;2)
- 1053 Durance, I., Ormerod, S.J., 2007. Climate change effects on upland stream  
1054 macroinvertebrates over a 25-year period. *Glob. Chang. Biol.* 13, 942–957.  
1055 <https://doi.org/10.1111/j.1365-2486.2007.01340.x>
- 1056 Džeroski, S., 2001. Applications of symbolic machine learning to ecological modelling.  
1057 *Ecol. Modell.* 146, 263–273. [https://doi.org/10.1016/S0304-3800\(01\)00312-X](https://doi.org/10.1016/S0304-3800(01)00312-X)
- 1058 Džeroski, S., Grbović, J., Walley, W.J., Kompare, B., 1997. Using machine learning  
1059 techniques in the construction of models. II. Data analysis with rule induction. *Ecol.*

1060 Modell. 95, 95–111. [https://doi.org/10.1016/S0304-3800\(96\)00029-4](https://doi.org/10.1016/S0304-3800(96)00029-4)

1061 European Council, 2000. Directive 2000/60/EC of the European Parliament and of the  
1062 Council of 23 October 2000 Establishing a Framework for Community Action in the  
1063 Field of Water Policy. Office for official publications of the European Communities,  
1064 Brussels., Office for official publications of the European Communities, Brussels.

1065 Everaert, G., Bennetsen , E., Goethals, P.L.M., 2016. An applicability index for reliable  
1066 and applicable decision trees in water quality modelling. *Ecol. Inform.* 32, 1–6.  
1067 <https://doi.org/10.1016/j.ecoinf.2015.12.004>

1068 Fabrègue, M., Braud, A., Bringay, S., 2013. OrderSpan: Mining Closed Partially Ordered  
1069 Patterns, in: Tucker A., Höppner F., Siebes A., Swift S. (Eds) *Advances in Intelligent  
1070 Data Analysis XII. IDA 2013. Lecture Notes in Computer Science, Vol 8207.* Springer,  
1071 Berlin, Heidelberg. Springer, Berlin, Heidelberg, pp. 186–197.  
1072 [https://doi.org/https://doi.org/10.1007/978-3-642-41398-8\\_17](https://doi.org/https://doi.org/10.1007/978-3-642-41398-8_17)

1073 Fabrègue, M., Braud, A., Bringay, S., Grac, C., Le Ber, F., Levet, D., Teisseire, M., 2014.  
1074 Discriminant temporal patterns for linking physico-chemistry and biology in hydro-  
1075 ecosystem assessment. *Ecol. Inform.* 24, 210–221.  
1076 <https://doi.org/10.1016/j.ecoinf.2014.09.003>

1077 Feld, C.K., Birk, S., Eme, D., Gerisch, M., Hering, D., Kernan, M., Maileht, K., Mischke, U., Ott,  
1078 I., Pletterbauer, F., Poikane, S., Salgado, J., Sayer, C.D., Van Wichelen, J., Malard, F.,  
1079 2016a. Disentangling the effects of land use and geo-climatic factors on diversity in  
1080 European freshwater ecosystems. *Ecol. Indic.* 60, 71–83.  
1081 <https://doi.org/10.1016/j.ecolind.2015.06.024>

1082 Feld, C.K., Segurado, P., Gutiérrez-Cánovas, C., 2016b. Analysing the impact of multiple

1083 stressors in aquatic biomonitoring data: A 'cookbook' with applications in R. *Sci.*  
1084 *Total Environ.* 573, 1320–1339. <https://doi.org/10.1016/j.scitotenv.2016.06.243>

1085 Forio, M.A.E., Van Echelpoel, W., Dominguez-Granda, L., Mereta, S.T., Ambelu, A., Hoang,  
1086 T.H., Boets, P., Goethals, P.L.M., 2016. Analysing the effects of water quality on the  
1087 occurrence of freshwater macroinvertebrate taxa among tropical river basins from  
1088 different continents. *AI Commun.* 29, 665–685. [https://doi.org/10.3233/AIC-](https://doi.org/10.3233/AIC-160712)  
1089 160712

1090 Fruget, J.-F., Centofanti, M., Dessaix, J., Olivier, J.-M., Druart, J.-C., Martinez, P.-J., 2001.  
1091 Temporal and spatial dynamics in large rivers : example of a long-term monitoring  
1092 of the middle Rhone River. *Ann. Limnol. - Int. J. Limnol.* 37, 237–251.  
1093 <https://doi.org/10.1051/limn/2001021>

1094 Fytilis, N., Rizzo, D.M., 2013. Coupling self-organizing maps with a Naïve Bayesian  
1095 classifier: Stream classification studies using multiple assessment data. *Water*  
1096 *Resour. Res.* 49, 7747–7762. <https://doi.org/10.1002/2012WR013422>

1097 George, A., Binu, D., 2012. DRL-Prefixspan: A novel pattern growth algorithm for  
1098 discovering downturn, revision and launch (DRL) sequential patterns. *Open*  
1099 *Comput. Sci.* 2, 426–439. <https://doi.org/10.2478/s13537-012-0030-8>

1100 Gibert, K., Izquierdo, J., Holmes, G., Athanasiadis, I., Comas, J., Sanchez-Marre, M., 2008.  
1101 On the role of pre and post-processing in environmental data mining Volume 3,  
1102 1937–1958.

1103 Giraudel, J.L., Lek, S., 2001. A comparison of self-organizing map algorithm and some  
1104 conventional statistical methods for ecological community ordination. *Ecol. Modell.*  
1105 146, 329–339. [https://doi.org/10.1016/S0304-3800\(01\)00324-6](https://doi.org/10.1016/S0304-3800(01)00324-6)

1106 Haury, J., Peltre, M.-C., Trémolières, M., Barbe, J., Thiébaud, G., Bernez, I., Daniel, H.,  
1107 Chatenet, P., Haan-Archipof, G., Muller, S., Dutartre, A., Laplace-Treyture, C.,  
1108 Cazaubon, A., Lambert-Servien, E., 2006. A new method to assess water trophy and  
1109 organic pollution - The Macrophyte Biological Index for Rivers (IBMR): Its  
1110 application to different types of river and pollution. *Hydrobiologia* 570, 153–158.  
1111 <https://doi.org/10.1007/s10750-006-0175-3>

1112 Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C.K., Heiskanen, A.S.,  
1113 Johnson, R.K., Moe, J., Pont, D., Solheim, A.L., de Bund, W. van, 2010. The European  
1114 Water Framework Directive at the age of 10: A critical review of the achievements  
1115 with recommendations for the future. *Sci. Total Environ.* 408, 4007–4019.  
1116 <https://doi.org/10.1016/j.scitotenv.2010.05.031>

1117 Lemm, J.U., Feld, C.K., 2017. Identification and interaction of multiple stressors in central  
1118 European lowland rivers. *Sci. Total Environ.* 603–604, 148–154.  
1119 <https://doi.org/10.1016/j.scitotenv.2017.06.092>

1120 Marzin, A., Archaimbault, V., Belliard, J., Chauvin, C., Delmas, F., Pont, D., 2012. Ecological  
1121 assessment of running waters: Do macrophytes, macroinvertebrates, diatoms and  
1122 fish show similar responses to human pressures? *Ecol. Indic.* 23, 56–65.  
1123 <https://doi.org/10.1016/j.ecolind.2012.03.010>

1124 MEDD, AE, 2003. Système d'évaluation de la qualité de l'eau des cours d'eau (SEQ-Eau),  
1125 version 2 - étude Inter-Agences N°52 réalisée par le Ministère de l'Ecologie et du  
1126 Développement Durable (MEDD) et les Agences de l'Eau (AE), France.

1127 MEEM, 2012. Guide technique Evaluation de l'état des eaux de surfaces continentales  
1128 (Cours d'eau, Canaux, Plans d'eau) édité par le Ministère de l'Environnement de

- 1129 l'Énergie et de la Mer, France.
- 1130 Meyer, A., Combroux, I., Schmitt, L., Trémolières, M., 2013. Vegetation dynamics in side-  
1131 channels reconnected to the Rhine River: what are the main factors controlling  
1132 communities trajectories after restoration? *Hydrobiologia* 714, 35–47.  
1133 <https://doi.org/10.1007/s10750-013-1512-y>
- 1134 Millenium Ecosystem Assessment Programm, 2005. *Ecosystems and human well-being :*  
1135 *synthesis*, Island Pre. ed.
- 1136 Mondy, C.P., Usseglio-Polatera, P., 2013. Using conditional tree forests and life history  
1137 traits to assess specific risks of stream degradation under multiple pressure  
1138 scenario. *Sci. Total Environ.* 461–462, 750–760.  
1139 <https://doi.org/10.1016/j.scitotenv.2013.05.072>
- 1140 Mondy, C.P., Villeneuve, B., Archaimbault, V., Usseglio-polatera, P., 2012. A new  
1141 macroinvertebrate-based multimetric index ( I 2 M 2 ) to evaluate ecological quality  
1142 of French wadeable streams fulfilling the WFD demands : A taxonomical and trait  
1143 approach. *Ecol. Indic.* 18, 452–467. <https://doi.org/10.1016/j.ecolind.2011.12.013>
- 1144 Motelay-Massei, A., Ollivon, D., Garban, B., Teil, M.J., Blanchard, M., Chevreuil, M., 2004.  
1145 Distribution and spatial trends of PAHs and PCBs in soils in the Seine River basin,  
1146 France. *Chemosphere* 55, 555–565.  
1147 <https://doi.org/10.1016/J.CHEMOSPHERE.2003.11.054>
- 1148 Nisbet, M., Verneaux, J., 1970. Composantes chimiques des eaux courantes: discussion et  
1149 proposition de classes en tant que bases d'interprétation des analyses chimiques.  
1150 *Ann. Limnologie* 6, 161–190. <https://doi.org/10.1051/limn/1970015>

- 1151 Oberdorff, T., Hughes, R.M., 1992. Modification of an Index of Biotic Integrity Based on  
1152 Fish Assemblages to Characterize Rivers of the Seine Basin . *Hydrobiologia* 228,  
1153 117–130. <https://doi.org/10.1007/BF00006200>
- 1154 Pringle, C.M., 1990. Nutrient spatial heterogeneity: effects on community structure,  
1155 physiognomy, and diversity of stream algae. *Ecology* 71, 905–920.  
1156 <https://doi.org/10.2307/1937362>
- 1157 Ren, J., Wang, L., Dong, J., Hu, C., Wang, K., 2009. A Novel Sequential Pattern Mining  
1158 Algorithm for the Feature Discovery of Software Fault, in: 2009 International  
1159 Conference on Computational Intelligence and Software Engineering. IEEE, pp. 1–4.  
1160 <https://doi.org/10.1109/CISE.2009.5367106>
- 1161 Reyjol, Y., Argillier, C., Bonne, W., Borja, A., Buijse, A.D., Cardoso, A.C., Daufresne, M.,  
1162 Kernan, M., Ferreira, M.T., Poikane, S., Prat, N., Solheim, A.L., Stroffek, S., Usseglio-  
1163 Polatera, P., Villeneuve, B., van de Bund, W., 2014. Assessing the ecological status in  
1164 the context of the European Water Framework Directive: Where do we go now? *Sci.*  
1165 *Total Environ.* <https://doi.org/10.1016/j.scitotenv.2014.07.119>
- 1166 Sallaberry, A., Pecheur, N., Bringay, S., Roche, M., Teisseire, M., 2011. Sequential patterns  
1167 mining and gene sequence visualization to discover novelty from microarray data. *J.*  
1168 *Biomed. Inform.* 44, 760–774. <https://doi.org/10.1016/J.JBI.2011.04.002>
- 1169 Schäfer, R.B., Von Der Ohe, P.C., Rasmussen, J., Kefford, B.J., Beketov, M.A., Schulz, R.,  
1170 Liess, M., 2012. Thresholds for the effects of pesticides on invertebrate communities  
1171 and leaf breakdown in stream ecosystems. *Environ. Sci. Technol.*  
1172 <https://doi.org/10.1021/es2039882>
- 1173 Schulz, R., Liess, M., 1999. A field study of the effects of agriculturally derived insecticide

1174 input on stream macroinvertebrate dynamics. *Aquat. Toxicol.* 46, 155–176.  
1175 [https://doi.org/10.1016/S0166-445X\(99\)00002-8](https://doi.org/10.1016/S0166-445X(99)00002-8)

1176 Serrano Balderas, E.C., Berti-Equille, L., Hernández, M.A.A., Grac, C., 2017. Principled data  
1177 preprocessing: Application to biological aquatic indicators of water pollution. *Proc.*  
1178 *- Int. Work. Database Expert Syst. Appl. DEXA 2017-Augus*, 52–56.  
1179 <https://doi.org/10.1109/DEXA.2017.27>

1180 Staentzel, C., Arnaud, F., Combroux, I., Schmitt, L., Trémolières, M., Grac, C., Piégay, H.,  
1181 Barillier, A., Chardon, V., Beisel, J.-N., 2017. How do instream flow increase and  
1182 gravel augmentation impact biological communities in large rivers: A case study on  
1183 the Upper Rhine River. *River Res. Appl.* <https://doi.org/10.1002/rra.3237>

1184 Stendera, S., Adrian, R., Bonada, N., Cañedo-Argüelles, M., Hugueny, B., Januschke, K.,  
1185 Pletterbauer, F., Hering, D., 2012. Drivers and stressors of freshwater biodiversity  
1186 patterns across different ecosystems and scales: a review. *Hydrobiologia* 696, 1–28.  
1187 <https://doi.org/10.1007/s10750-012-1183-0>

1188 Stoddard, J.L., Jeffries, D.S., Lükewille, A., Clair, T.A., Dillon, P.J., Driscoll, C.T., Forsius, M.,  
1189 Johannessen, M., Kahl, J.S., Kellogg, J.H., Kemp, A., Mannlo, J., Monteith, D.T.,  
1190 Murdoch, P.S., Patrick, S., Rebsdorl, A., Skjelkvale, B.L., Stainton, M.P., Traaen, T., Van  
1191 Dam, H., Webster, K.E., Wleting, J., Wllander, A., 1999. Regional trends in aquatic  
1192 recovery from acidification in North America and Europe. *Nature* 401, 575–578.  
1193 <https://doi.org/10.1038/44114>

1194 Teil, M.-J., Blanchard, M., Chevreuil, M., 2004. Atmospheric deposition of organochlorines  
1195 (PCBs and pesticides) in northern France. *Chemosphere* 55, 501–514.  
1196 <https://doi.org/10.1016/J.CHEMOSPHERE.2003.11.064>

- 1197 Trémolières, M., 1994. Changes in aquatic vegetation in Rhine floodplain streams in  
1198 Alsace in relation to disturbance. *J. Veg. Sci.* 5, 169–178. [https://doi.org/doi:](https://doi.org/doi:10.2307/3236149)  
1199 [10.2307/3236149](https://doi.org/doi:10.2307/3236149)
- 1200 Tsai, W.-P., Huang, S.-P., Cheng, S.-T., Shao, K.-T., Chang, F.-J., 2016. A data-mining  
1201 framework for exploring the multi-relation between fish species and water quality  
1202 through self-organizing map. *Sci. Total Environ.* 579, 474–483.  
1203 <https://doi.org/10.1016/j.scitotenv.2016.11.071>
- 1204 Van Looy, K., Piffady, J., Tormos, T., Villeneuve, B., Valette, L., Chandesris, A., Souchon, Y.,  
1205 2015. Unravelling River System Impairments in Stream Networks with an  
1206 Integrated Risk Approach. *Environ. Manage.* 55, 1343–1353.  
1207 <https://doi.org/10.1007/s00267-015-0477-1>
- 1208 Van Urk, G., Kerkum, F., Van Leeuwen, C.J., 1993. Insects and insecticides in the Lower  
1209 Rhine. *Water Res.* 27, 205–213. [https://doi.org/10.1016/0043-1354\(93\)90077-U](https://doi.org/10.1016/0043-1354(93)90077-U)
- 1210 Villeneuve, B., Piffady, J., Valette, L., Souchon, Y., Usseglio-Polatera, P., 2018. Direct and  
1211 indirect effects of multiple stressors on stream invertebrates across watershed,  
1212 reach and site scales: A structural equation modelling better informing on  
1213 hydromorphological impacts. *Sci. Total Environ.* 612, 660–671.  
1214 <https://doi.org/10.1016/J.SCITOTENV.2017.08.197>
- 1215 Villeneuve, B., Souchon, Y., Usseglio-Polatera, P., Ferréol, M., Valette, L., 2015. Can we  
1216 predict biological condition of stream ecosystems? A multi-stressors approach  
1217 linking three biological indices to physico-chemistry, hydromorphology and land  
1218 use. *Ecol. Indic.* <https://doi.org/10.1016/j.ecolind.2014.07.016>
- 1219 Webb, B.W., 1996. Trends in stream and river temperature. *Hydrol. Process.* 10, 205–

1220 226. [https://doi.org/10.1002/\(SICI\)1099-1085\(199602\)10:2<205::AID-](https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<205::AID-)

1221 [HYP358>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<205::AID-HYP358>3.0.CO;2-1)

1222 Webb, B.W., Hannah, D.M., Moore, R.D., Brown, L.E., Nobilis, F., 2008. Recent advances in

1223 stream and river temperature research. *Hydrol. Process.*

1224 <https://doi.org/10.1002/hyp.6994>

1225

1226 **Appendix 1: Set of thresholds used by SEQ (MEDD and AE, 2003) for**  
1227 **the physico-chemical parameters classified according to pressure**  
1228 **categories** (*here the threshold is always excluded from the better class, e.g.:*  
1229 *for nitrates, the threshold between the high and good classes is 2 mg/L,*  
1230 *therefore if  $[NO_3^-]=2$  mg/L, pressure category NITR is good)*  
1231

Parameter	unit	thresholds between the quality statuses			
		high/good	good/moderate	moderate/poor	poor/bad
MOOX (Oxidisable organic matter)					
O <sub>2</sub>	mg/L	8	6	4	3
%sat	%	90	70	50	30
BOD	mg/L O2	3	6	10	25
CDO	mg/L O2	20	30	40	80
C	mg/L	5	7	10	15
Trihalomethanes	mg/L	0.075	0.1	0.15	0.5
NH <sub>4</sub> <sup>+</sup>	mg/L	0.5	1.5	2.8	4
AZOT (Nitrogen excluding nitrates)					
NH <sub>4</sub> <sup>+</sup>	mg/L	0.1	0.5	2	5
NK	mg/L	1	2	4	10
NO <sub>2</sub> <sup>-</sup>	mg/L	0.03	0.3	0.5	1
NITR (Nitrate)					
NO <sub>3</sub> <sup>-</sup>	mg/L	2	10	25	50
PHOS (Phosphorous)					
PO <sub>4</sub> <sup>3-</sup>	mg/L	0.1	0.5	1	2
P	mg/L	0.05	0.2	0.5	1
EPRV (Effect of eutrophication)					
%sat	%	110	130	150	200
pH		8	8.5	9	9.5
PAES (Suspended solids)					
SS	mg/L	2	25	38	50
TEMP (Temperature)					
T	°C	24	25,5	27	28
ACID (Acidification parameters)					
pH		6.5	6	5.5	4.5
pH		8.2	9	9.5	10
MINE (Mineralisation)					
C	µS/cm	0	60	120	180
C	µS/cm	2500	3000	3500	4000
Cl <sup>-</sup>	mg/L	50	100	150	200
SO <sub>4</sub> <sup>2-</sup>	mg/L	60	120	190	250
Ca <sup>2+</sup>	mg/L	0	12	22	32
Ca <sup>2+</sup>	mg/L	160	230	300	500
Mg <sup>2+</sup>	mg/L	50	75	100	400
Na <sup>+</sup>	mg/L	200	225	250	750
Alkalinity	°F	0	3	5	8
Alkalinity	°F	40	58	75	100
Hardness	°F	0	4	6	8
Hardness	°F	40	70	90	125
MPMI (Heavy metals)					
As	µg/L	1	35	70	100
Cd	µg/L	0.009	0.09	0.85	3
Cr	µg/L	0.36	3.6	36	50
Cu	µg/L	0.27	2.7	27	40
CN <sup>-</sup>	µg/L	0.02	0.2	2	240
Sn	µg/L	1	10	100	55000
Hg	µg/L	0.007	0.07	0.7	1
Ni	µg/L	1.2	12	26	40
Pb	µg/L	1	10	30	50
Zn	µg/L	1.4	14	140	330
PEST (Pesticides)					
2,4-D-ester	µg/L	0.00001	0.0001	0.001	0.1
2,4-D-non-ester	µg/L	0.1	0.7	1.4	2
2,4-MCPA	µg/L	0.1	0.7	1.4	2

Aclonifen	µg/L	0.007	0.07	0.7	2
Alachlor	µg/L	0.1	0.7	1.4	2
Aldicarb	µg/L	0.005	0.05	0.5	2
Aldrin	µg/L	0.001	0.05	0.2	1
Aminotriazole	µg/L	0.1	0.7	1.4	2
Atrazine	µg/L	0.02	0.2	1	2
Atrazine desethyl	µg/L	0.02	0.2	1	2
Bentazone	µg/L	0.1	0.7	1.4	2
Bifenox	µg/L	0.007	0.07	0.7	2
Captan	µg/L	0.1	0.7	1.4	2
Carbendazim	µg/L	0.0007	0.007	0.07	2
Carbofuran	µg/L	0.0015	0.015	0.15	2
Chlorfenvinphos	µg/L	0.0003	0.003	0.03	2
Chlorothalonil	µg/L	0.0004	0.004	0.04	2
Chlorotoluron	µg/L	0.1	0.7	1.4	2
Chlorpyrifos-ethyl	µg/L	0.00005	0.0005	0.005	0.05
Cymoxanil	µg/L	0.006	0.06	0.6	2
Cyprodinil	µg/L	0.01	0.1	1	2
DDD op'	µg/L	0.0006	0.006	0.06	0.6
DDD pp'	µg/L	0.0006	0.006	0.06	0.6
DDE op'	µg/L	0.03	0.3	1.6	2
DDE pp'	µg/L	0.03	0.3	1.6	2
DDT op'	µg/L	0.0002	0.002	0.02	0.2
DDT pp'	µg/L	0.0002	0.002	0.02	0.2
Deltamethrin	µg/L	0.00002	0.0002	0.002	0.02
Dicamba	µg/L	0.1	0.7	1.4	2
Dichlorprop	µg/L	0.05	0.5	1.2	2
Dieldrin	µg/L	0.0005	0.005	0.05	0.5
Dinoterb	µg/L	0.0003	0.003	0.03	0.3
Diquat	µg/L	0.02	0.2	1	2
Diuron	µg/L	0.02	0.2	1	2
DNOC	µg/L	0.07	0.7	1.4	2
Endosulfan	µg/L	0.002	0.02	0.2	0.3
Endrin	µg/L	0.0003	0.003	0.03	0.3
Ethofumesate	µg/L	0.08	0.8	1.4	2
Fenpropidin	µg/L	0.0006	0.006	0.06	2
Fenpropimorph	µg/L	0.1	0.7	1.4	2
Flusilazole	µg/L	0.1	0.7	1.4	2
Folpel	µg/L	0.002	0.02	0.2	2
Fosetyl-aluminium	µg/L	0.1	0.7	1.4	2
Glyphosate	µg/L	0.04	0.4	1.2	2
Imazamethabenz-methyl	µg/L	0.1	0.7	1.4	2
Ioxynil	µg/L	0.04	0.4	1.2	2
Iprodione	µg/L	0.02	0.2	1	2
Isodrin	µg/L	0.0003	0.003	0.03	2
Isoproturon	µg/L	0.02	0.2	1	2
Lindane	µg/L	0.001	0.01	0.1	1.1
Linuron	µg/L	0.05	0.5	1.3	2
Mancozeb	µg/L	0.1	1	1.5	2
Maneb	µg/L	0.01	0.1	1	2
Methabenzthiazuron	µg/L	0.1	0.7	1.4	2
Methomyl	µg/L	0.03	0.3	1.1	2
Metolachlor	µg/L	0.1	0.7	1.4	2
Norflurazon	µg/L	0.01	0.1	1.2	2
Oxadixyl	µg/L	0.1	0.7	1.4	2
Oxydemeton-methyl	µg/L	0.003	0.03	0.3	2
Paraquat	µg/L	0.1	0.7	1.4	2

1233

1234

Parathion ethyl	µg/L	0.000003	0.00003	0.0003	0.03
Parathion methyl	µg/L	0.0002	0.002	0.02	2
Pendimethalin	µg/L	0.03	0.3	1.1	2
Prochloraz	µg/L	0.01	0.1	1	2
Prosulfocarb	µg/L	0.01	0.1	1	2
Simazine	µg/L	0.002	0.02	0.2	2
Simazine-desethyl	µg/L	0.02	0.2	1	2
Tebuconazole	µg/L	0.1	1	1.5	2
Terbumeton	µg/L	0.1	0.7	1.4	2
Terbuthylazine	µg/L	0.02	0.2	1	2
Terbutryn	µg/L	0.03	0.3	1.1	2
Tridemorph	µg/L	0.1	1.3	1.6	2
Trifluralin	µg/L	0.02	0.2	1	2
Vinclozolin	µg/L	0.1	0.7	1.4	2
HAP (Polycyclic aromatic hydrocarbons)					
Acenaphthene	µg/L	0.07	0.7	160	
Acenaphthylene	µg/L	0.04	0.4	99	
Benzo(a)anthracene	µg/L	0.0005	0.005	1.2	
Benzo(a)pyrene	µg/L	0.00003	0.0003	0.08	
Benzo(b)fluoranthene	µg/L	0.0001	0.001	0.3	
Benzo(g,h,i)perylene	µg/L	0.0003	0.003	0.6	
Benzo(k)fluoranthene	µg/L	0.0003	0.003	0.8	
Chrysene	µg/L	0.0006	0.006	1.5	
Dibenzo(a,h)anthracene	µg/L	0.000006	0.00006	0.014	
Fluoranthene	µg/L	0.0024	0.024	6	
Fluorene	µg/L	0.03	0.3	77	
Indeno(1,2,3-cd)pyrene	µg/L	0.00016	0.0016	0.4	
Naphtalene	µg/L	0.19	1.9	460	
Phenanthrene	µg/L	0.011	0.11	27	
Pyrene	µg/L	0.0024	0.024	6	
PCB (Polychloro biphenyls)					
PCB 28	µg/L	0.0001	0.001	0.01	2
PCB 52	µg/L	0.0001	0.001	0.01	2
PCB 101	µg/L	0.0001	0.001	0.01	2
PCB 118	µg/L	0.0001	0.001	0.01	2
PCB 138	µg/L	0.0001	0.001	0.01	2
PCB 153	µg/L	0.0001	0.001	0.01	2
PCB 180	µg/L	0.0001	0.001	0.01	2
PCBs sum	µg/L	0.0001	0.001	0.01	2
MPOR (Other organic hydrocarbons)					
Benzene	µg/L	0.5	5	7.5	10
C10-C13-Chloroalkanes	µg/L	0.05	0.5	5	14
1,2-Chloroaniline	µg/L	0.001	0.01	0.1	6
1,3-Chloroaniline	µg/L	0.001	0.01	0.1	6
1,4-Chloroaniline	µg/L	0.001	0.01	0.1	6
Chloroform	µg/L	1.2	3	6	10
1,2-Chloronitrobenzene	µg/L	3	30	90	150
1,3-Chloronitrobenzene	µg/L	3	30	90	150
1,4-Chloronitrobenzene	µg/L	3	30	90	150
Meta-cresol	µg/L	0.2	0.8	1.4	2
Ortho-cresol	µg/L	0.2	0.8	1.4	2
Para-cresol	µg/L	0.2	0.8	1.4	2
Phtalate DEPH	µg/L	0.03	0.3	3	
Dibutyltin	µg/L	0.09	0.9	3	6
3,4-Dichloroaniline	µg/L	0.003	0.03	0.3	2
1,2-Dichlorobenzene	µg/L	2	20	200	740
1,3-Dichlorobenzene	µg/L	2	20	200	740

1235

1236

1,4-Dichlorobenzene	µg/L	2	20	100	200
1,2-Dichloroethane	µg/L	3	4.5	6	60
1,2-Dichloroethylene	µg/L	50	200	350	500
Dichloromethane	µg/L	6.8	17	28	40
2,3-Dichlorophenol	µg/L	1	4	7	10
2,4-Dichlorophenol	µg/L	1	4	7	10
2,5-Dichlorophenol	µg/L	1	4	7	10
2,6-Dichlorophenol	µg/L	1	4	7	10
3,4-Dichlorophenol	µg/L	1	4	7	10
3,5-Dichlorophenol	µg/L	1	4	7	10
EDTA	µg/L	4	41	410	2000
Hexachlorobenzene	µg/L	0.0007	0.007	0.02	0.1
Hexachlorobutadiene	µg/L	0.01	0.1	1	6
4-para-nonylphenol	µg/L	0.033	0.33	3.3	21
Para-ter-octylphenol	µg/L	0.1	1	10	100
Pentabromodiphenylether	µg/L	0.02	0.2	2.4	
Pentachlorobenzene	µg/L	0.01	0.1	1	2
Pentachlorophenol	µg/L	0.01	0.1	1	2
1,1,2,2-Tetrachloroethane	µg/L	14	42	70	100
Tetrachloroethylene	µg/L	5	12	20	200
Tetrachloromethane	µg/L	2	2.3	2.5	20
Toluene	µg/L	10	100	1000	1500
Tributyltin compounds	µg/L	0.00004	0.0004	0.004	0.05
Tributyltin oxyde	µg/L	0.0002	0.002	0.02	2
1,1,1-Trichloroethane	µg/L	13	130	250	500
Trichloroethylene	µg/L	1.8	12	20	200
1,2,3-Trichlorobenzene	µg/L	0.3	3	25	50
1,2,4-Trichlorobenzene	µg/L	0.3	3	25	50
1,3,5-Trichlorobenzene	µg/L	0.3	3	25	50
2,3,5-Trichlorophenol	µg/L	0.05	0.5	2.2	4
2,3,6-Trichlorophenol	µg/L	0.05	0.5	2.2	4
2,4,5-Trichlorophenol	µg/L	0.05	0.5	2.2	4
2,4,6-Trichlorophenol	µg/L	0.05	0.5	2.2	4
3,4,5-Trichlorophenol	µg/L	0.05	0.5	2.2	4
Triphenyltin acetate	µg/L	0.002	0.02	0.2	2
Triphenyltin chloride	µg/L	0.002	0.02	0.2	2
Triphenyltin hydroxyde	µg/L	0.002	0.02	0.2	2
Meta-xylene	µg/L	0.1	1	10	1000
Ortho-xylene	µg/L	0.1	1	10	1000
Para-xylene	µg/L	0.1	1	10	1000

1238 **Appendix 2: Set of thresholds used by the WFD guide (MEEM, 2012) for**  
1239 **physico-chemical parameters classified according to pressure**  
1240 **categories** (*here for TEMP, ACID, BILO2, NUTRI: the threshold is always*  
1241 *excluded from the better class, e.g.: for nitrates, the threshold between the*  
1242 *high and good classes is 10 mg/L, therefore if  $[NO_3^-]=10$  mg/L, pressure*  
1243 *category is good; in contrast, for POSPE and SDP the threshold is always*  
1244 *included in the better class, e.g.: for arsenic, the threshold between the high*  
1245 *and bad classes is 4.2 $\mu$ g/L, therefore if  $[As]=4.2\mu$ g/L, pressure category is*  
1246 *high)*

Parameter	unit	thresholds between the statuses			
		high/good	good/moderate	moderate/poor	poor/bad
TEMP (Temperature)					
T	°C	24	25,5	27	28
ACID (Acidification parameters)					
pH		6,5	6	5,5	4,5
pH		8,2	9	9,5	10
BILO2 (Oxygen balance)					
O <sub>2</sub>	mg/L	8	6	4	3
%sat	%	90	70	50	30
BOD	mg/L O <sub>2</sub>	3	6	10	25
C	mg/L	5	7	10	15
NUTRI (Nutrients)					
PO <sub>4</sub> <sup>3-</sup>	mg/L	0,1	0,5	1	2
P	mg/L	0,05	0,2	0,5	1
NH <sub>4</sub> <sup>+</sup>	mg/L	0,1	0,5	2	5
NO <sub>2</sub> <sup>-</sup>	mg/L	0,1	0,3	0,5	1
NO <sub>3</sub> <sup>-</sup>	mg/L	10	50	<i>not yet defined</i>	<i>not yet defined</i>

Parameter	unit	status thresholds	status thresholds
		for annual average	for maximum concentrations
		high/bad	high/bad
POSPE (Specific pollutants)			
As	µg/L	4,2	
Cr	µg/L	3,4	
Cu	µg/L	1,4	
Zn	µg/L	7,8	
Chlortoluron	µg/L	5	
Oxadiazon	µg/L	0,75	
Linuron	µg/L	1	
2,4-D-ester	µg/L	1,5	
2,4-MCPA	µg/L	0,1	
SDP (Priority substances and priority hazardous substances)			
Alachlor	µg/L	0,3	0,7
Anthracene	µg/L	0,1	0,4
Atrazine	µg/L	0,6	2
Benzene	µg/L	10	50
Brominated dephenylethers	µg/L	0,0005	
Cd	µg/L	0,25	1,5
Carbon tetrachlorid	µg/L	12	
C10-C13-Chloroalkanes	µg/L	0,4	1,4
Chlorfenvinfos	µg/L	0,1	0,3
Chlorpyrifos (ethylchlorpyrifos)	µg/L	0,03	0,1
Cyclodiene pesticides sum	µg/L	0,01	
Total DDT	µg/L	0,025	
Para-para-DDT	µg/L	0,01	
1,2-Dichloroethane	µg/L	10	
Dichloromethane	µg/L	20	
Diuron	µg/L	0,2	1,8
Endosulfan	µg/L	0,005	0,01
Fluoranthene	µg/L	0,1	1
Hexachlorobenzene	µg/L	0,01	0,05
Hexachlorobutadiene	µg/L	0,1	0,6
Hexachlorocyclohexane	µg/L	0,02	0,04
Isoproturon	µg/L	0,3	1
Pb	µg/L	7,2	
Hg	µg/L	0,05	0,07
Naphtalene	µg/L	2,4	
Ni	µg/L	20	
Nonylphenol	µg/L	0,3	2
PAH	µg/L		
Benzo(a)pyrene	µg/L	0,05	0,1
Benzo(b)fluoranthene and benzo(k) fluoranthene sum	µg/L	0,03	
Benzo(g,h,i)perylene and Indeno(1,2,3-cd)pyrene	µg/L	0,002	
Simazine	µg/L	1	4
Tetrachloroethylene	µg/L	10	
Trichloroethylene	µg/L	10	
Tributyltin	µg/L	0,0002	0,0015
Trichlorobenzenes	µg/L	0,4	
Trichloromethane	µg/L	2,5	
Trifluraline	µg/L	0,03	