Application of a multidisciplinary and integrative weight-of-evidence approach to a 1-year monitoring survey of the Seine River

Iris Barjhoux, Lise C. Fechner, Jérémie D. Lebrun, Adriana Anzil, Sophie Ayrault, Hélène Budzinski, Jérôme Cachot, Laëtitia Charron, Arnaud Chaumot, Christelle Clérandeau, et al.

Application of a multidisciplinary and integrative Weight-of-Evidence approach to a one-year monitoring survey of the Seine River

Iris Barjhoux[1,✉], Lise C. Fechner[2,3], Jérémie D. Lebrun[2], Adriana Anzil[4], Sophie Ayrault[5], Hélène Budzinski[6], Jérôme Cachot[6], Laetitia Charron[1], Arnaud Chaumot[7], Christelle Clérandeau[6], Odile Dedourge-Geffard[1], Juliette Faburé[2,3], Adeline François[7], Olivier Geffard[7], Isabelle George[4], Pierre Labadie[6], Yves Lévi[8], Gabriel Munoz[6], Patrice Noury[7], Lucie Oziol[8], Hervé Quéau[7], Pierre Servais[4], Emmanuelle Uher[2], Nastassia Urien[2], Alain Geffard[1]

[1] Université de Reims Champagne-Ardenne, UMR-I 02 Stress Environnementaux et Biosurveillance des milieux aquatiques (SEBIO), UFR SEN, Moulin de la Housse, BP 1039, Reims, France

[2] Irstea, UR Hydrosystèmes et bioprocédés Antony (HBAN), 1 rue Pierre-Gilles de Gennes, CS 10030, 97261 Antony Cedex, France

[3] AgroParisTech, F-75005 Paris, France

[4] Laboratoire Ecologie des Systèmes Aquatiques (ESA), Université Libre de Bruxelles, Bruxelles, Belgique

[5] UMR 8212 CNRS CEA UVSQ Laboratoire des Sciences du Climat et de l'Environnement (LSCE) / Institut Pierre Simon Laplace (IPSL), Gif-sur-Yvette, France

[6] UMR 5805 Environnements et Paléoenvironnements Océaniques et Continentaux (EPOC), Université de Bordeaux, Pessac, France

[7] UR Milieux Aquatiques, Ecologie et Pollutions (MAEP), Laboratoire d'écotoxicologie, Irstea, Villeurbanne, France

[8] Ecologie Systématique Evolution, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

✉ **Corresponding author**: Iris Barjhoux (PhD)

University of Reims Champagne-Ardenne,
UMR-I 02 INERIS-URCA-ULH SEBIO,
UFR SEN, Moulin de la Housse, BP 1039,

51687 Reims Cedex 2, France.
E-mail contact: iris.barjhoux@univ-reims.fr / irisbarjhoux@hotmail.com (permanent)
Phone/Fax: +33 (0)326913719/+33 (0)326913342

Acknowledgements

This work was supported by the French research program PIREN-Seine. The authors would also like to thank the Aquitaine Region, the European Union (CPER A2E project) and the European Regional Development Fund for their financial support. This study was also carried out within the framework of the COTE Cluster of Excellence (ANR-10-LABX-45). IdEx Bordeaux (ANR-10-IDEX-03-02) provided the PhD grant allocated to G. Munoz. Finally, the authors would like to thank Annie Buchwalter for English revision of the manuscript.

Abstract

Quality assessment of environments under high anthropogenic pressures such as the Seine Basin, subjected to complex and chronic inputs, can only be based on combined chemical and biological analyses. The present study integrates and summarizes a multidisciplinary dataset acquired throughout a one-year monitoring survey conducted at three workshop-sites along the Seine River (PIREN-Seine program), upstream and downstream of the Paris conurbation, during four seasonal campaigns, using a Weight-of-Evidence (WOE) approach. Sediment and water column chemical analyses, bioaccumulation levels and biomarker responses in caged gammarids, and laboratory (eco)toxicity bioassays were integrated into four lines of evidence (LOEs). Results from each LOE clearly reflected an anthropogenic gradient, with contamination levels and biological effects increasing from upstream to downstream of Paris, in good agreement with the variations in the structure and composition of bacterial communities from the water column. Based on annual average data, the global hazard was summarized as ‘moderate’ at the upstream station and as ‘major’ at the two downstream ones. Seasonal variability was also highlighted; the winter campaign was least impacted. The model was notably improved using previously established reference and threshold values from national-scale studies. It undoubtedly represents a powerful practical tool to facilitate the decision-making processes of environment managers within the framework of an environmental risk assessment strategy.

Keywords: Seine River, Environmental risk, Weight-of-Evidence, Pollutants, Bioavailability, Biomarkers, Bioassays, Bacterial community

1. Introduction

The Seine river basin represents a catchment area of around 78,600 km² from its source at Seine-Source near Dijon in north-eastern France to its mouth in the English Channel in the northwestern city of Le Havre. It is supplied by a fairly regular network of tributaries. The central zone of the watershed is the convergence area of the main tributaries of the Seine River, and is occupied by the large Paris conurbation. In total, just over a quarter of the French population (~17.5 million) lives in this watershed, mostly (85%) in urban areas. The Seine watershed also harbors very intensive agricultural activities resulting in substantial diffuse sources of nutrients and pollutants such as pesticides (Billen et al. 2007). The heavy urbanization and industrialization of the Paris area also result in significant inputs of contaminants into the Seine River basin, including metals and persistent toxic substances such as polycyclic aromatic hydrocarbons (PAHs) and polychlorinated biphenyls (PCBs) (Blanchard et al. 2007; Thévenot et al. 2007). The wide variety of anthropogenic pressures that affect the Seine watershed makes this area an ideal case study of contemporary environmental problems in developed countries: the river's water quality and ecological status integrate and reflect the complex functioning of the watershed, especially the ways humans have shaped and exploited land- and water-scapes (Billen et al. 2007). However, the complexity and diversity of exogenous inputs result in an equally complex, diffuse, and chronic pressure on the ecosystems of the Seine continuum. The biological effects and long-term impacts of this pressure on the biota still remain difficult to evaluate directly.

The contamination level at a particular site can be quite readily determined through chemical analysis as defined by the presence of '*substances that would not normally occur or at concentrations above the natural background*'. However, 'pollution status' assessment additionally integrates chemical bioavailability and the biological impacts of contaminants on the environment (Chapman 2007). Consequently, it is now widely admitted that an efficient environmental risk assessment (ERA) should be conducted through an integrated and multidisciplinary strategy to provide answers to all these concerns. Moreover, such approaches are clearly recommended and even required by the European Water Framework Directive 2000/60/CE (European Commission (EC) 2000).

Between 2011 and 2012, eight research teams collaborated on a synchronous and integrative multi-marker approach aiming at a global assessment of the chemical and ecological/ecotoxicological status of three workshop-sites along the Seine axis situated upstream and downstream of Paris (PIREN-Seine program) (Fig. 1). This one-year monitoring program consisted of four field measurement campaigns corresponding to distinct seasons. During each sampling period, a wide panel of biological and chemical analyses was performed to characterize in detail the quality of the aquatic environment at each sampling location based on six fundamental aspects: (*i*) the physico-chemical quality of the water column and sediment, (*ii*) a comprehensive analysis of metal and organic contaminants in the same two compartments, (*iii*) the bioaccumulation of contaminants of concern in field-transplanted

gammarids and river biofilms, (*iv*) the biological responses in these gammarids exposed *in situ*, (*v*) the spatio-temporal variations in autochthonous bacterial community composition and metal tolerance acquisition, and (*vi*) the (eco)toxicity of water and sediment samples in laboratory bioassays. In total, about 550 parameters were monitored *per* site and *per* sampling period. Thus, one of the main challenges was to find a way to summarize and interpret the large dataset issued from this multi-marker study.

To achieve this objective, the concept of Weight of Evidence (WOE) appeared as an adequate strategy for a global and integrative multidisciplinary assessment of environmental quality in the area, because it is based on the packaging of a wide variety of data within several lines of evidence (LOEs). In each LOE the contamination level assessed through chemical analyses is combined with bioavailability analyses and biological responses from key species and/or model organisms at different levels of biological organization (Chapman et al. 2002; Dagnino et al. 2008). The resulting environmental diagnosis is based on the calculation of a hazard index for each LOE, which is then plotted on an evaluation grid allowing for clear and rapid hazard classification (Chapman et al. 2002; Dagnino et al. 2008; Piva et al. 2011). We also propose a global hazard evaluation that compiles all calculated LOE indices within a single one that is also finally assigned to a hazard class. This approach was successfully applied to the assessment of the health status in multi-contaminated environments such as harbors, urbanized and/or industrial areas (*e.g.*, Piva et al. 2011; Benedetti et al. 2012; Bebianno et al. 2015). These studies generally focused on sediment hazard assessment. However, the WOE approach is also applicable to other matrices such as effluents, water, and soils (Chapman et al. 2002; Chapman 2007; Dagnino et al. 2008), and to more global environmental diagnosis such as aquatic and terrestrial hazard assessment (Chapman et al. 2002; Chapman 2007; Piva et al. 2011).

Relying on the above-mentioned promising applications of the WOE procedure to sediment hazard assessment, the present study implements this multicriteria-based environmental diagnosis to a one-year monitoring survey of the Seine River axis. The three model sites investigated in the present work are situated along the Seine River continuum and characterized by a strong contamination gradient from upstream to downstream of the Paris urban area (Priadi et al. 2011; Fechner et al. 2012; Teil et al. 2014). As the WOE model described by Piva et al. (2011) relies on the calculation of Ratio-to-Reference (RTR) values, only part of the data obtained during the 2011-2012 campaigns was selected. The upstream station is often considered as a 'reference' site in similar case studies. This station was expected to be relatively unaffected by direct inputs from the Paris conurbation, but it was probably impacted by the intense agricultural activities surrounding the densely urbanized area, as well as by domestic or industrial inputs from the relatively small cities located upstream. The use of 'external reference' values established from several field-monitoring campaigns and physiological studies on the selected sentinel organisms would make it possible to classify all sites, including the upstream one. To that purpose, biomarkers and bioaccumulation levels were analyzed through an active approach in

transplanted *Gammarus fossarum* crustaceans, a common species in the field of ecotoxicology and biomonitoring. Gammarids have been reported as efficient accumulators of organic compounds and metals, whether essential or not (Besse et al. 2013; Lebrun et al. 2014). Besides, they are commonly used for the development of exposure biomarkers because they are easily sampled in the field and handled (Besse et al. 2013; Dedourge-Geffard et al. 2013; Lebrun et al. 2014). Moreover, translocated gammarid populations have been fully characterized, and reference levels are (or could be) established for bioaccumulation levels and biomarker responses (Xuereb et al. 2009; Geffard et al. 2010; Coulaud et al. 2011; Besse et al. 2013; Charron et al. 2013). According to the existence of reference levels or the possibility to derive them for each endpoint, the dataset selected for WOE integration was the following:

(1) Chemical hazard (LOE#1) was characterized through pesticide (PEST), alkylphenol (AKP), metal element (ME), and perfluoroalkyl substance (PFAS) analysis in the water column. MEs and PFASs were also measured in composite sediment samples together with more hydrophobic and (very) persistent compounds such as PAHs, PCBs, polybromodiphenyl ethers (PBDEs), and organochlorine pesticides (OCPs);

(2) Bioavailability (LOE#2) of the chemicals of concern, including PAHs, PCBs, PBDEs, OCPs, and MEs, was assessed by measuring bioaccumulation levels in caged gammarids;

(3) Biological responses (LOE#3) in the same population of transplanted gammarids were assessed using validated biomarkers such as digestive enzyme activity, feeding rate, reproductive toxicity, and acetylcholinesterase (AChE) activity;

(4) (Eco)toxicological responses at the organism/cellular level were also investigated using laboratory bioassays (LOE#4) performed on water column and sediment samples. They included genotoxicity, cytotoxicity, and endocrine disruption (ED) *in vitro* bioassays, as well as a fish embryo toxicity test, the Medaka embryo-larval assay (MELA).

The overall aim of this study was to confirm the importance and relevance of a multidisciplinary survey of aquatic environment quality. Such an approach is not realistically applicable in an ERA strategy without a practical tool to integrate and fruitfully interpret the resulting large dataset within a global environmental context. The present work applies the WOE model, adapted from Piva et al. (2011), to a practical case study on the Seine River continuum. The aim of this integrative approach is to assess the overall quality of the aquatic environment and prioritize hazards at each of the three sites. Such an approach could represent a promising decision-making tool for environmental managers.

2. Materials and Methods

2.1. Studied area and sampling procedure

The three sampling sites were previously described by Lebrun et al. (2015) and Faburé et al. (2015). Briefly, these sites are situated along the Seine River in the north of France (Fig. 1). Marnay (48°31'35.8" N, 3°33'29.6" E) is located approximatively 200 km upstream of Paris, in a non-urbanized area, and therefore expected to be at least partially free from direct inputs from the Paris conurbation. Conversely, Bougival (48°52'11.2" N, 2°07'47.1" E) and Triel (48°58'55.5" N, 1°59'53.1" E) are both situated downstream of Paris and its conurbation, at respective distances of approximately 40 and 80 km (Fig. 1). These stations are affected by various contamination inputs in relation to intense anthropogenic activities (Priadi et al. 2011, Teil et al. 2014). Sampling was performed at these sites during four campaigns undertaken in fall (C1 campaign, from August 31st to September 27th, 2011), spring (C2 campaign, from March 2nd to April 3rd, 2012), summer (C3 campaign, from June 1st to July 3rd, 2012), and winter (C4 campaign, from November 13th to December 18th, 2012), corresponding to contrasted temperature and flow rate conditions.

Dissolved metal concentrations were determined at the three sites during each sampling period (Faburé et al. 2015; Lebrun et al. 2015). At the end of each seasonal campaign, water was collected at each station as follows: 1 L of raw water in amber glass bottles for endocrine disruption bioassays, 10 L of raw water in high-density polyethylene (HDPE) containers for MELA, 20 L of raw water in two 10-L-HDPE containers for microbial community analyses. All containers were rinsed three times with river water before being filled in the field; they were all brought back to the laboratory in a cool box and then kept at 4 °C until further use. In addition, two 250-mL-HDPE bottles were filled in a similar way and stored at -20 °C for organic contaminant analyses. Prior to analyses, samples were thawed and filtered through GF/F (0.7 µm) Whatman glass microfiber filters previously ignited at 450°C for 6 h.

For each campaign (except C3), one composite surface (0–2 cm) bed sediment sample was collected in an aluminum container, brought back to the laboratory in a cool box, and stored either at 4 °C for bioassays or at -20 °C until freeze-drying, grinding and 2 mm-sieving for chemical analyses.

2.2. Field-caged gammarid exposure

The procedures are detailed in previous studies (Coulaud et al. 2011; Besse et al. 2013; Lebrun et al. 2015). Briefly, gammarids (*Gammarus fossarum*) were collected by kick sampling at La Tour du Pin, upstream of the Bourbre River (France). This site displays good water quality according to the data records of the RNB (French Watershed Biomonitoring Network). After a 15-day acclimatization

158 period in the laboratory (conditions detailed in Besse et al. (2013)) and 24 hours before *in situ* caging,
159 8 groups of 20 adult gammarids (10-11 mm) were caged in polypropylene cylinders (10 cm length, 5.5
160 cm diameter) capped at the ends with pieces of net (1 mm mesh) to ensure free water circulation. To
161 assess the effects on reproduction, three supplemental experimental systems, each containing seven
162 precopulatory pairs with D2-molt stage females (*i.e.*, hatched juveniles in brood pouches and visible
163 oocytes) were set up. A temperature probe was placed in the water to record temperature every hour
164 throughout the experiment. During the tests, gammarids were fed with the same alder (*Alnus*
165 *glutinosa*) leaves as during the acclimatization period in the laboratory, pre-conditioned for at least
166 6 ± 1 days in groundwater.

167 After 7 days of exposure, two replicates were collected and brought to the laboratory for
168 bioaccumulation measurements in whole organisms (3 pools of 5 gammarids) for each site. After 15
169 days of exposure, three replicates *per* site were collected and brought to the laboratory. Gammarids
170 from the same site were collected, counted (for survival rate assessment), then male gammarids were
171 pooled, dried, weighed, flash frozen in liquid nitrogen and stored at −80 °C until digestive enzyme
172 activity and AChE activity were analyzed. Leaf consumption was used to estimate the feeding rate for
173 each site and campaign. After 30 days of exposure, the last three replicates were collected and brought
174 to the laboratory. Gammarids from the same site were pooled together and counted (for survival rate
175 assessment); females were then selected to analyze reproduction markers (molt delay, number of
176 embryos/oocytes *per* female).

177

178 ### *2.3. Chemical analyses*

179 #### 2.3.1. Metal element measurements

180 For metal determination in caged gammarids, 3 pools of 5 individuals were digested by HNO_3 and
181 H_2O_2, as detailed by Lebrun et al. (2015). A reference material (Mussel Tissue ERM-CE278, LGC
182 Promochem, Molsheim, France) was included in each digestion series to control the quality of
183 digestion.

184 About 0.1 g of sediment was mineralized in closed Teflon vessels under a hood using a heating block
185 (Digiprep, SCP Science). A three-step digestion was performed as described by Priadi et al. (2011). A
186 geostandard was included in each digestion series (IAEA lake sediment SL1) to control chemical
187 mineralization efficiency. All reagents used for the digestion processes were ultrapure reagents to
188 avoid contamination.

189 Major and minor element concentrations were determined in filtered acidified water and in digested
190 sediment and gammarid solutions by inductively coupled plasma quadrupolar mass spectrometry

(ICP-QMS; X-Series, CCT II+ Thermoelectron, France), as previously described (Faburé et al. 2015; Lebrun et al. 2015, Le Pape et al. 2012, Priadi et al. 2011). Accuracy checking (SRM 1640a, NIST, Gaithersburg, USA) and plasma fluctuation corrections were also performed as described in the same references.

2.3.2. Organic compound analysis

Organic micropollutants were determined using previously established methods. Briefly, dissolved (<0.7 µm fraction) pesticides and PFASs were extracted using solid phase extraction with polymeric sorbents (100-500 mL samples) followed by analysis by liquid chromatography coupled to tandem mass spectrometry (Dufour et al. 2015; Munoz et al. 2015), while alkylphenols were determined using solid-phase microextraction (SPME) and gas chromatography coupled to mass spectrometry (Belles et al. 2014). Freeze-dried sediment (1 g) or gammarid (0.2 g) samples were extracted using microwave-assisted extraction followed by solid phase extraction adsorption chromatography clean-up (Budzinski et al. 2000; Nouira et al. 2013; Munoz et al. 2015).

2.4. Biomarker analysis

2.4.1. Digestive enzyme activity

The enzymatic activity of two carbohydrases (amylase and cellulase) and a protease (trypsin) was determined as previously described by Charron et al. (2013), using starch (1%), carboxymethyl-cellulose (2%), and N-benzoyl-DL-arginine 4-nitroanilide hydrochloride (3 mM) as substrates, respectively.

2.4.2. AChE activity

AChE activity was analyzed as described in Xuereb et al. (2009) according to the colorimetric method initially developed by Ellman et al. (1961), with DTNB (5,5'-dithiobis(2-nitrobenzoic acid) as a substrate.

2.4.3. Feeding rate assessment

Feeding rates were calculated according to the method described by Coulaud et al. (2011). Calculations were based on leaf disc scanning and expressed as consumed surface *per* day *per* living gammarid ($mm^2\ d^{-1}\ organism^{-1}$).

2.4.4. Reproduction markers

At the end of the exposure period (30 days), the size, molting stage, number of oocytes and embryos *per* female were determined according to Geffard et al. (2010). To accurately assess the females' molt stages, the third and fourth periopod pairs (dactilopodite and protopodite) of females were cut off, mounted on a microscope slide with a coverslip, and their integumental morphogenesis was observed (x200) to discriminate among the five molt stages (AB, C1, C2, D1, and D2). The number of oocytes per female in C2/D1-molt stage was determined by *in vivo* observation of the two ovaries under a binocular microscope. In the same way, embryos of females bearing a brood in 2^{nd}, 3^{rd}, or 4^{th} embryonic stage were manually recovered from the marsupium, placed on a slide with water, and counted under a binocular microscope. Desynchronization between female molt stage and embryonic development stage was also recorded to assess delays in female molt cycle (Geffard et al. 2010).

2.5. Bioassays

2.5.1. Endocrine disruption *in vitro* bioassays

Endocrine disruption (ED) bioassays were conducted on organic extracts prepared in dimethylsulfoxide (DMSO) from water column samples (1 L) or from freeze-dried sediment samples (1 g) according to Jugan et al. (2009) and Kinani et al. (2010), respectively. Three luciferase reporter bioassays were used to evaluate the ED potential of organic extracts from sediment or water column: using MELN cells (Balaguer et al. 1999), PC-DR-LUC cells (Jugan et al. 2007), or MDA-kb2 cells (Wilson et al. 2002), we measured disruptions of the transcriptional activity of the estrogen receptor ERα (ER), of the thyroid receptor TRα1 (TR), and of the androgen (AR) and glucocorticoid (GR) receptors, respectively, by bioluminescence.

The results were expressed as fold induction in relative luminescence units (RLUs) as compared to luciferase activity of the solvent control (DMSO 0.1%). Only RLU values significantly different from that of the solvent control (Student's *t*-test, $p < 0.05$) were considered as above the LD. Any detectable RLU levels above the bottom value of the sigmoidal dose-response curves of reference ligands were considered as above the LQ. This threshold value of the sigmoid was obtained by nonlinear regression of the Hill equation (GraphPad Prism 5 Software, San Diego, CA, USA). Furthermore, only RLU

levels significantly different from that of the corresponding blank value (Student's *t*-test, $p < 0.05$) were taken into account.

2.5.2. Microtox® and SOS Chromotest procedures

The two bioassays were performed on sediment elutriates to measure the toxicity of water-extractable pollutants. After thawing overnight at 4 °C, 6 g wet weight of sediment were mixed with 24 mL of deionized water for 10 min at 300 rpm. The solid phase was pelleted at 1,800 × g for 10 min, and the supernatant was immediately collected and stored at 4 °C in the dark prior to toxicity testing within 24 h.

For the Microtox® assay, the standard procedure of the Acute Toxicity Basic Test was used (AZUR Environmental 1998; ISO 1999). Bioluminescence was measured after 30 min on duplicated series of elutriate serial dilutions using a Microtox Model 500 analyzer (Azur Environmental).

The SOS-Chromotest developed by Quillardet and Hofnung (1985) was miniaturized in microplates. Briefly, *E. coli* PQ37 strain was exposed to the elutriate (3% final, v/v) for 3 h at 37 °C, in triplicate, with and without the liver S9 fraction (10% final, v/v) from β-naphthoflavone- and phenobarbital-treated rats (Trinova-Biochem). Following exposure, beta galactosidase (BG) and alkaline phosphatase (AP) activity levels were measured colorimetrically at 420 nm (Fluo Star Optima, BMG Labtech). The SOS control-relative induction factor (IF) was calculated by dividing the BG/AP activity ratio of the sample by the solvent control BG/AP ratio, as described by Quillardet and Hofnung (1985). Results were expressed as mean induction factor ± standard deviation (three replicates).

2.5.3. Medaka embryo-larval assay (MELA)

Japanese Medaka (*Oryzias latipes*) embryos of the CAB strain were provided by the UMS Amagen (Gif-sur-Yvette, France) 1 day post fertilization (dpf).

Whole sediment toxicity was evaluated by the Medaka Embryo-Larval Assay in sediment contact (MELAc), using the protocol described by Barhoumi et al. (2016). Reference non-contaminated sediment (Yville-sur-Seine) was used as a negative control (Vicquelin et al. 2011). Briefly, 25 embryos *per* replicate were laid onto a Nitex® mesh at the sediment surface and immerged into egg-rearing solution (ERS). To avoid hypoxia at the sediment-water interface, ERS was thoroughly renewed and dissolved oxygen was measured daily.

The toxicity of water samples was evaluated by the Medaka Embryo-Larval Assay in 96-well microplates, adapted from Helmstetter and Alden (1995). Before testing, the water samples were filtered through 0.8 µm-filters (Millipore) to remove particles. Twenty-five embryos *per* condition were individually incubated in 300 µL of water sample. Water was renewed daily, and spring water (Cristaline) was used as a negative control.

The procedure was similar for the two assays, and followed previously published protocols (Vicquelin et al. 2011; Barjhoux et al. 2012; Barhoumi et al. 2016). In summary, exposure was performed at 26 ± 0.3 °C, and stopped at the first hatching peak in one of the test conditions (10-11 dpf). Hatchlings and unhatched embryos were transferred to clean water or ERS, respectively, for three additional days. Viability, time to hatch, hatching success, body and head length, and developmental abnormalities were recorded in embryos and larvae according to Barjhoux et al. (2012).

2.6. Bacterial community composition

An aliquot (from 0.8 L to 5.0 L) of each water sample was filtered through a 0.2 µm pore-size, 47-mm-diameter polycarbonate filter (Millipore, MA). All filters were stored at -20 °C until use. DNA was extracted using phenol-chloroform-isoamylic alcohol, following an enzymatic cell lysis stage in the presence of lysozyme, mutanolysine and sodium dodecyl sulfate. Bacterial community structure (number and relative abundance of the different taxa) was assessed by pyrosequencing of the V1-V3 region of the bacterial 16S rRNA gene, and downstream sequence analysis was performed using the software program MOTHUR (full procedure described in García-Armisen et al. 2014).

Using the PRIMER v6 software program, we compared bacterial community structures among samples based on Bray-Curtis coefficient matrix, after square-root transformation of the data. This coefficient evaluates the dissimilarity between each pair of samples in terms of species abundance. The resulting matrix was used as a basis for a graphic representation of dissimilarities in a non-metric multi-dimensional scaling (NMDS) graph, where each sample was represented by a dot; the more different the structures of two bacterial communities, the further apart the two corresponding dots on the graph.

2.7. Data integration within the Weight-of-evidence (WOE) approach

The data selected to characterize contamination levels (*i.e.*, chemical analyses) in the area, contaminant bioavailability (*i.e.*, bioaccumulation levels in caged gammarids), and *in situ* biological responses (*i.e.*, biomarkers in gammarids) and following laboratory exposure (*i.e*, bioassays) were

integrated into a WOE approach according to Piva et al. (2011). Slight modifications and/or adaptations were made and are described below.

2.7.1. Line of evidence 1: sediment and water column chemistry (LOE#1)

Among the 210 metal and organic compounds analyzed in the abiotic compartment, we selected chemicals to be included in the WOE approach according to their mention in reference studies (MacDonald et al. 2000; Piva et al. 2011), and French and European regulatory documents (EC 2000, 2013; French Ministry of Ecology Energy Sustainable Development and Planning (MEDAD) 2007, 2015). The reference values used in LOE#1 calculations were environmental quality standards (EQSs) or environmental guideline values (EGVs) of the Ineris (French National Institute for Environmental Technology and Hazards) when available. Otherwise, Predicted No Effect Concentrations (PNECs) were gathered from environmental institutes recognized at the European level (Environment Agency, Ineris, Anses, and European Chemical Agency) and key reports (MacDonald et al. 2000; European Union (EU) 2005; MEDAD 2006; Dulio and Andres 2014). Then, the geometric mean of the PNECs was used as the reference value. Details on the selected reference values and the list of chemicals are given in Table S1 (for the water column) and Table S2 (for the sediment) in the Electronic Supplementary Material (ESM). Note that for both metal and organic compound analysis, data below the limit of detection (LD) were set at LD/2 before being integrated into calculations. Moreover, when the reference value was lower than the corresponding LD/2, measured concentrations below the LD were removed from the dataset.

The detailed calculation procedure implemented in LOE#1 is presented in ESM Fig. S1. As described by Piva et al. (2011), the elaboration of the chemical data into the corresponding hazard quotient (HQ) was based on the calculation of a ratio-to-reference (RTR) for each chemical and its weighting (RTR_w) according to chemical status within the Water Framework Directive (WFD) 2013/39/EU (EC 2013) (see ESM Tables S1 and S2).

The global Chemical Hazard Quotient (ChemHQ) was calculated by averaging the RTR_w values for chemicals whose measured concentrations were below or equal to the reference level (*i.e.*, $RTR \leq 1$) and summing the RTR_w values for chemicals whose concentrations exceeded the reference (*i.e.*, $RTR > 1$). With this calculation procedure, the resulting ChemHQ value increases with the number and the magnitude of exceeding endpoints, but is not strongly influenced by the number of chemicals whose concentrations are below the respective reference levels (Piva et al. 2011). This quotient was calculated for each site and each campaign, for the water column ($ChemHQ_{water}$) and the sediment ($ChemHQ_{sed}$). A hazard class was finally assigned to each calculated ChemHQ value according to the hazard classification grid established by Piva et al. (2011).

The contribution of each chemical and class of substance to the ChemHQ value was also calculated.

2.7.2. Line of evidence 2: bioavailability (LOE#2)

Bioavailability was assessed through bioaccumulation measurements in whole gammarids. Concentrations measured in gammarids following the caging procedure with control food supply have to be regarded as mainly proceeding from the water column rather than from the trophic route. Besse et al. (2013) used the same experimental conditions to study the bioaccumulation levels of 11 MEs and 38 hydrophobic organic compounds (including PAHs, PCBs, PBDEs, and OCPs) in gammarids *G. fossarum* of the same geographical origin as those used in the present study. In particular, the authors established bioaccumulation thresholds for 35 substances permitting to reveal bioavailable contamination of the environment when values go beyond these reference levels. These threshold tissue concentrations were used to derive the reference levels for bioaccumulation data in the present study. The list of the selected chemicals for LOE#2 and their respective reference values are presented in ESM Table S3.

Based on the procedure of Piva et al. (2011), the calculation method applied to LOE#2 was quite similar to LOE#1, with the calculation of an RTR value for each chemical and the weighting of these values according to the status of that chemical within the WFD (see Section *2.7.1*). One of the differences between the two LOEs is related to the addition of a correction function (Z(i)) to take into account the significance of the deviations from the reference values (Piva et al. 2011). As three replicates were not available for every compound and reference value, we set the correction function Z(i) as a fixed factor, as proposed by the authors. A hazard class was attributed to each resulting RTR_w value, as described by the authors.

The Bioavailability Hazard Quotient (BioavHQ) for each site and each campaign was calculated by averaging the RTR_w values whose relative hazard was classified as ‘slight’ and summing the RTR_w values with a ‘moderate’ to ‘severe’ hazard class, following the same reasoning as applied in LOE#1 (see Section *2.7.1*). A global hazard class for bioavailability was attributed to each BioavHQ, as described by Piva et al. (2011). As in the case of LOE#1, the contribution of each chemical and class of substance to the BioavHQ value was calculated. Details on the complete calculation procedure implemented in LOE#2 are presented in ESM Fig. S2.

2.7.3. Line of evidence 3: biomarkers (LOE#3)

The calculation method described by Piva et al. (2011) for biomarker LOE required a reference value (or control value) and an effect (inhibition and/or induction) threshold for each biomarker.

Only unilateral differences in comparison to reference values were taken into account, in agreement with their biological significance. As a result, only inhibition responses were taken into account for AChE activity, feeding rate, digestive enzyme activity levels, and the number of oocytes and embryos *per* female. In contrast, only induction effects were taken into account for mortality and molt delay endpoints. However, bilateral differences could be taken into account in the case of biomarkers for which both induction and inhibition responses have an (eco)toxicological/biological significance, as in LOE#4 for the time to hatch of embryos during the MELA (see Section *2.7.4*).

The calculations of reference and threshold values for gammarid **AChE activities and feeding rates** were adapted from Xuereb et al. (2009) and Coulaud et al. (2011). They were based on gammarid weight for AChE activity, and on the size of encaged gammarids and the mean temperature during *in situ* exposure for feeding rates. The thresholds (Th) were calculated according to the unilateral lower limit of the 95% confidence interval (CI95%) of the corresponding reference value (A. Chaumot, *personal communication*):

$$Th\ (\%) = \frac{Ref.value - CI95\%lower\ limit}{Ref.value} \times 100$$

For **digestive enzyme activity levels**, mean reference values were taken from Charron et al. (2013). The corresponding thresholds were calculated as described above for AChE activity levels and feeding rates.

The reference value for the **number of oocytes/embryos *per* female** after normalization of female size was adapted from Geffard et al. (2010). The corresponding thresholds were calculated as described above.

For the percentage of females with a **molt delay**, we expected the reference value to be 0%, and calculated the threshold value as the percentage representing the presence of 2 asynchronous females within a batch of 15 females *per* site and *per* campaign (*i.e.*, 13.3%). As a value equal to '0' is not acceptable in the calculations of Biomarkers Hazard Quotient (BiomHQ; see details for calculations below), '100' was added to the reference value (thus equal to 100%) and each measured molt delay (the threshold remained unchanged).

The reference value, threshold, and effect retained for each biomarker are reported in ESM Table S4. The complete calculation procedure was adapted from Piva et al. (2011), with a few modifications, and is described in ESM Fig. S3.

Briefly, for each biomarker response, we calculated the percentage of variation relatively to the reference (%VAR). The %VAR is then supposed to be corrected according to the statistical

significance of the difference between the reference value and the mean biomarker value (Z(i) function; Piva et al. 2011), resulting in an effect value (E(i)) for each endpoint. However, as the threshold values (Th) were pre-established using a statistical approach (*vs*. evaluated by 'expert judgement' in Piva et al. (2011)), we manually set the Z(i) function, as we did for bioavailability data (ESM Fig. S3). As mentioned above, only unilateral differences were taken into account. In other words, when only inhibition was considered as 'ecotoxicologically relevant' for a given biomarker, any induction effect was considered as 'within the reference range', resulting in an effect E(i) set at '0', and *vice versa*. Moreover, to take into account the fact that the reference value of some biomarkers could vary depending on exposure conditions over the year, it appeared more accurate to evaluate the annual average response of a biomarker by averaging the E(i) values calculated for each campaign (*per* station) rather than averaging the biomarker responses directly.

A hazard class was attributed to each E(i) value (ESM Table S5) according to the gradation scale of Piva et al. (2011) (ESM Fig. S3). The effect value was then weighted ($E_w(i)$) against the biological significance of the biomarker response (ESM Table S6). The Biomarker Hazard Quotient (BiomHQ) for each site and each campaign was calculated by averaging the $E_w(i)$ values for which E(i) relative hazard was classified as 'moderate', and summing the $E_w(i)$ values for which E(i) belonged to a 'major' to 'severe' hazard class (ESM Fig. S3). The procedure was based on the reasoning applied in LOE#1 (see Section *2.7.1*) and LOE#2.

Finally, a global hazard class for biomarkers was attributed to the BiomHQ value for each site and campaign, as proposed by Piva et al. (2011).

2.7.4. Line of evidence 4: bioassays (LOE#4)

The calculation method applied to derive the Bioassay Hazard Quotient (ToxHQ) was quite similar to the one used for BiomHQ (Piva et al. 2011), and is described in ESM Fig. S4.

The reference values used in the present study were the measurements from the negative control treatment of each bioassay. Responses from *in vitro* bioassays are usually expressed as induction or inhibition factors in comparison to the control. Thus, these data were just slightly modified to correspond to the percentage of variation relatively to the control value (%VAR(i)) defined in ESM Fig. S4. Afterwards the effect E(i) was calculated for each endpoint as described for biomarkers, using threshold (Th) values and a correction factor Z(i) (ESM Fig. S4).

For **MELA** results, Th values were calculated by examining the variability of the data from the negative control treatment and the associated CI95%, as described above for biomarkers such as AChE

activity (see Section *2.7.3*). The Z(i) function was consequently set similarly to BiomHQ calculations (see Section *2.7.3*).

For ***in vitro* bioassays**, the Th values were established based on expert judgement. For **Microtox®** results, we set the threshold and the Z(i) function according to acute toxicity levels based on the inhibition percentage (= %VAR(i)) of the bioluminescence recorded at the highest concentration. Thus, the Th value set at 10% represented the 'not toxic'/'moderately toxic' limit, according to our laboratory expertise and adapted from Bennett and Cubbage (1992) and Brouwer et al. (1990). As a result, data below this value were weighted by a Z(i) factor equal to 0.2 in the effect E(i) calculation. Similarly, bioluminescence inhibition factors above 10% and below 50% resulted in a weighting Z(i) value equal to 0.5. When the %VAR(i) was above 50% (*i.e.*, considered as 'clearly cytotoxic'), the Z(i) function was equal to 1.

A similar methodology was implemented for **SOS Chromotest** data using the threshold value established by Mersch-Sundermann et al. (1992). As a result, the Th value was set at 50%, in agreement with the induction factor threshold of 1.5 established as the 'not genotoxic'/'marginally genotoxic' limit by the authors. Moreover, the Z(i) function was set at (*i*) 0.2 for a non-significant value in comparison to the blank and/or for an induction factor below 1.5, (*ii*) 0.5 for an induction factor above 1.5 but strictly below 2, and (*iii*) 1 for an induction factor equal or superior to 2 (value above which effects could be considered as clearly 'genotoxic' according to Mersch-Sundermann et al. (1992)).

For **ED bioassays**, Th values of 50% and 10% were set for agonist (ER, TR, AR/GR) and antagonist (anti-AR) activities, respectively, according to our laboratory expertise (Lucie Oziol, *personal communication*). The Z(i) function was also manually set at (*i*) 0.2 for data below the LD or LQ values, (*ii*) 0.5 for data not significantly different from the blank value (according to Student *t*-test results with a 5% risk), (*iii*) 1 in all other cases.

Similarly to the reasoning applied in LOE#3 calculations, only unilateral differences in comparison to the reference value were taken into account, in agreement with the biological significance of each endpoint. Thus, any response with an effect other than the one defined as 'ecotoxicologically relevant' led to an effect E(i) set at 0, except for the time to hatch of medaka embryos for which bilateral differences were taken into account.

The reference value, threshold, and effect of each endpoint are reported in ESM Table S7. As in the case of biomarkers, the annual average response of a bioassay endpoint was obtained by averaging the E(i) values calculated for each campaign at each station. Each effect value E(i) was then weighted (E_w(i)) according to the corresponding bioassay endpoint. The weight of each response was defined as proposed by Piva et al. (2011), with slight modifications. The WOE approach was first proposed by these authors to assess sediment hazard in particular, with a low coefficient (0.3) for bioassays using

the water column as a test matrix. In the present study, we apply the procedure to both sediment and water column hazard assessment. Consequently, we chose to set the coefficient for water column testing similarly to what was used for sediment testing (*i.e.*, equal to 1 for total water and 0.8 for the water-dissolved fraction). Moreover, as an 'ecotoxicologically relevant' effect was identified for each marker and as 'contrary' effects were discarded from analysis, it seemed superfluous to weight endpoints according to the possibility of hormetic responses. Details on weighting calculations for each endpoint are given in ESM Table S8. Finally, the cumulative ToxHQ was calculated as the sum of the $E_w(i)$, and a global hazard class for bioassays was attributed as described by Piva et al. (2011) (ESM Fig. S4). This hazard quotient was elaborated for each site and each campaign, for the water column ($ToxHQ_{water}$) and for the sediment ($ToxHQ_{sed}$).

2.7.5. Weight of Evidence integration

The complete calculation procedure implemented for WOE integration is detailed in ESM Fig. S5. As described by Piva et al. (2011), the first step of the integration of the HQs derived from the four LOEs within a global index (WOE index) consisted in normalizing HQ values to a common scale. The authors also proposed to ascribe different weightings to LOE results according to their environmental relevance. Thus, they chose to multiply BioavHQ indices by 1.2× to give greater importance to bioavailability data as compared to the presence of chemicals in the abiotic compartment (*i.e.*, ChemHQs, weighted by 1.0×). Similarly, they suggested to apply a 1.2×-coefficient to the data acquired using bioassays (ToxHQ indices) because they reflected acute effects at the organism level, whereas biomarker responses (BiomHQ indices) describing sublethal effects at the molecular scale remained weighted by 1×. The situation was somewhat different in the present study, since biomarkers included both responses at the molecular level (*e.g.*, enzyme activity levels) and life history traits (*e.g.*, feeding behavior and reproduction ability). As a result, it seemed more relevant to apply greater weightings to the results of the LOE related to disturbances of organisms exposed *in situ* than to organisms exposed under laboratory conditions. We thus chose to weight the BioavHQ and BiomHQ indices by 1.2×, whereas ChemHQs and ToxHQ indices were still weighted by 1×.

The resulting HQ indices from the four LOEs were summed up and normalized to 100% to yield an overall WOE hazard index. Finally, each WOE value was assigned to a hazard class, as described by Piva et al. (2011) (ESM Fig. S5).

3. Results and discussion

3.1. Water column and sediment chemistry: LOE#1

3.1.1. Chemical hazard quotient for the water column ($ChemHQ_{water}$)

The chemical hazard relative to water column contamination was evaluated according to the concentrations of 15 pesticides, 2 AKPs, 1 PFAS (PFOS), and 9 MEs measured in the dissolved fraction. They are listed in ESM Table S1.

The concentration of each chemical (ESM Table S9) was used to calculate an integrative contamination index, $ChemHQ_{water}$ (Table 1). The contribution of each class of compounds to the global chemical hazard is presented in Fig. 2a. These results show that perfluorooctanesulfonic acid (PFOS) was omnipresent in the area and contributed to 51% (Marnay C4) to 99% (Bougival C2) of the $ChemHQ_{water}$ values. Average PFOS concentrations were above the EQS value of 6.5×10^{-4} µg/L (EC 2013) at all sites and for all sampling campaigns as well as for calculated annual means; the RTR values increased along the anthropogenic gradient from 1.30-4.37 in Marnay, 4.85-22.8 in Bougival, and up to 5.88-35.2 in Triel (ESM Tables S1 and S9). At each site, the lowest RTR value was recorded during the C4 campaign, and the highest during the C1 campaign. In contrast, the contribution of pesticides increased at each site in the C4 campaign, with values around 48% at Marnay, 24% at Bougival and 22% at Triel. Among the pesticides, one compound –metazachlor, a chloroacetanilide herbicide- accounted for almost the total contribution of this class of chemicals (21% to 46%; data not shown). The contamination gradient previously mentioned for PFOS was also recorded in the case of metazachlor winter concentrations, with RTR values increasing from 1.55 in Marnay to 2.29 and 2.38 in Bougival and Triel, respectively (ESM Tables S1 and S9). Metazachlor is an herbicide commonly used in rapeseed crops and usually applied in late August/early September. This substance is considered as 'moderately sorbing', and several months might go by between its application date and its release in the surrounding waters, depending on the intensity of the rain events and the hydrodynamic characteristics of the watershed (Passeport et al. 2013).

As shown in Fig. 2a, metal elements only contributed noticeably to $ChemHQ_{water}$ values at the downstream sites, with the highest contributions recorded in the C4 campaign (around 11%), as shown for pesticides. Despite a clear contamination gradient along the Seine river, the dissolved concentrations of metals did not exceed, or only slightly exceeded their respective EQS at the two downstream sites, as previously described (Faburé et al. 2015; Lebrun et al. 2015), and in agreement with previous studies at the same sites (Fechner et al. 2012). Values exceeding the corresponding reference value were limited and almost strictly related to copper concentrations, with a maximal RTR value below 1.35 noted in Triel during the autumn campaign (ESM Tables S1 and S9).

Following the integration of overall contamination data measured in the water column, the resulting $ChemHQ_{water}$ values obviously reflected the anthropogenic gradient pressure along the Seine River axis, with values increasing from the Marnay upstream site to the Bougival and Triel downstream sites (Table 1). As a result, the chemical hazard for the annual average value was classified as 'moderate' at Marnay and as 'severe' at the two stations downstream of the Paris agglomeration. Seasonal variations in water column contamination were also evidenced, with lower $ChemHQ_{water}$ values at all sites in winter (C4), thus downgrading the hazard class for the most impacted sites from 'severe' to 'major'. In contrast, the highest $ChemHQ_{water}$ values were recorded in the fall season (C1) at each station (Table 1), possibly as a consequence of the lower dilution of point source discharges under low flow conditions in the River Seine.

3.1.2. Chemical hazard quotient for sediment ($ChemHQ_{sed}$)

Sediment contamination was assessed based on the concentrations of 1 PFAS (PFOS), 18 PAHs, 8 PCBs, 7 PBDEs, 12 OCPs, and 15 MEs. They are listed in ESM Table S2, and detailed in ESM Tables S10 and S11).

Considering that for some chemicals such as PAHs, PBDEs and DDTs, the reference values were available both for individual substances and for the total concentrations of the classes of compounds, it was possible to calculate $ChemHQ_{sed}$ values in two different ways (Table 1). The calculation method strongly influenced the relative contribution of each class of compounds to the calculated global index (Fig. 2b and 2c). When 'individual concentrations' were used, the most contributive chemical family was PAHs at each site, and sampling time (except for Bougival C1 and Triel C1/C4 samples) with total contributions varying from 75% to 99% in Marnay, 52% to 74% in Bougival, and 75% to 99% in Triel (Fig. 2b). Among this class of compounds, the corresponding reference values were widely exceeded for phenanthrene and pyrene, with RTR values respectively between 3.0-18.9 and 2.9-15.5 in Marnay, 15.6-57.3 and 18.6-75.4 in Bougival, and between 5.7-836 and 12.1-672 in Triel (ESM Tables S2 and S10). RTR values around 200 were also noted for anthracene and benzo[a]anthracene in Triel C3 samples (ESM Tables S2 and S10). As in the case of PAHs, OCPs were omnipresent in the area, with overall contributions reaching 25% at Marnay, 57% at Bougival, and 74% at Triel (Fig. 2b). These high contribution levels were mainly attributable to heptachlor concentrations that exceeded the reference value of 0.02 µg/kg dw 7.7- to 16.9-fold in Marnay, 46.7- to 136-fold in Bougival, and 109- to 487-fold in Triel (ESM Tables S2 and S10).

When PAH concentrations were summed (ΣPAHs) and compared to the reference value for total PAHs, the resulting RTR values were much lower than those described above for individual substances. They only varied from 0.35 to 1.70 in Marnay, 2.34 to 9.75 in Bougival, and 1.33 to 65.5

578 in Triel samples (ESM Tables S2 and S10). As a result, the contribution of PAHs to $ChemHQ_{sed}$
579 values was also low, with values around 10% or lower for all samples, except Triel C3 for which
580 ΣPAHs accounted for about 73% of the calculated hazard quotient (Fig. 2c). Consequently, the global
581 relative contributions of OCPs and MEs logically increased with this calculation method (Fig. 2c).

582 The reference values for MEs in sediment were only exceeded in downstream samples. These overruns
583 were systematic and particularly substantial for Cd, Cu, Pb, and Zn, with RTR values reaching 23.8,
584 34.1, 20.7, and 39.1, respectively (Triel and Bougival samples combined; ESM Tables S2 and S10).
585 Higher enrichment factors (in relation to the geochemical background) have been reported for these
586 elements in sediment cores sampled downstream of the Paris conurbation as compared to upstream
587 and Oise River sites (Le Cloarec et al. 2011). The sites downstream of Paris receive and integrate all
588 kinds of pollutants that affect the rest of the Seine Basin. They result from industrial (*e.g.*, foundries,
589 wire factories) and agricultural activities (*e.g.*, the use of $CuSO_4$ as a fungicide and bactericide in
590 vineyards), but also intense urbanization (*e.g.*, the use of leaded gasoline, leaching of old Zn roofs
591 following rainfalls, effluents from waste water treatment plants (WWTPs)) (Le Cloarec et al. 2011;
592 Ayrault et al. 2012). The situation is particularly worsened at sites such as Triel, situated downstream
593 of the Oise river confluence: Triel is not only affected by the inputs from the Paris suburbs and
594 upstream activities, but also by the high industrialization of the Oise basin and one of the most
595 important sewage plants of the Paris agglomeration, the Seine-Aval WWTP of Achères located on the
596 banks of the Seine River, between the Bougival and Triel sampling sites (Le Cloarec et al. 2011;
597 Ayrault et al. 2012).

598 Similarly to what was observed for the water column, the integrative $ChemHQ_{sed}$ values clearly
599 illustrated the expected contamination gradient; however, values calculated using individual substance
600 concentrations were significantly higher than those calculated using total concentrations (Table 1).
601 While the hazard class remained 'severe' for the two downstream stations using either calculation
602 method (for all campaigns and the annual average value), the hazard status at Marnay varied from
603 'severe' to 'absent' for the C2 sample and from 'moderate' to 'absent' for the C4 sample when the
604 $ChemHQ_{sed}$ value was calculated using individual substances or total concentrations, respectively
605 (Table 1). The annual average hazard quotient for sediment chemistry was calculated using the mean
606 concentration of each chemical from the three composite sediment samples collected in the field
607 during the C1, C3, and C4 campaigns. Depending on the calculation method, the annual average
608 hazard class at Marnay varied from 'severe' to 'major' (Table 1).

609 According to MacDonald et al. (2000), ΣPAHs can be efficiently used to predict sediment toxicity,
610 with no substantial difference with toxicity predictions based on individual PAH concentrations.
611 However, in the approach developed here, the use of the ΣPAHs did not allow us to get access to
612 information on the individual compounds involved in exceeding the reference value. Identifying them

could nonetheless be valuable to identify the specific chemicals involved in the biological effects highlighted in LOE#3 and LOE#4. The information could also be exploited during further investigations aimed at identifying the source(s) of this pollution. Moreover, we preferred to adopt the most conservative and protective approach for aquatic biota as regards the environmental hazard. As a result, the $ChemHQ_{sed}$ values calculated using individual concentrations were kept for the subsequent step, (*i.e.*, final integration into the WOE index).

3.2. Bioavailability of chemicals: LOE#2

Bioavailability Hazard Quotient (BioavHQ) values were calculated according to the bioaccumulated concentrations of 17 PAHs, 8 PCBs, 1 PBDE, 7 OCPs, and 4 MEs (listed in ESM Table S3, and detailed in ESM Table S12) in caged gammarids.

The relative contributions of each class of chemicals to the global BioavHQ are illustrated in Fig. 3a. Accumulation levels of organic compounds were not analyzed during the C4 campaign, and PAH and PBDE data are missing for the C1 campaign, so contributions to the BioavHQ values are only discussed based on annual average data.

Among trace metals, Ni was noticeably accumulated by gammarids exposed *in situ*, namely 1.7- to 5.4-fold higher than the reference level, with no specific variation among sites attributable to the anthropogenic gradient (ESM Tables S3 and S12). This could result from a non-identified diffuse contamination source, and/or Ni geochemical background differences between the native region of gammarids (the Rhône-Alpes region) and the Seine Basin. In the water column, the Ni background seemed to be slightly lower in the Rhône watershed (mainly between 0.58 and 2.51 µg/L, and more locally up to 3.93 µg/L) than in the Seine Basin (2.51-3.93 µg/L), according to the FOREGS Geochemical Baseline Mapping Program (http://weppi.gtk.fi/publ/foregsatlas/). Thus, while Ni locally reached similar background levels in the Rhône watershed as in the Seine Basin, the globally lower Ni geochemical background in the Rhône Basin may partially explain the higher Ni bioaccumulation levels in transplanted gammarids. Yet excess Ni as compared to the reference value reflected an increase in the Ni bioavailable fraction between the two areas whatever the exact origin (higher geochemical background and/or anthropogenic activities).

In contrast, whereas RTR values for Pb accumulation at Marnay remained around or below 1 (annual average: 0.94), they ranged between 1.7 and 7.8 in Bougival, and between 1.8 and 3.8 in Triel (ESM Tables S3 and S12), reflecting a significant increase in Pb bioaccumulation in gammarids downstream of Paris. Nevertheless, metals contributed only little (< 10%; Fig. 3a) to the annual average BioavHQ values of the downstream sites. As a result, ME accumulation in gammarids exposed along the Seine axis represented a limited hazard in comparison to organic compounds, as showed by the RTR_w-based

hazard status mainly evaluated as ‘absent’ or ‘slight’ for MEs (only one ‘major’ status and one ‘moderate’ status were recorded for Ni accumulation; ESM Table S3).

The annual average RTR_w-based hazard classification was ‘absent’ for all organic compounds at Marnay (except for PCB 118 classified as ‘slight’), whereas bioavailability-related hazard reached the ‘major’ to ‘severe’ grade at the downstream stations (ESM Table S3). More specifically, PAHs accounted for up to 87% and 72% in the calculation of annual average BioavHQ values in Bougival and Triel, respectively. Still considering annual average data, all PAHs significantly accumulated (at least 2.5-fold; ESM Tables S3 and S12) in gammarids exposed at Bougival as compared to the reference levels. Among these compounds, respective reference bioaccumulation levels were exceeded around 10-fold or just above for acenaphtene, anthracene, benzo[e]pyrene and phenanthrene, nearly 20-fold for benzo[a]anthracene and chrysene/triphenylene, and an extreme >30-fold for fluoranthene and pyrene (ESM Tables S3 and S12). Annual average RTR values were lower in Triel-exposed gammarids, mainly falling between 1.2 and 9.2 (ESM Tables S3 and S12). Bioaccumulation levels more than 10-fold the reference values were only recorded for fluoranthene (11.1×) and pyrene (15.4×) at that site (ESM Tables S3 and S12), confirming that these compounds were the most bioaccumulated ones when compared to the reference levels.

These results are in overall agreement with the PAH concentrations measured in sediment, since particularly high RTR values were recorded for anthracene, phenanthrene, benzo[a]anthracene, and pyrene (see Section *3.1.2*). However, PAH bioaccumulation levels were higher at Bougival than at Triel, whereas PAHs were globally more abundant in Triel sediment than in Bougival sediment (organic carbon-normalized concentrations increased 1.7-fold to 5.4-fold between the two sites depending on the sampling campaign; ESM Table S11). This was likely related to variations in PAH bioavailability between the two sites.

PCBs ranked second among the chemicals contributing to the annual average BioavHQ values at the downstream sites, *i.e.* 11% in Bougival and 15% in Triel (Fig. 3a). The annual average RTR values ranged between 2.7 and 6.7 in Bougival, and between 2.5 and 5.1 in Triel, depending on the congener (ESM Tables S3 and S12). As a result, the RTR_w-based hazard was ‘moderate’ for PCB 50+28, PCB 52 and PCB 101, and ‘major’ for PCB 118 in Bougival (ESM Table S3). As regards Triel, only PCB 101 and PCB 118 accumulation levels represented a substantial hazard (*i.e.*, above the ‘slight’ status) evaluated as ‘moderate’ and ‘major’, respectively.

The overall LOE#2 results tend to suggest that PAHs were the main problematic class of compounds regarding their potential bioavailability/bioaccumulation in aquatic organisms in the two stations located downstream of Paris.

The integration of overall bioavailability data resulted in lower BioavHQ values for the winter campaign (C4) at all sites (Table 2). The hazard associated with bioaccumulation levels was consequently evaluated as 'slight' for this campaign. However, the conclusion on the C4 campaign should be interpreted carefully because it was only based on ME accumulation, and ME accumulation proved to be limited as compared to organic compounds at the other sampling periods.

With BioavHQ values varying between 1.8 and 13.4, the related hazard was also classified as 'slight' in Marnay for each sampling period as well as for the annual average value (Table 2). In contrast, the highest BioavHQ values were recorded in Bougival (except for C4) and the associated hazard was thus evaluated as 'major' (Table 2). The hazard was also identified as 'major' during the fall campaign (C1) at Triel; however, it decreased to 'moderate' for the other campaigns (C2 and C3) as well as for the annual average BioavHQ (Table 2).

3.3. Biomarker responses in gammarids exposed in situ*: LOE#3*

Biological responses in gammarids exposed *in situ* along the Seine axis were investigated using several biomarkers of various physiological impairments, including neurotoxicity, energy acquisition disturbance, feeding behavior impairment, reproduction dysfunctioning/failure, and survival (ESM Table S13). These markers have been studied in our laboratories for several years and are commonly used in laboratory and field experiments (Dedourge-Geffard et al. 2013; Xuereb et al. 2009; Geffard et al. 2010; Coulaud et al. 2011; Charron et al. 2013; Chaumot et al. 2015). The substantial insights gained from these numerous studies more particularly allowed us to *(i)* fully determine the basal level and variation range of each marker, *(ii)* identify and characterize the confounding factors that may modulate biomarker responses, especially under field conditions, and thus *(iii)* determine specific reference levels and effect thresholds for these biomarkers adapted to *in situ* deployment conducted in the context of an environmental survey (ESM Table S4). These reference values and thresholds were used to calculate biomarker Hazard Quotients (BiomHQs) applied to the Seine axis case study (Table 3).

The contribution of each category of markers to the calculated BiomHQ values is illustrated in Fig. 3b. This analysis is completed by the hazard class attributed to each biomarker response, presented in ESM Table S5.

No sign of neurotoxicity was highlighted: no significant inhibition of AChE activity was recorded in comparison to the established reference value (ESM Tables S4 and S13); the hazard relative to the AChE marker was 'absent' for all sites and all sampling campaigns (ESM Table S5).

Unlike AChE activity, the decrease in gammarid survival represented a 'severe' hazard at all sites and all sampling periods (ESM Table S5), except in the C4 campaign during which no significant mortality was noted (<4% in all stations; ESM Table S13). This acute toxicity endpoint was the only that showed an effect E(i) above 1 at Marnay (representing a 1.8- to 3.2-fold increase in comparison to the reference value, ESM Tables S4 and S13), therefore it contributed 100% to the calculated BiomHQ value for that station (Fig. 3b). The impact on gammarid survival at Marnay raises the question whether some early (sublethal) responses of other physiological functions, not addressed in the present study, could potentially exist. For instance, studying the impacts of exposure on the immune system and the inflammatory mechanisms could be of great interest by bringing supplemental data to the currently investigated biomarkers. Due to their direct implications in individual fitness, population and ecosystem health (Bols et al. 2001), immunomarkers are considered as attractive non-specific markers that could be consistently integrated into ERA and biomonitoring surveys (Bado-Nilles et al. 2015).

Mortality was modulated in the same order of magnitude at the downstream sites as at Marnay, with a 1.5- to 2.5-fold rise at Bougival, and a 1.6- to 3.0-fold rise at Triel (ESM Tables S4 and S13). However, it resulted in contributions to BiomHQ values that were more limited than at Marnay and ranged between 43% to 61% at Bougival, and 32% to 45% at Triel (Fig. 3b). Several other markers were significantly modulated in the gammarids exposed at the two downstream sites.

Digestive enzyme activities were particularly down-regulated in the C1 and C3 sampling periods, but were not significantly modulated during the winter (C4) campaign ('absent' hazard for all markers; ESM Table S5). Overall, inhibition of enzymatic activities ranged between 17% and 37% at Bougival and Triel during the fall (C1) and summer (C3) campaigns (ESM Tables S4 and S13). Special cases were recorded at Bougival: no significant inhibition of cellulase activity was noted during the C3 campaign, whereas trypsin inhibition increased by 60% of the reference value at the same time (ESM Tables S4 and S13). Still based on the digestive enzyme activities, the annual hazard (assessed by averaging the effect E(i) of the three sampling periods) related to energy acquisition parameters was evaluated as 'moderate' at Bougival, and ranked from 'moderate' to 'severe' at Triel (ESM Table S5). However, this class of markers only slightly contributed to the global hazard related to biomarker responses, with a maximal contribution barely exceeding 20% for the Triel C3 sampling point (Fig. 3b).

In contrast to energy acquisition markers, feeding rates were the most severely inhibited in gammarids exposed to the downstream stations during the winter (C4) campaign, with values representing 92% and 79% decreases in comparison to the corresponding reference value at Bougival and Triel, respectively (ESM Tables S4 and S13). These effects accounted for 100% of the calculated BiomHQ at the downstream sites during the C4 sampling period, and represented a 'severe' hazard (Fig. 3b,

747 ESM Table S5). Gammarid feeding activity was also repressed during the summer campaign, up to
748 21% at Bougival ('moderate' hazard) and 49% at Triel ('severe' hazard'). As a result, a 'major' annual
749 hazard was attributed to the effects on feeding behavior in gammarids exposed downstream of Paris
750 (ESM Table S5).

751 Among the reproductive impairment biomarkers, the number of oocytes *per* female did not decrease in
752 gammarids exposed along the Seine axis. Inversely, *in situ* exposure yielded values up to 2-fold higher
753 than the expected value (ESM Tables S4 and S13). Therefore the environmental hazard associated to
754 this marker was 'absent' for all sites and all campaigns (ESM Table S5). However, the number of
755 embryos *per* female decreased by 33% at Bougival and 71% at Triel during the C1 campaign (ESM
756 Tables S4 and S13), representing a 'severe' potential hazard to aquatic organisms (ESM Table S5).
757 Even more pronounced impacts were highlighted on molt delay, which noticeably increased during all
758 campaigns, by 27% to 57% at Bougival and by 40% to 50% at Triel (ESM Tables S4 and S13). The
759 overall environmental hazard was classified as 'severe' for both stations for this reproductive marker
760 (ESM Table S5). Reproductive impairments consequently represented the second most contributive
761 class of biomarkers to BiomHQ values for the C1 and C3 campaigns and the annual average value,
762 with contributions varying between 20% and 40-45% at Bougival and Triel (Fig. 3b).

763 All biomarkers were finally integrated into a global hazard quotient. The resulting BiomHQ values
764 clearly depicted the expected anthropogenic gradient, with values increasing from Marnay to Bougival
765 and then Triel, for each sampling campaign as well as for the annual average value (Table 3). The
766 lowest BiomHQ was recorded in the C4 campaign at all sites; the hazard was classified as 'slight'
767 (Marnay) or 'major' (Bougival, Triel). However, as in the case of bioavailability, these results should
768 be interpreted carefully because reproduction markers and cellulase activity were not investigated
769 during that campaign. The C4 data did not fulfil the minimum requisite advised by Piva et al. (2011) to
770 calculate the cumulative BiomHQ: only two markers had a weighting above 1.2, while the
771 recommended number is 3. Conversely, the highest BiomHQ values were noted in the fall (C1)
772 campaign (Table 3). This observation is consistent with the water column and sediment ChemHQ
773 values (using 'total concentrations'), which were also higher in C1 samples. In agreement with the
774 hazard class established from the other sampling periods, the overall level of risk (based on annual
775 average estimation) was identified as 'moderate' at Marnay, 'major' at Bougival, and 'severe' at Triel
776 (Table 3), reflecting a noticeable increase of physiological disturbances in gammarids exposed along
777 the Seine axis. A similar gradient of biological effects was reported in caged zebra mussels following
778 exposure at the same sampling sites in winter, spring, and summer (Michel et al. 2013). Genotoxicity
779 markers (DNA strand breaks and micronucleus frequency) significantly increased from Marnay to
780 Bougival and then Triel. Seasonal variations of the responses were also highlighted for DNA strand
781 breaks, with the lowest levels recorded in winter as compared to summer and spring (Michel et al.
782 2013). All these observations are in good agreement with the conclusions drawn from our LOE#3

integration results, suggesting that the effect gradient and seasonal trends (with lower biological disturbances in winter) are constant year in, year out.

Overall, our results demonstrate that the selected set of biomarkers efficiently reflected biological disturbances in gammarids following exposure along the Seine axis. The use of markers at different levels of biological organization, from the molecular level to life history traits, allowed us to discriminate among sites and observations. It also highlighted that *in situ* exposure differentially affected various physiological mechanisms depending on the level of anthropogenic pressure and the sampling period, thus demonstrating the complementarity of the selected endpoints. The reference values and biomarker thresholds used in the present study were specifically established according to the characteristics of the transplanted gammarid population (*e.g.*, size and weight) and some abiotic exposure parameters (*e.g.*, temperature). The use of such reference values and thresholds improved the reliability of the environmental diagnosis by integrating response variations due to the physiological state of gammarids and/or to field exposure conditions other than chemical pressure. Such finer characterization of the reference state is clearly needed when biological responses following seasonally varying contamination are studied. It can therefore be assumed that site qualification in terms of environmental hazard/risk is more relevant and robust using these adaptive references rather than more generic ones. For example, a 20-30% inhibition threshold for AChE activity is generally admitted in the literature for freshwater and marine invertebrates (Escartín and Porte 1996; Owen et al. 2002), but the value established and used in the present study was substantially lower (12%) (Xuereb et al. 2009). This type of methodology could be applied to responses analyzed within other LOEs such as bioavailability and bioassays, to refine and adjust the conclusions of the strategy (*e.g.,* a WOE approach) implemented to describe the ecological state of an aquatic environment.

3.4. Laboratory bioassays: LOE#4

The ecotoxicological diagnosis of the area was completed using a battery of laboratory bioassays performed on water column and sediment samples whose responses were integrated into the WOE approach to calculate cumulative hazard indices (ToxHQs) in the two abiotic compartments.

3.4.1. Bioassay hazard quotient in the water column ($ToxHQ_{water}$)

Aqueous samples were tested for embryotoxicity and teratogenicity using MELA (dissolved fraction) and for endocrine-disrupting potency using cellular *in vitro* bioassays on organic extracts. Responses were selected among the various endpoints monitored during the MELA according to their reliability, relevance and sensitivity to characterize survival (embryonic and larval survival rates), *in ovo*

development (hatching success and time to hatch) and growth (total body length and head size at hatching), and teratogenicity (total percentage of abnormal larvae). These responses (ESM Table S14) were integrated into the WOE approach to calculate $ToxHQ_{water}$ indices for each site and each sampling period (Table 4). The contributions of the different classes of endpoints to the global hazard quotient are shown in Fig. 4a.

Exposure of medaka ELS early life stages (ELS) to the dissolved fraction of the water samples collected at Marnay did not induce strong deleterious effects. Only a slight decrease (around 7%) in embryonic and larval survival rates exceeding the established thresholds was recorded in the C1 campaign (ESM Tables S7 and S14). Similarly, a slight reduction of the total body length of larvae at hatching was noted in medaka exposed to C1 and C3 samples, representing less than a 4% fall in comparison to the reference value (ESM Tables S7 and S14). The greatest modulations as compared to the reference at Marnay were highlighted by ER induction factors exceeding the corresponding Th value by 4.3- to 5.6-fold (ESM Tables S7 and S14). However, ER agonist activity was only evaluated as significant (in comparison to the blank) in the C3 samples (ER induction factor: 3.26×; ESM Table S14). As a result, survival, growth, and endocrine disruption were the most contributive classes of endpoints to the global hazard quotient in bioassays on water column samples from Marnay (Fig. 4a). In agreement with this limited effect, the $ToxHQ_{water}$ values were low (< 5) at Marnay, and the resulting hazard class for bioassays was ‘absent’ for all sampling periods as well as for the annual average value (Table 4).

Similar results were obtained for the Bougival C3 and C4 samples, with only slight effects on survival, growth, and estrogenic potency (ESM Table S14), so that the resulting hazard was classified as ‘absent’ for these periods (Table 4). In contrast, the C1 water sample from that station proved much more problematic as it strongly increased embryonic mortality (46%), and to a lesser extent larval mortality (14%) (ESM Table S14). Similarly, hatching success was reduced by 50% in comparison to the control. Teratogenicity also increased 4.4-fold as compared to the reference value, so that more than 80% of the hatchlings exhibited developmental abnormalities (ESM Tables S7 and S14). Additionally, ER activity was between 6.1 and 9.2 times higher than the established Th value (ESM Tables S7 and S14). However, these endocrine-disrupting effects did not contribute much to the global hazard (2%) because survival, development and teratogenicity accounted for 63%, 16% and 12% of the calculated $ToxHQ_{water}$, respectively (Fig. 4a). These effects account for the downgrading of the hazard associated to bioassay responses to the ‘moderate’ status (very close to the ‘major’ class) for the C1 campaign at Bougival. Nevertheless, in relation to the biological responses recorded in the fall (C1) campaign, the annual average hazard class was classified as ‘slight’ at Bougival (Table 4).

The Triel response profile was quite similar to that of Bougival, yet more pronounced. Bioassays on water samples from the winter (C4) campaign only revealed an impact on ED parameters, with

induction of ER activity 6.5-fold higher than the reference value (11times the Th level; ESM Tables S7 and S14). A slight induction of the AR/GR agonist activity was also noted, but it only represented a 1.83-fold increase as compared to the reference level (ESM Tables S7 and S14). These responses accounted for 74% of the cumulative $ToxHQ_{water}$ in Triel-C4 (Fig. 4a), but resulted in a global hazard classified as 'absent' (Table 4). ER and AR/GR agonist activities were induced with the same order of magnitude following cell exposure to Triel-C1 water organic extract (10.4× and 1.9× the corresponding Th value, respectively; ESM Tables S7 and S14). However, endocrine-disrupting effects only accounted for 7% of the calculated $ToxHQ_{water}$ at that site (Fig. 4a), as stronger impacts were detected in medaka ELS following exposure to the dissolved fraction of the water sample collected at Triel during the fall campaign. As observed for Bougival during the same sampling campaign, exposure led to a strong increase of the mortality rate of exposed embryos (only one third of the embryos were still alive at the end of the experiment; 13.3 times the Th value; ESM Tables S7 and S14). Moreover, hatching success was reduced by more than 70% in comparison to the control (~7 times the Th value), and the mean time to hatch was delayed by about 25.5% (1.6 times higher than the Th value) as compared to the reference (ESM Tables S7 and S14). An impact on medaka *in ovo* development was also reflected by the total body length of hatchlings, which was reduced by 5.4% in comparison to control organisms (about 4 times the Th value; ESM Tables S7 and S14). However, this effect on *in ovo* growth was quite limited as it only accounted for 7% of the global hazard quotient $ToxHQ_{water}$, whereas survival and development alterations respectively accounted for 61% and 25% (Fig. 4a). The resulting hazard attributed to the $ToxHQ_{water}$ index for Triel-C1 was identified as 'major' (Table 4). The growth of medaka embryos exposed to the Triel-C3 water sample was also slightly reduced: larvae were 3.4% shorter than the controls (2.4 times the Th value; data not shown). Nevertheless, the summer (C3) sample from Triel was particularly marked by a strong induction of both ER and TR activities, exceeding the corresponding Th values by 21.8× and 25.6× respectively (equivalent to respective induction factors of 11.9 and 13.8 as compared to the control; ESM Tables S7 and S14). As a result, endocrine-disrupting effects contributed to 80% of the cumulative $ToxHQ_{water}$ calculated for Triel-C3 (Fig. 4a), and the global hazard evaluated using bioassays was summarized as 'slight' (Table 4). Annual averaging of the effects observed during bioassays resulted in a global $ToxHQ_{water}$ classifying the hazard as 'slight' (but very close to the 'moderate' class) for Triel (Table 4). The $ToxHQ_{water}$ value was mainly due to impacts on survival (44%) and *in ovo* development (19%) of medaka ELS, as well as to endocrine-disrupting effects (27%) on specific cell lines (Fig. 4a).

3.4.2. Bioassay hazard quotient of sediment ($ToxHQ_{sed}$)

As in the case of the water column, ED assays were performed on sediment organic extracts, and the MELAc (MELA adapted to sediment testing by direct contact with particles) was implemented on whole sediment samples. In addition, cytotoxicity and genotoxicity of sediment elutriates were investigated using the Microtox and SOS Chromotest procedures, respectively. As previously mentioned, the responses of all the bioassays (ESM Table S14) were integrated into a common hazard quotient, $ToxHQ_{sed}$ (Table 4), and the contribution of each endpoint class to the hazard index was calculated (Fig. 4b).

Medaka exposure to Marnay sediments only resulted in limited sublethal effects on fish ELS. A slight delay in the average time to hatch of embryos was noted following exposure to the sediment sampled during the three campaigns (C1, C3 and C4), representing 2 to 5% variation in comparison to the control values (ESM Tables S7 and S14). Moreover, a significant increase in the percentage of abnormal larvae at hatching (around 2 times higher than the control, *i.e.*, around 35.5% of abnormal individuals; ESM Tables S7 and S14) was noticed after exposure to Marnay sediment collected during the fall (C1) period. *In ovo* growth of embryos exposed to C3 and C4 samples also slightly decreased: biometric measurements were lower than the reference values, especially hatchling head size for which the highest percentage of variation relative to the control was recorded (~8%; ESM Tables S7 and S14). The cytotoxicity of sediment elutriates remained moderate for Marnay C3 and C4 samples (less than 30% inhibition of bioluminescence), but was greater during the fall (C1) campaign with 50% inhibition (ESM Tables S7 and S14). The strongest effects were noted for ER agonist activity of sediment organic extracts, which noticeably increased by 8.5- and 9-fold,following exposure to C3 and C1 organic extracts, respectively, and to a lesser extent following exposure to the C4 sample (4×) (ESM Tables S7 and S14). As a result, development, growth, endocrine disruption, and cytotoxicity were the most contributive endpoint classes to the global hazard quotient $ToxHQ_{sed}$. They accounted for 26%, 28%, 18%, and 18% of the annual average calculation of $ToxHQ_{sed}$, respectively (Fig. 4b). The resulting hazard was summarized as ‘slight’ at Marnay for all the sampling periods as well as for the annual average value (Table 4).

The bioassay response profiles at the Marnay and Bougival stations were very similar, as illustrated by the endpoint contributions to the $ToxHQ_{sed}$ values (Fig. 4b). However, the response was clearly greater downstream of Paris. For example, the time to hatch of embryos exposed to Bougival C3 sediment was 24% longer than in the control treatment, and more than 50% of the larvae showed developmental abnormalities on average following exposure to Bougival C1 sediment (ESM Tables S7 and S14). Whereas elutriates from the Bougival fall (C1) and winter (C4) samples inhibited bioluminescence by around 40%, the Bougival C3 elutriate reduced it by up to 96% as compared to the reference value (ESM Tables S7 and S14). Similarly, TR and AR/GR agonist activities as well as AR antagonist activity were enhanced with Bougival samples as compared to Marnay samples: they exceeded the established respective Th values by 2.1× to 4.3×. As observed for Marnay, ER agonist activity showed

the greatest induction as compared to the reference level: between 7.7-fold in the winter (C4) campaign and 32-fold in the summer (C3) campaign (ESM Tables S7 and S14). As a result, the $ToxHQ_{sed}$ values for Bougival were higher than for Marnay. While the global hazard remained classified as 'slight' for the C1 and C4 campaigns (although close to the 'moderate' hazard limit; Table 4), it reached the 'moderate' grade for Bougival C3 due to the strong effects highlighted by bioassays. In agreement with these observations, the annual global hazard associated with bioassay responses was summarized as 'moderate' for Bougival (Table 4).

According to bioassay endpoints, the greatest toxicity was recorded at Triel. Signs of acute toxicity were highlighted in medaka ELS following exposure to C1 and C3 sediment: embryonic survival decreased by 13% as compared to the reference value in Triel C1, and hatching success did not exceed 1.3% and 0% in Triel C1 and C3, respectively (ESM Tables S7 and S14). Moreover, the few resulting hatchlings finally died before the end of the experiment (data not shown). Consequently, teratogenicity and *in ovo* growth were only evaluated for the Triel winter (C4) sample. Forty-nine percent of the larvae exposed to this sediment displayed developmental deformities, and their growth was reduced by 8% to 9% depending on the endpoint (ESM Tables S7 and S14). The cytotoxicity of Triel elutriates was very high in the C1 and C3 samples: bioluminescence was inhibited by 96% and 94%., respectively. In contrast, cytotoxicity was very limited in the Triel C4 sample: only 25% inhibition of bacteria bioluminescence was noted (ESM Table S14). Modulation of endocrine-disrupting responses following cell line exposure to Triel sediment organic extracts was of the same order of magnitude as in Bougival sediment. TR, AR/GR and Anti-AR activity induction exceeded the respective Th values by 1.9- to 5.3-fold, depending on the endpoint and the campaign (ESM Tables S7 and S14). In addition, ER agonist activity modulation was the strongest, with induction factors ranging between 4.1-fold in the winter (C4) campaign and 30-fold in the summer (C3) campaign (ESM Tables S7 and S14). Unsurprisingly, the hazard for toxicity assessed through laboratory bioassays was evaluated as 'severe' for Triel C1 and C3 sediment (Table 4). The survival and development endpoint classes respectively accounted for 51% and 28% of the global $ToxHQ_{sed}$ value in Triel C1. Development (47%), cytotoxicity (23%), and endocrine disruption (20%) were the most contributive endpoint classes to the Triel C3 $ToxHQ_{sed}$ value, whereas growth (41%) and teratogenicity (30%) accounted for the main part of the Triel C4 $ToxHQ_{sed}$ value (Fig. 4b). Due to the limited effects of this latter sample, the corresponding hazard was evaluated as 'slight'. Finally, based on bioassay responses we concluded that Triel sediment represented a 'major' yearly hazard (Table 4).

Overall, the ToxHQ values calculated for the water column and the sediment clearly reflected the anthropogenic gradient between the upstream and downstream sites, with values increasing from Marnay to Bougival and Triel (Table 4), similarly to biomarker responses (see Section *3.3*). Moreover, the yearly variations of bioassay responses identified the C1 and C3 water and sediment samples as the most toxic, whereas C4 samples only induced limited effects. These observations are in good

agreement with the conclusions drawn from bioavailability and biomarker LOEs (see Sections *3.2* and *3.3*). Moreover, the independent exposure of model organisms/cell lines to water column or sediment samples gave new insights into environmental contamination along the Seine River. The main part of the toxic effects we noted appeared to be associated with the sediment compartment since the hazard class indices in sediment were systematically higher than in water samples (Table 4).

Our results clearly show that the sediment compartment should be integrated into environmental quality assessment procedures as a potential non-negligible source of toxicity for aquatic organisms. Such procedures only based on water column analysis could underestimate the (eco)toxicological risk for ecosystems.

The accuracy and environmental significance of toxic responses based on laboratory bioassays (and of the resulting hazard assessment) could be improved by working on the reference and threshold values. Many bioassays, such as *in vitro* tests, presently use 'too clean to be real' standards as negative controls (*e.g.*, extraction blanks, ultra-pure water, etc.). The establishment of 'truly environmental' negative controls should be closely investigated to make these bioassay responses more realistic and relevant in an environmental context. The calculation of reference and threshold values from the analysis of environmentally 'clean' water and sediment samples could be a valuable alternative to characterize the basal levels and variations of very specific responses such as hormone-mimetic, genotoxicity, and cytotoxicity endpoints. We applied this kind of approach in some bioassays such as the MELA (and the MELAc), with drinking water and Yville-sur-Seine sediment (a pristine site in the vicinity of the Seine River axis) as negative controls for water and sediment samples, respectively. The integration of other reference matrices from various geographical localizations could also make it easier to understand response variability by taking into account the natural diversity and heterogeneity of the reference environments.

Such a strategy would undoubtedly improve the accuracy and the relevance of environment quality assessment using laboratory bioassays in large-scale studies as well as in more geographically-restricted contexts with locally contrasted areas.

3.5. WOE integration

The results of each LOE were integrated into a global WOE index, and a hazard class was attributed to each site and campaign (Fig. 5). Moreover, the contribution of each HQ index to the WOE value was calculated, and is presented in Fig. 6.

WOE levels clearly reflected the anthropogenic gradient along the Seine River, with values increasing from upstream to downstream of Paris (Fig. 5). They were systematically lower at Marnay (15% to

30%), intermediate at Bougival (38% to 64%) and the highest at Triel (39% to 70%) for all campaigns and for annual average values. The only exception was the spring (C2) sampling period for which the highest WOE index was noted at Bougival (Fig. 5). The resulting hazard classes varied from 'slight' to 'moderate' at Marnay, and from 'moderate' to 'major' at the two stations located downstream of Paris, and the WOE index of Triel C1 was very close to the 'severe' hazard level. The year-round hazard was assessed using the annual average LOE data. It summarized the overall hazard class as 'moderate' at Marnay and 'major' at Bougival and Triel (Fig. 5).

Additionally, the seasonal variations identified when examining the results of each LOE were also reflected in the WOE levels. The lowest value was consistently recorded for the winter (C4) campaign, at each site. In contrast, the fall (C1) and summer (C3) campaigns had previously been identified as the most impacted ones, and coherently resulted in the highest WOE levels (Fig. 5). The only exception was the Bougival C2 sampling point; however, it should be borne in mind that the hazard assessment during the spring campaign was only based on bioaccumulation and water contamination data. For this reason, contributions of HQ indices are not discussed below because they are biased by the missing C2 campaign data.

The contribution profiles from Marnay revealed that chemical contamination in the abiotic compartment was the main component of the WOE index, with a global contribution ($ChemHQ_{water}$ plus $ChemHQ_{sed}$ contribution) around 80% for each sampling period and the annual average value (Fig. 6a). At the upstream site, the greatest risk was attributable to the contaminants analyzed in the sediment: the contributions of the $ChemHQ_{sed}$ indices reached 60% (*vs*. around 20% for the $ChemHQ_{water}$ indices). Similar observations were made for the downstream stations during the winter (C4) campaign. The global contribution of chemicals in the abiotic compartment also represented around 80% for Bougival and Triel during the C4 campaign; however, the contributions of water column and sediment contamination were almost the same, indicating that the chemical hazard was governed by contaminants analyzed in the water column as well as in the sediment at both downstream sites (Fig. 6b and 6c). The cumulative contributions of ChemHQs were also substantial at the downstream stations during the C1 and C3 campaigns, and the annual average values ranged between 48% and 62% (Fig. 6b and 6c). However, biological effects recorded *in situ* (BiomHQs) and under laboratory conditions (ToxHQs) also contributed to the calculated WOE value in a non-negligible way. Biomarker responses thus accounted for 11% to 17% of the WOE indices, and bioassays contributed between 16% to 36% (Fig. 6b and 6c). Among the bioassays, the ones using sediment as a test phase usually yielded the highest contributions, suggesting again that the sediment compartment represents a noticeable hazard in terms of both contamination levels and biological effects.

The increase of the biological effects noted at some particular sites/sampling dates, such as Triel C1/C3, could be attributed to seasonal variations in the contamination levels, as revealed by LOE#1

results. The concentrations of some contaminants such as PFOS and pesticides are clearly influenced by seasonal factors such as hydrological conditions (Tamtam et al. 2008; Labadie and Chevreuil 2011) and/or by the seasonal use of certain chemicals such as pesticides, including metazachlor (Passeport et al. 2013).

Moreover, the levels of metal elements in gammarids revealed that bioaccumulation was influenced not only by the contamination levels but also by seasonal variables like temperature, especially concerning essential elements (Lebrun et al. 2015). A more accurate characterization and understanding of these variations would make it possible to refine the reference values and/or to define specific thresholds according to the substance and the influence of confounding factors on its accumulation levels in organisms. For instance, determining whether the temperature influenced bioaccumulation levels by modulating metabolic rates or contaminant bioavailability would be of great interest. These adjustments could improve the conclusions of the WOE procedure in a relevant and reliable way, by monitoring the impact of external abiotic factors likely to modulate the time-course of accumulation by exposed organisms, especially in close but contrasted areas.

Overall, the conclusions on the hazard represented by 'chemical' and 'biological' LOEs (*i.e.*, LOEs#1-2 *vs.* LOEs#3-4) are relatively coherent, although a shift in hazard severity was evidenced between the two types of LOEs. In fact, the hazard level is generally lower with 'biological' LOEs than with 'chemical' LOEs (Fig. 5). The table at the bottom of Fig. 5 also reveals a few exceptions (*e.g.*, Triel C4) for which the conclusion of the chemical and biological LOEs were not fully consistent, suggesting that the two approaches are complementary. These observations also emphasize the usefulness of WOE integration, as the class-based hazard ranking of the sites differed when considering the results of each LOE independently, especially at the downstream sites. For instance, when referring to the annual hazard class associated to $ChemHQ_{sed}$ values, we failed to discriminate among the three sites because the environmental hazard was evaluated at the highest level ('severe') in all cases (Table 1), indicating that all sites represented the maximal hazard level in terms of sediment contamination. In contrast, the annual average hazard classes associated to sediment bioassays ($ToxHQ_{sed}$) identified a hazard level increasing from 'slight' at Marnay to 'moderate' and 'major' at Bougival and Triel, respectively (Table 4). This suggests that only Triel was faced with a high level of environmental hazard. However, only Bougival was faced with a high ('major') hazard level according to annual average bioavailability measurements (Table 2), while both downstream stations were classified as particularly impacted by biomarker responses ('major' and 'severe' hazard at Bougival and Triel, respectively, based on annual average data; Table 3). The environmental diagnosis of the three sites would have been substantially different if based on one or another LOE output, under- or over-estimating the environmental risk at each site depending on the LOE. This would also have led to contradictory conclusions on the impacts of the Oise River inputs between the two downstream stations. These two latter aspects would complexify the decision-making process for environmental

managers. The solution lies in final WOE integration, which compiles the results from each LOE into a global hazard index associated to an integrative hazard class translating the overall environmental risk at each site.

As a result, it seems relevant to further analyze the environmental hazard in the studied area, basing the diagnosis on various aspects including contamination levels, bioavailability, and biological responses. Although the biological effects recorded in the present study can be considered as quite limited in comparison to contamination levels, remobilization of contaminants (especially those trapped in sediment) and/or variations of some controlling factors (*e.g.*, temperature, flow rates, or physico-chemistry of the surrounding environment) would result in a strong increase in bioavailable contamination levels, and this in turn could induce more severe biological impacts.

To avoid such critical situations for the health of aquatic ecosystems, efforts should focus on decontamination and remediation procedures in the most impacted sites (Bougival and Triel) before more adverse effects occur at the population level, as suggested by some biological responses evidenced in the present study (*e.g.*, acute toxicity, altered microbial communities, reproductive impairments etc.).

3.6. Bacterial communities

The bacterial community composition of the water column was studied using high throughput sequencing of bacterial 16S-rRNA genes during three sampling campaigns (C1, C3, and C4). Unfortunately, due to a technical problem, the data from the Triel sampling station were not available for the winter campaign (C4). The dissimilarity in bacterial community structure (number and relative abundance of different $OTU_{0.03}$) among the different samples are presented as a three-dimension NMDS graph (Fig. 7). This figure clearly shows that samples from Marnay were grouped together, while obviously separated from the Bougival and Triel samples that were close (except for the December 2012 Bougival sample). Similar differences were previously observed on ARISA (Automated Ribosomal Intergenic Spacer Analysis) profiles of river biofilms collected from the same sampling sites (Fechner et al. 2012). This suggests that the bacterial communities of the water samples collected upstream of Paris were different from those sampled downstream. These observations on bacterial communities are also in good agreement with the global hazard assessment illustrated by WOE indices, which were the lowest and varied between 15% to 30% for Marnay. In contrast, the integrative HQ calculated for the two downstream stations covered the same range of higher values (54% to 70%), except for the C4 campaign. For this campaign, WOE indices from Bougival were close to the range of values from Marnay; in parallel, bacterial community compositions were similar in the two stations during the winter campaign. Globally, this study shows that responses at the

bacterial community level reflect the global disturbance of the environment particularly well. Results from river biofilms collected during the same campaigns also corroborate this observation (Faburé et al. 2015). The use of microbial communities in an ERA context might be very powerful in the future; however the conclusions from such an approach (*i.e.*, the classification of sites in relation to one another) can only be relative since they are drawn from inter-site qualitative comparisons, as reference levels are nowadays still lacking.

4. Conclusion

The WOE approach applied in the present study proved efficient and relevant in terms of both global environmental hazard diagnosis and seasonality analysis. The procedure was particularly improved using external reference levels integrating natural variations of responses and confounding factors, especially in LOEs#2 and #3. This improved the reliability of WOE integration results, which better reflected the level of disturbance of organisms at each sampling time, without any interference related to acclimation or adaptation mechanisms likely to occur in chronically exposed populations. The establishment of reference values and thresholds from numerous studies conducted at the national scale also eliminated the need for a reference site in the study area, which could be very problematic in large rivers subjected to multiple and diffuse pressures. Our results reveal that at the upstream site, generally used as a relative reference or control site in previous investigations in the area, the low contamination levels nonetheless resulted in low but significant biological effects.

This approach should be pursued and further developed at larger spatial scales. Bioaccumulation and biological responses to pollutants as well as baseline levels may be modulated and altered by long-term variations and trends in some key endpoints, *e.g.*, growth and reproduction, themselves governed by global factors and large-scale processes (*e.g.*, climate trends and changes, oceanographic cycles, etc.) (Garmendia et al. 2015). The in-depth characterization of the baseline levels and relevant effect thresholds for such environmentally relevant endpoints is thus a challenge to ensure their relevance within ERA purposes.

Another strength of the present work lies in the use of gammarids from the same population for bioaccumulation measurements and biomarker analyses. A direct and strong connection was thus established between bioaccumulation levels and biological responses, strengthening the conclusions from the LOEs based on these data. Moreover, the use of these amphipods is entirely appropriate and relevant in the context of ecological/ecotoxicological field studies. Gammarids are widespread in European freshwaters, and are key actors in the functioning of these ecosystems as litter degraders and as a food source for fish and amphibian species. As a result, multiple biomarkers and bioassays using gammarids are available for field-testing of contaminant impacts. Moreover, modelling developments

quantifying the natural variability of these markers in relation to abiotic factors enhance the reliability of the *in situ* methodology and allow for its implementation at large spatial and temporal scales in monitoring programs (Coulaud et al. 2011; Chaumot et al. 2015).

The assessment of environmental quality was also improved by integrating water column and sediment analyses. In the water column, contamination (according to the selected compounds analyzed in this matrix) and toxicity (assessed by bioassays) remained relatively limited, but in the sediment variable stocks of pollutants accumulated and locally reached very high levels, with various impacts on laboratory-exposed organisms. These compounds are likely (at least partially) involved in some of the biological responses detected in gammarids through biomarker analyses, since they were significantly accumulated by exposed gammarids. This also proves that although we did not trace all contaminants in the water, a fraction of these contaminants is bioavailable to organisms through the water column. Overall observations suggest that a non-negligible ecological risk in the area could threaten benthic biota as well as pelagic organisms through the release and/or remobilization of the sediment-bound chemicals into the water column.

Other investigations were performed within the framework of the PIREN-Seine program (*e.g.*, bacterial community analyses, bioaccumulation of metal and organic compounds, metal tolerance acquisition in biofilms), but were not integrated into the WOE model because reference levels and thresholds still remain difficult to set. However, it should be mentioned that these approaches gave similar results, clearly differentiating the upstream (Marnay) station from the downstream (Bougival and Triel) sites, in relation to the contamination gradient. There is no doubt that these experiments have to be further developed, but even in their current state they can be valuable tools with successful *in situ* deployment in a biomonitoring context, and it would be relevant to integrate them into an ERA procedure as they provide information at the community level.

Finally, the WOE approach applied in the present study was based on the integration of each response into a global hazard index in a similar way, including pseudo normalization of the data. Thus the results from various sampling times remained comparable and were reported on a common grid of hazard classification. Such a procedure represents an advantageous and practical tool in the diagnosis of environmental hazard because it yields relevant information classifying the most problematic substances and effects, and gradually identifies the most impacted sites (comparing HQ values) with an associated hazard level. The most remarkable strong point of the approach lies in the ability of the model to integrate a large amount of endpoints characterizing various aspects of the environmental risk and to generate very ‘simple’ and ‘comprehensible’ integrative outputs from this large dataset, *i.e.*, the WOE index and the relative hazard class.

This WOE model may be very helpful for environment managers in decision-making processes to plan remediation procedures and/or actions to reduce emissions and/or uses of problematic substances.

Combined with the available biomonitoring tools used in the present study, this approach could also be implemented on a long-term basis to monitor the potential improvement of environmental quality following environmental management measures.

References

Ayrault S, Roy-Barman M, Le Cloarec MF, Priadi CR, Bonté P, Göpel C (2012) Lead contamination of the Seine River, France: Geochemical implications of a historical perspective. Chemosphere 87(8):902-910. doi: 10.1016/j.chemosphere.2012.01.043

AZUR Environmental (1998) Microtox basic solid-phase test (Basic SPT). Carlsbad, CA, USA

Bado-Nilles A, Jolly S, Porcher JM, Palluel O, Geffard A, Gagnaire B, Betoulle S, Sanchez W (2015) Applications in environmental risk assessment of leucocyte apoptosis, necrosis and respiratory burst analysis on the European bullhead, *Cottus* sp. Environ Pollut 184:9-17. doi: 10.1016/j.envpol.2013.07.049

Balaguer P, François F, Comunale F, Fenet H, Boussioux A-M, Pons M, Nicolas J-C, Casellas C (1999) Reporter cell lines to study the estrogenic effects of xenoestrogens. Sci Total Environ 233(1-3):47-56. doi: 10.1016/s0048-9697(99)00178-3

Barhoumi B, Clérandeau C, Landi L, Pichon A, Le Bihanic F, Poirier D, Anschutz P, Budzinski H, Driss MR, Cachot J (2016) Assessing the toxicity of sediments using the Medaka Embryo-Larval Assay and two other bioassays. Environ Toxicol Chem *(In press)*. doi: 10.1002/etc.3388

Barjhoux I, Baudrimont M, Morin B, Landi L, Gonzalez P, Cachot J (2012) Effects of copper and cadmium spiked-sediments on embryonic development of Japanese medaka (*Oryzias latipes*). Ecotoxicol Environ Saf 79:272-282. doi: 10.1016/j.ecoenv.2012.01.011

Bebianno MJ, Pereira CG, Rey F, Cravo A, Duarte D, D'Errico G, Regoli F (2015) Integrated approach to assess ecosystem health in harbor areas. Sci Total Environ 514:92-107. doi: 10.1016/j.scitotenv.2015.01.050

Belles A, Pardon P, Budzinski H (2014) Development of an adapted version of polar organic chemical integrative samplers (POCIS-Nylon). Anal Bioanal Chem 406(4):1099-1110. doi: 10.1007/s00216-013-7286-2

Benedetti M, Ciaprini F, Piva F, Onorati F, Fattorini D, Notti A, Ausili A, Regoli F (2012) A multidisciplinary weight of evidence approach for classifying polluted sediments: Integrating

sediment chemistry, bioavailability, biomarkers responses and bioassays. Environment International 38(1):17-28. doi: 10.1016/j.envint.2011.08.003

Bennett J, Cubbage J (1992) Review and Evaluation of Microtox Test for Freshwater Sediments. Washington State Department of Ecology, Washington, 28 pp

Besse JP, Coquery M, Lopes C, Chaumot A, Budzinski H, Labadie P, Geffard O (2013) Caged *Gammarus fossarum* (Crustacea) as a robust tool for the characterization of bioavailable contamination levels in continental waters: Towards the determination of threshold values. Water Res 47(2):650-660. doi: 10.1016/j.watres.2012.10.024

Billen G, Garnier J, Mouchel J-M, Silvestre M (2007) The Seine system: Introduction to a multidisciplinary approach of the functioning of a regional river system. Sci Total Environ 375(1-3):1-12. doi: 10.1016/j.scitotenv.2006.12.001

Blanchard M, Teil MJ, Guigon E, Larcher-Tiphagne K, Ollivon D, Garban B, Chevreuil M (2007) Persistent toxic substance inputs to the river Seine basin (France) via atmospheric deposition and urban sludge application. Sci Total Environ 375(1-3):232-243. doi: 10.1016/j.scitotenv.2006.12.012

Bols NC, Brubacher JL, Ganassin RC, Lee LEJ (2001) Ecotoxicology and innate immunity in fish. Dev Comp Immunol 25(8-9):853-873. doi: 10.1016/s0145-305x(01)00040-4

Brouwer H, Murphy T, McArdle L (1990) A sediment-contact bioassay with *Photobacterium phosphoreum*. Environ Toxicol Chem 9(11):1353-1358

Budzinski H, Letellier M, Thompson S, LeMenach K, Garrigues P (2000) Combined protocol for the analysis of polycyclic aromatic hydrocarbons (PAHs) and polychlorobiphenyls (PCBs) from sediments using focussed microwave assisted (FMW) extraction at atmospheric pressure. Fresen J Anal Chem 367(2):165-171

Chapman PM, McDonald BG, Lawrence GS (2002) Weight-of-evidence issues and frameworks for sediment quality (and other) assessments. Hum Ecol Risk Assess 8(7):1489-1515. doi: 10.1080/20028091057457

Chapman PM, Wang F, Janssen CR, Goulet RR, Kamunde CN (2003) Conducting ecological risk assessments of inorganic metals and metalloids: Current status. Hum Ecol Risk Assess 9(4):641-697

Chapman PM (2007) Determining when contamination is pollution - Weight of evidence determinations for sediments and effluents. Environment International 33(4):492-501. doi: 10.1016/j.envint.2006.09.001

Charron L, Geffard O, Chaumot A, Coulaud R, Quéau H, Geffard A, Dedourge-Geffard O (2013) Effect of water quality and confounding factors on digestive enzyme activities in *Gammarus fossarum*. Environ Sci Pollut Res 20(12):9044 - 9056. doi: 10.1007/s11356-013-1921-5

Chaumot A, Geffard O, Armengaud J, Maltby L (2015) Chapter 11 - Gammarids as Reference Species for Freshwater Monitoring. In: Mouneyrac C, Amiard-Triquet J-C, Amiard C (Eds) Aquatic Ecotoxicology: Advancing Tools for Dealing with Emerging Risks. Academic Press, pp. 253-280

Coulaud R, Geffard O, Xuereb B, Lacaze E, Quéau H, Garric J, Charles S, Chaumot A (2011) *In situ* feeding assay with *Gammarus fossarum* (Crustacea): Modelling the influence of confounding factors to improve water quality biomonitoring. Water Res 45(19):6417-6429. doi: 10.1016/j.watres.2011.09.035

Dagnino A, Sforzini S, Dondero F, Fenoglio S, Bona E, Jensen J, Viarengo A (2008) A "Weight-of-Evidence" approach for the integration of environmental "Triad" data to assess ecological risk and biological vulnerability. Integr Environ Assess Manag 4(3):314-326. doi: 10.1897/ieam_2007-067.1

Dedourge-Geffard O, Charron L, Hofbauer C, Gaillet V, Palais F, Lacaze E, Geffard A, Geffard O (2013) Temporal patterns of digestive enzyme activities and feeding rate in gammarids (*Gammarus fossarum*) exposed to inland polluted waters. Ecotoxicol Environ Saf 97:139-146. doi: 10.1016/j.ecoenv.2013.07.016

Dufour V, Cruz J, Chambolle M, Granger D, Le Menach K, Pardon P, Litrico X, Budzinski H (2015) Study of water contamination by pesticides, comparison of grab sampling and passive sampling – Application to natural water and an industrialzone sewage near Bordeaux. Revue des Sciences de l'Eau 28(3):223-228. doi: 10.7202/1034011ar

Dulio V, Andres S (2014) Recommandations du CEP auprès du MEDDE pour la sélection des Substances Pertinentes à Surveiller dans les milieux aquatiques pour le second cycle de la DCE (2016-2021). Rapport AQUAREF 2013, Final document, pp. 102. http://www.onema.fr/IMG/pdf/Aquaref2013_F1a_INERIS_Rapport%20CEP_subst%20pertinentes_VF.pdf (In French)

Ellman GL, Courtney KD, Andres Jr V, Featherstone RM (1961) A new and rapid colorimetric determination of acetylcholinesterase activity. Biochem Pharmacol 7(2):88-90

Escartín E, Porte C (1996) Acetylcholinesterase inhibition in the crayfish *Procambarus clarkii* exposed to fenitrothion. Ecotoxicol Environ Saf 34(2):160-164. doi: 10.1006/eesa.1996.0058

European Commission (EC) (2000) Directive 2000/60/EC of the European parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Official Journal of the European Union, pp. 72

European Commission (EC) (2013) Directive 2013/39/EU of the European parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy. Official Journal of the European Union, pp. 17

European Union (EU) (2005) Common implementation strategy for the water framework directive: envionmental quality standards (EQS), substance data sheet, priority substance No.28: Polycyclic Aromatic Compounds, Brussels, pp. 30

Faburé J, Dufour M, Autret A, Uher E, Fechner LC (2015) Impact of an urban multi-metal contamination gradient: Metal bioaccumulation and tolerance of river biofilms collected in different seasons. Aquat Toxicol 159:276-289. doi: 10.1016/j.aquatox.2014.12.014

Fechner LC, Gourlay-Francé C, Bourgeault A, Tusseau-Vuillemin MH (2012) Diffuse urban pollution increases metal tolerance of natural heterotrophic biofilms. Environ Pollut 162:311-318. doi: 10.1016/j.envpol.2011.11.033

French Ministry of Ecology Energy Sustainable Development and Planning (MEDAD) (2006) Arrêté du 9 août 2006 relatif aux niveaux à prendre en compte lors d'une analyse de rejets dans les eaux de surface ou de sédiments marins, estuariens ou extraits de cours d'eau ou canaux relevant respectivement des rubriques 2.2.3.0, 4.1.3.0 et 3.2.1.0 de la nomenclature annexée à l'article R. 214-1 du code de l'environnement. Journal Officiel de la République Française - Texte 15 sur 48 - pp. 3 (In French)

French Ministry of Ecology Energy Sustainable Development and Planning (MEDAD) (2007) Circulaire n°2007/23 du 7 mai 2007 définissant les « normes de qualité environnementale provisoires (NQEp) » des 41 substances impliquées dans l'évaluation de l'état chimique des masses d'eau ainsi que des substances pertinentes du programme national de réduction des substances dangereuses dans l'eau. Bulletin officiel du MEDAD 2007/15 - Texte 17/36 - pp. 13 (In French)

French Ministry of Ecology Energy Sustainable Development and Planning (MEDAD) (2015) Arrêté du 27 juillet 2015 modifiant l'arrêté du 25 janvier 2010 relatif aux méthodes et critères d'évaluation de l'état écologique, de l'état chimique et du potentiel écologique des eaux de surface pris en application des articles R. 212-10, R. 212-11 et R. 212-18 du code de l'environnement. Journal Officiel de la République Française - Texte 4 sur 125 - pp. 95 (In French)

García-Armisen T, İnceoğlu Ö, Ouattara NK, Anzil A, Verbanck MA, Brion N, Servais P (2014) Seasonal variations and resilience of bacterial communities in a sewage polluted urban river. PLoS ONE 9(3):art. no. e92579. doi: 10.1371/journal.pone.0092579

Garmendia L, Izagirre U, Soto M, Lermen D, Koschorreck J (2015) Combining chemical and biological endpoints, a major challenge for twenty-first century's environmental specimen banks. Environ Sci Pollut Res 22(3):1631-1634. doi: 10.1007/s11356-014-2925-5

Geffard O, Xuereb B, Chaumot A, Geffard A, Biagianti S, Noël C, Abbaci K, Garric J, Charmantier G, Charmantier-Daures M (2010) Ovarian cycle and embryonic development in *Gammarus fossarum*: Application for reproductive toxicity assessment. Environ Toxicol Chem 29(10):2249-2259. doi: 10.1002/etc.268

Helmstetter MF, Alden RWI (1995) Toxic responses of Japanese medaka (*Oryzias latipes*) eggs following topical and immersion exposures to pentachlorophenol. Aquat Toxicol 32(1):15-29. doi: 10.1016/0166-445x(94)00059-y

ISO (1999) Determination of the inhibitory effect of water samples on the light emission of *Vibrio fischeri* (luminescent bacteria test). Part 3: method using freeze dried bacteria. European Standard ISO 11348-3:1998. European Committee for Standardization, Bruxelles

Jugan ML, Lévy-Bimbot M, Pomérance M, Tamisier-Karolak S, Blondeau JP, Lévi Y (2007) A new bioluminescent cellular assay to measure the transcriptional effects of chemicals that modulate the alpha-1 thyroid hormone receptor. Toxicol In Vitro 21(6):1197-1205. doi: 10.1016/j.tiv.2007.03.020

Jugan ML, Oziol L, Bimbot M, Huteau V, Tamisier-Karolak S, Blondeau JP, Lévi Y (2009) *In vitro* assessment of thyroid and estrogenic endocrine disruptors in wastewater treatment plants, rivers and drinking water supplies in the greater Paris area (France). Sci Total Environ 407(11):3579-3587. doi: 10.1016/j.scitotenv.2009.01.027

Kinani S, Bouchonnet S, Creusot N, Bourcier S, Balaguer P, Porcher JM, Aït-Aïssa S (2010) Bioanalytical characterisation of multiple endocrine- and dioxin-like activities in sediments

from reference and impacted small rivers. Environ Pollut 158(1):74-83. doi: 10.1016/j.envpol.2009.07.041

Labadie P, Chevreuil M (2011) Biogeochemical dynamics of perfluorinated alkyl acids and sulfonates in the River Seine (Paris, France) under contrasting hydrological conditions. Environ Pollut 159(12):3634-3639. doi: 10.1016/j.envpol.2011.07.028

Le Cloarec MF, Bonte PH, Lestel L, Lefèvre I, Ayrault S (2011) Sedimentary record of metal contamination in the Seine River during the last century. Phys Chem Earth 36(12):515-529. doi: 10.1016/j.pce.2009.02.003

Le Pape P, Ayrault S, Quantin C (2012) Trace element behavior and partition versus urbanization gradient in an urban river (Orge River, France). J Hydrol 472-473:99-110. doi: 10.1016/j.jhydrol.2012.09.042

Lebrun JD, Uher E, Tusseau-Vuillemin MH, Gourlay-Francé C (2014) Essential metal contents in indigenous gammarids related to exposure levels at the river basin scale: Metal-dependent models of bioaccumulation and geochemical correlations. Sci Total Environ 466-467:100-108. doi: 10.1016/j.scitotenv.2013.07.003

Lebrun JD, Geffard O, Urien N, François A, Uher E, Fechner LC (2015) Seasonal variability and inter-species comparison of metal bioaccumulation in caged gammarids under urban diffuse contamination gradient: Implications for biomonitoring investigations. Sci Total Environ 511:501-508. doi: 10.1016/j.scitotenv.2014.12.078

MacDonald DD, Ingersoll CG, Berger TA (2000) Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. Arch Environ Contam Toxicol 39(1):20-31. doi: 10.1007/s002440010075

Mersch-Sundermann V, Mochayedi S, Kevekordes S (1992) Genotoxicity of polycyclic aromatic hydrocarbons in *Escherichia coli* PQ37. Muta Res-Genet Tox 278(1):1-9. doi: 10.1016/0165-1218(92)90279-9

Michel C, Bourgeault A, Gourlay-Francé C, Palais F, Geffard A, Vincent-Hubert F (2013) Seasonal and PAH impact on DNA strand-break levels in gills of transplanted zebra mussels. Ecotoxicol Environ Saf 92:18-26. doi: 10.1016/j.ecoenv.2013.01.018

Munoz G, Giraudel JL, Botta F, Lestremau F, Dévier MH, Budzinski H, Labadie P (2015) Spatial distribution and partitioning behavior of selected poly- and perfluoroalkyl substances in

freshwater ecosystems: A French nationwide survey. Sci Total Environ 517:48-56. doi: 10.1016/j.scitotenv.2015.02.043

Nouira T, Risso C, Chouba L, Budzinski H, Boussetta H (2013) Polychlorinated biphenyls (PCBs) and Polybrominated Diphenyl Ethers (PBDEs) in surface sediments from Monastir Bay (Tunisia, Central Mediterranean): Occurrence, distribution and seasonal variations. Chemosphere 93(3):487-493. doi: 10.1016/j.chemosphere.2013.06.017

Owen R, Buxton L, Sarkis S, Toaspern M, Knap A, Depledge M (2002) An evaluation of hemolymph cholinesterase activities in the tropical scallop, *Euvola (Pecten) ziczac*, for the rapid assessment of pesticide exposure. Mar Pollut Bull 44(10):1010-1017. doi: 10.1016/s0025-326x(02)00139-x

Passeport E, Tournebize J, Chaumont C, Guenne A, Coquet Y (2013) Pesticide contamination interception strategy and removal efficiency in forest buffer and artificial wetland in a tile-drained agricultural watershed. Chemosphere 91(9):1289-1296. doi: 10.1016/j.chemosphere.2013.02.053

Piva F, Ciaprini F, Onorati F, Benedetti M, Fattorini D, Ausili A, Regoli F (2011) Assessing sediment hazard through a weight of evidence approach with bioindicator organisms: A practical model to elaborate data from sediment chemistry, bioavailability, biomarkers and ecotoxicological bioassays. Chemosphere 83(4):475-485. doi: 10.1016/j.chemosphere.2010.12.064

Priadi C, Bourgeault A, Ayrault S, Gourlay-France C, Tusseau-Vuillemin M-H, Bonte P, Mouchel J-M (2011) Spatio-temporal variability of solid, total dissolved and labile metal: passive *vs.* discrete sampling evaluation in river metal monitoring. J Environ Monit 13(5):1470-1479. doi: 10.1039/c0em00713g

Quillardet P, Hofnung M (1985) The SOS Chromotest, a colorimetric bacterial assay for genotoxins: procedures. Mutation Research - Environmental Mutagenesis and Related Subjects 147(3):65-78. doi: 10.1016/0165-1161(85)90020-2

Tamtam F, Mercier F, Le Bot B, Eurin J, Tuc Dinh Q, Clément M, Chevreuil M (2008) Occurrence and fate of antibiotics in the Seine River in various hydrological conditions. Sci Total Environ 393(1):84-95. doi: 10.1016/j.scitotenv.2007.12.009

Teil M-J, Tlili K, Blanchard M, Labadie P, Alliot F, Chevreuil M (2014) Polychlorinated biphenyls, polybrominated diphenyl ethers, and phthalates in roach from the Seine River basin (France): Impact of densely urbanized areas. Arch Environ Contam Toxicol 66(1):41-57. doi: 10.1007/s00244-013-9955-8

Thévenot DR, Moilleron R, Lestel L, Gromaire M-C, Rocher V, Cambier P, Bonté P, Colin J-L, de Pontevès C, Meybeck M (2007) Critical budget of metal sources and pathways in the Seine River basin (1994-2003) for Cd, Cr, Cu, Hg, Ni, Pb and Zn. Sci Total Environ 375(1-3):180-203. doi: 10.1016/j.scitotenv.2006.12.008

Vicquelin L, Leray-Forget J, Peluhet L, LeMenach K, Deflandre B, Anschutz P, Etcheber H, Morin B, Budzinski H, Cachot J (2011) A new spiked sediment assay using embryos of the Japanese medaka specifically designed for a reliable toxicity assessment of hydrophobic chemicals. Aquat Toxicol 105(3-4):235-245. doi: 10.1016/j.aquatox.2011.06.011

Wilson VS, Bobseine K, Lambright CR, Gray Jr LE (2002) A novel cell line, MDA-kb2, that stably expresses an androgen- and glucocorticoid-responsive reporter for the detection of hormone receptor agonists and antagonists. Toxicol Sci 66(1):69-81. doi: 10.1093/toxsci/66.1.69

Xuereb B, Chaumot A, Mons R, Garric J, Geffard O (2009) Acetylcholinesterase activity in *Gammarus fossarum* (Crustacea Amphipoda): Intrinsic variability, reference levels, and a reliable tool for field surveys. Aquat Toxicol 93(4):225-233. doi: 10.1016/j.aquatox.2009.05.006

Figure captions

Fig. 1 Map of the three sampling stations (Marnay, Bougival, and Triel) along the Seine River (in the north of France) adapted from Faburé et al. (2015). Arrows indicate the water flow direction. The city of Paris is represented in dark grey, and the densely urbanized area surrounding the French capital is colored in light grey. Source (France map): www.histgeo.ac-aix-marseille.fr, © Daniel Dalet

Fig. 2 Contribution of each class of chemicals to ChemHQwater values (a), and $ChemHQ_{sed}$ values using reference values for individual substances (b), or for each class of compounds (c). See Table 1 footnote for details. PESTs, pesticides; AKPs, alkylphenols, PFAS, perfluoroalkyl substance (PFOS); MEs, metal elements; PAHs, polycyclic aromatic hydrocarbons, PCBs, polychlorinated biphenyls; PBDEs, polybromodiphenyl ethers; OCPs, organochlorine pesticides; CX, X^{th} campaign; AA, annual average value (mean of concentrations from the C1 to C4 campaigns)

Fig. 3 Contribution of each class of chemicals to BioavHQ values (a), and of each class of biomarkers to BiomHQ values (b). Only the contributions of chemicals to the annual average (AA) BioavHQ values calculated for each sampling site are shown. PAHs, polycyclic aromatic hydrocarbons, PCBs, polychlorinated biphenyls; PBDEs, polybromodiphenyl ethers; OCPs, organochlorine pesticides; MEs, metal elements. Within LOE#3 (BiomHQ values), neurotoxicity was investigated by assessing AChE activity. Energy acquisition markers included cellulase, trypsin, and amylase enzymatic activities. Survival was assessed using mortality rates. Feeding rate measurements were used to track feeding behavior. Reproduction was studied through molt delay and the number of embryos and oocytes *per* female. Please note that reproduction markers and cellulase activity were not investigated in the C4 campaign. CX, X^{th} campaign; AA, annual average value (mean of effects calculated from the C1 to C4 campaigns); N/A, not applicable (as BiomHQ was equal to 0 for Marnay C4)

Fig. 4 Contribution of each class of bioassay endpoints to $ToxHQ_{water}$ (a) and $ToxHQ_{sed}$ (b) values. Survival endpoints included embryonic and larval viability (MELA). Development was characterized by recording hatching success and time to hatch (MELA). The percentage of abnormal larvae was selected to illustrate teratogenicity (MELA). Biometric measurements of larvae at hatching, including total body length and head size, were used to evaluate *in ovo* growth (MELA). ER, TR, AR/GR, and Anti-AR induction factors were gathered to study endocrine-disrupting potency (ED *in vitro* bioassays). Cytotoxicity and genotoxicity potencies of sediment elutriates were evaluated using *in*

vitro bioassays, using Microtox and SOS Chromotest procedures, respectively. CX, X^{th} campaign; AA, annual average value (mean of effects E(i) calculated from the C1 to C4 campaigns)

Fig. 5 Weight of Evidence indices (WOEs) and associated hazard classes integrating the results of each LOE calculated for the three stations during the four sampling campaigns (C1 to C4), and annual average (AA) values. The hazard class attributed to each LOE hazard quotient (HQ) is summarized in the table below. Please note that in the C2 campaign, only $ChemHQ_{water}$ and BioavHQ were evaluated

Fig. 6 Contribution of each LOE hazard quotient (HQ) to the global WOE values calculated for the Marnay (a), Bougival (b), and Triel (c) stations. CX, X^{th} campaign; AA, annual average value. Please note that as only $ChemHQ_{water}$ and BioavHQ were evaluated in the C2 campaign, details on the contributions for that campaign are not presented

Fig. 7 3D-Non-metric Multi-Dimensional Scaling graphical representation of the dissimilarity between bacterial communities of the different samples. Mar, Marnay; Bou, Bougival; Tri, Triel; Sept11, C1 campaign; Jul12, C3 campaign; Dec12, C4 campaign

Supplementary Material captions (ESM_1.pdf)

Supplementary Tables

Table S1 List of chemical parameters included into the WOE approach for the water column (LOE#1), and associated weightings, CAS numbers and selected reference values

Table S2 List of chemical parameters included into the WOE approach for the sediment (LOE#1), and associated weightings, CAS numbers and selected reference values

Table S3 List of chemical parameters included into the WOE approach for bioavailability (LOE#2), selected reference values and hazard classes attributed to $RTR_w(i)$ values

Table S4 Reference values (Ref.) and thresholds (Th(%)) assigned to biomarker responses (LOE#3) analyzed in the study

Table S5 Details of the hazard classes attributed to the effect E(i) values for biomarker responses (LOE#3)

Table S6 Assigned weightings for the biomarker line of evidence (LOE#3) according to Piva et al. (2011)

Table S7 Reference values (Ref.) and thresholds (Th(%)) assigned to each endpoint for *in vivo* (MELA) and *in vitro* (ED, Microtox, SOS Chromotest) bioassays analyzed in the study (LOE#4)

Table S8 Details of the weighting calculations according to endpoint, matrix and exposure time, adapted from Piva et al. (2011)

Table S9 Concentrations of organic micropollutants and metal elements measured in the water column (<0.7 µm) at the three stations during each sampling campaign, and integrated into the Weight-of-Evidence approach (LOE#1)

Table S10 Concentrations of organic micropollutants and metal elements measured in sediment (<2 mm) at the three stations during each sampling campaign, and integrated into the Weight-of-Evidence approach (LOE#1)

Table S11 Main physico-chemical characteristics and organic contamination levels in sediments sampled along the Seine River

Table S12 Bioaccumulation levels of organic micropollutants and metal elements measured in gammarids at the three stations during each sampling campaign, and integrated into the Weight-of-Evidence approach (LOE#2)

Table S13 Biomarker responses in gammarids at the three stations during each sampling campaign, integrated into the Weight-of-Evidence approach (LOE#3)

Table S14 Bioassay endpoint values observed at the three stations during each sampling campaign, and integrated into the Weight-of-Evidence approach (LOE#4)

Supplementary Figures

Fig. S1 Details on the calculation procedure implemented within LOE#1 (sediment and water column chemistry), adapted from Piva et al. (2011)

Fig. S2 Details on the calculation procedure implemented within LOE#2 (bioavailability), adapted from Piva et al. (2011)

Fig. S3 Details on the calculation procedure implemented within LOE#3 (biomarkers), adapted from Piva et al. (2011)

Fig. S4 Details on the calculation procedure implemented within LOE#4 (bioassays), adapted from Piva et al. (2011)

Fig. S5 Details on the calculation procedure implemented within WOE integration, adapted from Piva et al. (2011)

Figures

Fig. 1

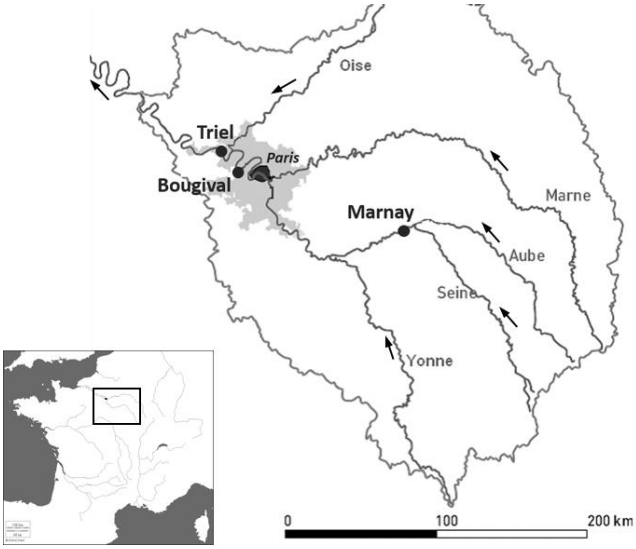
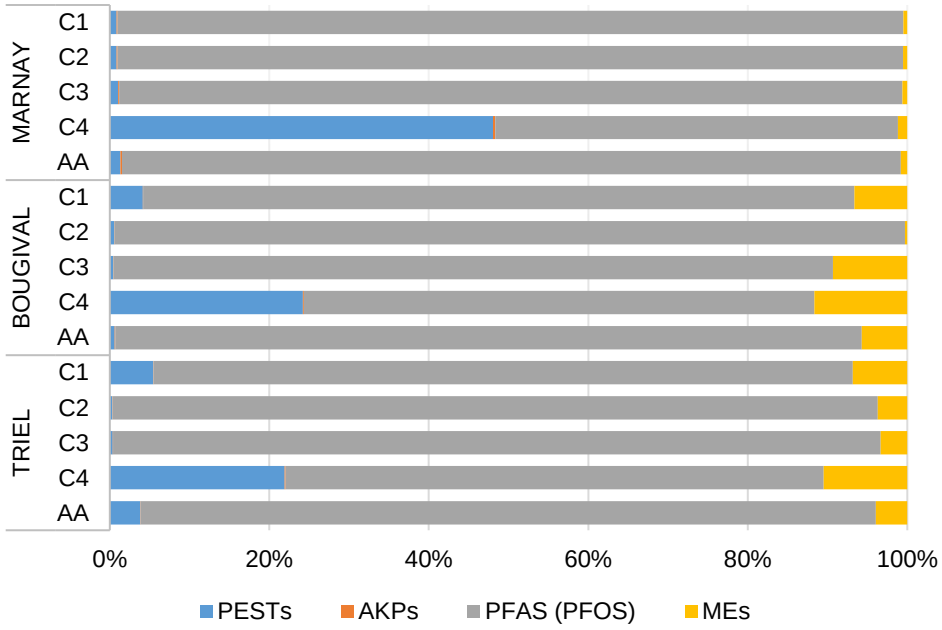
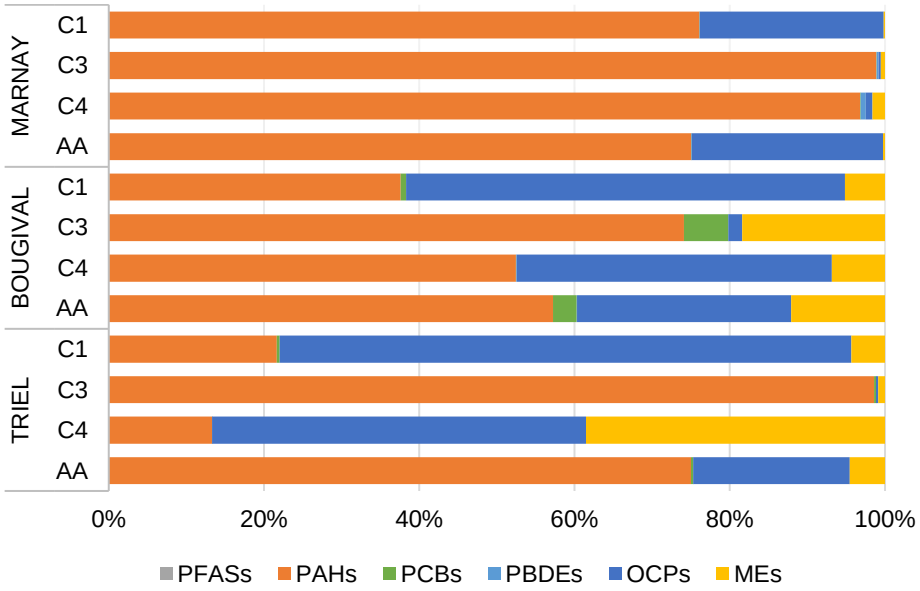


Fig. 2

(a)


MARNAY
BOUGIVAL
TRIEL
C1
C2
C3
C4
AA
0%
20%
40%
60%
80%
100%
PESTs
AKPs
PFAS (PFOS)
MEs

(b)


MARNAY
BOUGIVAL
TRIEL
C1
C3
C4
AA
0%
20%
40%
60%
80%
100%
PFASs
PAHs
PCBs
PBDEs
OCPs
MEs

(c)

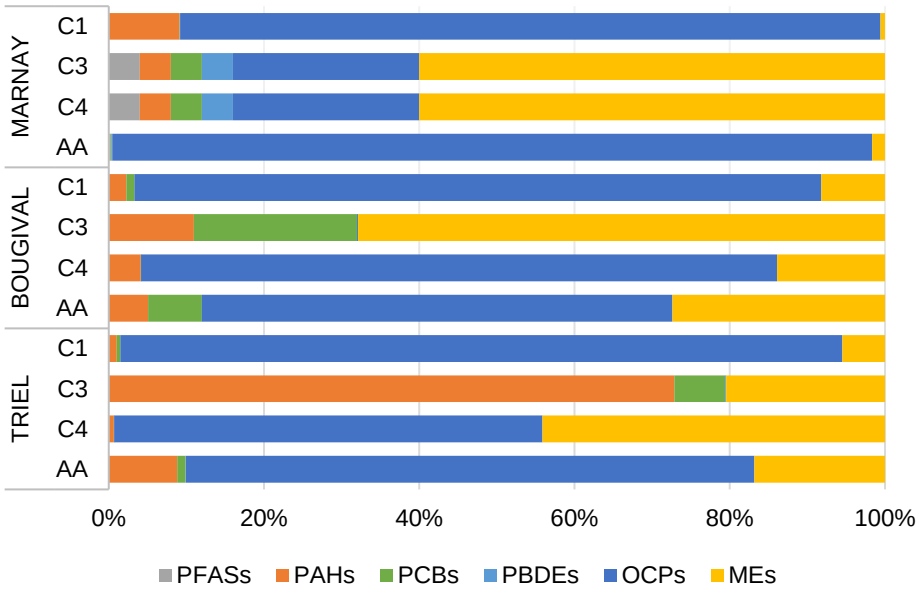
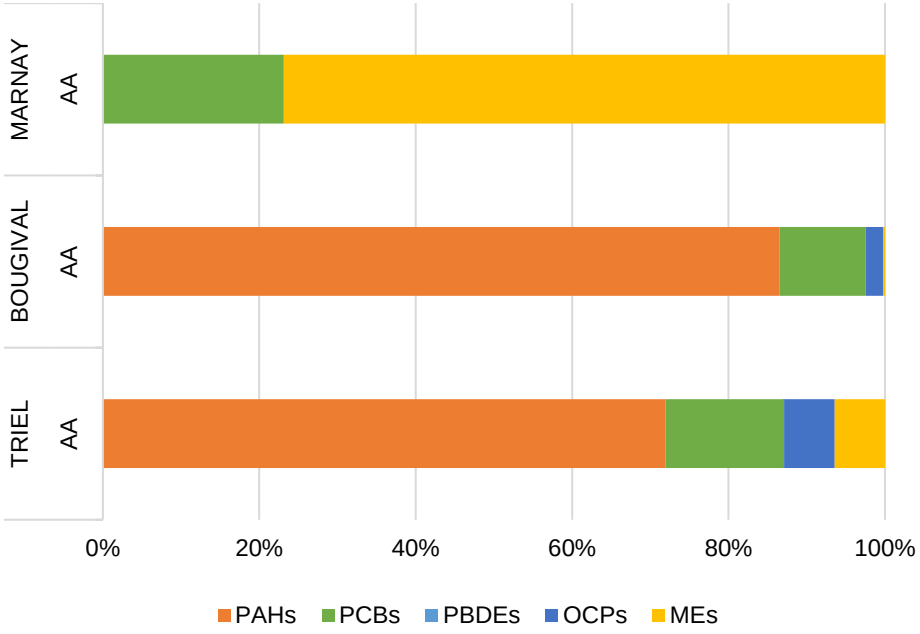

MARNAY
C1
C3
C4
AA
BOUGIVAL
C1
C3
C4
AA
TRIEL
C1
C3
C4
AA
0%
20%
40%
60%
80%
100%
PFASs
PAHs
PCBs
PBDEs
OCPs
MEs

Fig. 3

(a)


MARNAY
AA
BOUGIVAL
AA
TRIEL
AA
0%
20%
40%
60%
80%
100%
PAHs
PCBs
PBDEs
OCPs
MEs

(b)

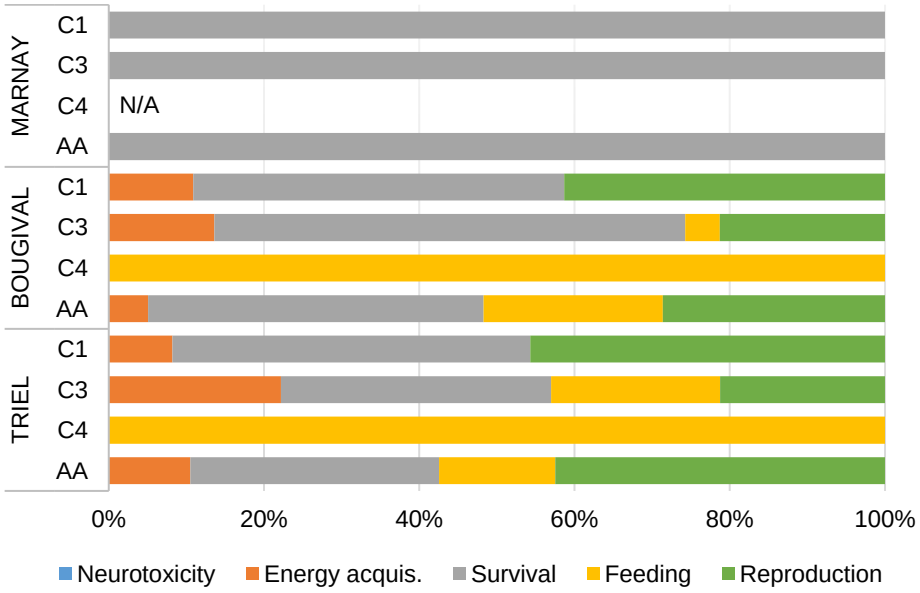
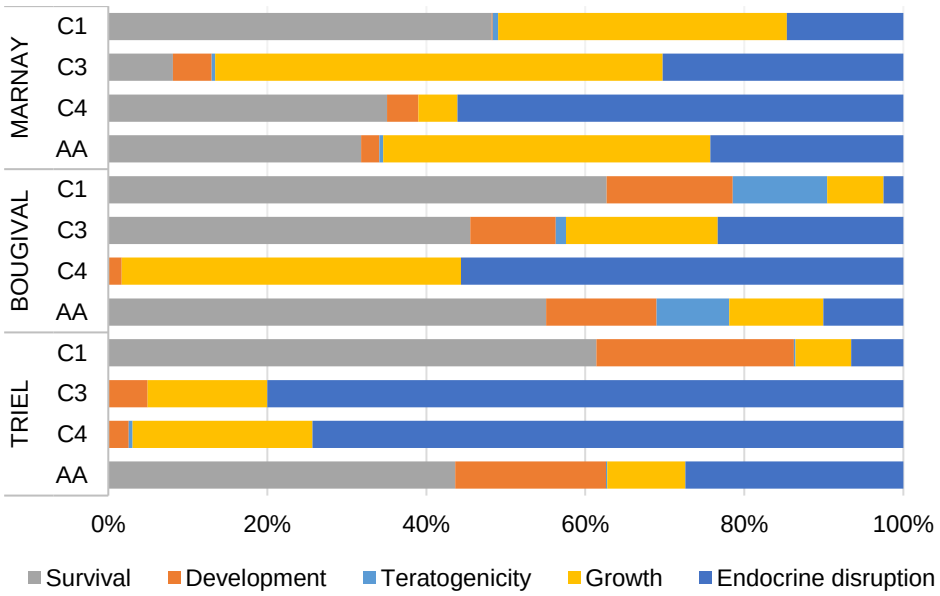

MARNAY
C1
C3
C4
N/A
AA
BOUGIVAL
C1
C3
C4
AA
TRIEL
C1
C3
C4
AA
0%
20%
40%
60%
80%
100%
Neurotoxicity
Energy acquis.
Survival
Feeding
Reproduction

Fig. 4

(a)


MARNAY
BOUGIVAL
TRIEL
C1
C3
C4
AA
0%
20%
40%
60%
80%
100%
Survival
Development
Teratogenicity
Growth
Endocrine disruption

(b)

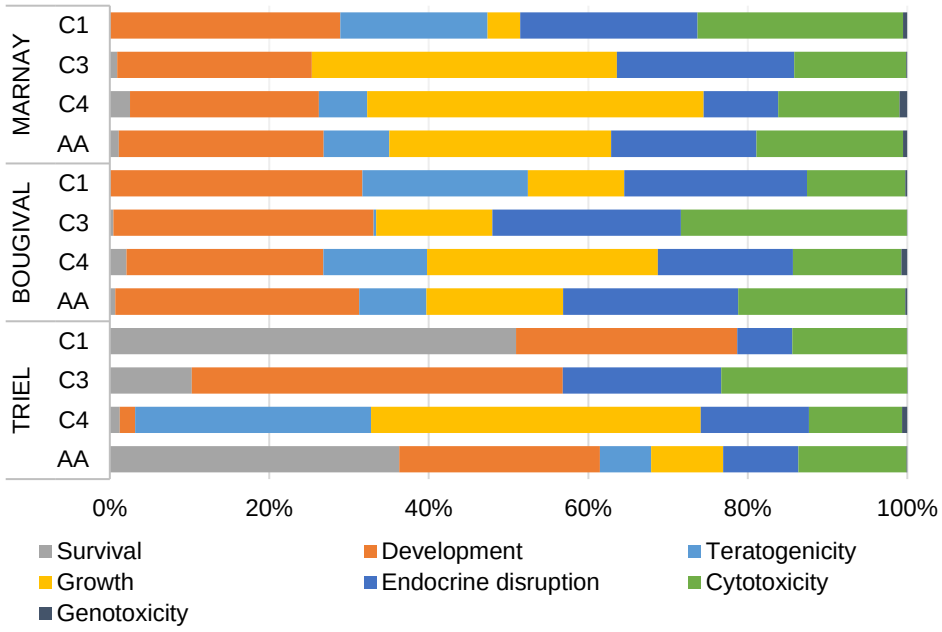
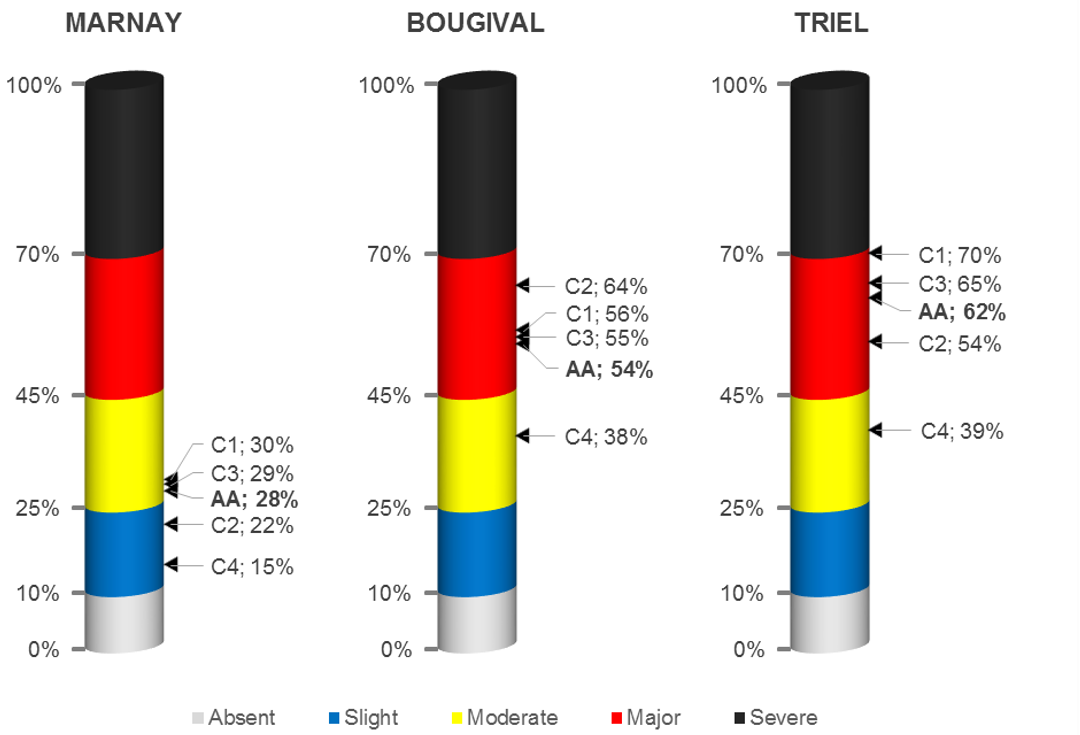

MARNAY
BOUGIVAL
TRIEL
C1
C3
C4
AA
0%
20%
40%
60%
80%
100%
Survival
Development
Teratogenicity
Growth
Endocrine disruption
Cytotoxicity
Genotoxicity

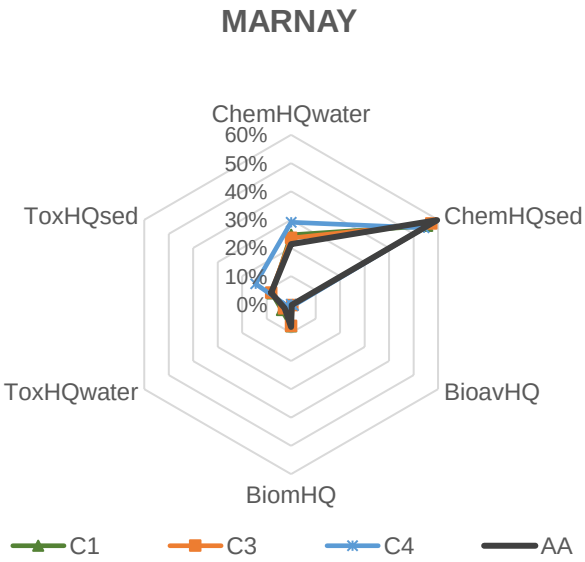
Fig. 5



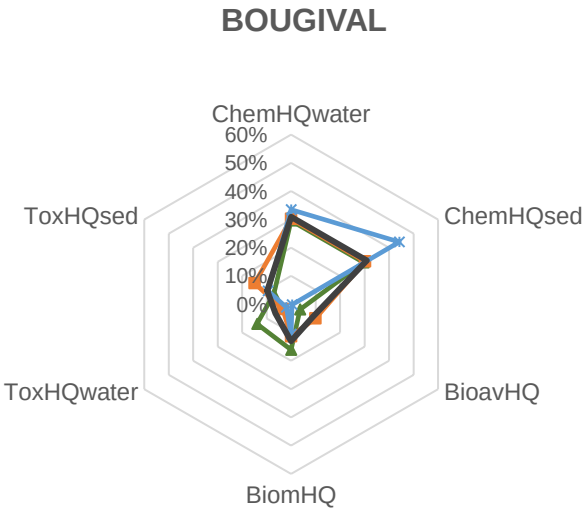
	Summary of the hazard classes attributed to the HQs values within each LOE														
HQs	MAR_C1	MAR_C2	MAR_C3	MAR_C4	**MAR_AA**	BOU_C1	BOU_C2	BOU_C3	BOU_C4	**BOU_AA**	TRI_C1	TRI_C2	TRI_C3	TRI_C4	**TRI_AA**
$ChemHQ_{water}$	Moderate	Moderate	Moderate	Moderate	Moderate	Severe	Severe	Severe	Major	Severe	Severe	Severe	Severe	Major	Severe
$ChemHQ_{sed}$	Severe	*N/A*	Severe	Moderate	Severe	Severe	*N/A*	Severe	Severe	Severe	Severe	*N/A*	Severe	Severe	Severe
BioavHQ	Slight	Slight	Slight	Slight	Slight	Major	Major	Major	Slight	Major	Major	Moderate	Moderate	Slight	Moderate
BiomHQ	Moderate	*N/A*	Moderate	Slight	Moderate	Severe	*N/A*	Major	Major	Major	Severe	*N/A*	Severe	Major	Severe
$ToxHQ_{water}$	Absent	*N/A*	Absent	Absent	Absent	Moderate	*N/A*	Absent	Absent	Slight	Major	*N/A*	Slight	Absent	Slight
$ToxHQ_{sed}$	Slight	*N/A*	Slight	Slight	Slight	Slight	*N/A*	Moderate	Slight	Moderate	Severe	*N/A*	Severe	Slight	Major

Fig. 6

(a)


MARNAY
ChemHQwater
60%
50%
40%
30%
20%
10%
0%
ChemHQsed
BioavHQ
BiomHQ
ToxHQwater
ToxHQsed
C1
C3
C4
AA

(b)


BOUGIVAL
ChemHQwater
60%
50%
40%
30%
20%
10%
0%
ChemHQsed
BioavHQ
BiomHQ
ToxHQwater
ToxHQsed

(c)

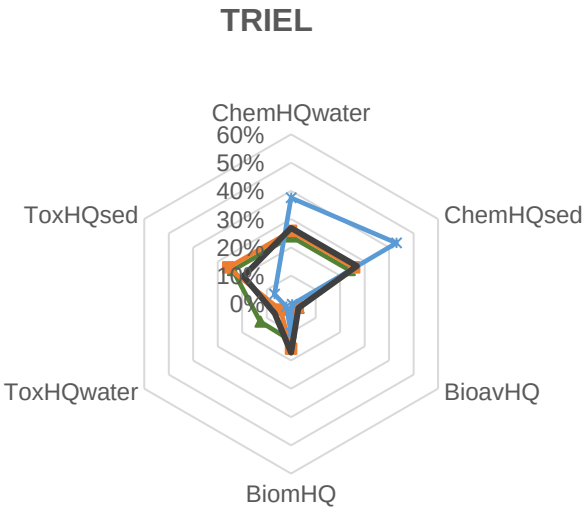

TRIEL
ChemHQwater
60%
50%
40%
30%
20%
10%
0%
ChemHQsed
BioavHQ
BiomHQ
ToxHQwater
ToxHQsed

Fig. 7

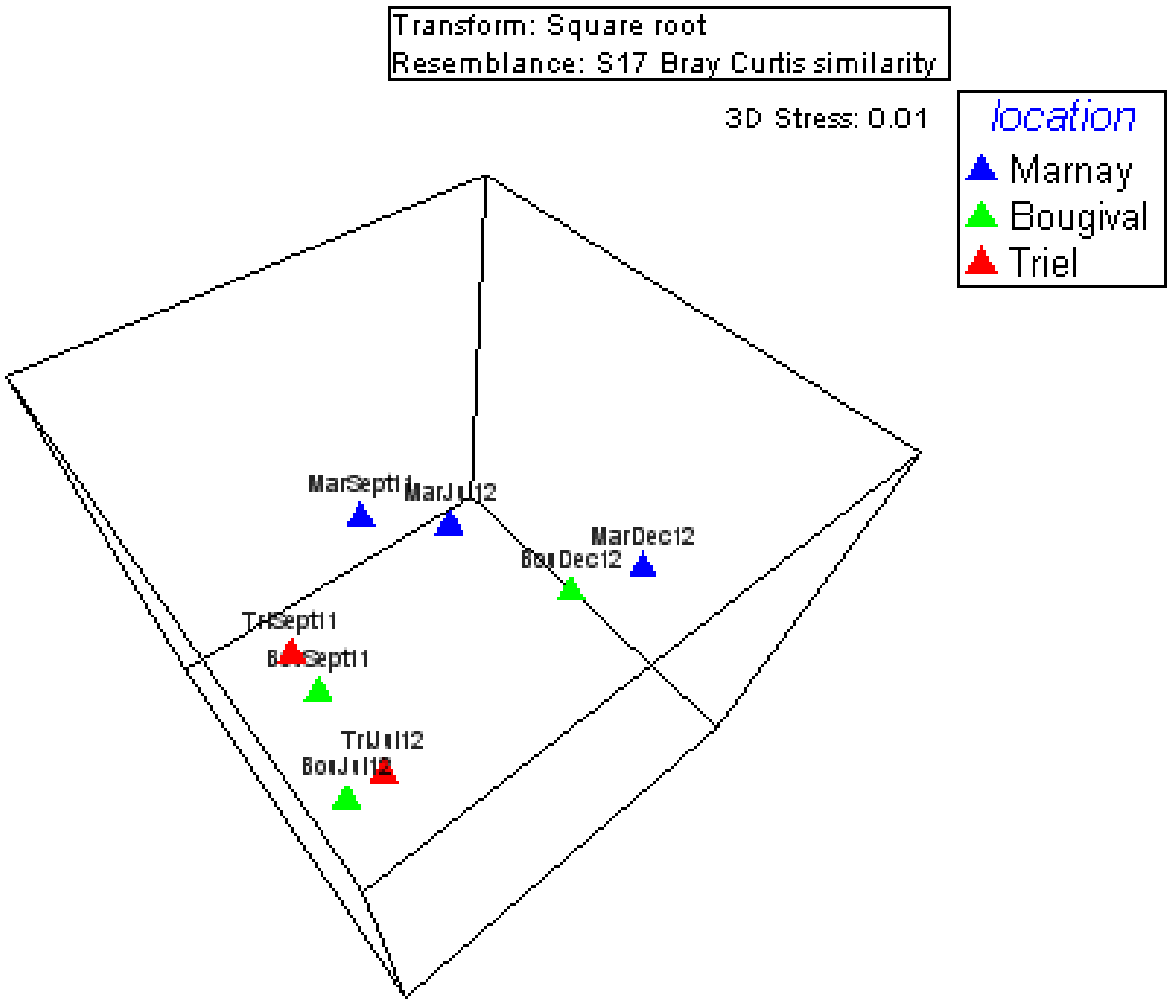

Transform: Square root
Resemblance: S17 Bray Curtis similarity
3D Stress: 0.01
location
Marnay
Bougival
Triel
MarSept11
MarJul12
MarDec12
BouDec12
TriSept11
BouSept11
TriJul12
BouJul12

Table 1 Chemical Hazard Quotients calculated for the water column ($ChemHQ_{water}$) and for sediments ($ChemHQ_{sed}$)

	Water column		Sediment			
	$ChemHQ_{water}$	**Hazard class**	**$ChemHQ_{sed}$[a]**	**Hazard class**	**$ChemHQ_{sed}$[b]**	**Hazard class**
MAR_C1	5.8	Moderate	92.6	Severe	24.3	Severe
MAR_C2	5.6	Moderate	*N/A*	*N/A*	*N/A*	*N/A*
MAR_C3	5.3	Moderate	17.1	Severe	0.3	Absent
MAR_C4	3.4	Moderate	6.2	Moderate	0.3	Absent
MAR_AA	4.7	Moderate	41.2	Severe	10.4	Major
BOU_C1	33.3	Severe	318	Severe	200	Severe
BOU_C2	14.9	Severe	*N/A*	*N/A*	*N/A*	*N/A*
BOU_C3	25.1	Severe	427	Severe	116	Severe
BOU_C4	9.9	Major	150	Severe	74.3	Severe
BOU_AA	19.6	Severe	301	Severe	133	Severe
TRI_C1	52.2	Severe	866	Severe	681	Severe
TRI_C2	29.1	Severe	*N/A*	*N/A*	*N/A*	*N/A*
TRI_C3	32.2	Severe	2,756	Severe	117	Severe
TRI_C4	11.3	Major	296	Severe	258	Severe
TRI_AA	30.5	Severe	1,309	Severe	355	Severe

a, Hazard quotient calculated using reference value (and measured concentration) for individual substance when available (excepted for ΣPBCs, ΣDDT, ΣDDE and ΣDDD; see ESM Table S2 for definition)

b, Hazard quotient calculated using reference value (and measured concentration) for a class of substances when available (for ΣDDTs, ΣPBDEs and ΣPAHs; see ESM Table S2 for definition)

N/A, not applicable; MAR, Marnay; BOU, Bougival; TRI, Triel, CX, X^{th} campaign; AA, annual average (mean of concentrations for C1 to C4 campaigns) value

Table 2 Bioavailability Hazard Quotients (BioavHQs) calculated using bioaccumulation levels in *Gammarus fossarum* following a 7 days-*in situ* exposure

	BioavHQ	Hazard class index[a]	Hazard class
MAR_C1	5.8	147%	Slight
MAR_C2	6.0	135%	Slight
MAR_C3	13.4	165%	Slight
MAR_C4	1.8	150%	Slight
MAR_AA	6.5	129%	Slight
BOU_C1	58.6	924%	Major
BOU_C2	189	2006%	Major
BOU_C3	262	2300%	Major
BOU_C4	1.8	150%	Slight
BOU_AA	233	2012%	Major
TRI_C1	66.0	924%	Major
TRI_C2	64.7	665%	Moderate
TRI_C3	75.7	788%	Moderate
TRI_C4	1.8	150%	Slight
TRI_AA	77.4	829%	Moderate

a, Index used to attribute the hazard class (see Section *2.7.2* for details on calculation).
MAR, Marnay; BOU, Bougival; TRI, Triel, CX, Xth campaign; AA, annual average (mean of the measurements for C1 to C4 campaigns) value

Table 3 Biomarker Hazard Quotients (BiomHQs) calculated using biomarker responses in *Gammarus fossarum* following *in situ* exposure along the Seine River

	BiomHQ	Hazard class index[a]	Hazard class
MAR_C1	26.4	154%	Moderate
MAR_C3	9.3	151%	Moderate
MAR_C4	0.0	70%	Slight
MAR_AA	11.9	151%	Moderate
BOU_C1	37.4	401%	Severe
BOU_C3	17.8	298%	Major
BOU_C4	12.5	216%	Major
BOU_AA	22.2	327%	Major
TRI_C1	51.2	401%	Severe
TRI_C3	26.9	446%	Severe
TRI_C4	9.4	216%	Major
TRI_AA	34.2	460%	Severe

a, Index used to attribute the hazard class (see Section *2.7.3* for details on calculation).
MAR, Marnay; BOU, Bougival; TRI, Triel, CX, Xth campaign; AA, annual average (estimated from the mean of the effects E(i) calculated for C1 to C4 campaigns) value

Table 4 Bioassay Hazard Quotients calculated for the water column ($ToxHQ_{water}$) and for sediment ($ToxHQ_{sed}$)

	Water column			Sediment		
	$ToxHQ_{water}$	**Hazard class index[a]**	**Hazard class**	**$ToxHQ_{sed}$**	**Hazard class index[a]**	**Hazard class**
MAR_C1	4.8	0.6	Absent	16.2	1.1	Slight
MAR_C3	3.4	0.4	Absent	16.5	1.1	Slight
MAR_C4	0.8	0.1	Absent	14.9	1.0	Slight
MAR_AA	3.0	0.3	Absent	15.9	1.1	Slight
BOU_C1	32.7	3.8	Moderate	28.9	1.9	Slight
BOU_C3	7.8	0.9	Absent	57.6	3.9	Moderate
BOU_C4	3.2	0.4	Absent	25.0	1.7	Slight
BOU_AA	14.5	1.7	Slight	37.2	2.5	Moderate
TRI_C1	36.5	4.2	Major	113	10	Severe
TRI_C3	11.6	1.3	Slight	68.6	8.5	Severe
TRI_C4	3.2	0.4	Absent	18.2	1.3	Slight
TRI_AA	17.1	2.0	Slight	84.1	5.8	Major

a, index used to attribute the hazard class (see Section *2.7.4* for details on calculation).
MAR, Marnay; BOU, Bougival; TRI, Triel, CX, X^{th} campaign; AA, annual average (estimated from the mean of the effects E(i) calculated for C1 to C4 campaigns) value