The SOLUTIONS project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 603437

Solutions for present and future emerging pollutants in land and water resources management

THEME

ENV.2013.6.2-2

Toxicants, environmental pollutants and land and water resources management

Start date of project: 1st October 2013 Duration: 5 years

External Deliverable 13.1

Diagnostic toolbox for ecological effects of pollutant mixtures, including bio-tests, trait-based database and detection tool and WoE studies at hot-spot sites

---

Due date of deliverable: 31.07.2017

Actual submission date: 28.08.2017

Author(s)

Thomas Backhaus

(University of Gothenburg, UGOT, Gothenburg, Sweden)

in cooperation with

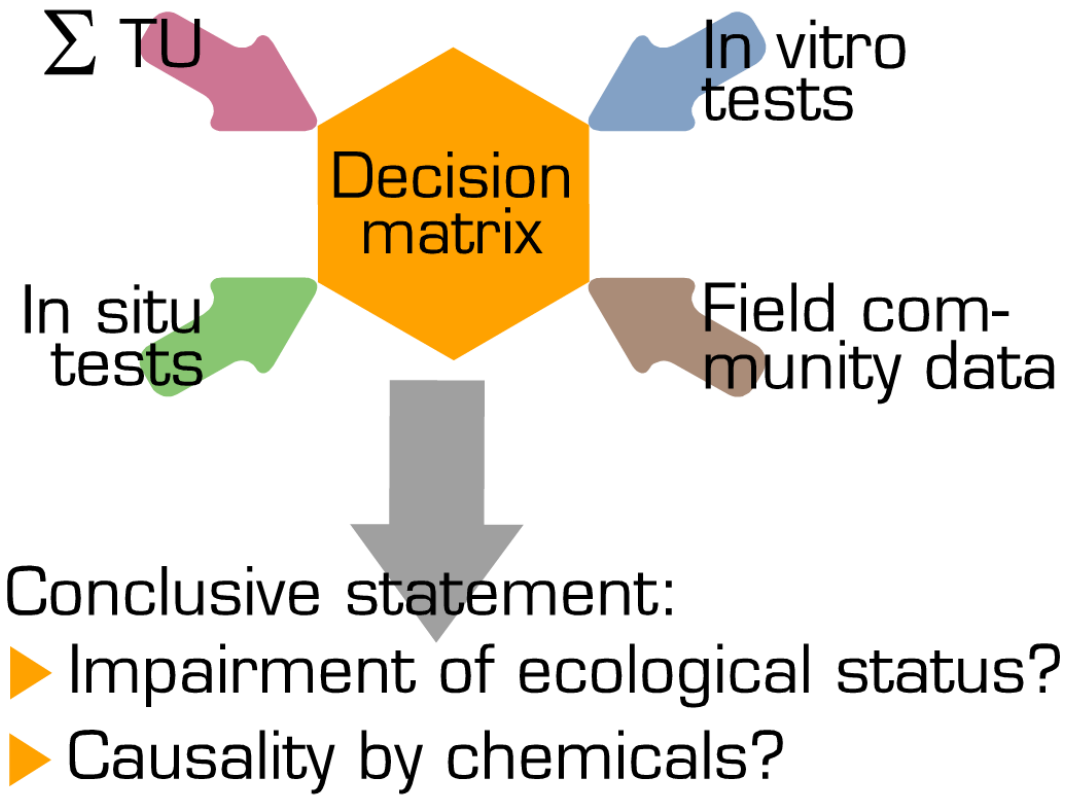
Helmut Segner, Henner Hollert, Björn Deutschmann, Paul van den Brink, Thomas-Benjamin Seiler, Ivana Teodorovic, Andreas Focks

Dissemination Level		
PU	Public	
PP	Restricted to other programme participants (including EC)	X
RE	Restricted to a group specified by the consortium (including EC)	
CO	Confidential, only for members of the consortium (including EC)	

I. Summary

The toolbox for the detection of the ecological impact of chemicals uses a statistically supported, transparent and formalized weight of evidence (WOE) approach that integrates four main individual lines of evidence (LOEs), (i) predictive mixture modelling, (ii) effect-directed analysis (EDA), (iii) *in situ* tests, and (iv) field-based monitoring studies. A systematic and quantitative method was developed for the aggregation of multiple *in situ* test results into one LOE, resulting in the definition of the average biomarker response (ABR). Integration of single LOE in a weight of evidence approach was defined in form of a decision matrix. The main idea of the approach is to systematically integrate these four LOEs, so that their strengths complement each other and allow a transparent site-specific assessment with particular attention to the establishment of links between chemical exposure and ecological impacts, identification of data gaps and management options. The focus for the development was to keep the methodology simple enough to enable routine use by non-scientists. Three practical weight of evidence examples are presented in addition, illustrating specific aspects of weight of evidence studies. The developed toolbox was applied to the Danube case study, to facilitate evaluation of the very comprehensive data set from Joint Danube Survey 3. The Rhine and the Holtemme cases are smaller scale studies focused on site specific toolbox application in an upstream/downstream set-up. The toolbox concept proved to be practical, simple and promising for further studies, with fairly high diagnostic power.

II. Graph


∑ TU
In vitro tests
Decision matrix
In situ tests
Field community data
Conclusive statement:
Impairment of ecological status?
Causality by chemicals?

III. List of Contents

III.1 Table of Contents

III.2 Tables and Figures

III.3 Glossary

ABR: Average biomarkers response. An index for the overall strength of response from a set of biomarkers, providing a statistically sound and meaningful aggregation of single biotest results (definition in chapter 3.2). This index can be weighted by specificity of the single methods for chemical impact (→ IoC) or by the ecological relevance of the biotest (→IoEEI)

EDA: Effect-Directed Analysis. A methodology that is based on the repeated chemical fractionation of a complex environmental sample into simpler and simpler sub-samples, accompanied by their ecotoxicological characterisation in a suite of small-scale ecotoxicological assays. The aim is to simplify the original sample with complex chemical contamination sufficiently, so that the identification of priority substances becomes feasible. Largely synonymous terms are EDF (Effect-Directed Fractionation) and TIE (Toxicity Identification and Evalution). For detailed overviews, see Brack (2011, 2016).

EDF: Effect-Directed Fingerprinting. The use of a battery of complementary bioassays (often *in vitro* or small-scale), in order to obtain characteristic response patterns that allow to draw conclusions on causative agents. Sometimes termed "effect-based screening". Details and application examples can be found in e.g. Schulze (2015), Neale (2017).

EQS: Environmental Quality Standard, as defined by the WFD as "the concentration of a particular pollutant or group of pollutants in water, sediment or biota which should not be exceeded in order to protect human health and the environment."

***In situ* tests**: Ecotoxicological experiments that are conducted in the field using caged organisms in order to measure the effects of real-world pollution scenarios, under realistic environmental conditions. Details and examples can be found in e.g. Baird (2007).

IoC: Index of Causality. A measure for the overall strength of evidence that a given chemical or chemical mixture is responsible for an observed effect. See chapter 3.2.

IoEEI: Index of Expected Ecological Impact. A measure for the overall expected ecological impact (as distinguished from e.g. isolated effects on individuals or populations of laboratory test organisms). See chapter 3.2.

LOE: Line Of Evidence. In the context of this report a term that denotes an individual (type of) ecotoxicologal study that provides a certain dataset and evidence type in the context of a broader study aim. The more independent LOE's provide similar results, the stronger the resulting overall conclusion of a study (→WoE). In this report the following LOE's are distinguished: →EDF, →EDA, →*in situ* tests, →STU.

NOEC: No Observed Effect Concentration. The highest tested concentration of a study in which no statistically significant effects of a chemical were observed.

PNEC: Predicted No Effect Concentration, as defined by REACH as "the concentration of the substance below which adverse effects in the environmental sphere of concern are not expected to occur"

STU: Arithmetic sum of → Toxic Units.

Toxic Unit: The ratio between the concentration of a compound found or estimated to occur in an environmental compartment and either a measure of its ecotoxicity (e.g. EC50, NOEC) or its environmental threshold (e.g. PNEC).

WoE: Weight Of Evidence. Concluding evaluation of the overall support that a suite of →LOE's lends a certain conclusion. A systematic WoE evaluation is suggested in this report in table 13.

1 Introduction

Toxic chemicals from point and diffuse sources might impact the ecological status of aquatic ecosystems. Appropriate strategies are therefore needed to identify impacted sites, quantify impacts, evaluate the causative involvement of chemical contaminants, identify trends and rank sites for the implementation of management measures. Since environmental compartments usually contain mixtures of chemicals with low, possibly non-toxic concentrations of the individual compounds, any approach to identify casual links between ecological impacts and chemical contamination has to involve concepts for mixture toxicity. However, In addition to toxic chemicals, other stressors such as habitat degradation and –fragmentation, nutrient pollution, oxygen depletion, pH shifts, temperature changes, invasive species, and hydromorphological changes, either alone or in combination with chemical stressors, may also cause a site to fail achieving good ecological status. Since the EU Water Framework Directive (WFD) aims at a good ecological status of all European water bodies through addressing water pollution, for water quality monitoring and assessment under WFD it is necessary to discriminate the impact of such non-chemical stressors from the effects of toxic chemicals – which often occur in complex mixtures of low individual concentrations –. This is challenging, and no single "one size fits all" strategy exists. Therefore, multiparametric approaches, so-called "toolboxes", are often used. A critical question in this context is whether the focus of a study is on i) impact description, identification & quantification or ii) on the analysis of causality & stressor identification.

Several mutually supporting approaches and sets of data are used in order to identify ecological impacts and to link them to the presence of chemicals at a site. The fundamental classes are:

1. Chemical monitoring profiles, amended with published ecotoxicological information and environmental thresholds (e.g. PNEC's, EQS values etc).
2. Effect-driven methods that combine chemical analytics with small-scale ecotoxicological assays and in-*vitro* test systems.
3. *In situ* approaches for the identification of chemical exposure or effects.
4. Community indices based on field surveys of community composition.

Each approach provides indicators for site conditions that allow to draw conclusions on, chemical and non-chemical stressors which might be the underlying cause of ecological status impairment, which biological elements are impacted, and which management measures might be useful in order to improve the status of an impacted site. However, the approaches have different focuses, strengths and weaknesses. Chemical monitoring and effect-driven methods reveal the chemical burden and allow to identify key toxicants, but their link to ecological outcomes is weak. In contrast, while *in situ* approaches and community indices provide increased ecological relevance, linking the effects to the causative stressors is difficult.. A weight-of-evidence approach (WOE) is often used to tie the different strands of evidence together in order to provide a more reliable and meaningful characterization of the environmental status and maximize the chance to diagnose the cause of an environmental disturbance (Suter 1993, Martinez-Haro et al. 2015). Any such investigation will always be a compromise between

legislative requirements, scientific knowledge, available experimental tools and resources (which, in turn, are determined by political and societal developments and value judgements).

The following text provides a short overview of each of the four approaches, briefly highlighting their strengths and limitations. The main focus is then to provide a "toolbox" of approaches that can be used for assessing the ecological impact of chemical mixtures at an exposed site, based on published information, previous experiences as well as the available ecotoxicological and ecological knowledge collected within SOLUTIONS. Particular emphasis will be put on *in situ* biotests, as an experimental approach that investigates the impact of site-specific pollution patterns on well-defined biological entities (most individuals, but also whole communities in the case of microorgansisms), under environmentally realistic conditions.

The novelty of the suggested toolbox is that we utilize statistically supported, transparent and formalized WOE approaches for establishing links between chemical exposure and ecological impacts, and instrumentalize mechanistic data and information for substantiating such WOE-derived linkages. By using a WOE approach, the developed toolbox informs whether (and which) biota at a site actually responds to existing exposures, and whether this compromises ecological structures and functions.

2 The principal approaches for ecological status assessment and analysis of causality

Table 1 provides an overview of the four main approaches for ecological status assessment and causality analysis as discussed below. The outlined approaches are, apart from field surveys, not yet elements for the ecological status assessment under WFD. Especially Effect-directed assessment and biomarkers are, however suggested and discussed to be considered in this context, so that in this text the term ecological status is used in a slightly wider sense as coined for the WFD.

2.1 Predictive chemical mixture toxicity modelling

Multi-residue target and non-target screening techniques provide a constantly improving overview of the chemicals occurring and co-occurring in aquatic ecosystems (e.g. Petrie et al., 2014; Gosetti et al., 2016; Richardson and Kimura, 2016; Inostroza et al., 2016a; Moschet et al., 2017). However, analytical profiles in themselves do not provide information about environmental hazards or risks. For this purpose, chemical-analytical fingerprints are combined with information on environmental hazards, either in the form of ecotoxicological data from various bioassays (No Observed Effect Concentrations (NOECs) or EC50 values) or in the form of environmental thresholds (environmental quality standards (EQSs), predicted no effect concentrations (PNECs), regulatory acceptable concentrations (RACs), ecotoxicological assessment criteria (EACs)). Most often the sum of the individual risk quotients (i.e. the ratio of measured environmental concentrations to NOECs/EC50s or environmental thresholds) is then used to describe the potential risk of the chemical mixture identified at a site. This methodology is rooted in the classical mixture toxicity concept of Concentration Addition (CA). For example, Backhaus & Karlsson (2014) used this approach for characterizing the overall risk of the compounds found in STP effluents, and Gustavsson et al. (2017) used a similar approach to characterize the potential ecotoxicological impact at pesticide-exposed sites in Southern Sweden on various trophic levels and on the ecosystem as a whole.

Combination of statistical and knowledge-based approaches to data integration can offer efficient means to generate additional lines of evidence that can inform subsequent research, monitoring, or decision-making as appropriate. Existing computational approaches can be used to build network models based on *a priori* knowledge about chemical exposures and biological effects. *A priori* knowledge can be first used to generate a large network of potential cause and effect relationships, i.e., a Knowledge Assembly Model (KAM). Smaller networks, termed hypotheses (HYPs), can be derived from the KAM for the specific site / location or exposure profile of concern (Schroeder et al., 2017). A number of available on-line resources have assembled and organized information about chemical-gene and chemical-protein interactions into computationally-accessible databases (Schroeder et al., 2016). For example, the Search Tool for Interactions of Chemicals (STITCH; Kuhn et al., 2012) and the Comparative Toxicogenomics Database (CTD; Davis et al., 2013) provide information about the impacts of chemicals on biological responses utilizing experimental data from controlled laboratory studies. When only chemical monitoring data are available, the KAMs (and HYPs) could be useful tool for identifying contaminants of

concern and hypothesizing the potential biological impacts i.e., perturbed genes or pathways (Schroeder et al., 2016) of chemical mixtures deriving from point or non-point sources at hot spots or sites of concern.

Critical issues of this methodology are related to (a) the chemical-analytical techniques used (in particular their sensitivity and the classes of chemicals considered *a priori*) and the temporal and spatial resolution of the chemical monitoring campaign, (b) the bias of the available toxicological information on acute toxicities or molecular responses rather than ecological responses, (c) the (non-) availability of reliable ecotoxicological information in particular for local species, and (d) the application of CA for mixtures of non-similar, potentially interacting compounds. Obviously, all conclusions on risks are confined to the compounds included in the analytical profile. As a consequence, any conclusion based on a chemical-analytical profile inherently misjudges the actual risk at a site. Site-specific physico-chemical factors such as water hardness, natural organic matter (NOM), dissolved organic carbon (DOC), pH and water temperature might introduce additional biases, which could lead to either over- or underestimations of risk. When utilizing ecotoxicological information produced in the laboratory, the difference between those optimized conditions and the situation on site can compromise an assessment. Organisms in the actual water body might experience severe deviations from a desired physiological and/or health status, due to suboptimal or fluctuating conditions. This can reduce resilience against chemical bioactivity and thus increase the deleterious impact of water pollutants. In summary, any assessment based solely on CA-based mixture modeling has to be considered a predictive screening-level assessment. It only pinpoints to potential risks, without providing any final assessment. Nevertheless, it is the best possible starting point for an assessment of potential toxic effects of chemicals.

The availability (or lack thereof) of robust ecotoxicological information is an issue that is often initially underestimated, but often turns out to be the major bottleneck for a TU analysis, in particular when so-called “emerging” pollutants are included in the chemical-analytical profile. This, for example, becomes obvious by the fact that TU analyses are usually restricted to the use of acute data on invertebrate and fish toxicity (Busch et al., 2016; Backhaus et al., 2014; Gustavsson et al., 2017), although information on chronic effects are obviously far more relevant in an ecological setting. It is a major challenge for the CA-based hazard characterization if well-researched compounds (e.g. WFD priority compounds or pesticides) are included in the monitoring profile next to compounds with a basically unknown ecotoxicity.

It has to be considered a main strength of the outlined predictive approaches that they make use of existing ecotoxicological information and chemical assessments in order to provide a first impression on whether a site might be impacted. That is, no experimental ecotoxicological work is required for the first risk conclusions. This is particularly helpful to analyze whether chemical pollution is an issue of potential relevance, warranting more in-depth follow-up studies. The approach also allows a risk-based ranking of the compounds detected in an analytical profile, points to areas in which improved chemical-analytical methods (with lower analytical detections limits or by considering a broader spectrum of compounds) are needed, and finally it allows an *if-then* analyses in order to assess and rank potential management measures.

Table 1: The four principal approaches for assessing whether the ecological status at a site is impacted by chemical pollution

	Predictive Mixture Modelling	Effect-Directed Analysis	*In situ* experiments	Field surveys/ biomonitoring
Chemical Assessment	▪Analytical profile, recorded previous to the assessment	▪Site-specific samples (often extracts in organic solvents) ▪In-depth analytical profiles of sample fractions.	▪Analytical profile, recorded in parallel to the experimentation	▪Not obligatory, often only available to a limited extent
Biological Assessment	▪No experimentation ▪Compilation of existing ecotoxicological information on identified chemicals, amended by QSAR modelling if needed ▪Compilation of existing environmental threshold of identified chemicals	▪Testing of site samples and adequate sub-samples (fractions) in an iterative process ▪Controlled exposure in the lab ▪Cellular assays, high-throughput screening assays ▪Multiple endpoints (molecular, physiological, population)	▪"Natural" uncontrolled exposure *in situ* over a limited timespan ▪Standard single species ▪Local single species ▪Local ecological communities ▪Multiple endpoints (molecular, physiological, histopathology, population, community)	▪"Natural" uncontrolled exposure *in situ* over the whole lifespan of the organism, in their natural environment ▪Calculation of ecological indices based on presence/absence data, individual counts, age structure, sex ratios of surveyed populations, trait based community structure
Impact Evaluation	▪Calculation of CA-based risk quotients	▪Comparison of total sample ecotoxicity and/or specific biological effects with concentration-response data recorded for (sub)fractions and identified chemicals	▪Comparison of endpoint patterns from the site of interest with reference site or reference conditions (values)	▪Comparison of ecological indices from the site of interest with reference conditions or site
Main advantages	▪No need for ecotoxicological experimentation	▪Strong causal link between the overall sample ecotoxicity and / or specific biological effects and the set of identified chemicals ▪Identification of new (group of) compounds of potential ecological relevance	▪Provides ecotoxicological information in a natural ecological context	▪Provides data directly relevant for ecological status assessments
Main disadvantages	▪Only screening-level assessment ▪Analytical methods set-up *a prior*, usually with only limited consideration of site-specific environmental risks	▪High resource demands ▪Limited ecological relevance, due to reliance on small-scale ecotoxicological and in vitro assays ▪Biases due to sample preparation	▪High resource demands ▪Need for reference site or reference conditions (values) ▪Causal link to chemical pollution not always clear	▪High resource demands ▪Need for reference conditions (site) ▪No causal link to chemical pollution

2.2 Effect-driven assessment methods (Effect Direct Analysis and Effect Driven Fingerprinting)

Different chemicals have different ecotoxicological profiles. That is, they affect different species, different endpoints and interact with different biomolecules at different concentrations. These patterns make an assessment of the ecological consequences of chemical exposure challenging, but at the same time they can be used in the context of Effect-Directed Analysis (EDA) and Effect-Driven Fingerprinting (EDF)[1] to gain insights into the status of an exposed system and to identify important pollutants.

Methods for the effect-directed analysis (EDA) of aquatic systems have been recently reviewed in-depth in a series of papers by Brack and coworkers (Connon et al., 2012; Brack et al., 2016;). In a nutshell, EDA methods start with the ecotoxicological characterization (usually based on microscale assays) of a complex environmental sample, most often an organic solvent extract of the pollutants present in a defined volume of sediment or water. Bioactive extracts are then repeatedly fractionated according to lipophilicity, pKa, molecular size and other physico-chemical parameters. All bioactive fractions are again characterized for their ecotoxicity and are also chemically increasingly well characterized. The iterative process of fractionation – biotesting – chemical analysis is repeated until the fractions are sufficiently well characterized (i.e. simple enough in their chemical composition), so that toxic effects can be linked to the presence of defined chemical classes or even individual chemicals. A final confirmation step then ties the toxicity profiles of the (sub-)fractions together in an attempt to describe the overall toxicity of the initial sample as a function of the identified (groups of) chemicals. That is, the overall EDA-aim is to provide causal links between the overall toxicity of the initial sample, which assumedly reflects the toxic pressure that organisms are experiencing at a site, and the final set of identified individual chemicals.

Critical issues in this context are potential loss of chemicals during sample preparation, analytical toxicant identification and confirmation, as well as the sensitivity of the analytical techniques in relationship to the toxicity of the analytes. The issue of bioavailability changes during sample preparation might warrant particular attention. Obviously, the success or failure of any EDA strategy critically hinges on the bioassays used for characterizing the toxicity of the (sub)samples and identified chemicals. Depending on sample type and volume and the specific study question, a variety of different bioassays are used (Brack et al., 2011; Connon et al., 2012; Hong et al., 2016; Brack et al, 2016 and references therein;). Given the usually limited sample amount available for testing, assays are usually miniaturized and confined to a small number of cell-based *in vitro* assays, single species assays with unicellular organisms, and/or to short-term tests with small invertebrates or fish embryos. EDA often uses a battery of assays that inform on specific biological effects as a response to specific chemical class or individual compounds detected in the (fractions of the)samples. *In vitro* reporter gene assays based on mammalian cells have been increasingly used during the last years for this purpose. They provide information on adaptive stress / oxidative / inflammation responses (ARE, NFkB), genotoxicity (p53), xenobiotic

[1] While EDA and its closely related sibling TIE (Toxicity Identification and Evaluation) are well established abbreviations in the literature, the term “Effect-Directed Fingerprinting” (EDF) is newly coined, in order to be able to summarize the corresponding techniques under one umbrella.

metabolism pathway (peroxisome proliferator activity (PPARg) and aryl hydrocarbon receptor activity (AhR-CAFLUX), estrogenicity (ERa GeneBLAzer, ESCREEN, BG1Luc4E(2) assays), androgenicity (AR GeneBLAzer, ARMDA-KB2), progestagenic activity (PR GeneBLAzer), glucocorticogenic activity (GR GeneBLAzer) and retinoic and retinoid acid activity (RAR and RXR GeneBLAzer), see e.g. Neale et al, 2015, 2017, König et al, 2017.

The use of microscale assays allows the recording of accurate and precise concentration-response relationships, but provides data of only limited direct ecological relevance. In particular, local species are usually not tested, larger organisms and multi-species approaches can rarely be used, the exposure time is limited and tests are not conducted in the presence of natural confounding factors. Also, bioavailability of the compounds in an extract largely differs from the situation on-site, notwithstanding the strong enrichment that is typical for this kind of samples. The strategy to use highly specific assays is a double-edged sword. On the one hand the response patterns facilitate to pinpoint specific groups of chemicals. On the other hand, the approach is inherently non-holistic and might bear the risk of overlooking modes of action relevant for a specific site. EDA approaches also seem to be commonly used to analyze the impact of organic pollutants solely.

The facts that *in-vitro* assays usually have a huge capacity and that they have different sensitivity profiles is increasingly used to record complex response profiles from various environmental samples. We term this approach “effect-driven fingerprinting” (EDF), and it is currently primarily used for prioritizing and ranking (groups of) chemicals, for example in the context of the ToxCast of the US EPA, a program which uses more than 700 high-throughput assays to characterize the response profiles of more than 1.800 defined chemicals (https://www.epa.gov/chemical-research/toxicity-forecasting). Similar approaches now start to appear for characterizing the (eco)toxicology of complex environmental samples, often in combination with toxic unit analyses and/or enrichment of the chemical cocktails from water, sediment or sludge (Connon et al. 2012, Hamers et al. 2013, Escher et al. 2014) .

The used assays and endpoints still have a considerable focus on the evaluation of human-health related impacts, which are used, e.g., for the assessment within the drinking-water cycle (e.g. Stalter et al., 2016; Neale et al., 2017). However, first papers are starting to appear in the scientific literature that used these fingerprinting approaches, often in combination with toxic unit analyses, also for the ecotoxicological and environmental evaluation of complex exposure situations (e.g. di Paolo et al., 2016;), see also the discussions by Brack et al. (2017) and Hunting et al. (2017).

2.3 *In situ* tests

In situ methods provide a direct measure of the integrated biological response in individuals exposed to a complex mixture of chemicals and non-chemical stressors at a site. Responses measured may be highly integrative, as in the case of apical endpoints in whole animals (e.g. survival, reproduction, growth, physiological condition etc.), or they may be specific to certain chemical classes and/or biological effects as in the case of assays / biomarkers or OMIC studies focused on a single signalling pathway.

The different *in situ* methods and approaches seek to balance the degree of control with environmental realism and ecological relevance (Baird et al, 2007). For example, assays with fish (and sometimes with mussels or other invertebrate species) may be conducted in the lab with field-collected water samples (e.g. Garcia-Reyero et al, 2011) which provides control of confounding environmental factors and low logistic cost, but fail to consider uncertainties due to fluctuating chemical exposures, the possible degradation of contaminants in collected/stored samples, or changing environmental conditions. At the other end of the spectrum, direct evaluation of responses of feral fish (e.g. Deutschmann et al, 2016), mussels or other invertebrates (e.g. Kolarević et al, 2016), aquatic plants (e.g. Dranguet et al, 2017) or biofilms (Munoz et al, 2009) considers chemical impacts in a realistic exposure scenario. However, these approaches often suffer from a limited capacity to derive causality between observed effects and stressors. In addition, large within-sample variability often leads to low statistical power and consequently the (statistical) inability to detect ecologically important changes when comparing exposed vs. control individuals. That said, finding control individuals is a challenge. Comparability of individuals cultured (or communities in a laboratory at optimized condition with the specimens caught in the field is questionable, and hence reference sites with "unexposed" controls should be used.). Finally, collection efforts and costs of monitoring studies can be high, particularly in the case of fish.

An alternative approach is to expose organisms or communities *in situ* using caging systems (e.g. Palace et al. 2005, Oikari 2006; Jasinska et al, 2015; Schroeder et al, 2017) or other controlled exposure systems such as bypass units (e.g., Triebskorn et al. 2003). Another option are transplant studies in which organisms or communities are transferred between non-polluted and polluted sites. This can offer a cost-effective middle-ground between controlled laboratory exposure and field monitoring. Although the exposure duration of caged individuals will typically be less than that of wild individuals, both experience similar fluctuating chemical exposures as well as exposure to the cumulative impacts of multiple chemical and non-chemicial stressors. Caging experiments are typically done downstream of WWTP discharge points, so fluctuation means daily fluctuations and working days vs. weekend. The duration of such experiments is limited to few weeks. If resources allow, caging can be repeated in different seasons to account for annual fluctuations and peak exposures. Often, organisms used in caging studies are from laboratory cultures so they have a known chemical exposure history and health status, which is not the case with field-collected animals. On the other hand, caged organisms cannot escape extreme pollution events or constant exposure to overall unfavourable environmental conditions at the study site, as their free-living conspecifics can do, so the caging experiments present a worst-case exposure scenario. *In situ* tests - most often implemented with microbiota, invertebrates or fish – take a middle ground between field surveys and laboratory tests with isolated species (Figure 1). They might make use of transplanted organisms, caging or flow-through systems. They can be run with either standard ecotoxicological test species or with indigenous species. *In situ* test systems that use standard assays are also called "biomonitors" or "biological early warning systems". *In situ* studies with standard test species have the advantage that life-cycle, physiology, biochemistry and genetic make-up of standard species are usually well characterized, which aids data interpretation and comparison with existing ecotoxicological information from the literature. It also facilitates genomic and epigenetic assessments, in order to

understand intra- and inter-species sensitivity distributions. *In situ* assays with local biota, on the other hand, are inherently more realistic, but often not an option if local species are not available for testing, are not stationary or do not tolerate the experimental conditions, especially caging.

All *in situ* approaches integrate the effects of the natural environment, the fluctuating chemical exposures, non-chemical stressors and physico-chemical conditions, into the observed responses of the exposed biota. Any sensible interpretation of *in situ* experiments therefore requires that the actual chemical exposure is recorded in parallel to the experiment, together with data on the general physico-chemical conditions at the site (water hardness, pH, temperature, oxygen saturation, etc.).

In situ experiments can analyze function, structure and fitness of the exposed organisms and communities. In the following we provide an overview of the main approaches for the main organisms groups relevant for the aquatic environment, i.e. micro-organisms, invertebrates, macrophytes and fish.

2.3.1 Micro-organisms

Micro-organisms drive crucial matter cycles and energy flows, such as nitrogen-, sulfur- and phosphorous cycles as well as primary production. Assessing impacts on their function, fitness and community structure is hence an important part of ecosystem status assessments. Bacteria, fungi and microalgae alike have been used in order to monitor and assess chemical pollution *in situ*. Table 12 provides an overview of approaches and endpoints used in order to elucidate chemical impacts on micro-organisms *in situ*.

Several assays are described in the literature that use populations of selected species, sometimes genetically modified to detect specific toxicants (e.g. Eltzov et al, 2011; Jouanneau et al., 2015; Hassan et al., 2016). Typical test-species include various species of microalgae (e.g. *Scenedesmus spec.*, *Chlorella spec.*, *Monoraphidium* spec, *Raphidocelis subcapitata* (formerly known as *Pseudokirchneriella subcapitata* or *Selenastrum capricornutum*) and bacterial species (e.g. *Pseudomonas putida*, *Aliivibrio fischeri*). Isolated fungal populations are currently only rarely used for *in situ* experimentation, despite their important role in the degradation of organic matter.

The most commonly used endpoints on such a population level of biological complexity are growth and reproduction. But also physiological parameters such as respiration, bioluminescence, photosynthetic activity and the capacity to biotransform especially sulfur- and nitrogen-containing molecules are extensively analyzed. On a community level, microbial biodiversity is commonly estimated in relation to chemical exposure, either determined via direct microscopic species counts, described using molecular fingerprints (patterns of photosynthetic pigments or fatty acids) as proxy and/or estimated via genetic methods, such as ARISA, TRFLP and next-generation (amplicon- and shotgun-) sequencing. Commonly used community-level functional parameters include the degradation of organic matter, gross primary production and community-level respiration. A broad range of biomarkers is investigated on all levels of biological complexity, as summarized by Amiard-Triquet et al. (2012).

On a community level, autotrophic biofilms (periphyton) are most commonly used, which do not only play a fundamental ecological role in aquatic ecosystems (Feckler et al., 2015), but which are also

convenient experimental entities that can be analyzed using various ecotoxicological endpoints on all levels of biological complexity. Other microbial communities used for *in situ* experimentation include planktonic communities (mainly phytoplankton) and complex heterotrophic communities.

Additionally, the concept of "Pollution Induced Community Tolerance" (PICT), reviewed in detail by Blanck (2002) and Tlili et al. (2016), has become increasingly popular for the *in situ* identification of ecological chemical impacts on microbial communities. PICT is based on the observation that an ecological community (biocoenosis) under chemical stress differs systematically from a community that originates from a pristine site, even if all site-specific physico-chemical parameters are similar: a community from a polluted site is tolerant to the chemicals present, while a community from a similar, but un-exposed site may not. This tolerance development takes place on an ecological level (sensitive species will be absent, as they are outcompeted by more tolerant species), a physiological level (by the elevated expression of physiological defence mechanisms such as cytochrome P450s) and a genetic level (increased prevalence of resistant genotypes) at the same time. The overall tolerance level can be quantified in a range of bioassays and is causally coupled to the chemical stress present at a site. By analysing the sensitivity profile of a community that originates from a specific site it is hence possible to determine which pollutants are present at sufficiently high concentrations to exert an ecological effect.

2.3.2 Invertebrates

Many different species are used to evaluate the effects of chemicals on the fitness of aquatic invertebrates (Table 2). Mostly *Daphnia* is used (Malaj et al., 2014; Altenburger et al., 2015) but also gastropods, rotifers, Echinodermata (sea urchins), insects and crustaceans (Martinez-Haro et al., 2015). These tests can evaluate whole individual endpoints (e.g. mortality, immobilisation, feeding inhibition) as well as sub-individual ones (Figure 1). The advantages are that the endpoints have a causal link to exposure, incorporate bioavailability and can act as early warning signals. Disadvantages are that their ecological relevance is often difficult to assess, especially for some sub-individual level responses (biomarkers) (Forbes et al., 2006). For individual level tests standardised protocols are available from OECD, ISO and ASTM, but these are not available for biomarker approaches[2] . Sub-organismal responses can on the one hand be used as early warning signals (biomarkers of effects) and on the other hand as assays to indicate the presence of certain mode of actions (biomarkers of exposure) (Van den Brink et al., 2008; Colin et al., 2016).

In situ approaches with invertebrates are well suitable to assess ecologically relevant functional parameters. The advantages of functional parameters are that they are easy to measure and integrate the response of structural elements (Table 2, Dolbeth et al., 2015; Peters et al., 2013; Johnston et al., 2015). Parameters like pH and dissolved oxygen can be measured continuously and remotely. These endpoints can be used as early warning signals and are often directly related to the ecosystem services provided by aquatic ecosystems (e.g. primary and secondary production, decomposition) but are not

[2] http://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-2-effects-on-biotic-systems_20745761,https://www.iso.org/committee/52972/x/catalogue/p/1/u/0/w/0/d/0, https://www.astm.org/COMMIT/SUBCOMMIT/E5047.htm

directly related to the assessment endpoints of e.g. the WFD, which are often structural ones. Processes like functional redundancy makes the aquatic ecosystem first responding in a structural way to e.g. insecticides, than in a functional way (Van Wijngaarden et al., 2005). On the other hand, functional endpoints might be relatively sensitive to chemicals, like herbicides, directly affecting primary production (Van den Brink et al., 2006).

2.3.3 Macrophytes

Aquatic macrophytes can be used as one of the tests in a larger battery to evaluate the whole water sample toxicity at a certain hot spot or whole effluent toxicity at the discharge point. For those purposes a standardized test using species from genus *Lemna* has been developed (ISO, 2005). Relatively recently, a sediment contact test with rooted macrophyte *Miriophyllum aquaticum* has been designed (Feiler et al, 2014) and ISO standardised to be used for assessing the whole sediment toxicity in various field surveys. However, this test proved to be of low sensitive test in a large test battery of sediment contact tests (Höss et al, 2010) with practically no diagnostic power.

The classical endpoint - growth inhibition (calculated on relative growth rate and yield basis) is a typical apical endpoint in tests with macrophytes which cannot be used for diagnostics or environmental forensics. Direct causal link to chemicals is hampered by the presence of co-founding factors (e.g. nutrients, DOC, pH, NOM etc.) in environmental samples which might lead to the false negative or positive results. The tests are not able to identify or guide to identification of chemical classes or individual chemicals of concern at the hot spot, while positive laboratory test results do not directly imply toxic effects *in situ.* The tests can be performed *in situ*, in on-site flow-through testing systems based on standard guidelines. Schlüter-Vorberg et al (2017) have recently used such a system with *Lemna sp.* to evaluate efficacy of different advanced waste-water treatment technologies for reducing micropollutant discharge into the aquatic environment, and more importantly, occurrence of stable transformation products with unknown toxicity. *Lemna* test proved to be rather insensitive and unable to detect any adverse effect of conventionally treated waste-waters as well as the effluents deriving from different advance treatment processes.

Traditionally, aquatic macrophytes were used as biomonitors in a number of field surveys, most often for metal bioaccumulation. Sánchez-Quiles et al (2017) have recently made use of over 150 survey data to map geographic distribution of metal accumulation in macrophytes and identify some areas as hotspots of trace metal contamination.

Advances in genomic research have led recently to application of transcriptomic studies in aquatic macrophytes to assess the impact of environmental pollution *in situ*. Dranguet et al (2017) have exposed macrophyte *Elodea nuttallii in situ* at a reservoir impacted by chlor-alkali plant effluent (increased concentrations of Hg and NaCl). The response at the transcriptomic level was strong, resulting in 8700 dysregulated genes and congruent with the concentrations of Hg and NaCl in the water of the impacted reservoir. Genes involved in development, energy metabolism, lipid metabolism, nutrition, and Redox homeostasis were dysregulated, which is in line with adverse outcome pathways and transcriptomic studies reported after exposure to high concentrations of Hg and NaCl under controlled conditions in the

laboratory. The authors concluded that transcriptomic response in aquatic macrophyte provided a sensitive measurement of the exposure and thus might be seen as a promising early-warning tool to assess water quality degradation at hot spots.

2.3.4 Fish

Fish acute toxicity tests, using adults or embryo-larval stage (FET), have been traditionally used in whole effluent toxicity monitoring and assessment under various national regulations (i.e. USA, Germany). The classical endpoints -– mortality and several other morphological endpoints in FET – are typical apical endpoint in tests with fish. Although they are being integrative and of the highest ecological relevance, they cannot be used for diagnostics or detection of specific adverse biological and ecological effects of contaminants of emerging concern detected in numerous discharges from municipal wastewater treatment plants (WWTPs). This situation has stimulated the development and application of biomarkers, that are molecular, cellular, or physiological indicators of exposure and/or effects of toxicants (Huggett et al. 1992, van der Ost et al. 2003).

In situ assessment of fish, weather sentinel or caged (with various exposure duration, from several days to several weeks), including the analysis of a number of biomarkers can, in line well known adverse outcome pathways, detect the specific effects in fish and in line with well-known adverse outcome pathways. Effects which can then be causally linked to the exposure profile at a specific site.

Biological effects in caged or sentinel wild fish as a result for exposure to mainly point sources were traditionally assessed using biomarkers. Biomarkers, that are molecular, cellular, or physiological indicators of exposure and/or effects of toxicants (Huggett et al. 1992, van der Ost et al. 2003). Biomarkers of exposure indicate that the biological system is exposed to a stressor and they may inform on the identity and intensity of the stressors. An example would be the induction of vitellogenin in male fish, indicating that the organism is exposed to estrogen-active compounds, and provided that benchmarking data on vitellogenin induction by estrogens in the fish species under investigation are available, the vitellogenin induction observed in the field can be translated into “estrogen equivalents” as a measure of intensity (e.g., Tyler et al. 2005, Burki et al. 2006). The exposure to an environmental stressor may be associated with functional or structural impairment and damage of the fish, and responses indicating such chemical-induced damage are biomarkers of (adverse) effect. An example would be the impairment of the fish’ immunocompetence (Arkoosh et al. 2001, Rehberger et al. 2017). Importantly, biomarkers indicate exposure of organisms and changes in their fitness, but this does not necessarily translate in a linear way into adverse outcomes of population and community structures and functions (Forbes et al. 2006, Segner 2011). As formulated by Hutchinson et al. (2006), biomarkers are “signposts but not traffic lights” for ecological effects.

A number of general (chemical / chemical class non-specific) biomarkers (for review see van der Oost et al, 2003), such as enzymes of biotransformation (phase I - typically CYP1A and phase II enzymes and cofactors - mainly GSH/GSSG and GSTs), oxidative stress parameters (enzymes - SODs, CAT, GPOX and nonenzymatic GSH/GSSG, LPOX), stress proteins (often HSPs), multixenobiotic resistance (MXR), haematological parameters (ALT, AST, hematocrit), immunological parameters (differential blood cell

counts, macrophage function), physiological and morphological parameters (histopathology, gross indices). In general, all the listed biomarkers provide an integrated measure of exposure over time and may reflect the combined results of (simultaneous) exposure to a number of chemicals or complex mixture, indicating chemically - induced stress in wild or fish caged *in situ*. It is, however, not possible to determine which chemical has caused the observed effect as none of the changes can be attributed to a specific compound or class of chemicals.

In addition to the above listed general biomarkers, a number of more specific biomarkers are being traditionally used in biomonitoring or *in situ* experiments with fish. Specificity herewith reflects either response to particular chemical class or specific biological effect. Some of those biomarkers include (for a review again see van der Oost et al, 2003) fish bioaccumulation (mainly POPs and metals), biotransformation products (mainly PAH metabolites in bile), metallothioneins (MTs), neurotoxic parameters (typically ACHE) and genotoxic parameters (DNA adducts and secondary modifications).

The explanatory power of biomarker studies was occasionally high (i.e. Cazenave et al, 2014), but sometimes the power of a biomarker battery turned also out to be rather low (i.e.Traven et al, 2013). In most cases biomarker studies were of confirmatory nature, providing evidence of chemical stress *in situ*, but their diagnostic power, due to low specificity of the majority of studied biomarkers, was traditionally low.

During the last two decades, EDC substances (mainly pharmaceuticals, personal care products and estrogens from WWTP effluents) came into limelight as a class of ecologically most relevant xenobiotics, with a potential to directly affect fish reproduction and consequently population dynamics, which might cause changes on fish community level. However, population declines, as a direct result of endocrine disruption, have been reported mostly in mollusks (tributyl tin identified as the main cause), but due to a number of difficulties connected with population - level studies of wild fish, there is less direct evidence of adverse effects on fish populations and assemblages caused by EDCs (Johnson and Chen, 2017, Baldigo et al, 2015). However, thanks to several well established AOPs focusing on reproductive dysfunction (https://aopwiki.org) a number of reproductive / EDC related biomarkers have been identified and used with high confidence. So, we can say that the standard fish biomarker battery changed gradually over the years, to focus mainly on a priori selected reproductive and endocrine parameters. Also, traditional methods, such as, e.g., measurement of activity of pre-selected enzymes, are being gradually accompanied or even replaced by higher throughput methods, such as a priori selected gene expression studies. A number of studies of wild and caged fish endocrine responses to WWTP effluents using some very specific biomarkers, such as intersex (occurrence of ova-testes), induction of vitellogenin (Vtg) in males, altered gene expression and physiology (altered steroid production), have provided evidence of *in situ* effects of EDCs deriving from WWTPs effluents (Jobling et al, 2002, Bahamonde et al, 2014).

To expand the scope beyond targeted investigation of endpoints selected a priori, several approaches were proposed recently (Li et al, 2017, Schroeder et al 2017) to identify other potentially disturbed biological pathways and related chemical constituents. Schroeder et al. (2017) used known chemical-

gene interactions to develop site-specific knowledge assembly models (KAMs) and formulate hypotheses (HYP) concerning possible biological effects associated with chemicals detected in water samples from each location (HYP). Traditional biomarkers (histopathology, morphological endpoints, vitellogenin) as well as hepatic gene expression data collected for fish exposed *in situ* together with integrated analysis of the transcriptome data in the context of the site-specific KAMs allowed for evaluation of the likelihood of specific chemicals contributing to observed biological responses. The study has demonstrated how, when both chemistry and biological response data are available for a site, it is possible to use a KAM-based approach to evaluate involvement of specific chemicals in eliciting the observed biological responses, when both chemistry and biological response data are available for a site. Such an approach may provide a line of evidence for evaluating potential cause-effect relationships between components of a complex mixture of contaminants and biological effects data, which can inform subsequent monitoring, investigation and decision-making.

2.4 Field surveys and community indices

Overall ecosystem status assessments are often based on a comparative assessment of the biodiversity and community composition at a given site of interest and a reference site, or reference conditions. Causal links to chemical pollution may be identified by correlation analysis of the community composition at a series of sites and pollutant concentrations (and/or the intensity of other stressors). This correlation analysis can be done in two ways: using multivariate analysis or by summarizing the community composition into a single value, i.e. using an index. Multivariate analysis is able to assess the correlations between multiple species and multiple stressors in a single analysis, and is able to assess the significance of stressors for explaining the difference in species composition between different sites (Rico et al., 2016). It can also partition the variance into parts that are explained by groups of or single stressors, and assess their statistical significance, indicating the relative importance of individual and groups of stressors (e.g., Munoz et al. 2009).

Species diversity can be summarized into a single value by using indices, that can also serve as descriptors for the overall ecological status. Such indices use information on diversity and abundance of various taxa (family to species level) or the trait based metrics to provide a dimensionless metric suitable for comparing and ranking the community composition at the investigated sites.

Perhaps the most well-known example of such ecological indices is the Index on Biotic Integrity (IBI) used for describing the attributes and structure of fish communities (Karr 1981, 1991), see also below. A recent overview of the community indices used in the EU member states for assessing the status of fish, invertebrate, macrophyte and algal communities is provided by Birk and coworkers (2012). Annex I of the corresponding Commission Decision (EC, 2013) lists the results of the European intercalibration exercise and provides the values of the various indices that define the boundary between high and good ecological status as well as the boundary between good and moderate status.

Monitoring activities assessing the chemical and ecological status of surface waters are routinely carried out within the Water Framework Directive (WFD) (Lyche-Solheim et al., 2013). How to link the chemical

and ecological status of water bodies is, however, often discussed and no commonly agreed standardized methods are available for this. On the one hand ecological indicators lack the specificity to be able to point to certain chemicals and/or mode of actions, while a direct linkage using multivariate analysis is troubled by the lack of good reference sites (Rico et al., 2016). Traits-based methods are available but need further development, while also the availability of traits data needs to be enhanced (Culp et al., 2011; Kuzmanovic et al., 2017). Advantages are, however, that these methods assess the endpoints of concern at the sites of concern, and therefore have a strong linkage with the assessment endpoints. These methods can also integrate the effects of multiple stressors in one assessment and identify vulnerable species. But it is difficult to disentangle the contribution of the individual stressors, although new methods become available (Baird et al., 2016).

2.4.1 Micro-organisms

Field surveys of micro-organisms are part of the biomonitoring strategy under the Water Framework Directive in order to assess the ecological status of an aquatic ecosystem. In lake ecosystems, Annex V stipulates that benthic communities of microalgae are surveyed in order to assess the "*Composition and abundance of aquatic flora*" for lakes, coasts and river ecosystems alike. "*Composition, abundance and biomass of phytoplankton*" is, in addition, to be monitored for lakes and coastal waters. The focus is on taxonomic composition and relative abundance of benthic diatoms as well as absolute abundance of phytoplankton (magnitude and frequency of algal blooms). Although not directly specified in the legal documents, most member states assess macrophytes and phytobenthos as two discrete biological quality elements (Kelly, 2013).

It should be emphasized at this point, that the function, abundance and biodiversity of bacterial and fungal communities are not included in the ecological status assessment under the WFD – despite their crucial importance for ecosystem function, in particular the cycling of carbon, nitrogen, phosphorous, sulfur and other elements, and their relevance for human health (e.g. as environmental reservoirs of antimicrobial resistance genes).

Diatoms, a distinct algal class with transparent cell walls made of silicon dioxide (so-called frustules), are most commonly used for the assessment of phytobenthos and phytoplankton communties. Other algal classes are only rarely considered (Kelly, 2013). As for all biomonitoring efforts, the assessment of phytoplankton and phytobenthos is implemented in terms of deviation from an ideal state, i.e. the "reference condition", which is specified separately for each water type and represents the values of the biological quality elements at an undisturbed site. High ecological status is thus operationalized as "*the phytoplankton community will be indistinguishable from the type specific reference conditions*" (EU Guidance Document No. 10, 2003). The appropriate definition of these reference conditions is thus a critical task, as it defines the overall frame of reference and thus decides on whether specific management measures might be needed for a monitored site. In the context of phytobenthos monitoring, the use of palaeolimnological data is often suggested (see discussion in Kelly (2013). However, this approach seems to ignore that ecosystems are not static, but are dynamic entities that develop over time, even without human influence.

Biodiversity assessments in general and for micro-organisms in particular have a long standing history especially in conservation science. In principle, two different index types can be distinguished. First, there are the classic, taxonomy-based indices that aim to provide an overall numerical value that captures biodiversity, usually related to the number of species present, their relative abundance and the evenness of the resulting distribution (e.g. Magurran, 2003; Dorazio et al, 2011; Riddle et al., 2011). From the perspective of these indices, every species is of equal importance. However, in a second class of indices, different weights are given to different species, in relation to their perceived ecological importance or their sensitivity to environmental stressors. A huge plethora of different phytoplankton and phytobenthos assessment methods have been developed in this area in the European member states (Kelly, 2013). An EU-wide intercalibration exercise was therefore implemented, in order to reach a common delineation especially between the "good" and "moderate" status classes (Birk et al, 2012). This activity resulted in officially approved boundary definitions (European Commission, 2008, 2013).

Most common is the use of weighted averages of relative or absolute abundances, condensed into various diatom indices. In particular the IPS (Index on Pollution Sensitivity) has found widespread use. It should be emphasized here, that "pollution" in this context is basically understood as a synonym to "eutrophication" and/or "acidification". The IPS is calculated as

$$IPS = \frac{\sum_{j=1}^{n} A_j v_j i_j}{\sum_{j=1}^{n} A_j v_j}$$

Where *A* denotes the relative abundance of species *j*, *v* is its indicative value in the range of 1 to 5 (basically a class-based trait-characterization, such as planktonic lifestyle, motility, etc), and *i* indicates its pollution sensitivity (=nutrient requirements) in classes of 1 to 3 (Besse, 2007). More recently, the closely related BDI (Biological Diatom Index) was suggested and subsequently standardized in France (Coste et al, 2009), which is based on the abundances of 209 key diatom species, selected for being as uncorrelated as possible. The BDI is supposedly an improvement over the IPS, better reflecting the contribution of key species and more useful for brackish algal communities (Besse, 2007). These, and similar indices are routinely used in broad-scale European phytoplankton and –benthos biomonitoring efforts, see also the recent reviews and discussions by Poikane (2016), Morin (2016) and Wu et al (2017).

Given these conceptual bases, one has to conclude that surveys of algal communities are currently almost exclusively focusing on diatom species, with the main aim to assess the eutrophication status of an aquatic ecosystem. Index values are therefore often strongly correlated with basic water chemistry parameters, such as phosphorous and nitrogen concentrations. Whether and to what extent these indices are informative of pollution with toxic compounds is largely unknown. However, the almost exclusive focus on diatoms seems to imply an inherent bias.

2.4.2 Invertebrates

The sensitivity of macroinvertebrates to changes in environmental quality make them an essential part of any biomonitoring program. Depending on their lifespan, macroinvertebrates live within aquatic systems for several months to multiple years. Macroinvertebrate communities are hence thought to

reflect chronic effects of pollutants, being at the same time relatively immobile and so continuously exposed to potential pollutants, but also to other habitat-specific factors. A variety of methods exist to collect benthic macroinvertebrates from aquatic habitats to assess the ecological status of aquatic systems. Such collection of community impairment is often done in combination with the assessment of other environmental variables (hydro-morphology, habitat parameters, physico-chemical parameters).

Diversity and abundances of macroinvertebrates resulting from field monitoring can be analyzed by multi-variate statistics. This type of analyses statistically correlates changes in the community composition with changes in explanatory variables. Groups of these variables, e.g. the abiotic factors mentioned above, or also concentrations of metals and organic pollutions can be tested for their correlation with community alterations (e.g. Rico et al, 2016, Kuzmanovic et al., 2017). Although this type of analysis is a powerful tool to identify potential drivers of community changes on a sound statistical basis, it comes with the implicit limitation that 'absolute' quality indicators are not calculated. Hence, as alternative to the multivariate analysis of such datasets, indicators or metrics for the biological quality of the community at a specific sampling site are used to summarize the ecological status.

Based on a long history, there are literally hundreds of indicators and metrics that can be calculated based on macroinvertebrate community data (Birk et al., 2012). Especially, there is a large number of national indices, which take into account local specifics about taxonomic composition of communities. The most commonly used metrics of biological assessment for rivers based on macroinvertebrates, are taxonomic richness and composition (number of species, diversity indices, percentage of some taxa, etc.); or biological information on ecological functions (e.g., habits and species traits of the aquatic fauna) (Barbour et al. 1999). In a systematic evaluation of the different types of indicators, the result indicated that trait-based indicators showed a higher correlation with environmental parameters than taxonomy indicators or other metrics (Cortes et al., 2013).

For the community level line of evidence of the ecological toolbox, a combination of general and functional and trait based indicators is suggested. These were, three standard community indices: the total abundance of individuals, the total taxa richness, and the Shannon-Wiener diversity index (Shannon, 1949); three taxonomy-based indices regularly used in the ecological status evaluation under the WFD: the Saprobic Index (Zelinka and Marvan, 1961), the Biological Monitoring Working Party (BMWP) index and the Average Score Per Taxon (ASPT) index (Armitage et al., 1983). Finally, the % of EPT and of chironomid species are suggested as functional indicators.

The BMWP index is based on the concept that different aquatic invertebrates have different tolerances to organic pollutants. Tolerance scores between 1 and 10 are associated to all families, where a higher value stand for less tolerance, hence for species which are more impacted by organic pollution. Pollution is here meant more in the sense of nutrient pollution. BMWP values consists of the sum of the tolerance scores of all macroinvertebrate families in the sample. A higher BMWP score is considered to reflect a better water quality. The Average Score Per Taxon (ASPT) score is derived from the BMWP. The ASPT equals the average of the tolerance scores of all macroinvertebrate families found, and ranges from 0 to 10. The main difference between both indices is that ASPT does not depend on the family richness

2.4.3 Macrophytes

Considerable efforts have been invested to develop appropriate biological methods for macrophytes in response to WFD monitoring demands, but European standard EN 14184:2014 - Guidance for the surveying of aquatic macrophytes in running waters - has been published only in 2014. It is developed to be applicable to all kinds of surface running water bodies, like natural brooks, streams and rivers and their heavily modified equivalents, as well as to artificial water bodies like canals or run-of-river reservoirs. The general principles of the approach may also be applied for monitoring water bodies in the fluvial corridor of a river, such as side channels and oxbows.

In the meanwhile, almost all EU countries have already developed national and regionally specific methods and a number of biotic indices for macrophyte assessment have been in use since year 2000 (Birk et al, 2006, 2010) so the wider application of the European standard is rather unlikely. Furthermore, most of the indices are designed to cover catchment scale, not particular local site or river section study. The information provided by all these methods include the composition and abundance of the aquatic macrophyte flora, using mainly taxonomy - based approach, which inevitably imply some regional and local considerations. Typically, indices are heavily burdened with individual indication value of particular species, while trait based information is still rather underused (Birk et al, 2006, 2010).

Ongoing criticism of WFD oriented assessment methods addressed, among other issues (for review see Wiegleb et al, 2016) the lack of relationships between anthropogenic pressures and macrophyte response. Most of the national indices are expected to respond to and detect hydromorphological alternations, eutrophication and organic pollution. Not a single existing method is expected to detect chemical pressure. Pressure - impact relationships have been properly statistically tested for only a few among a number of national indices in use (Birk et al, 2010) and typically correlate to some extent to nutrients / eutrophication level, but mainly across the pressure gradient, meaning that they are not designed for comparative studies such as typical upstream - downstream or hot spot studies.

Wiegleb et al (2016) evaluated several methods based on macrophyte community structure in respect to dividing human impact from the impact of stochastic variation caused by natural disturbance and concluded that none of the methods performed well, mainly because the reference to ‘species composition and abundance’ may not be an appropriate approach. Low number of species increases the uncertainty in taxonomy-based methods. In particular, the number of truly rheophytic or at least rheophilous species is very low. Most macrophyte species have their ecological optimum outside water courses, which may explain the partly unspecific indicator values in rivers in contrast to lakes. The gradient spanning from ‘near-natural’ to ‘impaired’ is non-linear. Pristine conditions often feature no or only few macrophytes. High macrophyte status is reached only as long as (slight) anthropogenic impact is taking place. They concluded that the macrophytes are not good bioindicators in rivers, as they do not react to the officially recognized stress factors in a predictable manner. The statement generally applies to all taxonomy - based assessment methods currently in use.

Bioindication value of macrophyte community assessment might slightly increase with a shift from taxonomy - based to trait- based approaches. Wiegleb et al (2016) showed that the NRW method (North-

Rhine-Westphalian Assessment System for Macrophytes in Water Courses) performed better than the other nationally accepted, intercalibrated, taxonomy - based metrics dominated methods. Baattrup-Pedersen et al (2016) found clear evidence that habitat degradation in the studied lowland streams mediated selective changes in the functional trait composition of the aquatic plant community. Trait composition clearly responded to hydromorphological alterations and eutrophication in both case studies. However, to this end, there is no evidence that either taxonomy or trait - based composition of aquatic plants responds in a predicted manner to chemical pressures and therefore we must conclude not only that there is no ready - made tool for detecting ecological impact of chemicals using aquatic plant communities, but that assessment of macrophyte communities (taxonomy or trait - based community assessment, regardless) is not likely to yield many valuable information about potential (*in situ*) ecological impact of (toxic) chemicals on aquatic ecosystems.

2.4.4 Fish Community Structure

The composition and structure of fish community, as it is the case with any other aquatic community, reflects the overall ecological conditions and informs about ecological integrity. Therefore it has been included as one of the four mandatory biological quality elements (BQE) necessary to access ecological status of European rivers under the WFD.

A number of authors have provided evidence that alterations of fish community structure and composition mainly happened as a consequence of physical pressures – hydromorphological alterations of the rivers, such as longitudinal interruptions (dams), loss of lateral connectivity, modification of canal morphology, various river engineering measures, loss of wetlands / flood plains/ riparian vegetation, pressure from non-native and invasive species, as well as pouching, overexploitation but also restocking of commercially valuable species (for review see Kautza et al, 2015).

Different fish-based (multi-metric) indices have been developed worldwide for assessing the ecological status of rivers. Most indices incorporate a reference condition approach and relevant biological variables or metrics (for the review see Noble et al. 2007), to describe the fish assemblage characteristics and to quantify the impact of human activities on the biota. The Index of Biotic Integrity (IBI) is the generic name retained to describe this general framework after the work of Karr (1981) - the first multi-metric index based on fish community structure.

The index of biotic integrity (IBI) was conceived to provide a broadly based and ecologically sound tool to evaluate biological conditions in streams (Karr 1981). IBI incorporates many attributes of fish communities to evaluate human effects on a stream and its watershed. Those attributes cover the range of ecological levels from the individual through population, community, and ecosystem. IBI uses three groups of metrics: species richness and composition, trophic composition, and fish abundance and individual condition. The value for each metric is based on comparison to a regional reference site with little or no influence from human society. So, all sites of interest must be either evaluated against the similar undisturbed site, or, regionally, least disturbed site. Assessment of biotic integrity explicitly incorporates biogeographic variation into evaluation of biological systems. IBI scores can be used to (1) evaluate current conditions at a site, (2) determine trends over time at a site with repeated sampling, (3)

compare sites from which data are collected more or less simultaneously, and (4) to some extent, identify the cause of local degradation (Karr et al 1986).

IBI was slightly modified by a number of authors to reflect the regional / local conditions and to be applicable outside the narrow biogeographic area to which it was originally designed for. Inspired by IBI, a number of new biotic indices were created, mainly relying on taxonomy, but also on ecological traits (and their states), such as ecological groups, fish trophic guilds, migration behaviour and flow velocity preferences (Frimpong and Angermeier, 2010, Birk et al, 2010, Herman and Nejadhashemi, 2015).

In the WFD context, the most crucial assessment parameters are composition, abundance and age structure of fish fauna, which have to be implemented in the evaluation index of each EU member state. There is a wide range of fish indices (based on different metrics) from various EU member states, e.g. France developed »French Fish based Index« (FBI) (Oberdorff et al. 2002), Germany »German fish/based assessment system« (FiBS) (Dussling et al. 2004), Austria »Fish Index Austria« (FIA) (Haunschmid et al. 2006), while the FAME consortium developed the »European fish index« (EFI) (Pont et al. 2007; https://fame.boku.ac.at/main_results.htm). In 2009, as a follow up activity, the New European Fish Index – EFI+ – was developed (http://efi-plus.boku.ac.at/software). Two European indices were developed to overcome regional or local differences and be potentially applicable throughout Europe. Regardless of differences in metrics, all indices have been developed in such a way to evaluate mainly the pressure from hydromorphological alterations and non-native / invasive species (Birk et al 2010, 2012), except EFI, which is mainly sensitive to water quality pressures and therefore seems to be more fit to purpose than other available indices for the assessment of ecological impact on fish community at pollution hot spots.

Fish community assessment, using taxonomy or trait-based approach, regardless, have rarely been successfully used to discriminate between chemical and non-chemical stress in multi - stressed aquatic ecosystems on a larger scale (e.g river basin, river, or significant stretch of a river) (Hall et al, 2009; Nõges et al, 2016). Also, to the best of our knowledge, there is no published success case study which proved the applicability of fish community assessment, as a stand-alone tool for identification and characterisation of toxic pressure to aquatic ecosystems. However, there are examples of regional or local hot spot studies in which fish community assessment, coupled with chemical analyses, (whole effluent) toxicity tests, bioassay and / or *in situ* fish biomarker studies was used for identification of overall pollution / toxic pressure as well as for the monitoring of overall water quality (e.g. An et al, 2002; Ra et al, 2007, Flinders et al, 2009). One of the most recent and the most pragmatic hot spot studies by Azimi and Rocher (2016) used 20 + years long monitoring results to evaluate if the overall improvement of the water quality downstream Paris had any positive impact on structure and composition of fish community. The results are promising, in terms of applicability of multi-metric fish index but more striking - sensitivity of the pragmatically selected taxonomy and trait-based metrics to changes of water quality which came as a result of the pollution abatement measures . The pragmatic approach was to group all species according to habitat preferences (limnophilic vs. rheophic), as well as feeding patterns (carnivorous vs. omnivorous). Before the implementation of abatement measures, the prevailing species were limnophilic and omnivorous, hence undemanding in terms of water quality and diets. Enhanced abatement measures led to gradual increase of limnophilic and carnivorous, as well as rheophilic and

omnivorous species. The study is an example how complex (full application of multi/metric fish index) but more importantly, simplified community structure analysis (based on key guilds / trait analyses) might be a pragmatic yet suitable tool to show *in situ* ecologial impact of pollution hot spot.

In summary, the ecological indices that are calculated on the basis of field surveys are not able to identify the underlying causes of an ecological impact, as observed changes might be driven by any number of combination of chemical and non-chemical stressors. However, the fact that biological receptors differ in their inherent sensitivity to stressors can provide information on the dominant stressors in an ecosystem (Segner et al. 2014). For instance, Species Sensitivity Distributions (SSD) map the statistical variation of the sensitivity of species to stressors, and this opens perspectives for its use in multiple stressor assessment. The multi-substance potentially affected fraction (ms-PAF) is the consequent adaptation of mixture toxicity calculations based on SSDs. The disadvantage of the SSD approach is its descriptive character, i.e. it does not explain why a certain species is sensitive or tolerant to a given stressor. Here, moving from a species to a trait perspective can provide deeper insight. Traits are intrinsic physiological (e.g. detoxification capabilities), and ecological (e.g., feeding types, reproductive strategy) properties of species or communities, which drive their sensitivity to stressors (van den Brink et al. 2010). Trait-based approaches have successfully been implemented as indicator systems for toxic impacts, e.g., the Species At Risk Index (SPEAR) (von der Ohe et al. 2004). Metrics based on receptor traits are by no way restricted to chemical stressors, but can be developed for other stressors as well, and are therefore particularly valuable for multiple stressor assessment (Verbrugge et al. 2012). Importantly, traits are not invariant, but they show phenotypic plasticity, e.g., through changes in morphology, behaviour or physiological acclimation, and they can vary over life history (Verschoor et al. 2004).

Measurements of ecological indices require substantial amounts of field work to record, and often become meaningful only if longer time-series are recorded or if a whole series of sites along stressor gradients can be comparatively assessed and compared to sites under reference conditions. The assessment is further complicated by the fact that increasing levels of stress might cause either a decrease or an increase in biodiversity, depending on the initial status of the exposed communities, the type and magnitude of stressors present and their interactions. The fact that several organisms, especially fish, are very mobile and might be migratory, further hampers site-specific assessments. Finally, by their very nature, observed changes in community structure cannot be used for prospective purposes and the predictive evaluation and comparison of management options.

However, field surveys are the only method to provides a real-world snapshot of the ecological status of communities in their real environment. , and can hence be regarded as the “gold standard” of the ecological status evaluation under the Water Framework Directive.

3 Application of weight-of-evidence approaches (WOE) for assessing chemical impacts on ecosystem status

3.1 Overview

Establishing causal links between chemical pollution and ecological impacts in the field is challenging, given the high variability of natural systems and the plethora of confounding variables potentially present (Burton et al., 2002, Hull and Swanson, 2006). The different approaches outlined above and summarized in Table 1 complement each other, with each one having its particular strengths and weaknesses. Each approach might produce one or several so-called “lines-of-evidence” (LOEs). In order to reach sound, consensual and actionable conclusions, all available LOEs need to be tied together, which is usually done in a process commonly known as a weight-of-evidence (WOE) evaluation. By combining various LOEs, the causal link between chemical contamination and observed or predicted ecological impacts is strengthened (Wolfram et al., 2012).

WOE in environmental risk assessment (ERA) has been defined by Linkov and colleagues as a “framework for synthesizing individual lines-of-evidence (LOE), using methods that are either qualitative (examining distinguishing attributes) or quantitative (measuring aspects in terms of magnitude) to develop conclusions regarding questions concerned with the degree of impairment or risk” (Linkov et al., 2009). That is, the WOE approach considers the strength and weaknesses of various types of data for selecting between several, often competing, management alternatives (Hull and Swanson, 2006; Hope and Clarkson, 2013). In the context of assessing the impact of chemicals in the environment, the WOE approach contrasts the strength of the evidence that adverse effects are caused by chemical exposure against the null hypothesis of no effects being present (Smith et al., 2002). The Sediment Quality Triad (SQT), originally conceived by Peter Chapman and colleagues (Long and Chapman, 1985; Chapman, 1990), was perhaps the first implementation of the WOE approach for assessing ecological impacts of chemical contamination. It provides a WOE framework for assessing site-specific ecological impacts by combining information on sediment chemistry, sediment toxicity, and benthic community structure (Chapman and Hollert, 2006). Since then, the WOE concept has been frequently used in environmental assessment. However, its application is often criticized as being inconsistent (Weed, 2005; Linkov et al., 2009; Krimsky, 2005; Suter and Cormier, 2011; Ågerstrand and Beronius, 2016) and lacking transparency (Hull and Swanson, 2006). A particularly critical and sometimes controversial issue in this context is the relative weighing of LOEs, which is often dependent on expert judgement (Suter and Cormier, 2011; Good, 1991; Linkov et al., 2009). Of further importance is the object or aim of protection. Acritical issue is the question about the object and aim of protection for a WOE application, since WOE approaches can come to different results when the aim is to protect, e.g. against community collapse, population impairment, death of individuals or against sublethal chronic effects.

Another aspect is that WOE can come to different results when the aim is to protect against, e.g., community collapse, population impairment, death of individuals, or sublethal chronic effects.

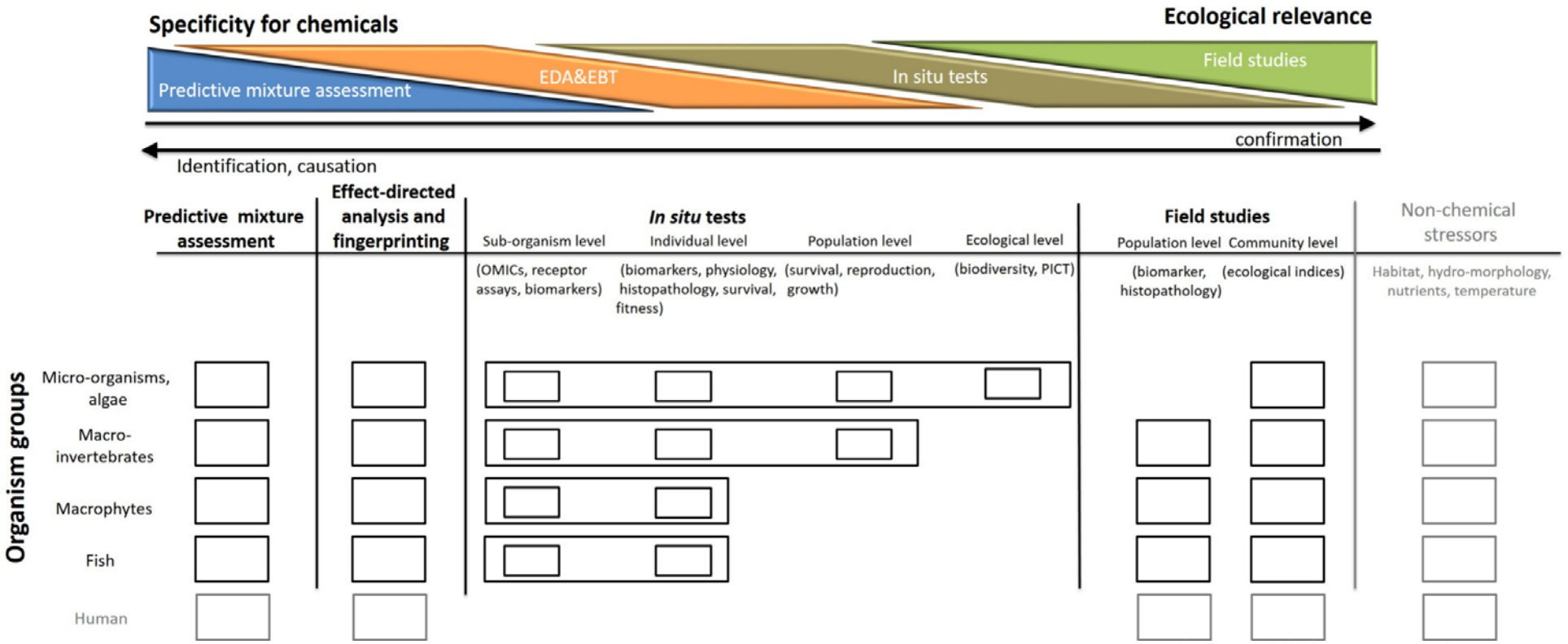


Figure 1: Combining WOE for identifying chemical impacts (top) Spectrum between chemistry (toxicity potential) and ecology (effects at community levels). *In situ* tests can have a linking function between those extremes. (Bottom) Four main lines of evidence, including the elements of evidence that can be included. The *in situ* LOE includes tests from sub organism to the community level. Mechanistic information is conserved by differentiation between the lines of evidence for the different organisms groups. Interpretation of results for the main lines of evidence are given in section 3.3 and Table 13.

WOE approaches are explicitly mentioned in regulatory frameworks such as the guidance documents of the WFD (EU Commission, 2011), which have been criticized because of their lack of explicit guidance (Hope and Clarkson, 2013; Ågerstrand and Beronius, 2016). However, the European Chemicals Agency (2015) argues that no firm rules can be established for the application of WOE approaches, as such assessments depend on expert judgement.

It is therefore not surprising that the explanatory power of diagnostic toolboxes that incorporate WOE approaches is discussed controversially (Suter 1993, Vos et al. 2000, Cormier and Suter 2013). A number of factors influence the diagnostic power of any toolbox. An important question is which parameters to include. Often the approach is to measure an extensive list of parameters („laundry list“) but this approach may in fact divert more from the detection of ecological impacts than improving its diagnostic power. Instead, a conceptual framework should be based on clearly defined questions which then allow the targeted selection of parameters for the toolbox (Lindemeyer and Likens 2010, van den Brink et al. 2013). When aiming for identifying the causative factor of ecosystem disturbance, it is important to include parameters which enable to go beyond correlative analyses, and which provide “mechanistic” information (Segner 2011, van den Brink et al. 2013). Furthermore, the desirable characteristics of the metrics used for environmental assessment need to be considered – which again depends on the purpose the toolbox is used for. Possible characteristics include the sensitivity of the indicator parameters to environmental stressors, their specificity for stressors, their cost-effectiveness, the availability of historical data, etc.

Critical issues for the selection of parameters for diagnostic toolboxes are “noise” and “benchmarking”. Noise refers to the variation of the parameters or indices if a given site is sampled repeatedly. The noise can arise from methodological limitations, but it can also reflect the biological and ecological stability at the study site. This noise affects the ability of a toolbox to detect differences between natural and human-made variation. Benchmarking refers to the problem of finding reference values of the measurement parameters, indicating an intact, biologically and ecologically “good” status (Sanchez et al. 2010).

Conclusions from multiple LOE have typically been derived implicitly from qualitative methods (Weed, 2005; Linkov et al., 2009), and also the sediment quality triad is based on correlation rather than causation (Chapman and Hollert, 2006). Several studies developed quantitative approaches for WOE assessment such as statistical methods (e.g. ordination, principal components analysis), Bayesian techniques, multi-criteria decision analysis (MCDA) (Good, 1991; Smith et al., 2002; Exponent, 2009; Hope and Clarkson, 2013; Schleier et al., 2015) or Fuzzy Logic and Hasse Diagram techniques (Hollert et al. 2002). The use of response curves assessing the divergence of impacted sites from reference conditions depending on exposure concentrations was suggested by Lowell et al. (2000). Whether qualitative or quantitative, a WOE approach can provide a framework for rigorous consideration of the strengths and weaknesses of various LOEs (Hope and Clarkson, 2013). In the best case, all LOEs promote

the null or the alternative hypothesis, but all other cases are in between and need expert judgement on causation.

Many aquatic ecosystems are exposed to a mixture of different anthropogenic impacts (e.g. Ormerod et al., 2010), so that the establishment of causality between environmental stressors and effects on aquatic ecosystems is difficult (Adams, 2003; Segner et al., 2014). Nevertheless, considering the need for environmental assessment tools, causation is recommended to indicate information gaps and to determine appropriate management actions (Chapman and Hollert, 2006). Causation by the integration of various LOE can both examine whether a site is ecologically impacted and identify the stressors which contribute to the impairment.

The dynamics of cause and effect are interwoven with fluctuating system characteristics, such as habitat quality, food quality and other natural physico-chemical parameters (Burton et al., 2002). Hence, differences between the LOE of chemistry, toxicity and field data can also be interpreted as indicative of additional factors that may control or mask effects of pollutants on the biota (Chapman, 2002; Chapman and Anderson, 2005; Wolfram et al., 2012). Therefore, the inclusion of co-factors, which might alter ecological effects and strengthen the alternative hypothesis, are recommended as additional LOE for a comprehensive WOE assessment based on causation.

In the following, we will describe the outline of a WOE approach that integrates four individual LOEs, namely (i) predictive mixture modelling, (ii) effect-directed assessments, (iii) *in situ* tests, and (iv) field-based monitoring studies, see Figure 1 and Table 1. The main idea of the approach is to systematically integrate these four LOEs, so that their strengths complement each other and allow a transparent site-specific assessment, including the identification of data gaps and management options. The suggested approach resembles the elements of the sediment quality triad approach, which is based on three LOEs (chemical analyses, bioassays, and community structure), see e.g. (Chapman & Hollert 2006).

In the following, the outlined approach will be specified in more detail. Emphasis will be on the formalized integration of the data generated within the *in situ* LOE (LOE 3, Figure 1), as this might encompass the most heterogeneous set of tools, covering a broad span from highly chemical-specific biomarkers of exposure (e.g. vitellogenin induction) to parameters that are directly ecologically relevant (e.g. observed changes in biodiversity of transplanted microbial communities). The condensation of the information provided by these tools will therefore be discussed in more detail in the following . Finally, the overall integration of the four main lines of evidence into a site-specific assessment will be discussed.

The analysis focusses on the ecological impacts of chemical pollution. Although such an analysis also provides input for the analysis of potential effects on human health (e.g. the results of chemical-specific biomarker studies might add to the assessment of indirect human health impacts via fish and drinking water consumption), the issue of human health impact assessment is beyond the scope of the present discussion.

An additional fifth LOE concerns potential impacts of non-chemical stressors, which will not be discussed in detail in the present text, although the final evaluation takes care to differentiate between chemical

and non-chemical impacts at a site. It is important to underline that the impairment of the ecological status, meaning changes in composition and structure of aquatic communities, often results from physical (hydromorphological alterations of the rivers) or biological pressure (non-native, invasive species, restocking) (for an overview see Kautza et al, 2015). All those pressures can be recorded and do not change rapidly, so information of local conditions at the site of interest should be easily collected from hydromorphological field studies or existing databases. Any observed adverse effects on any biological quality elements which cannot be directly associated with pollution induced pressure (e.g. scenarios 7, 13, 14, 15 in the decision matrix; Table 13) should be checked against the known or suspected physical and biological pressure, particularly in heavily modified water bodies. Special attention should be paid if biotic indices heavily relying on specific traits are used as LOE 4. Some examples include indices dependant on a) presence and abundance of migratory fish species (important to consider in case of longitudinal interruptions); b) loss of rheophil and increase of stagnophil fish species or floating macrophyte species (in impoundments); c) lithophilic fish species (in heavily canalized sections); d) high abundance of top predators (in case of regular restocking with commercially valuable species); e) ratio between native and exotic species (in case of biological invasions) etc. An example of an analysis how chemical stressors can be differentiated from non-chemical is given in Rico et al., 2016, where the authors used variance partition analyses to pinpoint what shares of variability in community and traits composition were correlated to changes in hydro-morphology, general water quality or pollutants.

Table 2: Common normalization methods for calculation of indices of biological observations

		Resulting range of N_i	**Comments**
1) Standardization via z-scores	$N_i = \frac{X_i - \bar{X}}{\sigma}$	$-\infty - +\infty$	- If X_i is normally distributed, then $\bar{N} = 0$ and *sd(N)=1* - Although the range of possible values is +/- ∞, 95% of the data will be in the interval *+/- 2sd = +/- 2* (assuming N_i is normally distributed)
2) Min-max	$N_i = \frac{X_i - \min(X_{i=1,\dots,n})}{\max(X_{i=1,\dots,n}) - \min(X_{i=1,\dots,n})}$	*0 – max*	- This is the most common normalization in ecotoxicological experiments, i.e. the normalization to 0-100% effect (see text). - If *min=0*, then the min-max normalization collapses to the ratio normalization, with $\text{Reference} = \max(X_{i=1,\dots,n})$
3) Ratio	$N_i = \frac{X_i}{\text{Reference}}$	$1 - \infty$	- Assuming that *0 > Reference <= min($X_{i=1,\dots n}$)*, otherwise the range is 0-1
4) Difference	$N_i = X_i - Reference$	*0 - X_i*	- Assuming that *0 > Reference <= min($X_{i=1,\dots n}$)*, otherwise the range is 0-1
5) Classification	$N_i = \begin{cases} 1 & if \quad X_i < T_1 \\ 2 & if \quad T_1 < X_i < T_2 \\ 3 & if \quad X_i > T_2 \end{cases}$	1 -3	- Example is using just 3 classes, but in principle any number of classes can be used

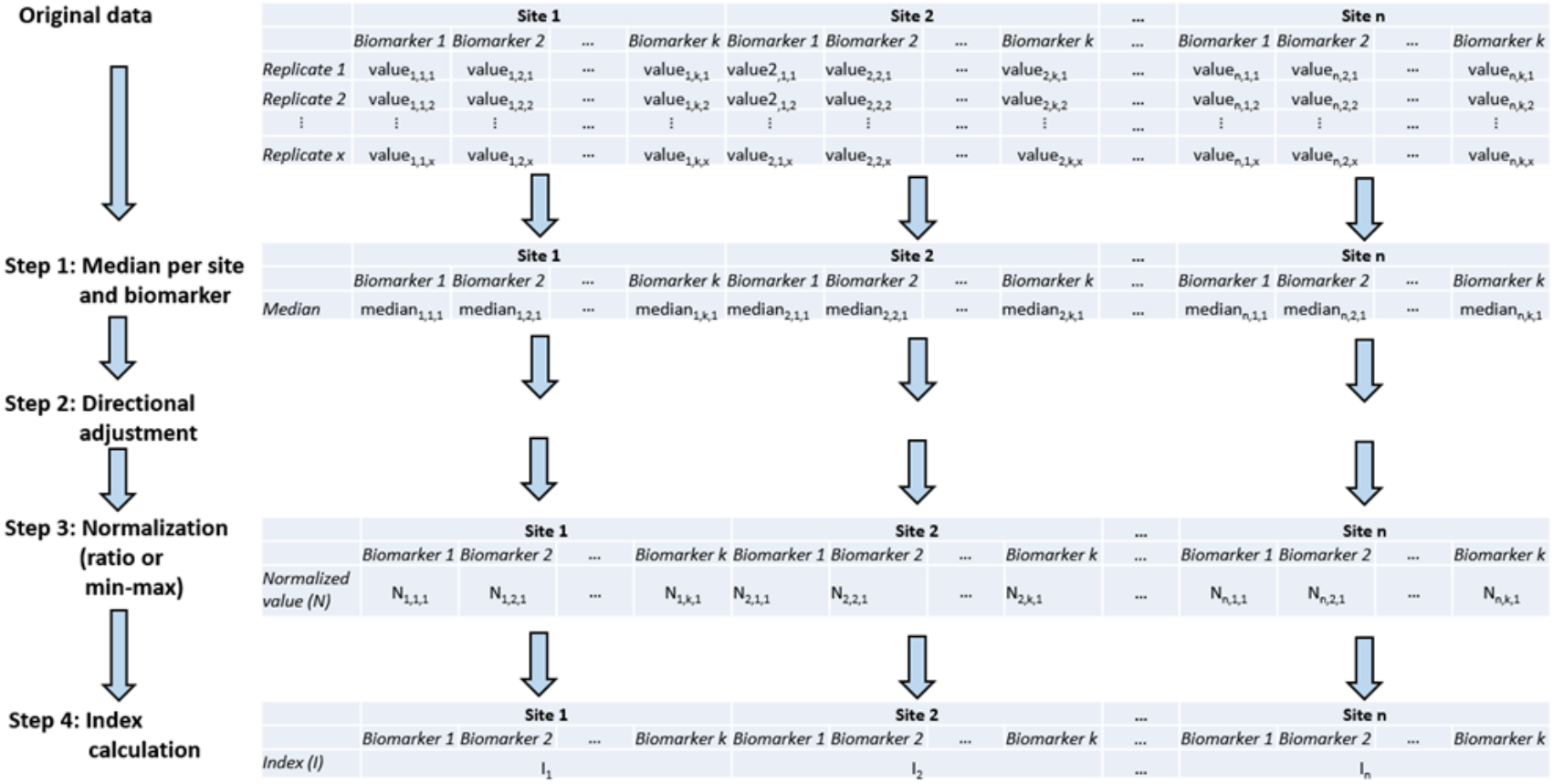


Figure 2: Flowchart for the calculation of the average biomarker response.

3.2 Integration of intermediate LOEs: Calculation of the average biomarker response (ABR)

While predictive mixture modelling and EDA approaches (LOE 1 & 2; Figure 1) suffer from limited ecological relevance, a lack of proven causality between effects and chemical pressure, as in case of data from field studies (LOE4; Figure 1) hampers the assessment of chemical impacts based on high-level LOEs. The intermediate *in situ* LOEs therefore play an important role for a meaningful assessment, linking the more "extreme" LOEs. Here, the main question is how information obtained on the various levels (structure, function, fitness, see Table 12) can be combined to reach a transparent and actionable conclusion about observed test responses that can then be integrated with the results from the other three LOEs.

Indices that condense biomarker responses have been previously suggested in the literature, especially in the form of the "integrated biomarker response" (IBR) by Beliaeff et al. (2002) and its second version (IBRv2) by Sanchez et al. (2013). However, both indices are riddled with conceptual problems. First of all, the overall value of the IBR is dependent on the specific sequence in which the individual responses are included in the index calculation. Both, the IBR and the IBRv2 are based on sums, which implies that the index increases simply when more individual biomarkers are recorded. These and other issues related to the unclear use of reference conditions, renders the final IBR/IBRv2 difficult to understand and interpret. In order to overcome those issues, we suggest the following simple data treatment and condensation, resulting in the calculation of an alternative index values, the average biomarker response:

Step 1: Averaging

Aim of this step is to average the response recorded for several individuals at a site. Assuming either interval- or ratio-scaled endpoints, this allows for the following possibilities:

1) Arithmetic mean
2) Median
3) Geometric mean
4) Harmonic mean
5) Mode

The arithmetic mean is too sensitive to outliers, so is the harmonic mean (especially for outliers in the lower end of the scale). This could be overcome by trimming either estimate, but given the often limited number of individuals (replicates) tested per site, this would quickly resemble a median calculation. The mode throws away too much valuable information, and would also require binning the values into classes prior to the averaging.

That leaves the geometric mean or the median as sound choices. The geometric mean would be preferable if we have a huge dynamic in the endpoint (basically, if it is inherently multiplicative, such as pH measurements). The median is even more stable against outliers, but also throws away more information.

Suggestion: use the median, and characterize the spread (variability within replicates per site) via interquartile ranges for aggregation over replicates from one site.

Step 2: Directional adjustment

Actual biological (biomarker) responses might increase (e.g. number of micronuclei) or decrease (e.g. acetylcholine-esterase activity) with increasing impacts. If not appropriately adjusted, averaging such values makes little sense. Aim of this step is therefore to ensure that higher values of the averages calculated in step 1 always indicate higher impacts. This can be achieved by a so-called directional adjustment. Depending on the data types at hand, this could be implemented by either multiplying the average from step 1 with (-1), taking the inverse (1/x), or subtracting it from one (1-x).

In the following it is therefore assumed that high values indicate high impacts, and that consequently a reference value from a pristine site is in general smaller than the measured value from the test site.

Suggestion: for the data at hand (see case-study examples), the inverse was taken in order to ensure that increasing values indicate increasing effects throughout the battery of biomarkers recorded.

Step 3: Normalization:

Aim is to rescale the responses of the various biomarkers into one common scale, in order to be able to generate an overall index without any implicit weighting or bias. Several common normalization methods are given in Table 2.

In ecotoxicology and toxicology, the most common normalization of raw values, in order to allow a comparison across treatments and across experiments, is the min-max normalization. This is usually implemented by referring to the arithmetic mean of a reference (aka “control”) and the equation then takes the form

$$N_i = \frac{X_i - \overline{\text{control}_{pre-treatment}}}{\overline{\text{control}_{post-treatment}} - \overline{\text{control}_{pre-treatment}}}$$

If $\overline{\text{control}_{pre-treatment}}$ is zero (e.g. when mortality is measured), then the equation collapses to a ratio-normalization, i.e.

$$N_i = \frac{X_i}{\overline{\text{control}_{post-treatment}}}$$

which scales N_i into the interval 0-1. Often, an inhibition value is calculated by subtracting the value from 1:

$$N_i = inhibition = 1 - \frac{X_i}{\overline{\text{control}_{post-treatment}}}$$

Translated into the situation at hand, i.e. the evaluation of biomarker responses from a series of field samples, this could be written as

$$N_i = \frac{X_i}{\text{Reference Site(s)}}$$

The min-max-transformation, respectively ratio-transformation, has the advantage that most evaluators are familiar with it, that the transformed values share a common range of 0 (maximum impact) to 1 (no impact), and that all transformed values are positive, as long as the reference shows the lowest impact. This tremendously helps the index calculation in step 4. It also directly follows the "Ecological Quality Ratio" philosophy that is embedded in the Water Framework Directive and which is calculated as

$$EQR = \frac{X_i - WORST}{\text{BEST} - WORST}$$

Where BEST and WORST give the values that indicate the best, respectively worst possible status / response.

The standardization via z-scores has the main drawback of re-scaling the values to its mean. That is, this approach strictly assumes a normal distribution of the underlying data and, which is perhaps more problematic, it does not allow to refer to any reference conditions. That is, changes in the absolute values of the recorded biomarker responses will go un-noticed. This transformation therefore has only very limited use for the task at hand.

The difference-normalization works in principle very similar to the min-max- and ratio-transformation, with the major drawback of the normalized values not sharing the same range, which would later lead to an implicit weighting of the different biomarkers in the overall index. Also this transformation is therefore of little use for the task at hand.

The final approach that is outlined in Table 2, i.e. the binning of the raw values into a series of classes, however, is another attractive approach. Whether the inherent loss of resolution (in comparison to the min-max- and ratio-transformations) is indeed of practical relevance, remains to be explored in more examples. Such a classification scheme would also allow to consider issues such as the spread of the raw data (by adjusting the class sizes accordingly). In analogy to the classes used in the WFD, five different classes could be defined, "pristine", "good", "moderately impacted", "severe impact" and "maximum impact". It should be noted that such a classification scheme also implies that appropriate reference conditions are available. The main drawback of such a classification-based normalization would be that it severely limits the possibilities to calculate an index value in step 4, which would then be supposed to reflect the "average class". The average of such class system, i.e. of data on an ordinal scale, could only be calculated as the mode value, which would substantially reduce the information content of the final index.

Suggestion: use ratio-normalization or min-max normalization

Step 4: Condensing the results from different biomarkers into one common index

After the results from the different biomarker investigations have been normalized, they can be condensed into one index value. This value should be robust against outliers, changing data situations,

and missing data. It should also be understandable, i.e. easy to communicate to stakeholders. Finally, it should allow a weighting of the different biomarkers that were included in the study, according to their perceived ecological relevance.

These features can only be fulfilled by some sort of averaging of the biomarker responses. Again, the median or the geometric mean seem to be the best choices (see discussion under Step 1), in view of their robustness and ease of interpretation.

Suggestion: use the (weighted) median or geometric mean of the biomarker responses to generate a final index value.

Overview and interpretation of the average biomarker response (ABR) index value

Figure 2 provides the technical flowchart for the calculation of the average biomarker response (ABR) index value. All calculations can be easily implemented in Excel or standard statistical software. The resulting index has an intuitively understandable meaning. A value of 0.5 for a given site would for example be interpreted as "the average biomarker response at the site is 50% of what would be expected for the reference site (resp. reference conditions)". Higher values indicate higher impacts. The overall uncertainty of the index values can be estimated via non-parametric resampling approaches (Bootstrapping).

The problem of defining reference conditions, resp. reference sites

Perhaps the biggest challenge in practical terms will be to define reference sites and reference conditions, i.e. the approach favoured under the WFD, which is strictly required to define what constitutes an impacted site. A reference site can be defined either from data from other LOEs - i.e. from toxic-unit analyses - effect-directed assessments, or from field surveys of biodiversity patterns. Also land-use data could be used for this purpose. An alternative approach would be to define reference conditions on a more fundamental biological (physiological and molecular level), by using data from healthy fish reared under laboratory conditions. This approach, however, is hampered by the fact that biomarker responses might be massively affected by the ecological conditions under which a fish lives, without those being necessarily directly detrimental, as long as the fish is in healthy conditions. Overall, the choice of an appropriate reference site and/or reference conditions will require a lot of expert knowledge in relation to the sites explored and the ecology and biology of the fish in which the biomarker responses are measured.

However, if a relative assessment is to be implemented only, e.g., in order to rank a series of sites that were sampled in a given monitoring campaign, the frame of reference can be developed from within a given series of data. Under these circumstances, "best" can be simply defined as "the least impacted site investigated", and "worst" as "the most impacted site investigated". This assessment can be done even by the series of biomarker responses itself, or any of the other lines of evidence, as outlined above. Care has to be taken if the biomarker responses are used for setting the frame of reference, in order to avoid an assessment based on circular logic (i.e. the biomarker responses are ranked in order to define the "worst" and "best" site of the campaign, in order to then allow a ranking of the sites).

Consideration of causality and ecological impact

In view of the complexity of the data produced by a potentially heterogeneous set of assays and ecotoxicological endpoints, it would be preferable to evaluate the issues "strength of causal link to chemical exposure" and "estimated size of ecologically relevant impact" separately. A suggestion was to calculate an index of causality (IoC) separately from an index of expected ecological impact (IoEEI). Each index could be supposed to reflect the overall weight of evidence of the used test battery for the respective issue. We suggest that the indices could be operationalized as a weighted average of the responses recorded. In contrast to summing up responses, such an estimate would have the major advantage of being fundamentally independent on the number of tests implemented, and in particular more data do not inherently increase the value of the index (which hampers the comparison of studies using different numbers of bioassays and makes it almost impossible to set criteria).

The IoC and IoEEI could thus be calculated as the median of the individual assays weighted by specificities for chemical exposure(eco)toxicologists. Therefore, it could be a function of the characteristics of the bioassays that were used at a given site and its value is independent of the actual outcome of the conducted experiments. Suggested values for the individual specificity of commonly used *in situ* bioassays for chemical exposure and their ecological relevance are given in Table 3.

The principal steps for condensing data from a suite of biomarker recorded from representative samples have been developed for for biomarker responses, as those data were recorded in the 3rd Joint Danube Survey 3, but the principal approach also works for condensations of data of other type in order to generate an appropriate index value.

Nevertheless, the IoC and IoEEI are not calculated in any case study, because the quantification of the weights appears at the moment as not transparent and objective enough to allow for a reliable use in the processing of weights average biomarker responses. It is, for example, very difficult to say that biomarker X has a 2-fold higher ecological relevance compared to another one. When weights could be quantified in a better way, the suggestions above give an outline how boomers could be aggregated under consideration of the specific aspects of specificity for chemical effects and ecological relevance.

Table 3: Weights for specificity of tests for the detection of chemical effects and for ecological relevance, summarizing expert opinions.

	Genotox		Enzyme tests					Gene expression			
	MN	Comet	AChE	EROD	GST	CES	CAT	mgst3a	nrf2a	erk2	gpx1
Chemical specificity (%)	70	35	80	75	62.5	35	25	57.5	40	45	30
Ecological relevance (%)	70	55	65	50	40	30	20	30	30	40	20

3.3 Final integration of the four lines of evidence

In a perfect world, the data produced in the different lines of evidence support each other and therefore allow to draw a clear conclusion on whether chemicals impact the ecological status of a site, and to quantify the impact present. However, in most real-world cases an assessor will struggle with incomplete data and conflicting results. Given the complex and widely divergent data situations that an assessor will have to handle, no strict numerical recipes can be provided for the final integration of LOEs 1- 4. Instead, Table 13 provides the complete decision matrix for the four LOEs (TU-calculation, EDA/EDF, *in situ* studies and field surveys). It outlines the fundamental conclusions that can be drawn from all 16 possible combinations of positive / negative findings in the individual LOEs. Such a systematic evaluation will maximize transparency and will also help to identify critical data gaps.

In order to be able to use the decision matrix, for each study a number of preparatory steps need to be done.

Step 1: Identification of the available LOE data.

While in smaller studies this step can be straight-forward, in large and complex studies, for example campaigns like the joint Danube surveys, the clarification of available test results for different sites requires substantial attention. Figure 1 gives a basis for the identification of the biological groups and possible LOE. By compiling all information, the number of sites for which a certain set of LOE is available need to be defined. These numbers might differ between the biological groups and sites.

Step 2: Definition of class borders

For study-specific results of single LOE, threshold values for the classification of possible effects must be defined. Threshold values allow for the translation of numerical results into the simple classes, which are the basis for using the decision matrix (Table 13). The definition of threshold values is necessary, but in many cases not straight-forward. In the best case, fish or macroinvertebrate indices are computed which are already associated with quality classes. The saprobic index is a good example for such an index, because, being based on macroinvertebrate abundances, it is used to categorise water quality since decades, and corresponding threshold values are available (Zelinka and Marvan, 1961). In many, if not most cases, however, there will be no clear thresholds given, so that for the classification of indicators choices have to be made in the process of a weight of evidence approach. Examples of such indicators are for example taxa richness, Shannon diversity or similar community metrics, for which it appears difficult to set absolute limits for division into categories. Aggregated biomarker responses belong, however, also to this category, especially when they are calculated in a relative way (see page 32). Here the suggestion is to find a good compromise between balancing the classes for a certain study (balancing the number of 'good' and 'bad' indicator results for the series of sites, so that 'good' and 'bad' can always be found), and setting absolute criteria which allows also to put LOE results for all sites into one category only. An example for the first rationale, balancing the classes, would be to always select the upper and lower quartile from a set of LOE results as class borders. Setting limits for the sum of toxic units, would be an example for the second rationale, what means to fix absolute thresholds. The

Danube case study will provide an example for both rationales. Most important is, however, that any selection is clearly documented to allow for the reproduction of results and for an understanding and interpretation of the results.

Step 3: Computing LOE matrices and interpretation of the results.

Often, indicator values and hence class information will be available for a set of sampling or monitoring sites, that can differ in number for different biological groups. Per set of LOE results, as compiled in step1 and as translated into categories in step2, the decision matrix (Table 13) can be evaluated.

To a different extent, the following principal conclusions can then be drawn from the data at hand:

1) Evidence for chemical-induced effects. Under these circumstances, the information provided by LOEs 1, 2 and/or 3 should also provide information on which (groups of) chemicals are suspects for "driving" the overall impact. This informs risk mitigation strategies.
2) There is an impact on the ecology at a site, but it is either unclear or even unlikely that chemicals are a main cause. This conclusion is, given the LOEs outlined in the present text, often based on the absence of clear response patterns in the more chemical-specific LOEs 1 and 2. In order to identify non-chemical causes, such as changes in hydromorphology, additional LOEs might have to be investigated.
3) No visible chemically-induced impact on the ecology at the investigated site However, effort should be invested to get at least a qualitative or semi-quantitative estimate on the power of the overall analysis. This basically means to analyze the following question: Which effect sizes and effect types on which organism groups (biological quality elements) would have been detected given the sampling strategy and the information available from the different LOEs? For example, chemical-analytical methods with an insufficient sensitivity could easily tempt to draw the conclusion that LOE1 (toxic unit summation) does not indicate the presence of toxic pressures – although in fact the individual analytical levels of detection do not warrant such a conclusion, see discussion in e.g. Gustavsson et al. (2017).
   Additionally, the results of all LOEs will always have fundamental knowledge gaps, i.e. "unknown unknowns" (e.g. compounds not included in the chemical-analytical monitoring profiles in LOE1 or important physiological processes not queried in LOE3). An overall assessment should be performed to identify, and if possible close such knowledge gaps, using ,e.g., historical data, data on land-use and chemical consumption patterns, etc.
4) It should be finally emphasized that also the conclusion "no conclusions on the presence/absence of chemical-induced ecological effects possible", is a perfectly valid study outcome. At first sight this might be disappointing for water managers and other stakeholders, as it does not provide any specific management options. However, it is clearly preferable to invest further into follow-up studies (e.g. prolonged or intensified monitoring efforts), instead of implementing potentially wrong but still costly management measures. However, even a study that does not provide hard-and-fast conclusions on pollution-induced ecological effects should be able to deliver sound suggestions for the next steps. Again, a reflection on which types of

effects and which effect sizes were detectable in a given study (and which were not), should provide guidance for the next steps.

The LOEs are integrated according to the strategy outlined in Table 13 separately for each organism group (micro-organisms, invertebrates, higher plants, fish), as outlined in Figure 1. In line with the strategy implemented under the Water Framework Directive, the overall impact on the ecological status at a site will be assessed based on the most sensitive group of organisms (following the "one out, all out" strategy). It should be emphasized that micro-organisms are currently only partly included as a biological quality element under the WFD. Here we suggest to expand from the WFD's sole focus on micro-algae (diatoms) and also include bacteria and fungi, in view of the fundamental ecosystem services that these two organism groups provide. Also, the possible indirect effects on human health via environmental reservoirs of antimicrobial resistance genes might warrant increased consideration of bacterial communities for water quality assessments.

The outlined approach for the integration of available LOE information is exemplarily described for data taken from the river Danube (JDS3), see chapter 4.1.6 for further details. The scenarios are analysed in view of the decision that is to be taken in the end, i.e. answering the question "is there an impact at the test site?".

Importance of documentation

It should be noted that the whole data analysis pipeline, from recording the raw data at the sites of interest to the final assessment involves a series of data reduction steps. This always includes an element of subjectivity and expert knowledge, which is why we would like to highlight the need for keeping the whole process transparent and retraceable. That is, given the complex evaluation and the fact that the assessment will in most cases be depending on expert knowledge, it is absolutely crucial to make all underlying raw data (exposure estimates as well as results from all experiments) available for independent scrutiny and for follow-up studies. It is an absolute minimum to simply reproduce data tables as PDFs in supporting information of scientific papers. Much preferable is to deposit all data in documented, numeric form in public repositories, see also the discussion in, e.g., Bechhofer et al. (2013). The "Registry of Research Data Repositories" for example provides an extensive list of the available scientific repositories, e.g., any of the repositories listed by http://www.re3data.org/. PLOS (Public Library of Science) also maintains a list of recommended data repositories[3]. Also the data-analysis pipeline (the specific process of analyzing the raw data) warrants specific documentation. Highly condensed "material and methods" sections in scientific papers and the sprinkling of formulae over a collection of excel sheets might be considered insufficient in this context. Instead, in the spirit of "reproducible research" (e.g. Mesirov, 2010), a combination of data-bases and documented data-analysis pipelines in, e.g., R, python, SAS or similar languages is recommended.

[3] http://journals.plos.org/plosone/s/data-availability#loc-recommended-repositories

4 Practical examples

The following chapters contain 3 examples for the application of weight of evidence approaches for the assessment of the ecological quality of rivers, and the identification of possible effects of chemical compounds. These examples are of a different nature and extent, some results have already been published while others are unpublished. It should be emphasized that the field studies that were used to illustrate the outlined approaches and their pros & cons in this chapter 4 were implemented as team efforts.

The case study about the Danube river (section 4.1) is based on a huge dataset that was in large parts collected during the Joint Danube Survey 3 (JDS3). Joint Danube Survey 3 was organized by the International Commission for the Protection of the Danube River (ICPDR). Many results were published already in the ICPDR JDS 3 official report (Liška et al, 2015), several scientific publications deriving from the Solutions project activities (Neale et al, 2015, Rico et al, 2016, Deutschmann et al 2016) and internal deliverable of WP 13 of Solutions project (Focks et al, 2015), but so far no systematic WOE evaluation was performed. Many people contributed to this case study, in particular the contributions of Andreu Rico (Wageningen/Madrid), Mirna Velki (Aachen), Dina Tenji, Sonja Kaisarevic, Sandor Sipos, and Vladimir Jovanovic (Novi Sad) should be mentioned. The data from JDS3 provided information for all 4 LOE, what makes this chapter the main application example for the toolbox. Focus was set on testing the aggregation methodology for the diverse *in situ* test results (fish biomarkers) and the overall integration of the LOE.

A next example from river Rhine reports on a targeted study focusing on algal toxicity and algal community structure (section 4.2). The study is evaluated using LOE1 (mixture toxicity modelling based on chemical analytics) and LOE4 (algae community data) were used. Results are discussed in reflection of the main decision table (Table 13). The Rhine case-study was conducted as a joint effort from EAWAG (Ahmed Tlili, Juliane Hollender, Bettina Wagner, Renata Behra) and the University of Gothenburg (Thomas Backhaus, Natalia Corcoll, Åsa Arrhenius).

For the final application example from the Holtemme river, a small creek in Mid-Germany, analyses are still on-going (section 4.3). For example macroinvertebrate data need still to be analysed, and results are not yet published. Nevertheless, first results from a small river with less dilution capacity as compared to huge streams like the Danube or the Rhine are reported here. This study was conducted as a joint effort from UFZ (Werner Brack, Rolf Altenburger, Matthias Liess, Markus Weitere, Ilona Bärlund, Martin Krauss, Pedro Inostroza, Wibke Busch, Tobias Schulze) and the RWTH Aachen University (Henner Hollert, Thomas Benjamin Seiler, Björn Deutschmann, Carolina DiPaolo, Nele Markert).

4.1 Case study Danube: Weight of evidence evaluation of the JDS3 data

4.1.1 Introduction

The aim of this case study was to identify, quantify and distinguish *in situ* ecological impacts of chemical stress from the other stressors present in the Danube River Basin. For reaching that aim, data from the different lines of evidence (Figure 1) was analysed.

1) Results from in depth chemical analyses of water samples were analysed by predictive mixture toxicity modelling (sum of TU, STU).
2) Results from a suite of *in vitro* bioassays, performed with extracts from high volume and passive sampling, were taken into account as published (Schulze at al, 2015; Neale et al., 2015).
3) Results from a battery of relevant *in situ* biomarkers in sentinel fish (*Alburnus alburnus* and *Neogobius* sp.) were analysed and aggregated using the methodology as developed in this deliverable (section 3.2)[4].
4) Taxonomy- and traits-based analyses of fish and macroinvertebrate community data were performed to identify possible ecological impacts[5].

By evaluating the single LOE in a weight of evidence approach, all the data were evaluated with the aim to identify pollution patterns along the River Danube or pollution hot-spots for further in-depth studies. The in situ bioassays, both *in vitro* and *in vivo*, and community data analyses differ concerning their targeted levels of biological organization but they have the potential to detect in situ exposure and effects to confirm or reject possible ecological impact of chemicals. The advantage of bioassays over chemical analytics is that they detect the integrated toxic potency being present at a given field site, even when the chemical composition of the mixture is unknown. While in vivo assays may be considered to be more relevant for the ecological outcome than in vitro assays, the latter provide insight on the modes of action that are active at a given field site, and it is this information that is essential to causatively link chemical exposure to ecological effects ,as illustrated in the Adverse Outcome Pathway (AOP) framework. The critical question is how (quantitative) links between the bioassay information and the ecological outcome can be defined in a WOE approach.

The BQE data and fish samples used in this study were gathered as part of the Third Joint Danube Survey (JDS3), which was organized by the International Commission for the Protection of the Danube River (ICPDR) and carried out in 2013 in cooperation with a large number of international scientific institutions. The tremendous amount of data and samples that was collected in a most coordinated way, provided an unique opportunity to investigate relations between the ecological and the chemical status of the second largest European river. The JDS3 campaign included 68 predetermined sampling sites for

[4] Parts of these results have been published (Deutschmann et al 2016)
[5] Parts of these results have been published (Rico et al, 2016)

WFD compliant chemical and biological monitoring, while subsets of sampling sites were covered by fish sampling (32 sites), high volume water sampling (22 sites) and passive sampling (8 stretches) for target chemical analyses and *in vitro* bioassays (Liška et al. 2015). For fish biomarker analyses, 26 out of 32 fish sampling sites - from JDS 2 (Germany) to JDS 67 (The Danube Delta) were selected.

4.1.2 LOE 1: Chemical analysis and predictive mixture modelling

The JDS3 data set on water concentrations is the result of the one of the most comprehensive analytical investigations done up to now in a single river basin. In addition to metals, basic water quality parameters and priority substances, multi-component target-analysis of water samples supplemented by non-target screening with the major goal to search for as many compounds as possible revealed the presence of more than 200 different organic compounds. JDS3 target screening of 654 substances in the Danube water samples resulted in 277 JDS3 substances actually determined above the limit of quantification in the samples. In general, a large number of substances were found in very low concentrations. Overall, concentration levels of most of these substances slightly decreased downstream the Danube to the Black Sea, although no real concentration gradient has been recognized, so it can be said that the chemical profile is rather flat. As far as hot spots identified solely based on water chemical composition, slightly elevated concentrations of target substances were measured downstream the municipal wastewater discharge points of major cities along the Danube (e.g. Budapest, Belgrade) while due to the relatively small discharge of most tributaries, the Danube itself hardly showed higher concentrations after their inflows. (Liška et al, 2015).

For the classification of the potential toxicity of the chemical mixture measured in each sampling site, the sample from a specific site was characterized as acutely toxic when the log-STUs was ≥−2, and chronically toxic when the log-STUs was <−2 and ≥−3. This classification assumes that an extrapolation factor of 100 and 1000 applied to the acute toxicity values suffices to protect for acute and chronic effects at the community level, respectively.

Sum of Toxic Units for fish

Fish acute toxicity data were available for about 2/3 of the detected chemicals (Busch et al, 2016) which provided a solid base for mixture prediction modelling, which should identify the main drivers of chemical stress at selected sites and enable ranking the sites along the River Danube based on overall toxic pressure. Preliminary results (data not published yet) show that the acute toxic pressure for fish is between 1 and 5% along the Danube. For the majority of the sites, the acute toxic mixture pressure is below 5% of the acute LC50 for fish (STU ranging from 0.007 to 0.05), with one exception only – STU at site JDS48 was 0.5, due to a high copper concentration value. Virtually all sites showed potential for acute toxicity, only for some of the sites in the Danube delta (JDS63, JDS64 and JDS68) and one additional site (JDS29) the STU was below 0.01. Metals (mainly Zn, Cu, Ni, occasionally Cd) and arsenic highly dominated the STU values for fish along the whole river.

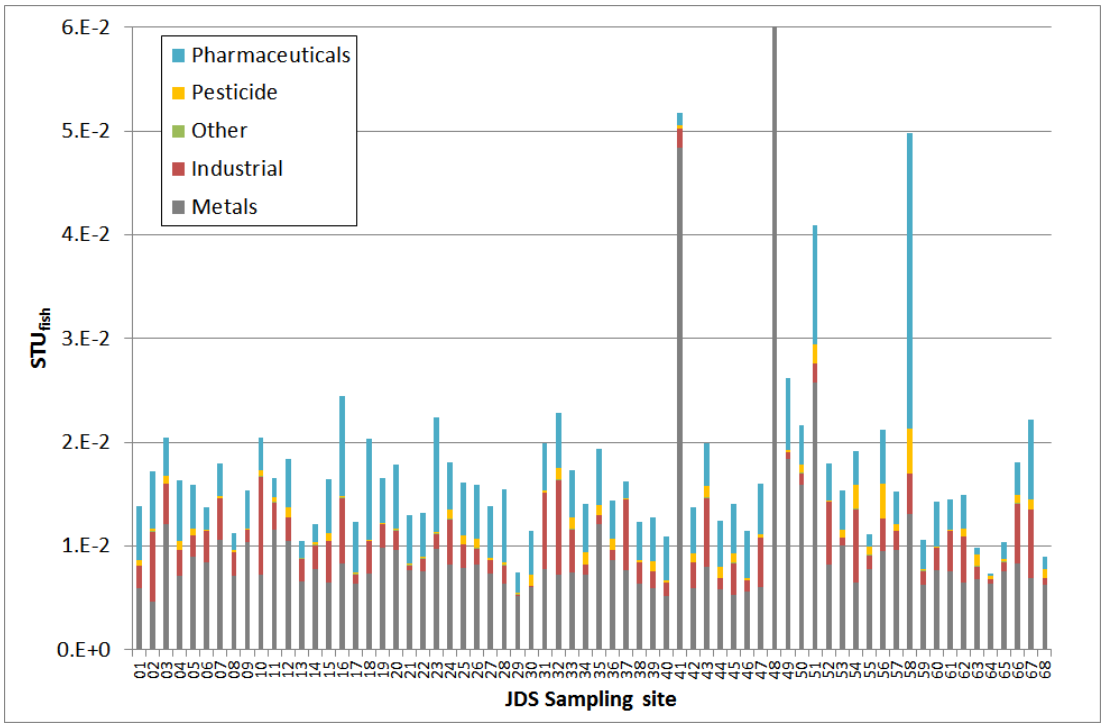


Figure 3: Sum of toxic units as calculated for fish in JDS3 , based on experimental or QSAR LC50, see text for more details. Color codes indicate different compound classes.

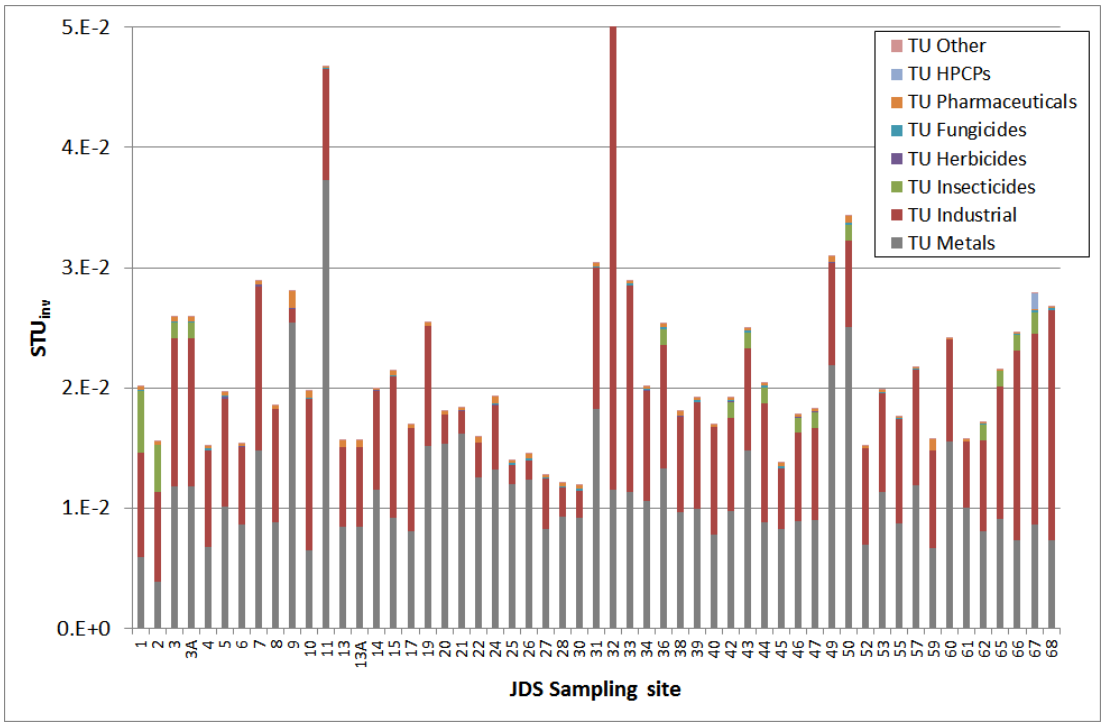


Figure 4: Sum of toxic units as calculated for macroinvertebrates in JDS3 , based on LC50 Daphnia magna. Color codes indicate different compound classes.

Nevertheless, also some organic chemicals were detected in concentrations which are not so far from potentially toxic effects, e.g. irbesartan, a pharmaceutical for treatment of hypertension, was detected at a toxic unit of 0.051 at site JDS58. Also indeno(1,2,3-c,d)pyrene, a PAH compound, valsartan acid and diclofenac were occasionally detected at TU of larger than 0.005.

Sum of Toxic Units for macroinvertebrates

For the evaluation of toxicity potential for macroinvertebrates, experimental acute toxicity data (*D. magna* EC50) were obtained from the E-Tox database (De Zwart, 2002). When experimental toxicity data were not available, they were obtained from the Quantitative Structure-Activity Relationships (QSARs) described in Dimitrov et al. (2000) and contained in the QSAR Toolbox (www.oasis-lmc.org/). After toxicity data collection, TUs could be calculated for 227 out of the 243 measured metals and organic contaminants, 27% of which were based on experimental data and 73% on QSARs. Apart from the total sum of TU, the data was used to calculate sum TU values for different chemical groups (a.o. metals, pesticides, industrial chemicals).

For all JDS3 sites, STU values were above 0.01, ranging between 0.012 and 0.16. According to the classification scheme using an assessment factor of 100 for potential toxic effects, all sites showed a risk for acute toxic effects on macroinvertebrates. The majority of the STU sites showed similar STU values, as these interval between the 5- and the 95-percentile was between 0.013 and 0.032. The potential toxicity at the sampling sites was primarily determined by contamination with heavy metals (mainly copper, nickel and zinc) and industrial contaminants, and to a lesser extent by insecticides, pharmaceuticals, and household and personal care products (HPCPs). The highest toxic pressure due to heavy metal pollution was found at the sampling sites 9, 11, 49 and 50, with log-STUs ranging between−1.4 and −1.7. Acute toxicity potential for industrial pollutants (log-TU from 0.83 to −2) was found in 13 sampling sites mainly due to relatively high concentrations of the PAH compound benzo(g,h,i)perylene. Chronic toxicity potential (log-STU>-3) for insecticides was found for 27% of the samples due to the occurrence of low diazinon concentrations. The toxicity potential of herbicide and fungicide concentrations, as well as the compounds included in the miscellaneous category (e.g. feed additives, sweeteners, tobacco constituents) was below 0.001 and hence can be considered to be insignificant.

Summary

Overall, LOE1 indicated a considerable risk for macroinvertebrates when using an safety factor of 100 based on *Daphnia magna* LC50 values. The risk is nearly exclusively based on heavy metals and PAH compounds, and the variability between the toxic pressure between the sampling sites is rather low. This finding is quite similar for fish, where a toxic pressure of 0.01 STU should be interpreted with care, considering that toxicity thresholds stem to good parts from QSARs and hence a safety factor should be applied here. Overall, the LOE1 indicated a low, but significant potential for (acute) effects of water pollutants, mainly caused by heavy metals.

4.1.3 LOE 2: Effect-based fingerprinting

Effect-based fingerprinting (EDF) in large rivers such as the River Danube requires significant pre-concentration and the extraction of large water volumes for subsequently applying a number of different bioassays and multi-target analysis. A mobile large-volume extraction device (LVSPE) was used to extract water samples of up to 1000 litres on-site during the JDS3 at 22 sampling sites. The extracts were used for a set of different *in vitro* and *in vivo* bioassays (Schulze at al, 2015; Neale et al, 2015). The bioassays covered a broad range of biological effects and endpoints including (for abbreviations see Neale et al., 2015):

- fish embryotoxicity (FET test - embryo coagulation and lack of heartbeat),
- algal growth and photosynthesis inhibition,
- cytotoxicity,
- (anti-)estrogen-like activity (reporter gene MELN assay),
- glucocorticoid-like activity (GR CALUX),
- mutagenic activity (Ames test with and without metabolic activation - Ames -S9/ +S9),
- adaptive stress responses (oxidative stress using ARE-bla, genotoxicity using p53RE-bla and inflammation using NF- κB-bla),
- induction of xenobiotic metabolism (AhR and PXR mediated activity using CAFLUX and HG5LN-hPXR assays), and
- neurotoxicity (acetylcholinesterase inhibition - AChE).

The results of *in vitro* assays were numerically expressed to follow the results published in JDS 3 report (Focks et al, 2015), where 1 stands for “No effect”, 1.5 for “weak effect” and 2 “effect” (Table 4). Despite the overall low concentrations of organic compounds in the Danube, particularly compared to other rivers in Europe (Liška et al, 2015), all extracts were effective in at least one or more bioassays with the endpoints mutagenicity, AhR and PXR mediated activity, oxidative stress responses, ER activation and green algae growth or photosystem II inhibition (Schulze at al, 2015). Although most samples did have a response in the ER, PXR, AhR and NF-κB assays, the effects were relatively low (Neale et al, 2015).

Sample JDS41 (Danube tributary - River Velika Morava) with the highest effect in the ER activation and oxidative stress response assays, cytotoxic in several other assays, was the most polluted site with the highest amount of total detected chemical concentration (Neale et al, 2015). However, due to the fact that the site JDS41 is at the tributary, the site was not sampled for fish and consequently has not been analysed for *in situin situ* biological effects by fish biomarker analyses. Along the River Danube, several sites (JDS 27, 33, 36, 39, 53, 60, 65 and 67) were effective in at least one or more bioassays (Schulze at al, 2015: Neale et al, 2015). All the listed sites were also JDS 3 fish sampling sites which enabled matching the results from the battery of bioassays with the JDS sites for which fish biomarker analyses were performed.

The assays indicative of activation of ER, PXR, AhR, and the NF-κB response tended to be the most responsive, followed by the oxidative stress response. The p53 response occurred only at higher effect concentrations. The least responsive assay was the FET test. Overall, there was no significant relationship between effect and the sum of detected chemicals at each site for the different assays. Contribution of the individual detected chemicals to the biological effects has been calculated based on comparison between BEQbio and BeQchem[6] values. Between 3 and 71% of the AhR activation and up to 80% of ER activation could be explained by detected chemicals. In contrast, the detected chemicals could explain less than 0.2% of the biological effect in the adaptive stress response assays, PXR assay, and the FET test (Neale et al., 2015). For toxicological profiling of passive sampler extracts, the same battery of bioassays as for LVSPE was used (Vrana et al, 2015). Except for the oxidative stress response, toxic equivalents, particularly for ER and AhR activation, were detected in a similar range for high volume and passive sampling (Hilscherova et al. 2017).

It needs to be emphasized that specifically the receptor-dependent bioassays such as AhR and ER were responsive to Danube samples. In addition, a large fraction of the chemical compounds being responsible for this toxic potential could be identified by chemical analytics. Finally, for a number of receptor-mediated toxic events, there exist AOPs, for instance, there are well established AOPs for ER-mediated ecotoxicity (see AOP WIKI). For these pathways, there exist "Key Effect Relationships" (KER) that are verified by Bradford-Hill criteria and which link key events from the molecular and cellular levels as measured in in vitro bioassays, to adverse organism- and population-level effects (Becker et al. 2016). Thus, they provide a strong LOE which ecologically relevant fitness parameters of organisms will be at risk. The drawback in this LOE linking *in vitro* bioassays and ecological outcomes is the insufficient quantitative information on the concentration-response relationships: which response intensity of the *in vitro* assay will translate into an adverse effect in the organism? Such information is partly available for laboratory model species, for which good toxicokinetic and toxicodynamic information does exist, but virtually non-existing for the field site species. However, it needs to be emphasized that this drawback is not specific for *in vitro* assays but applies principally for all biotest data including the LC50 data which form the basis for many of the modelling and prediction tools used in environmental risk assessment. Nevertheless, the results from the Danube case study, while highlighting the suitability of in vitro assays for characterizing toxicity profiles of environmental compartments and provide LOE for potential ecological outcomes, they also point to the urgent research needs for improving our knowledge on the concentration-response relationships between in vitro bioassay signals and in vivo toxic effects

Summary

Overall the LOE2 results indicate more within-site differences (different response between the single applied bioassays at a site) than between the single JDS sites. Aggregation of the applied class scores (Table 4) by taking the arithmetic mean results in very similar values for all sites reported here. This is

[6] BEQ: bioanalytical equivalent concentration, BEQbio: BEQ from bioanalysis, BEQchem: DEQ frm chemical analysis (Neale et al., 2015)

also because only very few sites that showed clear responses of single bioassays, for example JDS 41 and 63 for AhR, JDS 55 and 67 for oxidative stress response, JDS 41, 55, 57 and 63 for p53 response, and JDS 36 and 41 for NF-κB, were also sampled for fish biomarkers. One important conclusion that can be drawn from the evaluation of LOE2 is that a better matching between the sites for large volume sampling and fish biomarker analyses would have tremendously improved the interpretation of the results.

4.1.4 LOE 3: In situ effects - fish biomarkers

Fish species for the study of *in-situ* biomarkers were selected based on the occurrence and abundance at all selected sampling sites. To contrast between different ecological traits, the two most abundant and frequently caught fish species with different traits - eurytopic *Alburnus alburnus* Linneus 1758*, fam.* Cyprinidae (common bleak) and a typical bottom-dwelling *Neogobius melanostomus* Pallas *1814 fam.* Gobiidae (round goby) were chosen.

The following biomarker analyses were applied:

- Genotoxicity / DNA damage: blood samples from *A.alburnus* - Micronucleus test, Comet assay
- Enzyme activities - both fish species (*A.alburnus, Neogobius sp*.)
    - Activity of Phase I biotransformation enzymes:
        - ethoxyresorufin-O-deethylase - EROD
        - Carboxylesterase - CES
    - Activity of Phase II biotransformation enzymes:
        - glutathione-S-transferase - GST
    - Oxidative Stress: catalase - CAT
    - Neurotoxicity : acetylcholinesterase - AChE
- Gene expression analyses in liver samples from *A. alburnus*, normalised to the housekeeping gene glyceraldehyde-3-phosphate dehydrogenase (*gapdh*).
    - Activation of signaling ERK1/2 pathway (involved in the regulation of e.g cell survival, differentiation, metabolism, proliferation etc): extra-cellular signal regulated protein kinase 2 (*erk2*)
    - Antioxidant defense: glutathione peroxidase 1 (gpx1)
    - Activation of transcriptional regulator of detoxifying and antioxidant genes: nuclear factor erythroid 2-related factor 2A (*nrf2a*)
    - Detoxification of xenobiotics: membrane integral glutathione S-transferase 3a (*mgst3a*)
- Histopathology - liver samples of *Neogobius melanostomussp.*

Detailed descriptions of the materials and methods used for these biomarker analyses are given in the deliverable ID T5.1 “Report on WoE and trait-based results for JDS3 samples” (Focks et al., 2015).

Qualitative discussion of biomarkers

Multiple comparison tests showed no difference in CAT activity and relative expression of any of the selected genes between the fish samples from selected sites. Flat CAT and *gpx1* response comes in line with the findings from *in vitro* assays that oxidative stress assay (ARE) was not among the most responsive *in vitro* assays. The similar holds true for induction of xenobiotic metabolism, since the response of biotransformation enzymes (CES, GST) as well as the expression of *mgst3a* and *nrf2a* was also flat (with little and random differences in enzyme activities between sites), corresponding to the moderate response of AhR and PXR *in vitro*. The AChE activity *in situ* also followed the flat pattern of *in vitro* assays, with one exception only - extremely high *in situin situ* values at site JDS 33 (downstream Novi Sad, a city of 250000 inhabitants, no WWTP). The results of *in situin situ* EROD activity were the most scattered of all the biomarkers, with no consistency in response of two species or predictability according to known pressures. It is interesting that again site JDS 33 appears as the extreme site - the lowest EROD (as well as GST) acitvity in *N. melanostomus* was recorded there.

The highest EROD activity and the only case of *erk2* up-regulation was recorded, unexpectedly, in fish caught in Chilia arm in the Danube Delta (site JDS 66). Against expectations, no *in situin situ* effects were seen in fish from sites JDS 38 and 39 located downstream Belgrade (city of I million inhabitants, no WWTP) and Pancevo (heavily industrialized region downstream Belgrade).

In an assessment of the genotoxic potential (Deutschman et al, 2016), JDS60 (Chiciu/Silistra) was the only site where a significantly elevated MN formation was found. The site is located 50 km downstream of the confluence of the River Arges (impacted by the city of Bucharest). The results come in line with *in vitro* Ames test (after metabolic activation). The results from the comet assay indicated high genotoxic potential at the site JDS47 (upstream the confluence of the River Timok), but that was the only biological effect observed at that particular site (none of the studied enzymes showed any extreme activity).Since only *N. melanostomus* livers collected at 4 sampling sites (JDS 33, 48, 60 and 62) were examined, hystopathological analysis cannot offer more insight or present an additional line of evidence in this case study. It has to be noted, however, that at sites JDS60 and 62 ruptured membranes were observed, while at sites 33 and 48 only low levels of membrane ruptures were observed.

Aggregation of biomarker results

For the evaluation of the multiparametric response pattern of the *in situ* biomarkers, the approach outlined in section 3.2 was followed. Biomarker response value averages were calculated by using the geometric mean. For one of the tests, i.e. the acetylcholine-esterase inhibition, directional adjustment was necessary. This was performed by taking the reciprocal value of the geometric mean. After that, all average biomarker values were supposed to show increased values for increased chemical effects. Average biomarker responses were then transformed into the same interval by ratio-normalization to the maximum. Normalisation to the maximum was chosen, because it was not possible to use a reference site for the normalisation of the biomarker responses, due to the long term, diverse and heavy anthropogenic pressures along the whole river. No clear reference site could be identified, which showed consistently low biomarker responses for all tests. In addition, information on baseline enzyme activity or constitutive gene expression could be found for the chosen species. Results from all different

biomarkers per site were condensed into one index value by calculation of the geometric mean of all values per site (Table 4).

The average biomarker responses per site (column 'ABR' in Table 4) can't be interpreted directly in an absolute sense, because they were normalised to the maximum, so that the maximum possible response is always 1, irrespective of the actual intensity of the response. This means, a geomean of 1 for all biomarkers at a specific site would say that this site shows the maximum possible response for all biomarkers tested at this site (highest possible impact), while zero would mean the opposite. The index values allow for a relative ranking of the sites, and an analysis of the distributions of the overall biomarker response.

Results of biomarker aggregation

The aggregated biomarker index values were calculated for all 19 sites and ranked between 0.23 and 0.59. These values reflect what was already mentioned earlier: there were neither sites where the different biomarkers showed consistently lower reactions (reference sites), but also no sites with maximum response in all biomarker tests. Interestingly, site JDS33 showed the second-lowest index value of 0.30, despite the site is potentially heavily impacted by the wastewater of Novi Sad, a city with approximately 250,000 inhabitants. This fact could point at a high dilution potential in the Danube, which has been found in another study where the impact of wastewater was diminished already 7 km downstream the city (König et al, 2017). Calculation of the index values enabled a ranking of the sites, based on an integrated biomarker response index, and allow for the identification of most polluted sites. Highest index values were found for sites JDS 40, 66, and 58. Especially for site 66 this is surprising, because this site is located in the Danube delta.

The aggregation of the single biomarker values was achieved by normalisation to the maximum, since no reference site could be identified amongst all samples JDS3 sites. In consequence, no absolute interpretation of the aggregated indicator values (Table 4) can be achieved. Nevertheless, the values can be interpreted in a relative sense. When the aggregated indicator value ranges at 0.5, that means in average a response of 50% of the maximum response, and this can be interpreted as a clear signal amongst all tested biomarkers. A lower threshold could be defined at 0.1., which means that the (normalised) average response was below 10% of the maximum response. When applying this scheme, from the 19 sites with biomarker information, 10 sites can be considered to show medium-level effects, while 9 sites show a high average biomarker response.

Summary

In general, the calculated index values confirm results from LOE1 and 2: the sites which have been sampled for biomarkers in JDS3 do not differentiate very much with respect to their toxic potential, at least for fish.

4.1.5 LOE 4: Taxonomy- and traits-based assessment of aquatic communities

Macrophyte, macroinvertebrate and fish community structure and composition was analysed and reported in the JDS3 in a WFD compliant way (Stankovic et al, 2015; Graf et al, 2015; Bammer et al, 2015). The raw BQE data were kindly provided by ICPDR. A number of taxonomy- and traits-based indices were used to summarise the status of the fish and macroinvertebrate communities. Macrophyte data were not evaluated as part of the WoE study, because of the earlier reported bias in the sampling of macrophyte communities during JDS3 (see e.g. Focks et al., 2015).

Basically, all community assessment is based on the reported species at a sampling site and the corresponding abundances. For such information, there is a plethora of diversity measures for ecological communities, which, depending on the question at hand, calculate measures of species richness (e.g. number of taxa, Margalef, Menhinick), for heterogeneity of the community (Shannon, Brillouin, Simpson) and other taxonomic or functional indicators (for an overview, see e.g. Gotelli & Chao, 2013 or Magurran 2004). Values of many metrics depend on the sample size, which makes them less useful for the evaluation of data sets such as the JDS3. For a relative assessment of the JDS3 sites, a set of 14 indices was calculated, i.e. total abundance, species number, Shannon index, Dominance_D, Simpson_1-D, Shannon_H, Evenness, Brillouin, Menhinick, Margalef, Equitability_J, Fisher_alpha, Berger-Parker and Chao-1. The meaning of absolute values of these indices is not discussed here, but for the used indices threshold values are defined in section 4.1.6. A comparison of the average ranks of the 32 fish sites, based on all 14 or only the 3 basic indices revealed for the fish data that the ranking of all 32 fish sites did only change marginally when being based on 3 or 14 indices. Cross correlation analysis showed also that indeed 7 of the 14 indices correlated with the total abundance with coefficients larger than 0.9 (see Table 5). Therefore the most basic measures, i.e. the total abundance and the species number (richness , number of taxa), together with the Shannon-Wiener index as a measure for the heterogeneity of the communities at a site were selected. The advantage is, that those measures are readily accessible for fish and macroinvertebrates. Both for fish and macroinvertebrates, a large number of specific community indices have been developed that give information about the biological integrity (see also section 2.4). In addition to the selected 3 general diversity indices, for fish and macroinvertebrates each 3 additional indices were selected and used for the evaluation of potential community impairment.

Table 4 Summary of the results of *in vitro* assays and *in situ* fish biomarkers at selected sites along the River Danube.

	In vitro* tests (LVSPE extracts)														***Alburnus alburnus***													***Neogobius melanostomus***					
JDS3 ID	**Algae GI ***	**Algae SII ***	**FET #**	**Ames -S9 ***	**Ames +9 ***	**p53 #**	**ARE #**	**NFkB #**	**ER #**	**PXR #**	**AhR #**	**GR ***	**AChE ***	**Average All**	**GST**	**EROD**	**CAT**	**CES**	**AChE**	**MN§**	**Comet§**	**erk2**	**gpx1**	**nrf2a**	**mgst3a**	**GST**	**EROD**	**CAT**	**CES**	**AChE**	**ABR (sec. 3.2)**
27	1.5	1	1	1	2	1	1.5	1.5	2	2	1.5	1	1	**1.385**	0.6	0.7	0.6	1.0	0.5	0.1	0.2					0.8		0.8	0.7	0.5	**0.48**
28															0.2	0.2	0.3	0.5	0.9	0.1						0.7		0.8	0.8	0.4	**0.40**
31															0.5	0.3	0.6	0.7	0.8	0.5		0.2	0.6	0.5	0.2	1.0	0.2	0.8	0.9	0.4	**0.46**
33	1	1	1.5	1.5	2	1	1.5	1	1.5	2	2	1.5	1	**1.423**	0.2	0.5	0.4	0.5	0.8	0.4	0.3	0.2	0.6	0.6	0.9	0.1	0.0	0.4	0.4	0.1	**0.30**
36	2	1.5	1	1.5	2	1	1.5		2	2	1.5	2	1	**1.583**	0.2	0.1	0.4	0.4	1.0	0.4		0.3	0.7	1.0	0.5	0.8	0.3	0.8	0.8	0.2	**0.45**
38																				0.5						0.6	0.6	0.9	0.5	0.3	**0.55**
39	2	1.5	1	1.5	2	1	1	1.5	1	2	1.5	1.5	1	**1.423**	0.2	0.3										0.3	0.1				**0.23**
40															0.6	0.5	0.5	0.8	0.5		0.6										**0.59**
47															0.2	0.8	0.6	0.5	0.9		1.0					0.2	0.4	0.4	0.7	0.2	**0.46**
48															0.4	1.0	0.8	0.7	0.9	0.1		0.2	0.6	0.3	0.6	0.7	0.7	0.7	0.7	0.4	**0.52**
50																										0.6	0.3	0.7	0.7	0.5	**0.52**
53	2	1.5	1	1	2	1	1.5	1.5	1.5	2	1.5	1	1	**1.423**	0.6	0.7	0.7	0.7	0.7	0.1						0.9	0.3	1.0	1.0	0.3	**0.53**
54															0.8	0.7	0.6	0.5	0.8			0.3	0.2	0.4	0.2	0.7	0.3	0.7	0.9	0.4	**0.49**
58															0.8	0.3	0.9	1.0	0.6							0.6	0.2	0.5	0.8	0.4	**0.56**
60	2	1.5	1	1	2	1	1.5	1.5	2	1	1	1	1	**1.346**		0.1				1.0		0.3	1.0	0.8	1.0	0.3	0.2	0.7	0.3	1.0	**0.49**
62															0.7	0.4	0.5	0.6	0.7	0.2	0.2	0.3	0.7	0.9	0.6	0.6	0.4	0.7	0.7	0.4	**0.51**
65	2	1.5	1	1.5	2	1.5	2	1	2	2	1.5	1	1	**1.538**	0.6	0.7	0.6	0.6	0.5	0.3	0.2	0.2	0.3	0.5	0.1	0.4	0.2	0.5	0.7	0.5	**0.39**
66															1.0	0.9	1.0	0.8	0.6	0.2		1.0	0.4	1.0	0.2	0.6	1.0	0.5	0.7	0.2	**0.58**
67	1.5	1.5	1.5	1	1	1.5		1.5		2	1	1	1	**1.318**	1.0	0.3	0.9	0.8	0.6	0.4	0.1	0.8	0.4	0.9	0.8	0.4	0.2	0.7	0.4	0.5	**0.51**

***In vitro* assays**: Algae GI - Algae growth inhibition assay; Algae PSII - Algae photosystem II inhibition assay; FET - Fish Embrio test; Ames -S9 / +S9 Ames mutagenicity test without and with metabolic activation; p53 - genotoxicity assay p53RE-bla; ARE - oxidative stress assay ARE-bla; NF- κB - inflammation assay NF- κB-bla; ER - (anti-)estrogen-like activity reporter gene MELN assay; AhR - xenobiotic metabolism AhR mediated activity CAFLUX assay; PXR - xenobiotic metabolism PXR mediated activity HG5LN-hPXR assay; GR - glucocorticoid-like activity GR CALUX; AChE - acetylcholinesterase inhibition assay; ***In situ* biomarkers, aggregated and normalised (see text)**: GST - activity of glutathione-S-transferase; EROD - ethoxyresorufin-O-deethylase; CAT - activity of catalase; CES - carboxylesterase; MN- micronucleus frequency; Comet - genotoxicity - comet assay; Gene expression: *erk2* - extra-cellular signal regulated protein kinase 2; *gpx1* - *glutathione peroxidase 1; nrf2a* - nuclear factor erythroid 2-related factor 2A; *mgst3a* - membrane integral glutathione S-transferase 3a.

* Based on qualitative expressions from Schultze at al, 2015; # Recalculated from Neale et al, 2015; § Recalculated from Deutschman et al, 2016; The rest is unpublished data (in preparation for publication);

Table 5: Cross correlation table between 14 diversity indices, calculated on basis of raw data from 32 fish sites sampled in JDS3.

	Taxa_S	Individuals	Dominance_D	Simpson_1-D	Shannon_H	Evenness	Brillouin	Menhinick	Margalef	Equitability_J	Fisher_alpha	Berger-Parker	Chao-1
Chao-1	0.85	0.03	0.59	0.01	0.75	0.18	0.20	0.17	0.18	0.07	0.07	0.29	1.00
Berger-Parker	0.14	0.60	0.45	0.67	0.30	0.77	0.60	0.63	0.61	0.63	0.63	1.00	0.29
Fisher_alpha	0.08	0.98	0.73	0.98	0.62	0.82	0.97	0.91	0.98	1.00	1.00	0.63	0.07
Equitability_J	0.08	0.98	0.73	0.98	0.62	0.82	0.97	0.91	0.98	1.00	1.00	0.63	0.07
Margalef	0.17	0.95	0.80	0.98	0.70	0.80	1.00	0.91	1.00	0.98	0.98	0.61	0.18
Menhinick	0.16	0.94	0.58	0.96	0.42	0.86	0.90	1.00	0.91	0.91	0.91	0.63	0.17
Brillouin	0.19	0.95	0.80	0.98	0.71	0.79	1.00	0.90	1.00	0.97	0.97	0.60	0.20
Evenness	0.13	0.83	0.69	0.85	0.51	1.00	0.79	0.86	0.80	0.82	0.82	0.77	0.18
Shannon_H	0.67	0.59	0.97	0.58	1.00	0.51	0.71	0.42	0.70	0.62	0.62	0.30	0.75
Simpson_1-D	0.02	0.96	0.71	1.00	0.58	0.85	0.98	0.96	0.98	0.98	0.98	0.67	0.01
Dominance_D	0.53	0.71	1.00	0.71	0.97	0.69	0.80	0.58	0.80	0.73	0.73	0.45	0.59
Individuals	0.02	1.00	0.71	0.96	0.59	0.83	0.95	0.94	0.95	0.98	0.98	0.60	0.03
Taxa_S	1.00	0.02	0.53	0.02	0.67	0.13	0.19	0.16	0.17	0.08	0.08	0.14	0.85

Fish community

For the community status assessment of fish, three different indices were used for the official reporting of the JDS3 results (Liška et al, 2015): FIA (Fish Index Austria), which is mainly sensitive to hydromorphological pressure; FIS (Fish Index Slovakia) - one of the rare indices highly sensitive to biological pressures (e.g. introduced or invasive species); and EFI (European Fish Index) which is designed basically to respond to overall water quality and eutrophication. Focussing on the EFI values of the 32 selected fish sites (see

Table 6), for 21 sites EFI values were reported from the JDS3. The majority fall into the moderate status (12), 3 even to the poor status while EFI for one site - for site JDS38 bad ecological status is associated to the EFI. That is consistent with the values of the diversity indices, particularly with taxa richness as the lowest number of species among the selected species was recorded at JDS 38. Also the average ranking of site 38 was quite low, which is somehow expected, as the site is located in metropolitan area of Belgrade.

Table 6: Fish indices and metrics for community diversity , for the 32 JDS3 fish sites. FIA, EFI and FIS values taken from the JDS3 report (Liska et al., 2013). Averaged ranks for the Species numbers, Total abundance and Shannon index, corresponding single ranks in brackets. Sorted by the average rank. Rows for JDS sites JDS 47, 39, 60, 31, 38, 62, 33, 40, 28, 36, 65, 67, 66 and 27 (14 sites) are in bold, since for these sites also fish biomarker results are available. For sites JDS 50 and 53 biomarker results are available, but no fish index values were available.

JDS3 code	rkm	FIA	EFI	FIS	Species number	Total abundance	Shannon_H	Average Rank
JDS08	2008	Bad	Good	Good	40 (1)	2807 (17)	2.36 (6)	8
JDS46	926				35 (5)	7980 (6)	1.88 (14)	8
JDS62	**167**	**Moderate**			**37 (3)**	**10104 (3)**	**1.74 (20)**	**9**
JDS67	**21**	**Moderate**			**35 (5)**	**3323 (13)**	**2.09 (10)**	**9**
JDS53	**557**				**35 (5)**	**3918 (9)**	**1.87 (15)**	**10**
JDS27	**1434**	**Good**	**Moderate**	**Moderate**	**32 (9)**	**1678 (21)**	**2.66 (1)**	**10**
JDS40	**1107**	**Good**	**Moderate**	**Bad**	**38 (2)**	**5443 (8)**	**1.59 (22)**	**11**
JDS60	**378**	**Poor**	**Moderate**		**36 (4)**	**9967 (4)**	**1.12 (27)**	**12**
JDS20	1705	Good	Moderate	Moderate	29 (13)	1849 (20)	2.41 (5)	13
JDS31	**1303**	**Moderate**	**Moderate**	**Bad**	**33 (8)**	**6018 (7)**	**1.57 (23)**	**13**
JDS65	**130**	**Moderate**			**31 (11)**	**2864 (16)**	**1.97 (12)**	**13**
JDS22	1632	Good	Moderate	Poor	29 (13)	1853 (19)	2.02 (11)	14
JDS28	**1384**	**Good**	**Moderate**	**Moderate**	**32 (9)**	**3148 (15)**	**1.76 (19)**	**14**
JDS15	1807	Moderate	Moderate	Moderate	29 (13)	708 (29)	2.55 (3)	15
JDS02	2415	Good	Good	Poor	29 (13)	18077 (1)	0.90 (32)	15
JDS13	1876	Good	Moderate	Moderate	30 (12)	1674 (22)	1.93 (13)	16
JDS66	**18**	**Moderate**			**29 (13)**	**1194 (26)**	**2.23 (8)**	**16**
JDS57	488	*			27 (21)	902 (27)	2.61 (2)	17
JDS52	602	Poor			28 (19)	793 (28)	2.45 (4)	17
JDS04	2285	Good	Good	Bad	28 (19)	16180 (2)	0.93 (31)	17
JDS44	1040	Good	Poor		27 (21)	3803 (11)	1.61 (21)	18
JDS50	**685**	*****			**29 (13)**	**3227 (14)**	**1.15 (26)**	**18**
JDS47	849	Poor			27 (21)	9156 (5)	1.04 (28)	18
JDS06	2204	Bad	Good	Bad	27 (21)	3746 (12)	1.04 (29)	21
JDS36	**1200**	**Poor**	**Moderate**	**Moderate**	**24 (28)**	**550 (30)**	**2.26 (7)**	**22**
JDS33	**1252**	**Moderate**	**Moderate**	**Poor**	**26 (26)**	**1483 (25)**	**1.85 (16)**	**22**
JDS07	2120	Bad	Good	Poor	27 (21)	1589 (23)	1.19 (24)	23
JDS10	1895	Moderate	Moderate	Moderate	25 (27)	504 (32)	2.17 (9)	23
JDS14	1847	Moderate	Poor	Bad	24 (28)	3844 (10)	0.96 (30)	23
JDS38	**1159**	**Moderate**	**Bad**	**Poor**	**24 (28)**	**1530 (24)**	**1.77 (18)**	**23**
JDS39	**1151**	**Moderate**	**Poor**	**Bad**	**17 (32)**	**1971 (18)**	**1.16 (25)**	**25**
JDS51	637	*			20 (31)	544 (31)	1.81 (17)	26

Table 7: Macroinvertebrate community indices. Number show the index value per site and the rank of that value (in brackets). Sites are order by the average rank across total abundance, species number and Shannon index.

JDS3 site	Species number	Total abundance	Shannon H	ASPT index	SPEAR pesticide	Saprobic index	Avg rank
52	100 (4)	1626 (16)	3.2 (3)	4.0 (3)	6.3 (17)	2.4 (14)	8
17	86 (7)	1880 (12)	2.6 (13)	3.8 (7)	5.3 (27)	2.1 (41)	11
2	106 (2)	936 (29)	3.3 (2)	3.7 (9)	5.2 (31)	2.1 (29)	11
1	123 (1)	1478 (17)	2.5 (18)	5.0 (1)	15.5 (1)	2.1 (36)	12
4	94 (5)	1341 (21)	2.7 (11)	3.3 (23)	5.5 (26)	1.9 (53)	12
40	85 (9)	1058 (24)	3.0 (4)	3.4 (19)	4.6 (34)	2.6 (7)	12
3A	79 (12)	3452 (5)	2.4 (22)	3.0 (29)	2.8 (45)	1.9 (55)	13
5	86 (7)	4192 (2)	1.9 (37)	4.9 (2)	12.0 (4)	1.9 (54)	15
50	103 (3)	883 (31)	2.6 (12)	3.3 (24)	9.3 (5)	2.0 (46)	15
61	61 (25)	2708 (10)	2.4 (20)	3.8 (6)	5.8 (24)	2.5 (9)	18
47	71 (16)	993 (26)	2.6 (14)	2.8 (38)	3.0 (42)	2.4 (13)	19
49	61 (25)	1724 (14)	2.5 (17)	2.5 (51)	2.1 (47)	2.1 (36)	19
22	80 (11)	527 (39)	2.8 (9)	3.0 (32)	3.1 (41)	2.4 (10)	20
39	75 (15)	2613 (11)	2.1 (33)	3.5 (15)	6.4 (16)	2.4 (11)	20
53	69 (20)	721 (34)	3.0 (5)	2.5 (49)	5.6 (25)	2.3 (20)	20
19	90 (6)	1748 (13)	1.6 (43)	2.6 (46)	6.6 (14)	2.1 (30)	21
20	78 (14)	2846 (8)	1.7 (42)	3.7 (10)	7.8 (10)	2.1 (32)	21
43	70 (18)	1102 (23)	2.4 (23)	3.0 (30)	4.7 (33)	2.4 (14)	21
9	65 (21)	1468 (18)	2.2 (29)	3.4 (17)	8.1 (8)	2.1 (39)	23
45	62 (23)	932 (30)	2.6 (15)	2.8 (39)	1.0 (52)	2.7 (6)	23
25	62 (23)	1379 (19)	2.3 (27)	3.3 (21)	6.0 (21)	2.2 (23)	23
14	81 (10)	983 (27)	2.1 (35)	2.9 (37)	5.9 (22)	2.3 (20)	24
7	79 (12)	1374 (20)	1.8 (41)	2.9 (36)	5.2 (28)	2.2 (26)	24
10	70 (18)	134 (54)	3.4 (1)	3.2 (25)	3.8 (38)	2.0 (43)	24
68	44 (43)	1229 (22)	2.7 (10)	2.5 (51)	4.2 (37)	2.1 (36)	25
46	60 (28)	3463 (4)	1.5 (44)	3.5 (14)	5.1 (32)	3.0 (3)	25
3	52 (35)	2814 (9)	2.0 (36)	3.4 (16)	2.4 (46)	1.9 (52)	27
44	61 (25)	317 (47)	2.8 (8)	2.7 (41)	6.2 (18)	2.4 (16)	27
55	41 (44)	4481 (1)	1.9 (38)	2.8 (40)	14.0 (2)	2.0 (49)	28
59	59 (29)	294 (49)	2.9 (7)	2.6 (45)	5.2 (30)	2.1 (32)	28
42	59 (29)	4174 (3)	0.9 (55)	2.6 (48)	4.4 (36)	2.9 (4)	29
24	58 (31)	3298 (6)	1.0 (53)	2.9 (35)	6.1 (20)	2.1 (30)	30
6	71 (16)	630 (36)	1.9 (39)	2.7 (42)	3.2 (40)	2.3 (17)	30
60	50 (36)	958 (28)	2.2 (31)	3.2 (27)	8.2 (6)	2.0 (42)	32
8	64 (22)	339 (44)	2.2 (30)	3.9 (4)	8.1 (7)	2.0 (49)	32
66	49 (38)	830 (33)	2.3 (26)	3.0 (31)	6.6 (15)	2.7 (5)	32
33	53 (34)	348 (43)	2.4 (21)	2.9 (34)	1.2 (50)	2.3 (17)	33
38	50 (36)	1654 (15)	1.4 (49)	2.6 (46)	1.2 (51)	2.1 (32)	33

13A	54 (33)	674 (35)	1.8 (40)	3.6 (13)	6.7 (13)	2.3 (19)	36
26	46 (41)	334 (45)	2.4 (24)	2.5 (50)	0.7 (53)	2.1 (39)	37
11	32 (51)	18 (55)	2.9 (6)	3.8 (7)	3.0 (44)	2.0 (46)	37
30	45 (42)	193 (52)	2.5 (19)	3.6 (12)	3.6 (39)	2.5 (8)	38
21	58 (31)	868 (32)	1.3 (51)	3.0 (33)	6.2 (19)	2.2 (28)	38
57	37 (48)	258 (51)	2.5 (16)	3.6 (11)	6.8 (12)	2.0 (49)	38
28	20 (55)	3059 (7)	1.0 (54)	1.5 (55)	1.3 (49)	3.1 (2)	39
15	40 (45)	501 (40)	2.1 (34)	3.9 (5)	7.9 (9)	2.0 (43)	40
34	31 (52)	350 (42)	2.3 (25)	3.3 (21)	3.0 (43)	2.4 (11)	40
62	48 (40)	263 (50)	2.1 (32)	2.7 (44)	5.8 (23)	2.1 (32)	41
65	39 (46)	1056 (25)	1.1 (52)	2.5 (51)	1.5 (48)	2.2 (25)	41
31	49 (38)	304 (48)	1.5 (46)	2.7 (42)	4.4 (35)	2.3 (20)	44
32	39 (46)	563 (38)	1.4 (48)	2.2 (54)	0.5 (54)	3.3 (1)	44
36	37 (48)	615 (37)	1.5 (47)	3.4 (18)	5.2 (29)	2.0 (43)	44
27	25 (54)	163 (53)	2.3 (28)	3.2 (26)	0.0 (55)	2.2 (27)	45
67	33 (50)	321 (46)	1.5 (45)	3.1 (28)	6.9 (11)	2.0 (48)	47
13	28 (53)	365 (41)	1.3 (50)	3.4 (20)	12.8 (3)	2.2 (24)	48

The values of the diversity indices vary from 1.04 (JDS 47) to 2.66 (JDS 27, area of Danube- Drava nature reserve), for the Shannon-Wiener index, while the number of species span from 17 (JDS38) to 38 (JDS 40). Overall the average rankling based on the diversity measures seems at least not to contradict to most of the fish indices (EFI, FIA and FIS). Unfortunately, for a number of sites no fish indices are available.

Macroinvertebrates

For the macroinvertebrate field monitoring/community line of evidence, in addition to the general diversity metrics three indices were selected for the analyses: the Average Score Per Taxon (ASPT) index (Armitage et al., 1983); the saprobic index and finally, as trait-based indicator, the $SPEAR_{pesticide}$ index.

An overall ranking of the 55 sites was achieved by calculating ranks per each of the 3 general diversity indices, in accordance with the fish community assessment. Many of the top-ranked site were from the upper Danube, e.g. JDS1, 2, 4, 5. However, also more downstream sites ranked high in average: JDS 40, 52, and 17. Site JDS38 (Belgrade metropolitan area), which ranked very bad in the fish community assessment, is for the macroinvertebrates in the lower part, but not as bad as for the fish. High scores for JDS 40 and 52 prove that the tributaries do not have significant impact on overall quality of the Danube, due to low, in most cases, insignificant discharge vs. Danube. The ranking of sites concerning macroinvertebrates did not significantly correlate with the fish –based ranking (spearman coefficient = 0.12; p=0.59). Worst sites for the macroinvertebrates are found in most parts of the Danube, e.g. the lowest ranked site is from the upper part (JDS13 - metropolitan area of Bratislava - SK, which is to be expected), but also sites JDS 62, 65 and 67 from the Danube delta and sites 31, 32, and 36 ranked very low. Low rank of the mid (31, 32, 36) and lower section sites (62, 65, 67) mostly comes as the result of

low ASTP scores, which is somehow expected. Those are typical sandy - muddy bottom lowland sites, under constant pressure of partly or untreated waste waters from a number of agglomerations along the river (JDS31 Ilok - CRO and Backa Palanka - RS, two cities at the same rkm, across each other; JDS62 metropolitan area of Braila, RO), with natural and pressure - enhanced dominance of Oligochaeta and Chironomidae. Overall the calculated metrics indicate a moderate quality for many of the sampled sites, since e.g. the maximum ASPT value was 5, and the highest SPEARpesticide value was 15.5.

4.1.6 Integration of the lines of evidence

The JDS data set consists of one of the most comprehensive data sets in terms of sampled variables, and JDS3 was probably the largest river monitoring campaign in the world. Huge effort has been put into the study, by the organisers, ICPDR, but also by many of the participating institutions. This large effort resulted consequently in a very comprehensive data set.

In the previous 4 sections the four LOE have been presented and discussed. In this section, the integration of the LOE into a common weight of evidence evaluation is presented. This will be achieved by following the 3 steps as outlined in section 3.3: Identification of the available LOE data, computation of quality classes for the single LOE, and compiling LOE matrices and interpretation of the results for all relevant sites.

Step 1: As outlined above already, macrophyte data will not be considered because of the earlier reported bias in the sampling of macrophyte communities during JDS3 (see e.g. Focks et al., 2015). Microorganisms have not been considered at all for the Danube case study, hence the focus for the integration will be on macroinvertebrates and fish. Results of LOE2 (effect-based fingerprinting) were reported in section 4.1.3, but appear too similar in the results as compared between the single JDS3 sites to be used for the integration exercise. In addition, the number of sites in the intersection between LOE2 and LOE3 was rather small, which would reduce the number of sites for which 'complete' lines of evidence could be built even further, since for only 9 of the 19 sites for which LOE3 results were available also LOE2 data are there (Table 4), despite LOE2 was tested for in total 22 JDS3 sites. Therefore, focus in this section will be on LOE1 (toxic pressure evaluation), LOE3 (*in situ* effects), and LO4 (community level) for fish, and on LOE1 and LOE4 for macroinvertebrates, since no *in situ* test results for macroinvertebrates were done in JDS3 .

After the collection of the number of available data per LOE and biological group, two input matrices for the decision matrix will be compiled:

- Matrix 1 for 19 JDS sites, for which 5 LOE entries are available (fish: LOE1, LOE3, LOE4, and macroinvertebrates: LOE1, LOE4).
- Matrix 2 for 32 JDS sites, where LOE1 and LOE4 are available for each fish and macroinvertebrates.

Table 8: Overview of the number of sampling sites for which data for the single LOE and biological groups are provided. LOE2 is not used and does not count for the intersection. EFI : European fish index, see 2.4.4.

	Fish	Macro invertebrates
LOE1	68	68
(LOE2)	22	
LOE3	19	-
LOE4	32 (21 EFI)	55
Intersection	19 (LOE1+3+4) 32 (LOE1+4)	55

Step2: The values in the single lines of evidence has to be translated into quality classes in order to allow for the use of the decision matrix (Table 13).

For that, the elements of the single LOE are discussed with respect to the possibility to interpret them in an absolute way, and threshold values or class borders are collected and defined where not available from the literature. Three classes are differentiated: CLEAR, MID, NONE, which relate to clear signals, moderate signal and missing signal for impairment.

LOE1 fish and macroinvertebrates

All evaluations in LOE1 are based on the sum of toxic units (STU). Following the rationale of Rico et al. (2016), the sum of toxic units are translated into classes of potential impact by the assumption of assessment or uncertainty factors. For macroinvertebrates, the toxicity ranking is based on experimental or QSAR values for *D.magna*, hence a safety factor of 100 for acute and 1000 for chronic effect appear appropriate for such transformation. For fish, the situation is very similar due to large uncertainty because many of the toxicity threshold have been calculated by baseline-QSAR. Therefore, the following class borders are suggested:

CLEAR:	-2 <	STU	
MID:	-3 <	STU	< -2
NO:		STU	< -3

LOE3 for fish: average biomarker response

As mentioned above (section 4.1.4), the aggregation of the single biomarker values was achieved by normalisation to the maximum so that no absolute interpretation of the aggregated indicator values (Table 4) can be achieved. The rationale to achieve translation of the quantitative information as

obtained in the average biomarker response (ABR) into the three quality classes is as follows. When the aggregated indicator value ranges at 0.5, that means an *average* response of 50% of the maximum response, which can be interpreted as a threshold to clear signals for all tested biomarkers. A lower threshold can be defined at 0.1, assuming that the *average* response was below 10% of the maximum response. Applying this scheme, the following class borders for the 3 quality classes can be derived:

CLEAR:	0.5 <	ABR	
MID:	0.1 <	ABR	< 0.5
NO:		ABR	< 0.1

LOE4 for invertebrates and fish

For fish and invertebrates, three general diversity measures have been taken into account: total abundance, species number and Shannon-Wiener index (SWI). The first two measures depend heavily on the sample size, so that no reasonable class borders can be suggested. For the Shannon-Wiener index, values above 2.0 are in general thought to indicate a good diversity and heterogeneity of a sample, hence we take 2.0 as the border between NO and MID , and consider values below 1.5 as being a clear signal. Summarised, that means

CLEAR:	1.5 >	SWI	
MID:	2.0 >	SWI	> 1.5
NO:		SWI	> 2.0

For macroinvertebrates, three indices were calculated: the saprobic index (SAI), the average score per taxon (ASPT) and the SPEARpesticide (SPE).

For the SAI, class borders are taken from conventional water quality assessment:

CLEAR:	2.7 <	SAI	
MID:	2.3 <	SAI	< 2.7
NO:		SAI	< 2.3

where already a good quality is interpreted as no signal.

For the ASPT, class borders are defined assuming that an average score of 5 indicates a good quality, hence no signal, and 3 is the transition to clear effects. Hence, for ASPT the suggested borders are:

CLEAR:	3 >	ASPT	
MID:	5 >	ASPT	> 3
NO:		ASPT	> 5

For the SPE, class borders are difficult to define, pragmatically the class borders are defined as follows:

CLEAR:	5 >	SPE	
MID:	10>	SPE	> 5
NO:		SPE	>10

Where the borders probably have to be readjusted, especially following the question whether a SPE value of larger than 10 can be interpreted as 'no signal'.

For the community status assessment of fish, the FIA (Fish Index Austria) is mainly sensitive to hydromorphological pressure, the FIS (Fish Index Slovakia) is sensitive to biological pressures and the EFI (European Fish Index) is designed basically to respond to overall water quality and eutrophication. Therefore, we used the EFI for classification of LOE4, together with the Shannon diversity index (SDI). Values for the EFI have been taken from the JDS3 report in form of classifications, hence no class borders need to be set. The EFI class boundaries are set as follows: High 0.669-1; good 0.449-0.669; moderate 0.279-0.449; poor 0.187-0.279 and bad 0-0.187.

In accordance with the overall policy of WFD, the 'one out – all out' principle is used as well for the integration of LEO. In concrete terms, that means if either the EFI or the Shannon Diversity index (SDI) indicates a clear signal, LOE4 is marked as signal. If both classes show no signals, LOE4 is marked as NO signal, else it is shown as intermediate.

4.1.7 Results of weight of evidence integration

Raw values have been compiled into the LOE (see Figure 13 and Figure 14 for the details). Using 5 LOE (LOE1, LOE3, and LOE4 for fish and LOE1 and LOE4 for invertebrates), results for 19 sites could be obtained, still with some gaps. This number is reduced in comparison to the overall 68 sites, due to the lower number of sites with fish biomarker data available. Focussing on matrix 1 first (Figure 5 left) and on the sites for which complete information is available, there are some sites, e.g. 38, 48, 53, where all 3 fish LOE show a clear response. For sites 38 and 53, this is corroborated by clear signals also in both invertebrate LOE. For these sites, scenario 1 of the decision matrix (Table 13) appears to fit best, and in consequences chemical pollution seems to be proven to cause ecological impacts and chemical-oriented risk mitigation might be required. A second block of sites consists of sites 40, 62, 66, and 67, where LOE and LOE2 show clear singles, but on the community level (LOE4) the signal is only moderate. Hence, scenario 2 of the decision matrix appears to be the closest choice: Effects on individual species are present, i.e. the investigated site(s) are close to the manifestation of an ecological impact from chemical pollution and it could be that community impacts not visible due to gaps in the data (e.g. seasonality). This means, at these sites special focus could be put on assessment of community level effects.

Figure 5: Weight of evidence matrices for the Danube case study. Left: matrix 1; 19 sites x 5 LOE, see evaluation above. Right: matrix 2; 32 sites x 4 LOE. The values indicate 2: CLEAR – clear signal, 1: MID-moderate signal, 0: NO – no signal. Empty boxes indicate missing values. For the derivation of the LOE values see Figure 13 and Figure 14 at the end of the document.

Matrix 1

Weight of evidence

	Fish			Invertebrates	
	LOE1	LOE3	LOE4	LOE1	LOE4
JDS50	2	2		2	1
JDS54	2	2			
JDS38	2	2	2	2	2
JDS48	2	2	2		
JDS53	2	2	2	2	2
JDS40	2	2	1	2	2
JDS62	2	2	1	2	2
JDS66	2	2	1	2	2
JDS67	2	2	1	2	1
JDS58	2	1			
JDS39	2	1	2	2	1
JDS47	2	1	2	2	2
JDS60	2	1	2	2	1
JDS27	2	1	1	2	2
JDS28	2	1	1	2	2
JDS31	2	1	1	2	2
JDS33	2	1	1	2	2
JDS36	2	1	1	2	2
JDS65	2	1	1	2	2

Matrix 2

Weight of evidence

	Fish		Invertebrates	
	LOE1	LOE4	LOE1	LOE4
JDS47	2	2	2	2
JDS49	2	2	2	2
JDS50	2	2	2	2
JDS53	2	2	2	2
JDS55	2	2	2	2
JDS56	2	2	2	2
JDS34	2	2	2	1
JDS41	2	2	2	1
JDS46	2	2	2	1
JDS48	2	2	2	1
JDS57	2	2	2	1
JDS58	2	1		
JDS28	2	1	2	2
JDS29	2	1	2	2
JDS31	2	1	2	2
JDS32	2	1	2	2
JDS33	2	1	2	2
JDS36	2	1	2	2
JDS37	2	1	2	2
JDS38	2	1	2	2
JDS39	2	1	2	2
JDS51	2	1	2	2
JDS52	2	1	2	2
JDS54	2	1	2	2
JDS35	2	1	2	1
JDS40	2	1	2	1
JDS42	2	1	2	1
JDS43	2	0	2	2
JDS27	2	0	2	1
JDS30	2	0	2	1
JDS44	2	0	2	1
JDS45	2	0	2	1

The next block of sites is built by sites 27, 28, 31,33,36,and 65, where LOE1 indicates a clear signal, but LOE3 and LOE4 only moderate response. This pattern points to scenario 7 of the decision matrix, which basically states that the situation is somewhat unclear. In similar way, all sites can be evaluated step by step and compared with the conclusions in the decision matrix. Two facts about matrix 1 (Figure 5) are striking, however: 1) LOE1 signals indicate a clear effect for all sites. This is certainly due to the choice of the thresholds for the evaluation of the sum of toxic units. With another choice of the thresholds, all sites might have shown moderate or no signals. The choice of a 'safety factor' of 100 appears, however, not overcautious. The finding that all sites rank similar in the LOE evaluation is not caused by the choice of the threshold values, but by the very similar total sums of toxic units (see Figure 3 and Figure 4). This reflects the exposure situation in the Danube, but is reason to wonder how such LOE could then be useful for the differentiation of chemical causation of ecological impacts in such study design. 2) In all lines of evidence signals indicate moderate to clear signals. Here, it appears that the either the ecological and exposure situation in the complete Danube is serious, or that the choice of sampling sites could profit from local adaptation, e.g. finding local upstream/downstream settings where differences in e.g. toxic pressure could show larger differences than when sampling sites are far apart from each other as in case of the JDS3.

Considering matrix 2, with 4 LOE values for 32 sites, the situation is similar. For all sites, the LOE1 for both fish and invertebrates indicate clear signals. Those LOE1 signals are at the community level confirmed for sites 47, 49, 50, 53 55, 56, 34, 41,46,48 and 57, hence falling into case 1 of the decision matrix. For another set of sites (58,28, 29, 31, 32, 33, 36, 37, 38, 39, 51, 52, 54, 35, 40, 42), LOE4 for fish shows moderate signals, hence indicating that more information is needed . At least for some sites, i.e. 27, 30,43, 44, 45; LOE4 shows no signals, what falls into case 8 of the decision matrix. Consequently, for these sites, it appears that chemical pollution does not cause ecological impacts at the moment. What has not been considered neither in this conclusion nor in the toxic pressure assessment is the question of bioavailability of compounds, which can make a difference as some of the evaluated metals could not be corrected for bioavailability.

4.1.8 Summary and conclusion of the toolbox development and the Danube case study

The JDS3 data set provides a tremendous richness of data. Parts of the data that were produced or analysed by Solutions partners were already published (e.g. Deutschmann et al., 2016; Rico et al., 2016), but since an integrative evaluation of data in the whole data width and depth is challenging, a systematic approach needed to be developed. With the developed toolbox approach, including the definition of the average biomarker response (section 3.2) and the systematic combination of the LOE (section 3.3), a compromise was found between the complexity that is necessary to evaluate such data set and on the other hand the applicability and comprehensiveness of such approach, especially for non-scientist who work in national or regional water boards. The work on the toolbox is not yet finished, as it is intended to publish a scientific paper on its development and application for the Danube data set, including Spreadsheet calculation templates and R-scripts. By application to the Danube data set, the

toolbox approach could prove its practicality, simplicity and stringent definition, since it was possible to transform the multitude of data into comprehensible matrices (Figure 5), which summarise the overall evaluation without losing too much precision. The stringent definition and clear documentation of all defined steps provide transparency and enables an understanding of the results.

Considering the results of the Danube case study weight of evidence evaluation, it was possible to subdivide the sampling sites into classes of similar effect patterns, and to associate interpretations from the decision matrix. Also for such WOE evaluation, the results from the Danube showed the before mentioned 'flat' profile, meaning that the differences between the sites were not very pronounced. This is the drawback of such large-scale expedition like the Joint Danube Survey, which on the other hand provides an unseen richness of details for the restricted number of sampling sites. A suggestion for a next Danube survey would be to try to find for a selection of sampling sites local upstream-downstream settings, for example for important tributaries or for known point sources of pollution. This might help to combine the strength of such large scale expedition with the advantages of more focused, local experimental set-ups.

The toolbox application resulted in the identification of a number of sites where all LOE indicate impairment, from predictive toxicity modelling over biomarker responses up to community level indicators. In total, the picture emerged that many of the Danube sampling sites show clear anthropogenic impacts, and in all of them the toxic pressure suggests toxicants as potential cause. In that context, the biomarker response (LOE3) for half of the sites (see Figure 5, sites 58, 39, 47, 60, 27, 28, 31, 33, 36, 65) indicate that the link from toxic pressure to community effects is not always as clear as it might appear from only linking chemical pressure to community effects (e.g. for sites 39, 47, 60). Here, the biomarkers and their aggregation in form of the ABR show their potential to add another aspect to the overall evaluation of the chemical and ecological quality of water bodies.

For the compilation of the LOE results, some uncertainties are still remaining. First, concerning the mixture toxicity modelling, the summation of toxic units is a worst case approach that doesn't need to prove effectivity in practice, because not all compounds will add their effects to each other. Also the choice of the extrapolation factors from threshold values to community (100 resp. 1000 for acute and chronic effects) is a choice that can be challenged. On the other hand, the threshold values appear not too conservative when considering the QSAR-to-field and species to species extrapolation that is beneath the STU values. Another issue remains in the evaluation of metal toxicity. Macroinvertebrate toxicity threshold have been corrected for bioavailability for copper, lead and zinc (Rico et al., 2016), but for fish that was not done, hence it might be that the STU overestimate the metal toxicity. Calculation of the average biomarker responses (ABR) was hampered by missing references for the single biomarkers. Using the 'ratio to max' approach, reasonable values could be obtained, but it would be preferable to have either a clear reference site (e.g. in an upstream/downstream setting) or defined ranges of standard responses of the biomarker tests. Finally, for the evaluation of community level impairment, the choice of threshold values for transformation of diversity measures into effect classes is a big challenge.

4.2 Case Study Rhine[7]

The presented study used two lines of evidence:

1) the analysis of toxic unit distributions, based on standard ecotoxicity data from simple, single-species algal bioassays (see e.g. Gustavsson et al., 2017 for an example of this approach), in order to identify drivers of mixture toxicity and in order to analyze whether the recent STP upgrade at one of the sites was successful (see below), and
2) *in situ* experimentation with complex algal communities based on the concept of pollution-induced community tolerance (PICT) (see above, and Tlili et al. 2016), in order to explore the ecological impacts of the pollution scenarios found.

The PICT method relies on differences in the pollution tolerance of the different species that make up a community, lead to shifts in the community structure under chronic toxicant exposure. As a consequence, a community that was previously affected by an exposure to chemical pollution displays a lower sensitivity to those pollutants than a reference community that has never been exposed. Community tolerance to micropollutants is usually quantified by comparing the responses of physiological endpoints of the reference community with the chronically pre-exposed community (Tlili et al. 2016).

Increased community tolerance to single micropollutants has been demonstrated with periphyton, a consortium of microorganisms that grow on submerged substrata surfaces and which plays a crucial ecological role in aquatic ecosystems as a basis for the food-web (Battin et al. 2016). However, field studies examining tolerance of periphyton to micropollutant mixtures remain rare (Pesce et al. 2011, Tlili et al. 2016). In a recent study, Tlili et al. (2017) assessed periphyton tolerance to micropollutant mixtures extracted from passive samplers that have been deployed at multiple wastewater-impacted streams. Results from this study show that periphyton collected downstream of a WWTP discharge point has a higher tolerance towards the extracts than periphyton sampled upstream of the wastewater discharges. Most importantly, the study showed that the proportional increase of tolerance from upstream to downstream was strongly correlated to the intensity of contamination by micropollutants at the respective sites. These findings support the notion that PICT can be used as an effect-based tool, in combination with passive samplers, to investigate whether a site-specific exposure pattern causes ecologically relevant shifts in microbial diversity. This applies not only to assess impacts but also to monitor the recovery of impacted streams following for instance the upgrading or removal of the WWTPs (Tlili et al. 2016).

In Switzerland, the water protection act entered into force in March 2014. As a results, WWTPs are currently upgraded by applying additional treatment steps such as ozonation or powder-activated

[7] Parts of this work are currently submitted for publication as "Tolerance of stream biofilms to anthropogenic chemicals indicates causality and reflects ecosystem recovery" by Ahmed Tlili, Natalia Corcoll, Åsa Arrhenius, Thomas Backhaus, Juliane Hollender, Bettina Wagner and Renata Behra. A second paper based on the toxic unit analysis is currently (August 2017) under preparation and is expected to be submitted during autumn 2017.

carbon filtration to reduce the input of micropollutants into aquatic ecosystems (Eggen et al. 2014, Stamm et al. 2016). This development offered a unique opportunity to study the ecological consequences of substantially reducing the micropollutant loads from WWTP effluents. The main goals of this study were therefore (a) to analyse how powerful an analysis of toxic units is for identifying drivers of ecological mixture toxicity, and (b) to examine tolerance to micropollutant mixtures of *in situ* periphyton that has been sampled from one upgraded and two non-upgraded WWTPs. We hypothesised that following the upgrading of the WWTP, tolerance to the micropollutant mixtures of upstream and downstream periphyton will be similar. In contrast, an increased tolerance should be visible downstream of discharges from non-upgraded sites.

4.2.1 Study setup

This study was carried out from the 15th of March to the 30th of April 2016, upstream and downstream of three WWTPs located in north-eastern Switzerland and named Herisau, Buttisholz and Hochdorf. Those WWTP discharge points are all located in tributaries to the river Rhine. The sampling sites were previously investigated during a survey that has been conducted from the 15th of March to the 30th of April 2014 and designed to assess effects of micropollutant mixtures from the wastewater discharges on periphyton (Tlili et al. 2017). The WWTP at Herisau was upgraded in June 2015 with powder-activated carbon filtration, whereas no modifications occurred at Buttisholz and Hochdorf. This situation offered us the opportunity to compare the ecological impact of the chemical mixtures present in the WWTP effluent before and after the WWTP upgrade.

In order to compare the results of the present study with the earlier one performed in 2014, the experimental design that was used by Tlili et al (2017) was used basically unchanged. In short, 6 week old periphyton communities were sampled from locations up- and downstream of the three WWTP discharge points. In parallel, Chemcatcher® passive samplers – styrenedivinylbenzene (SDB) discs – were deployed at each discharge site of the WWTP to accumulate polar organic pollutants.

Chemicals in the organic extracts from the passive samplers as well as the pollution in the water were characterized by means of GC/MS. Biofilm biodiversity was analyzed via DGGE fingerprints based on PCR amplification of the algal 18S rRNA gene fragments (Tlili et al., 2008). Tolerance of the harvested periphyton to (i) the extracts from the passive samplers (for comparative purposes, periphyton was also exposed to extracts already collected in 2014, i.e. before the WWTP at Herisau was upgraded), (i) Diuron and (ii) a mixture of 8 PSII inhibitors was characterized as the EC50 values determined after 12 hours of exposure. The uptake of radiolabeled C14 as a measure of primary production was used as the endpoint to characterize the tolerance development of the algal part of the community. Bacterial productivity was measured as the incorporation of 14C-leucine into protein according to Buesing and Gessner (2003).

In order to characterize the toxic unit distribution up- and downstream of the STP discharge point, ecotoxicological data on the toxicity of the identified chemicals were collected from the US EPA data collection at www.epa.gov/ecotox. Several thousand data were initially retrieved for the 57 compounds included in the monitoring profile. These data were then confined to estimates on effects on growth, reproduction of defined species of eukaryotica algae, exposed for 1- 7 days. Only 276 relevant data were

finally included in the calculation of the toxic units, providing estimates for 30 of the 57 compounds. Data for the remaining chemicals were first filled by manually searching through (i) the peer-reviewed literature and (ii) grey reports, including safety-datasheets. Data for an additional 24 chemicals were found in these sources. QSAR estimates (from ECOSAR, vers. 1.11) were used for the remaining 3 chemicals (Amisulpride, 5-Methyl-2H-benzotriazole, Oxazepam). It should be pointed out here, that none of those three chemicals were identified as relevant (i.e. potentially contributing substantially to the mixture risk). That is, the uncertainty in those rough estimates does not impact the final assessment, even if it would be assumed that the QSAR estimates are off by 2 orders of magnitude.

4.2.2 Site characterization: chemical-analytical profiles

57 compounds were monitored at the sites. Complex chemical mixtures were found at each of them, with between 10 and 50 compounds being present simultaneously at concentrations above the analytical level of quantification. Clear concentration differences were detected between up- and down-stream sites (Figure 6). Unfortunately, the water samples from the Buttisholz discharge point were partially destroyed during sample storage. The analytical profiles from this site are therefore only semi-quantitative and will not be further discussed.

4.2.3 Line of Evidence 1: Analysis of toxic unit distribution

A toxic unit (TU) is simply the ratio between the concentration of a chemical and its toxicity to the organism or community of interest. It provides a convenient dimensionless measure of the absolute toxicity load that an organism or an ecological community experiences, as a result from being exposed to certain concentration of a pollutant. In the present study the TU was operationalized as the ratio between the site-specific concentration of a compound and its average EC50 in standard single-species algal bioassays (see details above). The sum of toxic units (STU) provides an estimate for the total toxicity load being present at a site (ignoring any contribution of chemicals not included in the monitoring profile, see discussion in chapter 1, above).

Figure 6 shows the toxic unit distribution up- and downstream of the discharge points at Herisau and Hochdorf, respectively, for the years 2014 and 2016. The successful upgrade of the STP in Herisau is clearly visible in Figure 6: in 2014 (prior to the STP upgrade), the sum of toxic units (STU) is 1.8×10^{-3} upstream and 4.9×10^{-3} downstream the STP discharge point. In other words, the total concentration of the 57 compounds monitored at those two sites equals 0.18%, respectively 0.49% of the mixture EC50 estimated by Concentration Addition. The effluent adds a total toxic load of 3.1×10^{-3} TU to the stream.

With 1.7×10-3, the STU measured upstream the discharge point in 2016 is basically identical to the value of 2014 (1.8×10-3). However, in 2016 the downstream STU is now reduced to only 2.1×10-3. That is, the STP discharge only adds a minute TU of 0.3×10-3 to the stream, which is a reduction by a factor of 10 compared to the situation in 2014.

A closer look at the TU distribution in Figure 6 reveals that especially Clarithromycin, a particularly algal-toxic antibiotic commonly used in human medicine to treat skin and respiratory tract infections, is particularly affected by the STP upgrade.

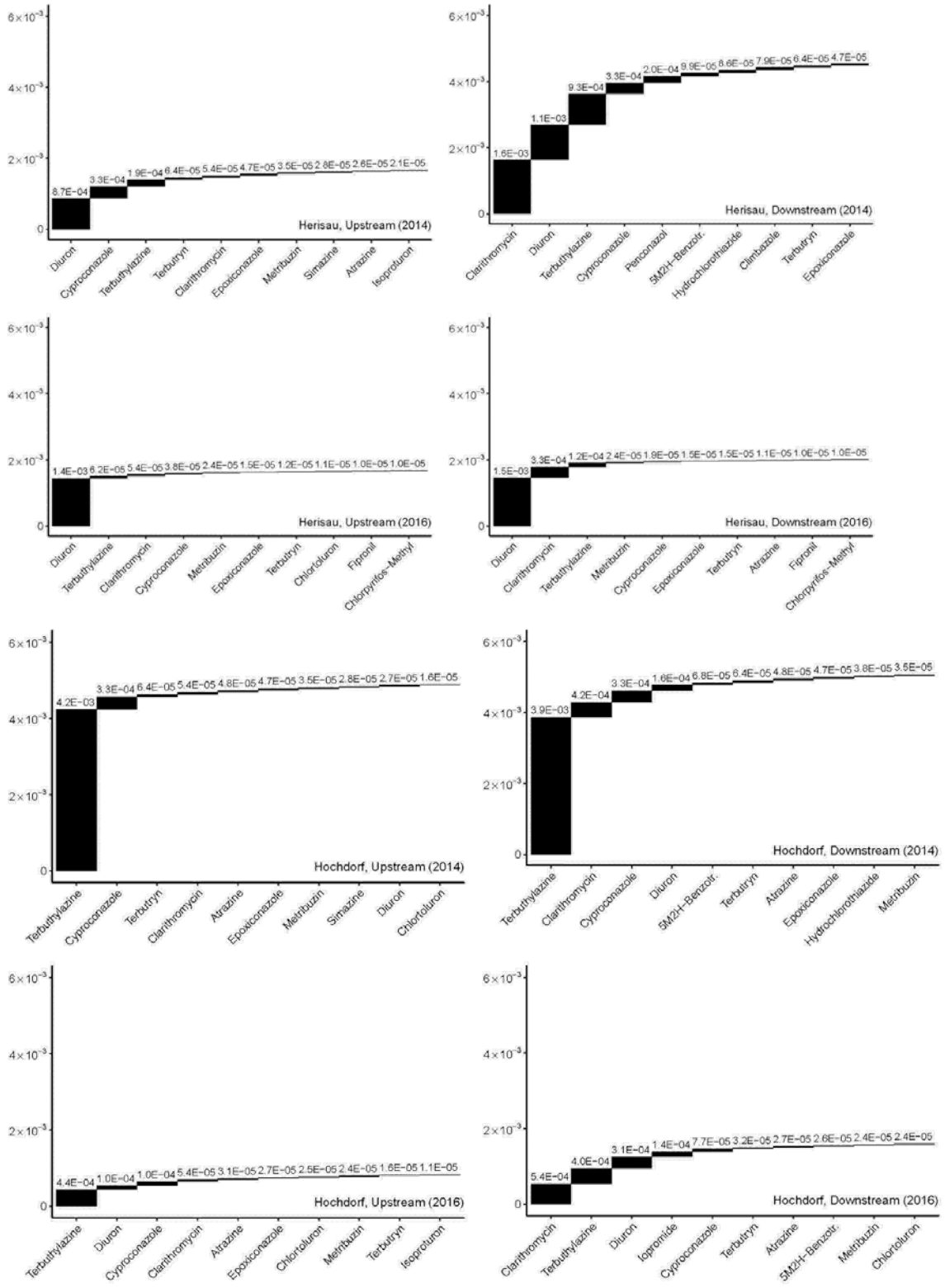


Figure 6: Distribution of toxic units up- and downstream of the STP discharge point at Herisau and Hochdorf in 2014 and 2016

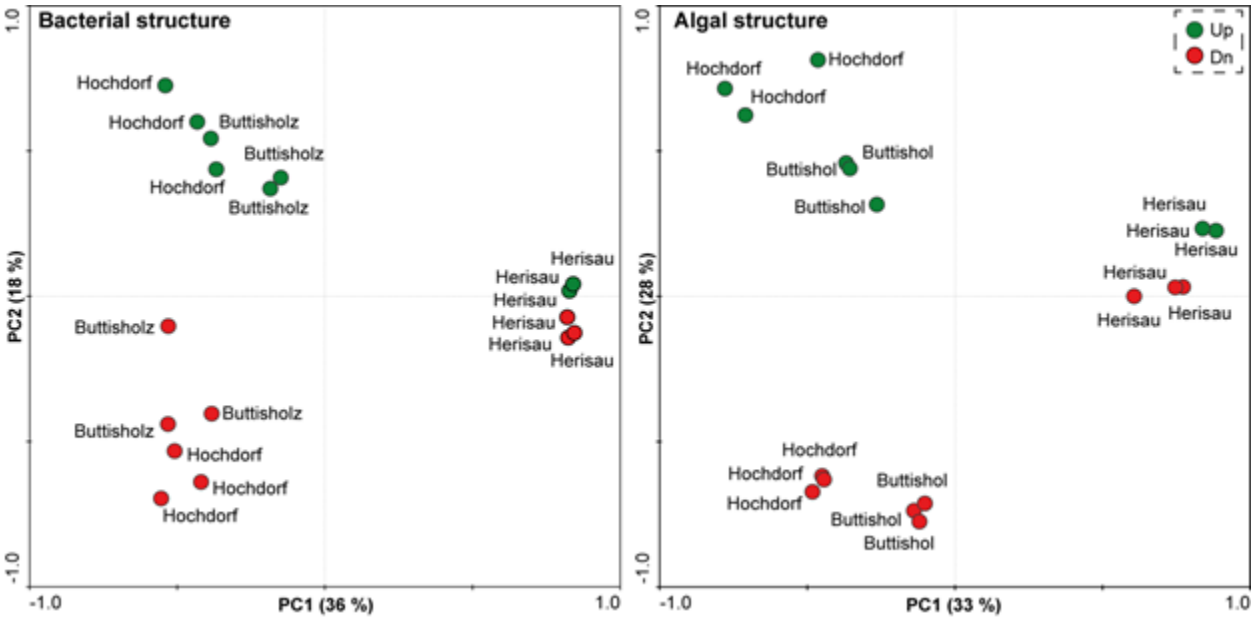


Figure 7: Microbial biodiversity as determined via DGGE fingerprints up- and downstream the selected WWTP sites

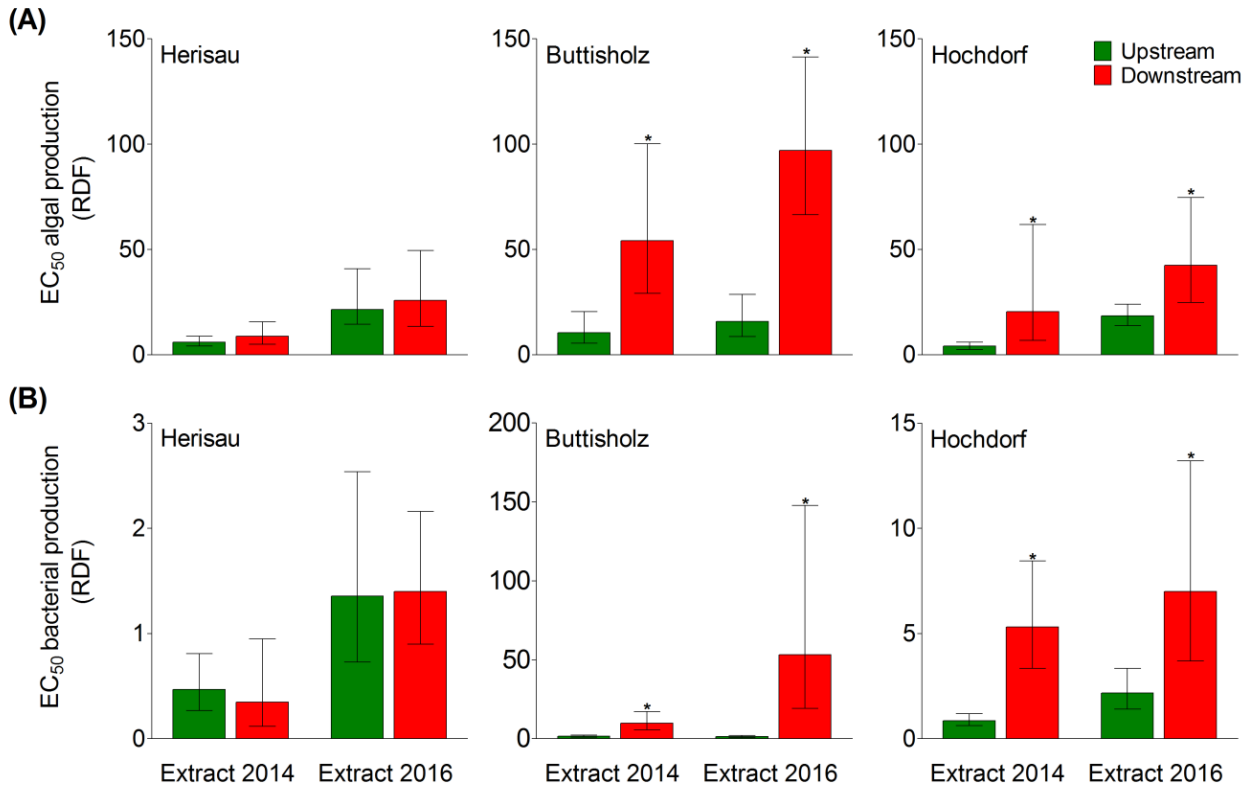


Figure 8: Tolerance pattern of periphyton up- and downstream the three WWTP discharge points

While the compound is present at a TU of 1.6×10^{-3} downstream the STP discharge point in 2014, it only adds a TU of 3.3×10^{-4} to the river in 2016, which is a reduction by a factor of almost 5.

Diuron and Terbuthylazine, two common pesticides and biocides from the group of photosynthesis-inhibiting herbicides, are two other mixture toxicity drivers. The Diuron TU is almost constant at all sites in both years (TU between 8.7×10^{-4} and 1.1×10^{-3}), which clearly indicates the compound is a water pollutant present in the stream already upstream the STP discharge site. In contrast, Terbuthylazine shows a pattern that is very similar to the one observed from Clarithromycin.

Terbuthylazine is also a mixture toxicity driver at the Hochdorf site, completely dominating the TU distribution in 2014 at both sites and upstream in 2016. The only reason why the compound does not also dominate the TU distribution of the water sample taken downstream in 2016 is because of a similarly high TU by Clarithromycin, which is obviously present in the wastewater stream entering the river at the Hochdorf site. The samples taken in 2014 and 2016 clearly differ in their total STU, which is caused by the different TU contributions of Terbuthylazine, which is already present at high loads in the water upstream the discharge point in 2014. This most likely reflects spraying events in nearby agricultural areas just prior to the sampling time.

In summary, the analysis of toxic unit distributions identified two mixture toxicity drivers (Terbuthylazine and Clarithromycin) amongst the 57 measured chemicals, and it provided clear evidence of the successful upgrade of the STP plant at Herisau. However, given that the TU distribution is based on ecotoxicity data from simple single species tests, this analysis does not directly allow us to directly infer on the ecological consequences of the pollution being present. The critical questions in this context are: (1) does the seemingly low STU of only a few percent of the estimated EC50 lead to ecologically relevant changes in the biodiversity at the investigated sites, and (2) does the STP upgrade lead to tangible, ecologically relevant improvements? These two questions were analyzed in a second line of evidence, using *in situ* periphyton experiments.

4.2.4 Line of Evidence 4: Results of in situ periphyton experiments

Principal component analysis applied to the DGGE data (Figure 7) showed that for algal and bacterial structures axes PC1 and PC2 together explained more than 61% and 54 % of the total variability, respectively. For both algae and bacteria, PC1 is related to the sampling location, separating Herisau from Buttisholz and Hochdorf. PC2 is correlated to the influence of the wastewater discharges on microbial community structure. All upstream periphyton were clearly separated from downstream periphyton at Buttisholz and Hochdorf along PC2, while differences were less pronounced at Herisau, the site at which the WWTP was upgraded in 2015. According to PC2, algae in upstream and downstream periphyton from Herisau were more similar to upstream than to downstream periphyton from Buttisholz and Hochdorf.

One year before upgrading the WWTP at Herisau, micropollutant concentrations at Herisau were 16 times higher downstream than upstream of the wastewater discharges, and algae in downstream periphyton show a clear pollution-induced tolerance (Tlili et al. 2017). Today, roughly one year after the

WWTP upgrade, the tolerance induction is gone: following short-term exposure assays to the passive sampler extracts, either obtained in 2014 or 2016, upstream and downstream of Herisau displayed similar EC50 values based on algal and bacterial production measurements, indicating no induced-tolerance at this site (Figure 8). Also the chemical-analytical fingerprints confirm the efficiency of the employed powder-activated carbon to retain substantial amounts of micropollutants.

Interestingly, critical physicochemical factors such as temperature, nutrient concentrations and organic matter continue to be higher downstream the Herisau effluent plume, which provides further support to the notion that the disappearance of algal tolerance was specifically due to the substantial decrease of micropollutant concentrations downstream.

In sharp contrast, downstream periphyton from Buttisholz and Hochdorf displayed significantly higher tolerance to the chemicals present in the passive sampler extracts, as compared to the periphyton sampled upstream. This is clearly reflected in the higher EC50 values recorded for downstream periphyton (Figure 8). However, depending on the used extract and measured endpoint, the increase of tolerance (i.e. the ratio between the EC50 values) from upstream to downstream was different between and within the two sites. On the one hand, algal tolerance to the extracts from 2014 and 2016 increased by 5 and 6 times at Buttisholz, respectively, and by 5 and 2 times at Hochdorf, respectively. On the other hand, bacterial tolerance at Hochdorf increased similarly by 6 times to both extracts but by 6 and 38 times to the extracts from 2014 and 2016, respectively.

4.2.5 Summary and Conclusions

In summary, our results from the non-upgraded sites correspond well to the earlier study by Tlili et al. 2017). They show that periphyton tolerance at the non-upgraded sites (i.e. Buttisholz and Hochdorf) is quite consistent over time and that the toxicants detected in the analytical survey indeed shape biodiversity and tolerance of exposed microbial communities.

The results recorded up- and downstream of Buttisholz furthermore provide the opportunity to shed some light on the relation between simply ecotoxicity and tolerance development. Figure 8 shows that the extracts of 2014 and 2016 have the same toxicity to the un-adapted up-stream communities.

However, despite this seemingly identical toxicity, the periphyton collected downstream of the Buttisholz discharge point is far better adapted to the 2016 extract than to the 2014 extract. It is still somewhat tolerant to the 2014 extract, i.e. the recorded EC50 is higher downstream than upstream. But the tolerance to the 2016 extract is clearly higher. That is, the downstream communities are definitely better adapted in general to handle pollution than the upstream communities. However, even those adapted communities are somewhat “surprised” if suddenly exposed to a new, different mixture, i.e. the 2014 extract. This clearly shows that the ecological consequences (the impact on microbial diversity) goes beyond mere summary ecotoxicity estimates, such as simple EC50 values recorded for the extract of a site.

A similar analysis could not be performed for the Herisau and Hochdorf sites, as the toxicities of the 2014 and 2016 extracts are different already for the un-adapted communities (in contrast to the

Buttisholz site, where the toxicities of both extracts was identical). The study explored and compared two of the four possible lines of evidence (LOE) (Figure 1), i.e. the TU analysis and the community, *in situ*, experimentation. In this case study both focused on effects on micro-organisms.

Following the decision matrix of Table 13, the results from both LOE's go together almost perfectly: while LOE2 shows that there are indeed ecologically relevant effects that are caused by the pollution cocktail present at the sampling sites, LOE1 pinpoints to the toxicity drivers (Terbuthylazine and Clarithromycin). It can therefore be concluded that chemicals (and not non-chemical stressors) are a cause for ecological changes in the sampled river systems.

Additionally, the study indicates that the mere toxicity evaluation, even if done *in situ*, might not be sufficient for characterizing chemical impacts in the environment. Additionally, adaptive processes warrant increased attention.

4.3 Case study Holtemme

The present study used two different lines of evidence to analyse whether pollution- induced impacts on aquatic organisms due to anthropogenic activities can be detected:

1) the analysis of toxic unit distribution, based on acute ecotoxicological data (EC50 values) from Busch et al. (2016) for algae, daphnids and fish and chemical data of Inostroza et al. (2016a)
2) *in situ* results of biomarker response in feral fish (*Salmo trutta*) and genetic structure and body burdens of *Gammarus pulex* (Inoztroza Inostroza et al., 2016b)

The river Holtemme, with a total length of about 47 km, is located in the Bode catchment area (Saxony-Anhalt, Germany) and is part of the TERENO observation platform. TERENO spans an Earth observation network across Germany that extends from the North German lowlands to the Bavarian Alps. This unique large-scale project aims to catalogue the long-term ecological, social and economic impact of global change at regional level. Scientists and researchers want to use their findings to show how humankind can best respond to these changes[8].

The source of the Holtemme is located in the mountain brook in the conservation area Harz with a high water quality, before becoming an increasingly polluted and channelized lowland river (Inostroza et al., 2016b). The Brocken Mountain partially consists of siliceous bedrock without any large buffer capacity. The Holtemme river catchment (277.8 km^2) is characterized by semi-natural forest in the upstream sections, and agricultural areas and medium-sized towns in the central and lower sections. Effluents of two wastewater treatment plants (WWTPs) of the medium-sized towns Wernigerode and Halberstadt serving approximately 150 000 inhabitants, together with agricultural landuse, represent the main source of pollution (Reuter et al., 2003). Due to its remarkable gradient of anthropogenic influences with clearly defined pollution sources, the Holtemme River represents typical features of Central European rivers in a close proximity (Inostroza et al., 2016) and was therefore implemented into the SOLUTIONS framework as a case study site. The Holtemme river receives significant loads of treated wastewaters,

[8] http://teodoor.icg.kfa-juelich.de/overview-en?set_language=en

together with other impacts such as agriculture, road runoff and small crafts and industries (Inostroza, 2016).

4.3.1 Study setup

In contrast to the Danube case study which investigated a long river section with high dilution effects, the Holtemme case study was designed as a study in a smaller catchment area, to test the hypothesis that in smaller catchment areas, a WOE concept using a large set of LOEs is suitable to give comprehensive insights into the effects of multiple stressors in aquatic systems. In terms of biomarker response analysis at the Holtemme river, a comparable battery of biomarkers as used during the Joint Danube Survey 3 (JDS3) was utilized to assess the impact of specific point sources. In contrast to the Danube, where pollution sources were often unknown, or high dilution effects occurred, the Holtemme River with its typical depth of 10 to 30 cm and the turbulent flow of the water column may be considered homogenous (Inostroza, 2016).

Therefore, sub-individual responses of sentinel fish where used to assess the impact of WWTP effluents and agricultural land use on aquatic organisms. Because of its high water quality in the upper stretch in the source area and differences in land use pattern (Figure 9) the basis for a reference site was given.

The sampling took place in October 2014 and was organized and conducted by the Helmholtz-Centre for Environmental Research (UFZ) in Leipzig and Magdeburg, Germany. The Institute for Environmental Research,RWTH Aachen Universtiy, Germany, contributed by collecting tissue and blood samples of brown trout (*Salmo trutta*). The comprehensive data collection consisted of biological and chemical information such as chemical target screening data, *in vitro* assay data and ecological community data of invertebrates and fish. Chemical data and data of genetic erosion and body burden as proxy for pollution-induced chemical stress in *Gammarus pulex* are available from a study by Inostroza et al. (2016a).

For *in situ* biomarker investigations samples of brown trout (*Salmo trutta*) were collected at four different sites along the river Holtemme (reference site, study site 2, 3 and 4) and conserved for further processing in the laboratory. Figure 10 gives an overview of the study sites in the Holtemme. The reference site (RS) was located in the forest area 2.7 km downstream of the source. Site 2 and 3 were located directly downstream of WWTPs of Wernigerode and Halberstadt, respectively. Site 4 with highest percentage of agricultural influence, was located approximately 38 km downstream of the source near the town Nienhagen (Saxony-Anhalt).

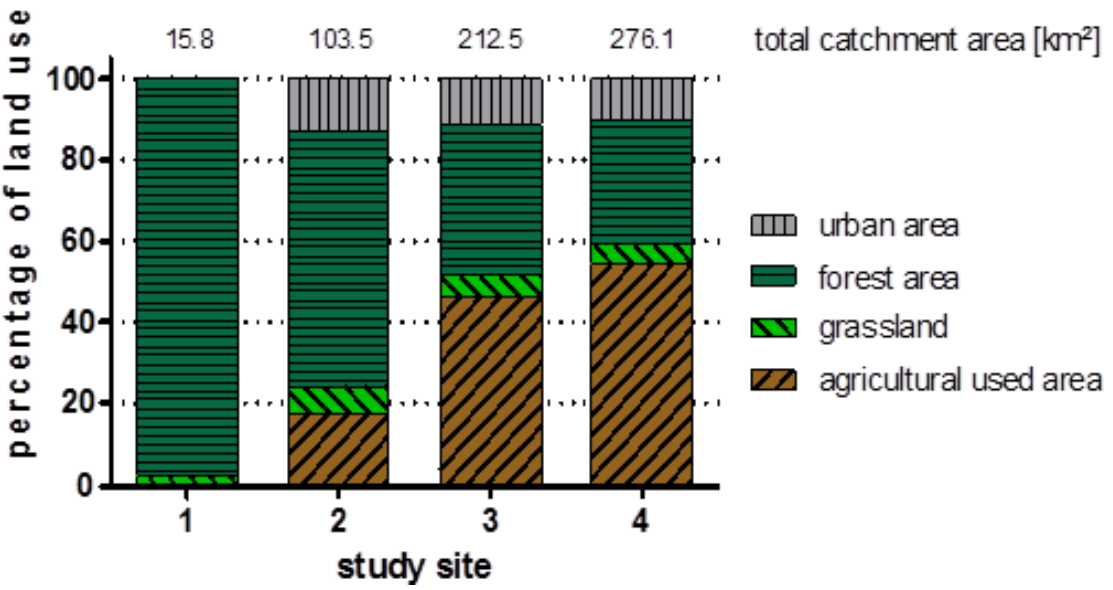


Figure 9: Landuse in the catchment area of the river Holtemme. Land use patterns expressed for each study site as a Different utilizations are stated for each study site in percent of the total catchment area noted above each plot.

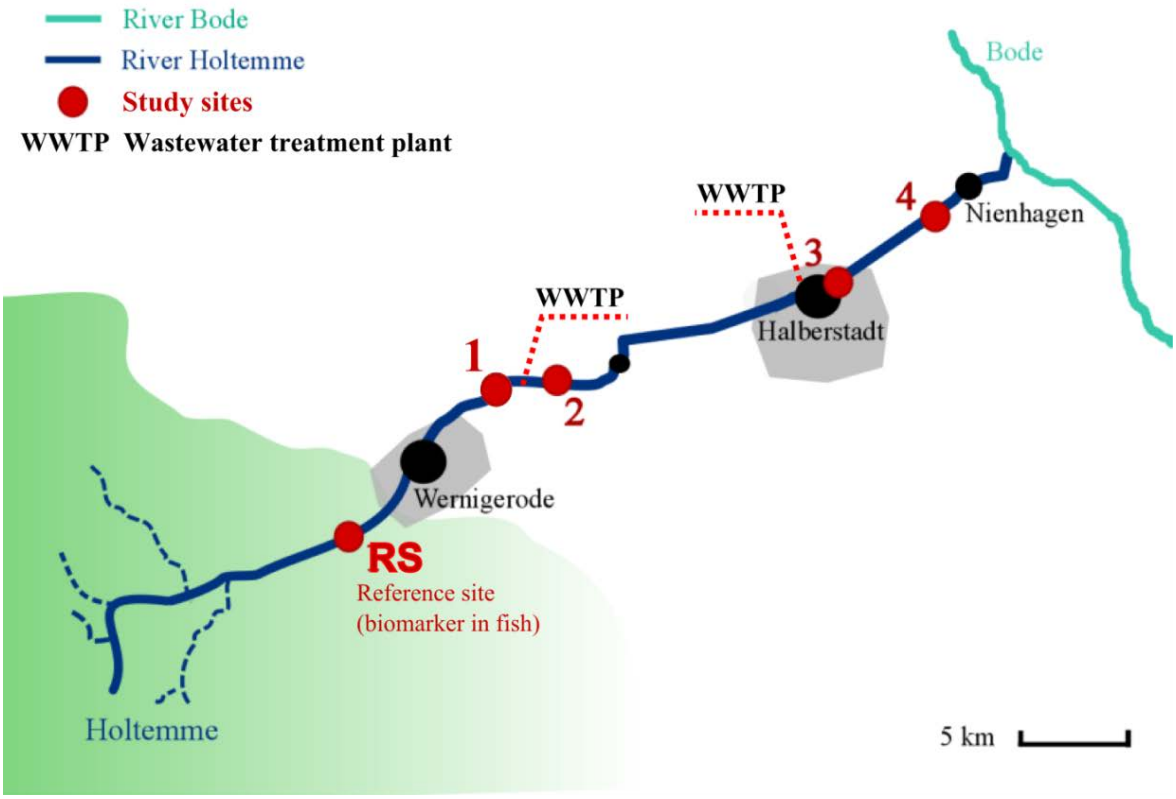


Figure 10: Overview of the sampling sites for the river Holtemme case study.

Between the sites 3 and 4 two weirs built anthropogenic barriers for migration of organisms, and a sink for pollution from the water phase into sediments. Location 1 was located approximately 15 km downstream from the source and upstream of the first WWTP and was used for chemical analysis of water samples and *G. pulex*. Unfortunately, for site 1 no biomarker data are available and RS provides no chemical data.

Fish species for the study of *in situ* biomarkers were selected based on the occurrence at the sampling sites. Brown trout (*Salmo trutta*) is the dominant fish species in the Holtemme river.

The following biomarker analyses were applied:

- Genotoxicity / DNA damage: blood samples - Micronucleus test
- Endocrine activity (17β-estradiol and testosterone concentrations): blood samples - Enzyme-linked immunosorbent assay (ELISA)
- Enzyme activities – liver samples
    - Activity of Phase I biotransformation enzymes:
        - ethoxyresorufin-O-deethylase - EROD
        - carboxylesterase - CES
    - Activity of Phase II biotransformation enzymes:
        - glutathione-S-transferase - GST
    - Oxidative Stress: catalase - CAT
    - Oxitative Stress/cell damage - TBARS
    - Neurotoxicity : acetylcholinesterase – AchE inhibition

In order to identify drivers of mixture toxicity, toxic units (TUs) were derived by using on-site substance concentrations and their lower 5% percentile of acute EC50 values. Toxicity values were obtained from Busch et al. (2016) for the different organism groups. If data were missing (namely Cyclamate, NAAP and Quinmerac) acute EC50 values were estimated using QSAR (ECOSAR, v1.1) for each compound and organism group, respectively.

4.3.2 Site characterization: analytical profile

Grab water samples collected during the key date sampling campaign in October 2014 were analyzed using LC-MS/MS analysis. At the sites corresponding to in situ data of fish up to 47 substances were detected. The compounds belonged to different classes of pollutants such as pesticides, pharmaceuticals, industrial chemicals and some other main transformation products. Table 9 lists the detected concentrations at the relevant sampling sites. Numbers and concentrations of anthropogenic substances clearly increased downstream of the first WWTP for all investigated downstream sites.

4.3.3 Line of Evidence 1: Analysis of toxic unit distribution

As mentioned in chapter 5.3.3. a toxic unit (TU) is the ratio of a concentration of a chemical and its toxicity to an organism or community of interest. The sum of TUs were calculated on the basis of three different approaches: (a) Assuming that all non-detects (i.e. concentrations below the LOD) are actually present at exactly the LOD. This is the worst case assumption that is still in line with the empirical data (highest sum of toxic units). (b) Assuming that all non-detects (i.e. concentrations below the LOD) are actually not present at all. This is the best case assumption that is still in line with the empirical data (lowest sum of toxic units). (c) Estimating the average toxic unit by the Kaplan-Meier approach. This approach estimates the average toxic unit in case an actual value is missing. Further details are given in Gustavsson (2017). This method constitutes the most realistic case which makes best use of the available information. The order of the three different sums of toxic units (STU) is always STU(a)>STU(c)>STU(b). Since differences of all of the three approaches were absolutely minute for most sites and organism groups this study focused on approach (b) assuming that all non-detects are set to zero. Figure 11 shows the distribution of the STUs for each organism group along the river course. For all organism groups and the "most sensitive tropic level" (MST; sum of the highest TU for each compound selected from the three organism groups fish, algae and dapnids) there was a clear increase of the STUs at the sites downstream of the first WWTP. This was particularly demonstrated for the organism group of daphnia at the sites 3 and 4. For fish and algae this effect was less clear but also for these organism groups the total concentration of the 48 detected compounds at site 4 equals more than 2% of the mixture EC50 estimated by concentration addition. For daphnia up to 17% of the effect were covered by the monitored micropollutants at site 3 downstream of the second WWTP.

The data behind the STU (Figure 11) suggests that overall toxicity within an organism group is mostly related to one main driver. For the algae the mixture toxicity at site 1 was driven by the metabolite of the herbicide Atrazine (Desisopropylatrazin; $5.2x10^{-3}$), toxic effects on daphnids and fish were negligible in comparison. The toxic unit distribution for algae and MST are therefore basically identical. For the algae the sites 2-3 were more equally distributed. Anyway, also at these sites herbicides (or their metabolites) contributed most to the overall toxicity. The STU for the daphnids was clearly driven by insecticides. At site 2 Fipronil was present at a TU of $1.3x10^{-2}$. At site 3 and 4 the highest TU can be related to the insecticide Diazinon with $1.7x10^{-1}$ (site 3) and $1.3x10^{-1}$ (site 4). In other words, Diazinon represents 17% and 13% of the mixture toxicity estimation, respectively. Although the STU for fish was relatively low the pharmaceutical Diclofenac with a TU up to $1.2x10^{-2}$ was the main toxicity driver of the STU.

Table 9: Detected organic micropollutants in grab water samples of the Holtemme River (concentrations in ng L-1). Method detection limits (MQLs) in ng L-1.

			Study sites			
Water concentration [ng/L]	**CAS Nr**	**MQL**	**1**	**2**	**3**	**4**
Insecticides						
Diazinon	333-41-5	0,3			1,7	1,3
Fipronil	120068-37-3	0,6		3,2		4,6
Imidacloprid	138261-41-3	2,4			7,5	5,7
Fungicides						
2-Aminobenzimidazole	934-32-7	1,0		1,3	2,7	2,2
Boscalid	188425-85-6	1,8		3,0	4,0	3,5
Carbendazim	10605-21-7	0,8	1,0	6,4	5,2	6,7
Propiconazole	60207-90-1	0,8		25,4	19,8	33,3
Tebuconazole	107534-96-3	0,7		20,9	18,1	21,3
Thiabendazole	148-79-8	0,8		1,6	2,8	2,2
Herbicides						
2,6-Dichlorbenzamide	2008-58-4	1,0		2,3	3,7	
Atrazine	1912-24-9	0,5	5,1	3,2	4,5	3,7
Clomazone	81777-89-1	0,8		2,6	1,5	4,6
Deisopropylatrazin	1007-28-9	1,0	3,5	3,0	4,1	3,8
Desethylatrazin	6190-65-4	1,5	5,4	3,9	5,5	6,3
Desethylterbuthylazine	30125-63-4	1,0				1,5
Diuron	330-54-1	1,5		2,7	5,0	4,4
Fenuron	101-42-8	1,0		2,9	2,4	4,1
Flufenacet	142459-58-3	1,0		1,5	1,9	5,3
Flurtamone	96525-23-4	0,7				0,8
Isoproturon	34123-59-6	0,5		1,9	2,4	2,8
MCPA[a]	94-74-6	1,0		2,2		19,2
Mecoprop	93-65-2	1,5				7,3
Metazachlor	67129-08-2	0,5		2,3	2,8	4,1

Pethoxamid	106700-29-2	0,7	1,3			1,0
Prometryn	7287-19-6	0,4		0,7	0,8	0,7
Prosulfocarb	52888-80-9	8,0			10,2	25,9
Quinmerac	90717-03-6	2,5		2,7		3,7
Simazine	122-34-9	0,5	4,3	4,2	5,5	5,1
MT13[b]	66753-07-9	0,6	1,2	11,5	4,9	12,1
Terbutryn	886-50-0	0,4	1,2	4,9	12,3	9,0
Terbutylazin	5915-41-3	0,4		0,5	0,6	0,8
Wastewater chemical						
CBZ-diol[c]	35079-97-1	2,5		379,7	771,5	633,9
1H-Benzotriazole	95-14-7	10,0	10,0	528,2	1370,1	975,8
5MBT[d]	136-85-6	2,5	17,8	402,4	1112,9	744,9
Acesulfame	55589-62-3	4,0	66,0	531,0		822,0
Acetylsulfamethoxazole	21312-10-7	3,0			8,1	7,0
Caffeine	58-08-2	5,0	68,6	218,2	99,1	81,1
Carbamazepine	298-46-4	0,5	2,0	219,2	503,6	373,4
Cotinine	486-56-6	2,0	33,6	42,2	34,0	57,0
Cyclamate	139-05-9	16,0	83,0	497,0		141,0
DEET[e]	134-62-3	0,4		3,8	14,8	15,8
Diclofenac	15307-86-5	2,5	15,2	522,1		880,7
NAAP[f]	83-15-8	1,5	21,9	612,6	839,0	717,3
p-Toluene-Sulfonamide	70-55-3	10,0	14,0	44,7	90,7	115,3
Saccharin	81-07-2	15,0		91,3		32,4
Sucralose	56038-13-2	18,0	56,7	891,6		1375,1
Sulfamethoxazole	723-46-6	1,5		13,5	43,2	27,2
Triethyl citrate	77-93-0	5,0	9,9	50,2	26,6	51,5

[a]2-Methyl-4-chlorophenoxyacetic acid, [b]Terbuthylazine-2-hydroxy, [c]10,11-Dihydroxy-10,11-dihydrocarbamazepine, [d]4-/5-Methyl-1H-benzotriazole, [e]N,N-Diethyl-meta-toluamide, [f]n-Acetyl-4-aminoantipyrine

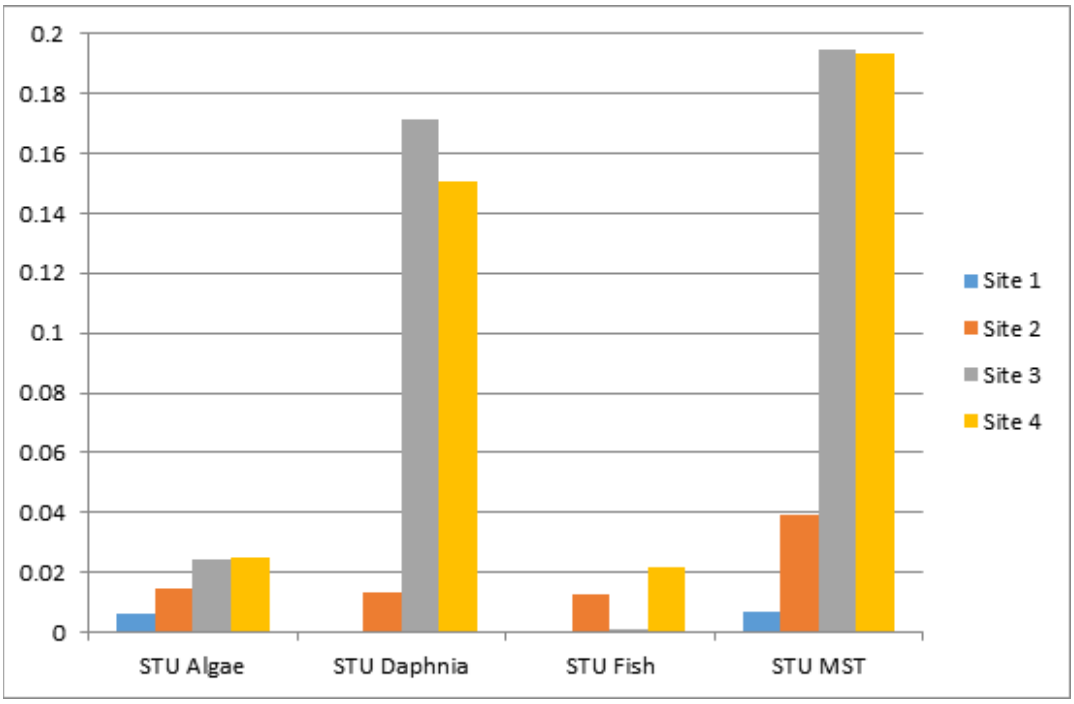


Figure 11: Sum of Toxic Units (STU) for the Holtemme and organism groups algae, daphnia, fish and the most sensitive species (MST; sum of the highest TU for each compound selected from the three organism groups fish, algae and daphnia) at the different study sites along the River Holtemme.

Table 10: Biomarker response in fish of anthropogenic impacted study sites of the Holtemme compared to a reference site (RS)

Biomarker	Study site		
	2	3	4
EROD	n.e.	↑	↑
GST	↓	n.e.	n.e.
CES	n.e.	↓	n.e.
CAT	↓	n.e.	n.e.
TBARS	↑	n.e.	n.e.
Micronucleus formation	↑	↑	↑

n.e.: no significant effect; ↑significantly elevated; ↓significantly lowered; for the analysis of the significance of variations between experimental data the t-test ($p \leq 0.05$) was used. In the case that data were not normally distributed ($p \leq 0.05$) and displayed no equal variances ($p \leq 0.05$) data were analyzed using Mann-Whitney U-test ($p \leq 0.05$).

Table 11: Calculated values for Index of Causality (IoC) and Index of Expected Ecological Impact (IoEEI) based on the biomarker response data of the Holtemme. Data were ranked in terms of "least/worst impacted site investigated".

Study site	ABR	Rank
1	0.54	"best"
2	0.66	
4	0.78	
3	0.86	"worst"

4.3.4 Line of evidence 2: In situ results

The biomarker results from fish indicated significant differences for almost all investigated biological endpoints (EROD, GST, CES, CAT, TBARS, micronucleus formation in fish erythrocytes) for at least one of the sampling sites downstream of the WWTP compared to reference site. Effects of inhibition or increased values were demonstrated. In contrast, concentrations of the steroid hormones 17β-estradiol and testosterone in blood plasma as well as activities of AChE did not show significant differences along the river stretch compared to the reference site. Table 10 summarizes biomarker response data in fish. Figure 12 visualizes the biomarker response pattern along the Holtemme river.

Specific trends of biomarker response pattern were observable (e.g. EROD activity and micronucleus formation) but there was no overall hotspot for all biomarkers. In summary, the most significant responses were identified for sampling sites directly downstream of the WWTPs Wernigerode and Halberstadt (site 2 and 3). For the study site 4, with no specific point sources of pollution, only the chronic DNA damage which impacts the erythrocytes of fish and the very sensitive EROD activity was significantly different in comparison to the reference site. Some of the biological endpoints (e.g. GST and CES) responded with inhibition instead of expected increased values. The variation of phase I and II enzyme activity and content levels by contamination are affected by several biotic and abiotic factors, such as age, size, maturity, health status, species, season, temperature and complexity of chemical mixture (Bonacci et al., 2003; van der Oost et al., 2003; Havelkova et al., 2008). Furthermore, specific inhibitors present in the environment, e.g. heavy metals, non-planar PCB congeners, organotins or other pesticide compounds, could alter enzyme activities and contents of CYP1A enzymes (Brüschweiler et al., 1996; Whyte and Tillit, 2000; Bozcaarmutlu and Arinc, 2004; Brammell et al., 2004) and GST (Al-Ghais and Ali, 1999); Letelier et al., 2006; Trute et al., 2007; Espinoza et al., 2012; Hernandez-Moreno et al., 2014). However, geogenic influences and historical background contamination for the reference site cannot entirely be excluded.

Aggregation of biomarkers

Biomarker response analyses often lack of causality of the heterogeneous results in terms of linking them to chemical exposure. Like suggested in chapter 3.2we calculated the average biomarker response

(ABR) to reflect the overall weight of evidence of the used biomarker battery. Results of these calculations including the ranking of the investigated sites are summarized in Table 11.

The data confirms the ranking of the sites as discussed above. ABR values site 3 (downstream the second WWTP) was classified “worst”, respectively was most impacted. In contrast, site 1, located closest to the reference site (see Figure 1) showed the lowest ABR values and hence indicates the increasing effect levels after WWTP influences (sites 2 and 3). Nevertheless, also site 1 shows already relative high values with an ABR of 0.54, which was driven by the GST, CAT and CES assays which had highest impact values here. In this evaluation, the calculation of the ABR provides an objective and integrative basis for the evaluation of the effect levels based on *in situ* biomarkers, which adds to the discussion of the single biomarker results. It helps here to identify that also site 1, close to the reference site shows significant impacts in some of the tests.

4.3.5 Body burdens and genetic structure of Gammarus pulex

Inostroza et al. (2016a) selected a list of 74 target analytes for body burden analysis of *Gammarus pulex* according to chemical analysis of the water grab samples mentioned above. Inostroza and co-workers found both increasing numbers and concentrations of wastewater chemicals in *G. pulex* along the course of the Holtemme River with strong peaks after WWTPs, whereas the number of the detected compounds and concentrations upstream of the first WWTP were low.

Chronic exposure to pollutions can result in the loss of genetic variation within a population and a decrease in fitness, a process referred to as genetic erosion as proposed by van Straalen and Timmermanns (2002). Naturally, genetic diversity increases in the course of a flowing aquatic system. Inostroza et al. (2016a) reported that the overall genetic diversity of *G. pulex* along the Holtemme increased from upstream to downstream sites but was lower after WWTP outlets.

Additionally, private alleles are commonly used as proxies for relative mutation rates (Theodorakis and Shugart, 1997; Nadig et al., 1998; Mengoni et al., 2000; Whitehead et al., 2003; Theodorakis et al., 2006; Inostroza et al., 2016a) and the researchers found a remarkable increase in private alleles in *G. pulex* downstream of the first WWTP, a significant reduction in the following river course, followed by a subsequent increase downstream of the second WWTP.

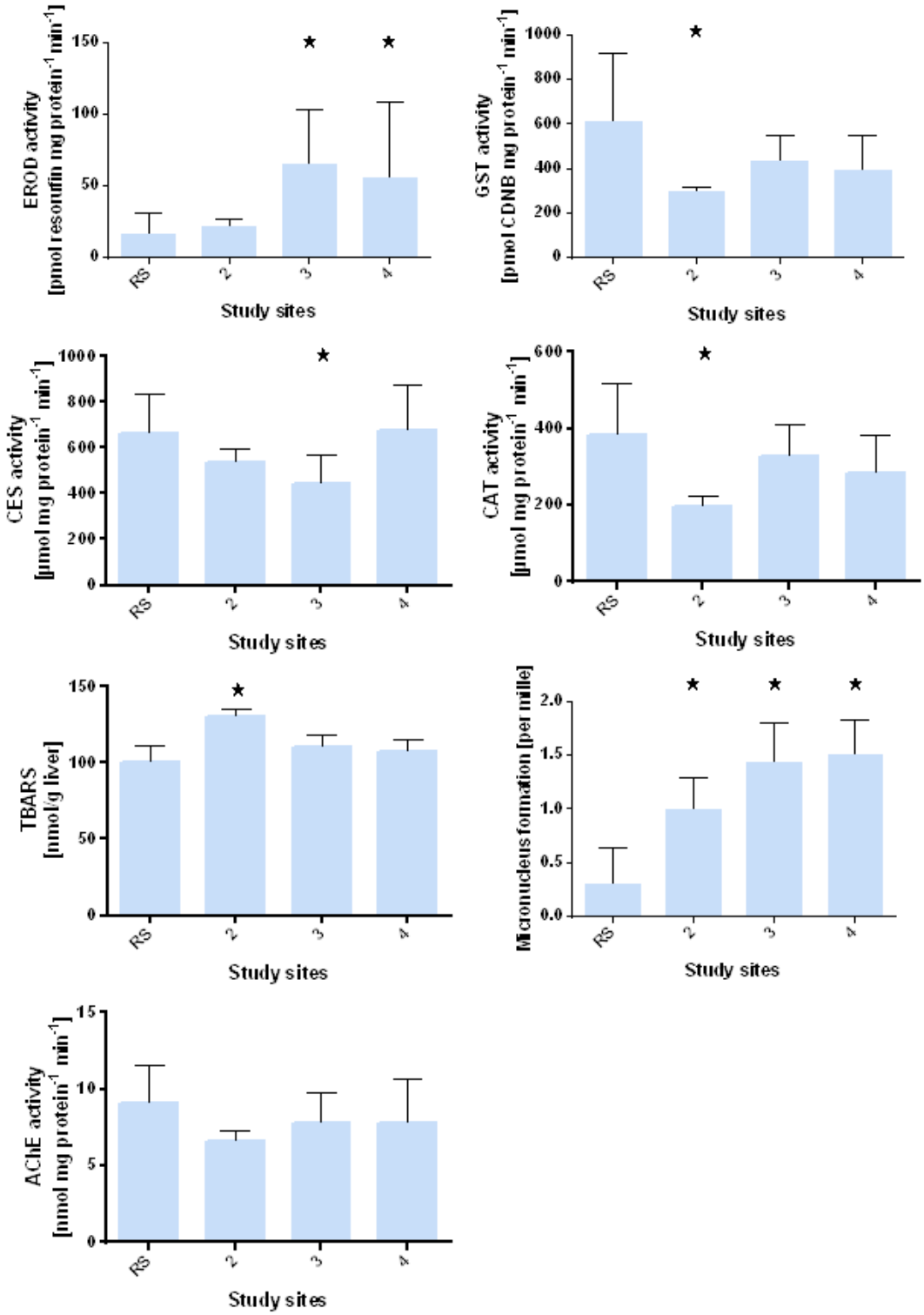


Figure 12: Results of biomarker response analysis of brown trout (Salmo trutta) of the River Holtemme collected during the key sampling in Oktober 2014. Each bar represents the mean value of 10 individuals per site. Error bars represent the standard deviation (SD). Asteriks depict significant differences compared to reference site (RS); t-test (p = 0.05). In the case that data were not normally distributed (p = 0.05) and displayed no equal variances (p = 0.05) data were analyzed using Mann-Whitney U-test (p = 0.05).

4.3.6 Summary and conclusions

In summary, our results suggest pollution-induced impacts in the River Holtemme in the lower part of the river downstream of WWTPs, and with increasing anthropogenic activities in the catchment area. Particularly downstream of the WWTP of the town Halberstadt and the area of intensive agricultural land use, STUs indicate an increase of pollutant pressure on biology. Most severe impacts were demonstrated by the STU for daphnids which were driven mainly by insecticides. Biomarker response in fish confirmed changes along the river course, although the different patterns of biomarkers did not show a monotonous increase along the river course, and some of the biomarkers did not respond as expected. Nevertheless, our study found significant changes in response in fish at the sites downstream the first WWTP in Wernigerode compared to a reference site which was chosen on the basis of land use information. Additionally, the values of the average biomarker response (section 3.2) showed corresponding trends to the results of the STU data, where sites 3 and 4 were identified as being most impacted. The reference site for this study showed lowest values amongst the studied sites, but the average biomarker response was not as low as expected here. These results underline the utility and practicability behind the presented approach of integrated biomarker response to reduce uncertainty in terms of unclear results and reference conditions in this case study.

Results of body burdens analysis and the genetic structure of *G. pulex* in the Holtemme published by Inostroza et al. (2016b) demonstrated a steady increase of concentrations of anthropogenic chemicals along the river course that most likely originated from wastewater. Strong peaks were found at sites where effluents of WWTPs enter the river and a remarkable increase in private alleles as proxies for relative mutation rates at these sites was detected. In contrast, the genetic diversity was lower downstream of WWTP outlets which should naturally increase along the course of a flowing river system. These results are in accordance with genotoxic effects assessed in fish which showed a steady trend towards increasing micronucleus formation. With respect to the decision matrix of Table 13, the results fit together reasonably. While LOE3 shows effects on sub-individual and individual level for anthropogenic impacted sites, LOE1 indicates the presence of pollution pressure downstream of WWTP effluents and in the agricultural catchment area and identified the main possible toxic drivers. These indicate effects due to chemical exposure on the aquatic organisms in the River Holtemme. Nevertheless, follow-up studies analysing community data can shed more light on possible ecological impacts of the toxic pressure and investigate whether the indicated biomarker response pattern is reflected at community levels.

5 Evaluation of the ecological toolbox development

This document contains an overview about strengths and limitations of four main lines of evidence that can be used to evaluate the ecological status of aquatic ecosystems and the causative role of toxic chemicals in ecological impairment. A systematic and quantitative aggregation method was developed, and the integration of the single LOE in a weight of evidence approach was defined in form of a decision matrix. Four weight of evidence case studies are given, illustrating specific aspects, with the Danube case study (section 4.1) as the main example involving a very comprehensive data set.

The aim of the toolbox development is a statistically supported, transparent and formalized WOE approach for the establishment of links between chemical exposure and ecological impacts, that allows for the use of mechanistic data and information for substantiating such WOE-derived linkages.

From the development of this approach and its case studies applications, a number of general findings emerged that are summarised as the final conclusions of the toolbox development.

1. Dedicated use of simple statistics within a rigorous evaluation scheme

Previous WoE studies have typically derived either from qualitative or correlative methods (Weed, 2005; Linkov et al., 2009; Chapman and Hollert, 2006), or used quantitative approaches such as statistical methods (e.g. ordination, principal components analysis), Bayesian techniques, multi-criteria decision analysis (MCDA) (Good, 1991; Smith et al., 2002; Exponent, 2009; Hope and Clarkson, 2013; Schleier et al., 2015) or Fuzzy Logic and Hasse Diagram techniques (Hollert et al. 2002). In the developed toolbox, we deliberately used only basic statistical methods such as geometric mean calculations, and no multivariate evaluations or other advanced techniques. The focus was on the development of a reasonable and robust order of basic data transformation and data compilation steps (sections 3.2 and 3.3). This rationale for this emphasis was that the use of the toolbox should be possible in a wide application range, e.g. national or regional water authorities can implement the toolbox in any spread sheet calculation environment. This deliverable explored initial steps and applications of the toolbox, but the development is not finalised yet. The next steps will be to provide standard sheets or R scripts together with examples to spread and ease the use of the toolbox. Also an implementation within the Solutions RiBaTox decision support system is anticipated.

2. Importance of clearly defined study objectives and fit-for purpose study design

Two out of three practical examples (the Rhine and the Holtemme) were typical upstream - downstream gradient studies. The objective of the Rhine study was clearly focused - to evaluate whether an upgrade of waste water treatment processes is effective in reducing the ecological impacts of effluents on the receiving water body. The Rhine project, in addition to being a gradient study, was also a BACI (Before-and-After-Impact) study as it enabled comparison of chemical profiles and ecological impacts before and after the upgrade of WWTP. This highlights the importance of keeping the study design constant in repeated campaigns. Both the Rhine and Holtemme studies examined the overall impact of WWTPs by a typical upstream - downstream approach, by using two LOE only. Both studies used predictive modelling

(LOE1), while the intermediate LOE (LOE 3 - *in situ* fish biomarkers) was used in the Holtemme case study, and community data (LOE 4) was used in the Rhine case study. The results from case studies demonstrate that when a clear research question is defined, the toolbox does not need an extensive list of parameters („laundry list“) for improving its diagnostic power.

Another scenario is provided by the Danube case study which represents an example of a very large set of data, containing all available LOE data for testing the toolbox. The JDS3 campaign, although complemented with a number of additional analyses and data collections beyond the scope of WFD compliant monitoring, was organised mainly to provide additional information to regular mandatory chemical and ecological status assessment. Therefore, the lack of consistency between sampling and analytical efforts per site narrows the scope of the evaluation to only a subfraction of the JDS3 sampling sites, and particularly, it excludes the most interesting sites concerning chemical stressors. In contrast to small scale studies (Rhine, Holtemme) which made use of a very focussed research question, huge campaigns such as the Joint Danube Surveys are often organised once in several years and basically represent a snapshot at the time of the campaign, which is far from investigative monitoring as such. Knowledge on local conditions when selecting sampling sites is crucial, particularly in large rivers, due to high mixing and dilution capacity. In case of strong indication of site specific pollution, it is worth considering *in situ* caging experiments using BQE of interest. In general, the choice of sampling sites could be guided by ‘local’ research questions and local upstream/downstream settings in order to improve the potential for finding differences in data analyses.

3. Importance of reference sites or conditions

Clearly, the problem of reference conditions or sites is more visible on large scales, e.g. the Danube Case Study than for typical smaller scale upstream - downstream type of studies for point sources. To differentiate, a reference site is a concrete site in a certain distance to the sampling site. A reference condition, however, might be a more theoretical construct for a given environmental scenario, a certain community, or a certain expression level of a biomarker – so it is not simply a concrete measurement, but can be of a more theoretical knowledge. The availability of good reference sites or the definition of reference conditions is necessary to enable the detection of potentially subtle differences in biological responses between the sites. The extremely low spatial resolution (67 sampling sites along the 1800 km long river section, including several tributaries) in the Danube example contrasts the very high spatial resolution of Holtemme case, where reference site could be identified in the most upper source region and additional 3 sites right downstream the point sources along the rest of the river of total length of less than 50 km were sampled. A flat chemical analytical profile and corresponding flat biomarker responses as well as a high percentage of heavily modified stretches of the Danube hampered the identification of a real reference site. On the other hand, the selection of species for *in situ* biomarker analysis (LOE 3) was based on occurrence and sufficient abundance, and not on the availability of reference conditions for *in situ* tests. The selected species (*Alburnus alburnus* and *Neogobius sp.*) were native species (which secure ecological relevance of the study), but were not studied before in the lab and also genetically not described. This missing information hampered the identification of deviations

from the natural variation in constitutive gene expression or enzyme activity. One approach could also be to rank results between 0 and 100 % effectiveness and use the locations with the minor effects in the catchment area or in site-independent classification systems (Keiter et al. 2009, Hollert et al. 2002). Lesson learned from the examples is that for the setup of *in situ* studies, reference sites should be defined beforehand. In case this is not possible, the choice of species for the testing should consider ecological relevance, but the species of choice should at the same time be well studied and preferably genetically described,. An increased number of specimens per site appears to be necessary to compensate for high natural variability, also more uniform sample in terms of age, preferably of mature specimens with determined sex, might allow for gender specific analyses.

4. Better use of existing monitoring data

Matching chemical profiles from a single sampling occasion (often grab water sample) with responses in long living species, such as macroinvertebrates and particularly fish appears problematic. On the other hand, extremely high monitoring efforts are invested EU-wide for the chemical profiling of water bodies. Planning monitoring campaigns but also the evaluation of such data sets might benefit from open databases with chemical analytical data allowing for a better characterisation of intensity and variability of chemical exposure at sites or river stretches of interest. Examples from literature show how repeated efforts in combination with the use of historical data (particular for the BQE of interest) provides a better ground for firm conclusions (Hollert et al., 2009; Keiter et al., 2006, 2009).

5. Increase the diagnostic power of the toolbox using complementary batteries of *in vitro* assays and *in situ* biomarkers

The outlined LOE are, apart from field surveys, not yet elements for the ecological status assessment under WFD. Currently effect-directed analysis and *in situ* biomarkers are, however, suggested and discussed in this context.

In the analyses of efforts in the JDS3 data set, it appeared that the selection of *in vitro* assays should be done in better alignment with the measured *in situ* biomarkers of toxicological impacts. This would enable to establish firmer links between chemical data, *in vitro* observed effects, and *in situ* toxicological responses.

Biomarkers serve two purposes: one is to support causative links between adverse effects and exposure to toxic chemicals. To this end, biomarkers must be specific. On the other hand, biomarkers should also pinpoint to potential adverse outcomes. Here less specific biomarkers, which, however, inform on the health status of the resident organism can be helpful, for instance, oxidative stress or histopathological markers. Despite of new developments, the predictive power of biomarkers for ecological outcomes is still limited. Here, growing integration of AOP information can be helpful. Even in a small river such as the Holtemme with relative high toxic pressure, biomarker responses are not linear, a fact which questions the use of traditional non-specific biomarkers in wild fish. Non-specific biomarkers need to be accompanied with more specific and more ecologically relevant ones. By means of *in vitro* assays it is possible to identify the diversity of modes of actions being present at field site (rather than the diversity

of chemicals), and this information can pinpoint the toxicity pathways through which the pollutants are likely to interfere with biology, hence which ecological functions of organisms and populations are at risk. However, this LOE is still of a more qualitative than quantitative one.

6. Toxic properties of typically detected substances urgently needed

The number of detected and quantified chemicals in the case studies is not balanced by the number of available experimental toxicity data. For the utmost major part of chemicals, only baseline toxicity data could be used for the evaluation of toxic pressure. In addition, the focus of experimental data lies in the acute domain, while for long-living organism like macroinvertebrates and fish the chronic exposure is of more relevance, especially for compounds which do not show high variability in time such as most household products, pharmaceutical or industrial chemicals. There is an urgent need for more fish and invertebrate chronic toxicity data for more relevant predictive modelling. This need could in parts already be satisfied when chronic test would not report NOEC values but rather EC10 or similar, because NOEC values are semi-quantitative and hence not useful for predictive modelling.

.

Table 12: *In situ* tests for ecological impact assessment on function, fitness and structure

Organism Group	Function	Fitness	Structure
Fish			
Communities	- nutrient and energy linkage between pelagic, benthic and littoral zones (Meador and Goldstein 2003) - top down control of foodwebs (Jeppesen et al. 2010)	- Growth (biomass) (Irons et al. 2007 - Reproductive output/rates (Irons et al. 2007)	Changes in taxonomic and/or trait (fish guilds) composition (Karr 1981, Suter 1993a,b; , Noble et al, 2007; Birk et al 2010, 2012, Azimi and Rocher 2016)
Individual species	- sex ratio - age structure - behaviour - Bioaccumulation - Adams et al. 1999	Disease, Condition factor (Cazenave et al, 2014), Behaviour (Garcia-Reyero et al, 2011)	- Phenology Cancer frequency (Shinn et al. 2015)
Sub-organismal level	- reproductive parameters (Amiard-Triquat et al 2012) - Histopathology (Van der Ost et al. 2003) - biomarkers of effect (Adams et al. 1999, Cazenave et al, 2014)	- scope for growth (Amiard-Triquat et al 2012,Enzyme biomarkers (Van der Ost et al. 2003, Cazenave et al, 2014) - Genomics, gene expression and profiles, transcriptomics, unsuperwised functional analysis (Garcia-Reyero et al, 2011, Beringer et al, 2014, Li et al, 2017, Schroeder et al 2016, 2017)	- body indices (Amiard-Triquat et al) - Histopathology (Van der Ost et al. 2003) - Biomarkers of exposure (Adams et al. 1999)
Macrophytes			
Communities	- Primary production, nutrient cycling, (Keruzoré et al, 2013, Forni, 2014)	- taxonomic and/or trait (life form) composition (Birk et al, 2006, Baattrup-Pedersen et al, 2016, Wiegleb et al, 2016)	
Individual species	- Bioaccumulation (Sánchez-Quiles et al (2017)	- Growth, Biomass production (ISO, 2005, Feileret al, 2014; Höss et al, 2010)	

Organism Group	Function	Fitness	Structure
Sub-organismal level		- Biomarkers: enzyme activity, gene expression, transcriptomics (Forni et al, 2012; Forni and Tommasi, 2016; Dranguet et al, 2017)	
Invertebrates			
Communities	- Secondary, tertiary production and consumption, decomposition, rate of change of biota, community respiration (Dolbeth et al., 2015; Johnston et al., 2015; Peters et al., 2013)	- Changes in community and/or trait composition assessed using diversity indices or multivariate analyses (Rico et al., 2016; Kuzmanovic et al., 2017; Beiras and Duran, 2014; Martinez-Haro et al., 2015)	
Individual species	- Secondary, tertiary production and consumption (Dolbeth et al., 2015)	- Immobility, mortality, growth, reproduction, feeding rates, (Malaj et al., 2014; Martinez-Haro et al., 2015; Beiras and Duran, 2014;)	- Changes in species abundance
Sub-organismal level		- biomarkers (molecular, biochemical and physiological responses) (Martinez-Haro et al., 2015; Colin et al., 2016)	
Microorganisms			
Complex heterotrophic microbial communities	- Leaf litter degradation (Artigas et al., 2012, Colas et al., 2016) - Respiration (Pringault et al., 2016) - Productivity (Pringault et al., 2016)		- Changes in overall species composition, measured via genomic markers (nextgen-seq, RFLP, DGGE), lipid composition, etc
Complex autotrophic communities (biofilms)	- Respiration (Rosi-Marshall et al., 2013; Artigas et al., 2014) - Primary production (Davis et al., 1988; Corcoll, 2011; Artigas et al., 2014) - Enzyme activities (Bonet et al., 2013, Bonet et al., 2014) - Bioaccumulation (Guasch et al., 2012;	- Recovery (Boivin et al., 2006; Dorigo et al., 2010; Dorigo et al., 2010a) - Growth (Rosi-Marshall et al., 2013) - Pollution tolerance (Boivin et al., 2006; Montuelle et al., 2010;	- Biomass (He et al., 2015) - Changes in algal species composition (Guasch et al., 2012) - Changes in algal biovolumes (Corcoll et al., 2012) - Changes in species composition,

Organism Group	Function	Fitness	Structure
	Bonet et al., 2013; Bonet et al., 2014) - Primary production - Respiration (Tlili et al., 2011) - Community-level physiological profiling (Montuelle et al., 2010)	Guasch et al., 2012; Fechner et al., 2012, Fechner et al., 2012a; Fechner et al., 2014; Foulquier et al., 2015; Tlili et al., 2017)	measured with genetic methods (nextgen-seq, RFLP, DGGE, ARISA) (Dorigo et al., 2010; Montuelle et al., 2010; Fechner et al., 2012a; Fechner et al., 2014) Changes in trait composition (Dunck et al., 2016)
Complex autotrophic communities (plankton)	- Respiration (Artigas et al., 2014) - Primary production (Davis et al., 1988; Artigas et al., 2014)	- Pollution tolerance (Larras et al., 2016)	- Changes in species composition (De La Broise and Palenik, 2007), - Changes in species measured as proxies (genomic composition, pigment composition, etc)
Individual algal species		- Growth (dos Santos et al., 2002; Marques et al., 2011; Wang et al., 2011; Bauer et al. 2012)	- Cell deformation Bauer et al. (2012)
Individual bacterial species	- Bioluminescence (Lopez-Roldan et al., 2012; Masner et al., 2017)		

Table 13: Decision matrix on how to combine the four main lines of evidence

Scenario	**LOE1:** Predicted mixture risk quotient	**LOE2:** Effect-direct analysis signal?	**LOE3:** *In situ* effects?	**LOE4:** Community alteration?	***Conclusion:*** *Pollution-driven ecological impacts?*	***Comments***
1	>1	yes	yes	yes	*present*	• *Chemical pollution proven to cause ecological impacts at the investigated site(s)* • *Chemical-oriented risk mitigation required*
2	>1	yes	yes	no	*likely*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact from chemical pollution* • *Hypothesis to be evaluated: community impacts not visible due to gaps in the data (e.g. seasonality)*
3	>1	yes	no	yes	*likely*	• *Chemical pollution likely causes ecological impacts at the site(s)* • *Hypothesis to be evaluated: in situ tests do not reflect all relevant modes of action*
4	>1	yes	no	no	*unclear, follow-up studies required*	• *No impacts at present, but indications that chemicals are a potential problem at the investigated site(s) in the long run, depending on the size of the predicted risk quotient and the specific EDA results. Follow-up in situ and community monitoring studies required.* • *Hypotheses to be evaluated: (i) contaminants might not be bioavailable in the field, (ii) check the validity of CA-based assessment, (iii) check consistency of ecotoxicological endpoints of EDA-studies and in situ tests and the organism groups included in the community sampling.*

Scenario	**LOE1:** Predicted mixture risk quotient	**LOE2:** Effect-direct analysis signal?	**LOE3:** *In situ* effects?	**LOE4:** Community alteration?	***Conclusion:*** *Pollution-driven ecological impacts?*	***Comments***
5	>1	no	yes	yes	*present*	• *Chemical pollution most likely causes ecological impacts at the investigated site(s)* • *Hypothesis to be evaluated: check consistency between risk drivers identified in the CA assessment, endpoints used in the in situ tests and the bioassays & endpoints used for the EDA assessment*
6	>1	no	yes	no	*unclear, follow-up studies required*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact. Unclear whether the impact is caused by chemicals* • *Hypotheses to be evaluated: (i) check consistency between risk drivers identified in the CA assessment, endpoints used in the in situ tests and the bioassays used for the EDA assessment and the organism groups included in the community sampling, (ii) check whether community impacts are not visible due to gaps in the data (e.g. seasonality)*
7	>1	no	no	yes	*unclear, follow-up studies required*	• *Ecological impacts present at the investigated site(s), but role of chemical pollution is unclear* • *Hypotheses to be evaluated: (i) in situ test suite and endpoints used in the EDA study do not reflect all relevant modes of action, compare with risk drivers identified in the CA-analysis, (ii) check validity of CA-based assessment (especially the used environmental thresholds), (iii) check for the presence of non-chemical stressors*
8	>1	no	no	no	*no impacts at present*	• *Chemical pollution does not cause ecological impacts at the investigated site(s) at the moment. But, depending on the size of the*

Scenario	**LOE1:** Predicted mixture risk quotient	**LOE2:** Effect-direct analysis signal?	**LOE3:** *In situ* effects?	**LOE4:** Community alteration?	***Conclusion:*** *Pollution-driven ecological impacts?*	***Comments***
						predicted risk quotient and its specific calculation, there are indications that chemicals are a potential problem in the long run • *Hypotheses to be evaluated: (i) contaminants included in the CA-assessment might not be bioavailable, (ii) validity of CA-based assessment needs to be checked (thresholds used)*
9	<1	yes	yes	yes	*present*	• *Chemical pollution causes ecological impacts at the investigated site(s)* • *Hypotheses to be evaluated: (i) check validity of CA-based assessment (chemicals included, environmental thresholds used), (ii) check for the presence of synergistic interactions*
10	<1	yes	yes	no	*unclear, follow-up studies required*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact, potentially caused by chemical pollution* • *Hypotheses to be evaluated: (i) check validity of CA-based assessment, in comparison with the results from the EDA assessment (chemicals included, environmental thresholds used), (ii) community impacts not visible due to gaps in the data (e.g. seasonality)*
11	<1	yes	no	yes	*likely*	• *Chemical pollution likely causes ecological impacts at the investigated site(s)* • *Hypotheses to be evaluated: (i) check validity of CA-based assessment (chemicals included, environmental thresholds used) and compare whether risk drivers identified in the EDA studies are appropriately considered, (ii) check consistency of endpoints used in the in situ tests*

Scenario	**LOE1:** Predicted mixture risk quotient	**LOE2:** Effect-direct analysis signal?	**LOE3:** *In situ* effects?	**LOE4:** Community alteration?	***Conclusion:*** *Pollution-driven ecological impacts?*	***Comments***
						and the EDA analysis
12	<1	yes	no	no	*unlikely*	• *Chemical pollution currently does not cause ecological impacts at the investigated site(s)* • *Hypotheses to be evaluated: (i) EDA-identified risk drivers are not bioavailable in the field, (ii) check whether EDA-identified risk drivers are included in the CA-based assessment*
13	<1	no	yes	yes	*no*	• *Ecological impacts present the investigated site(s), but the impacts are likely not caused by chemical pollution* • *Hypotheses to be evaluated: (i) non-chemical stressors present, (ii) contaminants present that are not included in the chemical monitoring and the EDA assessment, (iii) synergistic interactions, (iv) check validity of CA-based assessment, (v) check how specific the suite of in situ tests is for chemical pollution*
14	<1	no	yes	no	*no*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact, most likely not caused by chemical pollution* • *Hypothesis to be evaluated: check whether in situ test suite is dominated by endpoints that indicate general stress / non-chemical stressors*
15	<1	no	no	yes	*no*	• *Ecological impacts present the investigated site(s), but the impacts are most likely not caused by chemical pollution* • *Hypotheses to be evaluated: check for the presence of impacts from*

Scenario	LOE1: Predicted mixture risk quotient	LOE2: Effect-direct analysis signal?	LOE3: *In situ* effects?	LOE4: Community alteration?	***Conclusion:*** *Pollution-driven ecological impacts?*	***Comments***
						non-chemical stressors
16	<1	no	no	no	*no*	• *No impacts, neither from pollution nor from other stressors ("pristine" site)* • *No further action required*

	RAW values									Classes										Weight of evidence				
										LOE1	LOE3	LOE4		LOE1	LOE4					Fish			Invertebrates	
	STUfish	ABR	SDIfish	EFI	STUinv	SDIinv	SAI	ASPT	SPE	STUfish	ABR	SDIfish	EFI	STUinv	SDIinv	SAI	ASPT	SPE		LOE1	LOE3	LOE4	LOE1	LOE4
JDS50	0.0216	0.5	1.2		0.0343	2.6	2.0	3.3	9.3	CLEAR	CLEAR	CLEAR		CLEAR	NO	NO	MID	MID	JDS50	2	2		2	1
JDS54	0.0192	0.6								CLEAR	CLEAR	CLEAR							JDS54	2	2			
JDS38	0.0124	0.5	1.8	5	0.0181	1.4	2.1	2.6	1.2	CLEAR	CLEAR	MID	CLEAR	CLEAR	CLEAR	NO	CLEAR	CLEAR	JDS38	2	2	2	2	2
JDS48	0.5248	0.5		4						CLEAR	CLEAR	CLEAR	CLEAR						JDS48	2	2	2		
JDS53	0.0154	0.5	1.9	4	0.0198	3.0	2.3	2.5	5.6	CLEAR	CLEAR	MID	CLEAR	CLEAR	NO	NO	CLEAR	MID	JDS53	2	2	2	2	2
JDS40	0.0109	0.6	1.6	3	0.0170	3.0	2.6	3.4	4.6	CLEAR	CLEAR	MID	MID	CLEAR	NO	MID	MID	CLEAR	JDS40	2	2	1	2	2
JDS62	0.0149	0.5	1.7	3	0.0172	2.1	2.1	2.7	5.8	CLEAR	CLEAR	MID	MID	CLEAR	NO	NO	CLEAR	MID	JDS62	2	2	1	2	2
JDS66	0.0181	0.6	2.2	3	0.0246	2.3	2.7	3.0	6.6	CLEAR	CLEAR	NO	MID	CLEAR	NO	CLEAR	MID	MID	JDS66	2	2	1	2	2
JDS67	0.0222	0.5	2.1	3	0.0279	1.5	2.0	3.1	6.9	CLEAR	CLEAR	NO	MID	CLEAR	MID	NO	MID	MID	JDS67	2	2	1	2	1
JDS58	0.0498	0.5								CLEAR	MID	CLEAR							JDS58	2	1			
JDS39	0.0128	0.2	1.2	4	0.0192	2.1	2.4	3.5	6.4	CLEAR	MID	CLEAR	CLEAR	CLEAR	NO	MID	MID	MID	JDS39	2	1	2	2	1
JDS47	0.0160	0.5	1.0	3	0.0183	2.6	2.4	2.8	3.0	CLEAR	MID	CLEAR	MID	CLEAR	NO	MID	CLEAR	CLEAR	JDS47	2	1	2	2	2
JDS60	0.0143	0.5	1.1	3	0.0242	2.2	2.0	3.2	8.2	CLEAR	MID	CLEAR	MID	CLEAR	NO	NO	MID	MID	JDS60	2	1	2	2	1
JDS27	0.0139	0.5	2.7	3	0.0127	2.3	2.2	3.2	0.0	CLEAR	MID	NO	MID	CLEAR	NO	NO	MID	CLEAR	JDS27	2	1	1	2	2
JDS28	0.0155	0.4	1.8	3	0.0121	1.0	3.1	1.5	1.3	CLEAR	MID	MID	MID	CLEAR	CLEAR	CLEAR	CLEAR	CLEAR	JDS28	2	1	1	2	2
JDS31	0.0199	0.5	1.6	3	0.0303	1.5	2.3	2.7	4.4	CLEAR	MID	MID	MID	CLEAR	CLEAR	NO	CLEAR	CLEAR	JDS31	2	1	1	2	2
JDS33	0.0173	0.3	1.9	3	0.0289	2.4	2.3	2.9	1.2	CLEAR	MID	MID	MID	CLEAR	NO	MID	CLEAR	CLEAR	JDS33	2	1	1	2	2
JDS36	0.0144	0.4	2.3	3	0.0254	1.5	2.0	3.4	5.2	CLEAR	MID	NO	MID	CLEAR	CLEAR	NO	MID	MID	JDS36	2	1	1	2	2
JDS65	0.0104	0.4	2.0	3	0.0216	1.1	2.2	2.5	1.5	CLEAR	MID	MID	MID	CLEAR	CLEAR	NO	CLEAR	CLEAR	JDS65	2	1	1	2	2

Figure 13: Raw values and classes for WoE calculations, matrix 1 (19 sites x 5 LOE). For abbreviations of column headers see section 4.1.6.

	RAW values								Classes								Weight of evidence				
									LOE1	LOE4		LOE1	LOE4					Fish		Invertebrates	
JDS3 cod	STUfish	Shannor	EFI	STUinv	SDI	SAI	ASPT	SPE	STUfish	SDIfish	EFI	STUinv	SDIinv	SAI	ASPT	SPE		LOE1	LOE4	LOE1	LOE4
JDS44	0.0124	1.6	Poor	0.0205	2.8	2.4	2.7	6.2	CLEAR	MID	CLEAR	CLEAR	NO	MID	CLEAR	MID	JDS47	2	2	2	2
JDS47	0.0160	1.0		0.0183	2.6	2.4	2.8	3.0	CLEAR	CLEAR		CLEAR	NO	MID	CLEAR	CLEAR	JDS49	2	2	2	2
JDS06	0.0137	1.0	Good	0.0154	1.9	2.3	2.7	3.2	CLEAR	CLEAR	NO	CLEAR	MID	MID	CLEAR	CLEAR	JDS50	2	2	2	2
JDS07	0.0180	1.2	Good	0.0289	1.8	2.2	2.9	5.2	CLEAR	CLEAR	NO	CLEAR	MID	NO	CLEAR	MID	JDS53	2	2	2	2
JDS14	0.0122	1.0	Poor	0.0199	2.1	2.3	2.9	5.9	CLEAR	CLEAR	CLEAR	CLEAR	NO	NO	CLEAR	MID	JDS55	2	2	2	2
JDS38	0.0124	1.8	Bad	0.0181	1.4	2.1	2.6	1.2	CLEAR	MID	CLEAR	CLEAR	CLEAR	NO	CLEAR	CLEAR	JDS56	2	2	2	2
JDS60	0.0143	1.1	Moderate	0.0242	2.2	2.0	3.2	8.2	CLEAR	CLEAR	MID	CLEAR	NO	NO	MID	MID	JDS34	2	2	2	1
JDS02	0.0173	0.9	Good	0.0156	3.3	2.1	3.7	5.2	CLEAR	CLEAR	NO	CLEAR	NO	NO	MID	MID	JDS41	2	2	2	1
JDS04	0.0163	0.9	Good	0.0152	2.7	1.9	3.3	5.5	CLEAR	CLEAR	NO	CLEAR	NO	NO	MID	MID	JDS46	2	2	2	1
JDS50	0.0216	1.2		0.0343	2.6	2.0	3.3	9.3	CLEAR	CLEAR		CLEAR	NO	NO	MID	MID	JDS48	2	2	2	1
JDS39	0.0128	1.2	Poor	0.0192	2.1	2.4	3.5	6.4	CLEAR	CLEAR	CLEAR	CLEAR	NO	MID	MID	MID	JDS57	2	2	2	1
JDS51	0.0409	1.8							CLEAR	MID							JDS58	2	1		
JDS46	0.0115	1.9		0.0178	1.5	3.0	3.5	5.1	CLEAR	MID		CLEAR	MID	CLEAR	MID	MID	JDS28	2	1	2	2
JDS62	0.0149	1.7		0.0172	2.1	2.1	2.7	5.8	CLEAR	MID		CLEAR	NO	NO	CLEAR	MID	JDS29	2	1	2	2
JDS53	0.0154	1.9		0.0198	3.0	2.3	2.5	5.6	CLEAR	MID		CLEAR	NO	NO	CLEAR	MID	JDS31	2	1	2	2
JDS27	0.0139	2.7	Moderate	0.0127	2.3	2.2	3.2	0.0	CLEAR	NO	MID	CLEAR	NO	NO	MID	CLEAR	JDS32	2	1	2	2
JDS40	0.0109	1.6	Moderate	0.0170	3.0	2.6	3.4	4.6	CLEAR	MID	MID	CLEAR	NO	MID	MID	CLEAR	JDS33	2	1	2	2
JDS31	0.0199	1.6	Moderate	0.0303	1.5	2.3	2.7	4.4	CLEAR	MID	MID	CLEAR	CLEAR	NO	CLEAR	CLEAR	JDS36	2	1	2	2
JDS65	0.0104	2.0		0.0216	1.1	2.2	2.5	1.5	CLEAR	MID		CLEAR	CLEAR	NO	CLEAR	CLEAR	JDS37	2	1	2	2
JDS22	0.0132	2.0	Moderate	0.0159	2.8	2.4	3.0	3.1	CLEAR	NO	MID	CLEAR	NO	MID	CLEAR	CLEAR	JDS38	2	1	2	2
JDS28	0.0155	1.8	Moderate	0.0121	1.0	3.1	1.5	1.3	CLEAR	MID	MID	CLEAR	CLEAR	CLEAR	CLEAR	CLEAR	JDS39	2	1	2	2
JDS36	0.0144	2.3	Moderate	0.0254	1.5	2.0	3.4	5.2	CLEAR	NO	MID	CLEAR	CLEAR	NO	MID	MID	JDS51	2	1	2	2
JDS33	0.0173	1.9	Moderate	0.0289	2.4	2.3	2.9	1.2	CLEAR	MID	MID	CLEAR	NO	MID	CLEAR	CLEAR	JDS52	2	1	2	2
JDS10	0.0204	2.2	Moderate	0.0198	3.4	2.0	3.2	3.8	CLEAR	NO	MID	CLEAR	NO	NO	MID	CLEAR	JDS54	2	1	2	2
JDS20	0.0179	2.4	Moderate	0.0181	1.7	2.1	3.7	7.8	CLEAR	NO	MID	CLEAR	MID	NO	MID	MID	JDS35	2	1	2	1
JDS15	0.0164	2.6	Moderate	0.0214	2.1	2.0	3.9	7.9	CLEAR	NO	MID	CLEAR	NO	NO	MID	MID	JDS40	2	1	2	1
JDS13	0.0105	1.9	Moderate	0.0156	1.8	2.3	3.6	6.7	CLEAR	MID	MID	CLEAR	MID	MID	MID	MID	JDS42	2	1	2	1
JDS66	0.0181	2.2		0.0246	2.3	2.7	3.0	6.6	CLEAR	NO		CLEAR	NO	CLEAR	MID	MID	JDS43	2	0	2	2
JDS08	0.0113	2.4	Good	0.0185	2.2	2.0	3.9	8.1	CLEAR	NO	NO	CLEAR	NO	NO	MID	MID	JDS27	2	0	2	1
JDS67	0.0222	2.1		0.0279	1.5	2.0	3.1	6.9	CLEAR	NO		CLEAR	MID	NO	MID	MID	JDS30	2	0	2	1
JDS57	0.0152	2.6		0.0217	2.5	2.0	3.6	6.8	CLEAR	NO		CLEAR	NO	NO	MID	MID	JDS44	2	0	2	1
JDS52	0.0180	2.5		0.0152	3.2	2.4	4.0	6.3	CLEAR	NO		CLEAR	NO	MID	MID	MID	JDS45	2	0	2	1

Figure 14: Raw values and classes for WoE calculations, matrix 2 (32 sites x 4 LOE). For abbreviations of column headers see section 4.1.6.

Figure 15: All-in-one version of the decision matrix (see Table 13).

Scenario	LOE1: Predicted mixture risk quotient	LOE2: Effect-direct analysis signal?	LOE3: *In situ* effects?	LOE4: Community alteration?	*Conclusion: Pollution-driven ecological impacts?*	*Comments*
1	>1	yes	yes	yes	*present*	• *Chemical pollution proven to cause ecological impacts at the investigated site(s)* • *Chemical-oriented risk mitigation required*
2	>1	yes	yes	no	*likely*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact from chemical pollution* • *Hypothesis to be evaluated: community impacts not visible due to gaps in the data (e.g. seasonality)*
3	>1	yes	no	yes	*likely*	• *Chemical pollution likely causes ecological impacts at the site(s)* • *Hypothesis to be evaluated: in situ tests do not reflect all relevant modes of action*
4	>1	yes	no	no	*unclear, follow-up studies required*	• *No impacts at present, but indications that chemicals are a potential problem at the investigated site(s) in the long run, depending on the size of the predicted risk quotient and the specific EDA results. Follow-up in situ and community monitoring studies required.* • *Hypotheses to be evaluated: (i) contaminants might not be bioavailable in the field, (ii) check the validity of CA-based assessment, (iii) check consistency of ecotoxicological endpoints of EDA-studies and in situ tests and the organism groups included in the community sampling.*
5	>1	no	yes	yes	*present*	• *Chemical pollution most likely causes ecological impacts at the investigated site(s)* • *Hypothesis to be evaluated: check consistency between risk drivers identified in the CA assessment, endpoints used in the in situ tests and the bioassays & endpoints used for the EDA assessment*
6	>1	no	yes	no	*unclear, follow-up studies required*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact. Unclear whether the impact is caused by chemicals* • *Hypotheses to be evaluated: (i) check consistency between risk drivers identified in the CA assessment, endpoints used in the in situ tests and the bioassays used for the EDA assessment and the organism groups included in the community sampling, (ii) check whether community impacts are not visible due to gaps in the data (e.g. seasonality)*
7	>1	no	no	yes	*unclear, follow-up studies required*	• *Ecological impacts present at the investigated site(s), but role of chemical pollution is unclear* • *Hypotheses to be evaluated: (i) in situ test suite and endpoints used in the EDA study do not reflect all relevant modes of action, compare with risk drivers identified in the CA-analysis, (ii) check validity of CA-based assessment (especially the used environmental thresholds), (iii) check for the presence of non-chemical stressors*
8	>1	no	no	no	*no impacts at present*	• *Chemical pollution does not cause ecological impacts at the investigated site(s) at the moment. But, depending on the size of the predicted risk quotient and its specific calculation, there are indications that chemicals are a potential problem in the long run* • *Hypotheses to be evaluated: (i) contaminants included in the CA-assessment might not be bioavailable, (ii) validity of CA-based assessment needs to be checked (thresholds used)*
9	<1	yes	yes	yes	*present*	• *Chemical pollution causes ecological impacts at the investigated site(s)* • *Hypotheses to be evaluated: (i) check validity of CA-based assessment (chemicals included, environmental thresholds used), (ii) check for the presence of synergistic interactions*
10	<1	yes	yes	no	*unclear, follow-up studies required*	• *Effects on individual species present, i.e. the investigated site(s) are close to the manifestation of an ecological impact, potentially caused by chemical pollution* • *Hypotheses to be evaluated: (i) check validity of CA-based assessment, in comparison with the results from the EDA assessment (chemicals included, environmental thresholds used), (ii) community impacts not visible due to gaps in the data (e.g. seasonality)*
11	<1	yes	no	yes	*likely*	• *Chemical pollution likely causes ecological impacts at the investigated site(s)* • *Hypotheses to be evaluated: (i) check validity of CA-based assessment (chemicals included, environmental thresholds used) and compare whether risk drivers identified in the EDA studies are appropriately considered, (ii) check consistency of endpoints used in the in situ tests and the EDA analysis*
12	<1	yes	no	no	*unlikely*	• *Chemical pollution currently does not cause ecological impacts at the investigated site(s)* • *Hypotheses to be evaluated: (i) EDA-identified risk drivers are not bioavailable in the field, (ii) check whether EDA-identified risk drivers are included in the CA-based assessment*
13	<1	no	yes	yes	*no*	• *Ecological impacts present the investigated site(s), but the impacts are likely not caused by chemical pollution* • *Hypotheses to be evaluated: (i) non-chemical stressors present, (ii) contaminants present that are not included in the chemical monitoring and the EDA assessment, (iii) synergistic interactions, (iv) check validity of CA-based assessment, (v) check how specific the suite of in situ tests is for chemical pollution*

6 References

Adams, S.M. (2003). Establishing causality between environmental stressors and effects on aquatic ecosystems. Human and Ecological Risk Assessment 9, 17-35.

Adams, SM., Bevelhimer, MS., Greeley, MS., Levine, DA. and The, SJ. (1999). Ecological risk assessment in a large river reservoir. 6. Bioindicators of fish population health. Environmental Toxicology and Chemistry 18:628-640.

Ågerstrand, M. and Beronius, A. (2016). Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed. Environment International 92-93, 590-596.

Al-Ghais, S.M., Ali, B. (1999). Inhibition of Glutathione S-Transferase Catalyzed Xenobiotic Detoxication by Organotin Compounds in Tropical Marine Fish Tissues. Bulletin of Environmental Contamination and Toxicology 62, 207-213.

Altenburger, R., Ait-Aissa, S., Antczak, P., Backhaus, T., Barceló, D., Seiler, T.-B., Brion, F., Busch, W., Chipman, K., de Alda, M. L., de Aragão Umbuzeiro, G., Escher, B. I., Falciani, F., Faust, M., Focks, A., Hilscherova, K., Hollender, J., Hollert, H., Jäger, F., Jahnke, A., Kortenkamp, A., Krauss, M., Lemkine, G. F., Munthe, J., Neumann, S., Schymanski, E. L., Scrimshaw, M., Segner, H., Slobodnik, J., Smedes, F., Kughathas, S., Teodorovic, I., Tindall, A. J., Tollefsen, K. E., Walz, K.-H., Williams, T. D., Van den Brink, P. J., van Gils, J., Vrana, B., Zhang, X. and Brack, W. (2015). Future water quality monitoring — Adapting tools to deal with mixtures of pollutants in water resource management, Science of The Total Environment, 512–513, pp. 540–551.

Amiard-Triquet, C., Amiard, J.C. and Rainbow P. S.(2012). Ecological Biomarkers Indicators of Ecotoxicological Effects, 1st edition, CRC Press.

An KG., Park SS. and Shin JY. (2002). An evaluation of a river health using the index of biological integrity along with relations to chemical and habitat conditions, Environment International 28, 411–420.

Arkoosh MR, Clemons E, Hufman P, Kagley AN (2001) Increased susceptibility of juvenile Chinook salmon to vibriosis after exposure to chlorinated and aromatic compounds found in contaminated urban estuaries. J Aquat Animal Health 13:257-268.

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

Artigas, J., Majerholc, J., Foulquier, A., Margoum, C., Volat, B., Neyra, M. and Pesce, S. (2012). Effects of the fungicide tebuconazole on microbial capacities for litter breakdown in streams, Aquatic Toxicology, 122–123, pp. 197–205.

Artigas, J., Pascault, N., Bouchez, A., Chastain, J., Debroas, D., Humbert, J. F., Leloup, J., Tadonleke, R. D., ter Halle, A. and Pesce, S. (2014). Comparative sensitivity to the fungicide tebuconazole of biofilm and plankton microbial communities in freshwater ecosystems, Science of the Total Environment., 468–469, pp. 326–336.

Azimi, S. and Rocher, V., (2016). Influence of the water quality improvement on fish population in the Seine River (Paris, France) over the 1990–2013 period. Science of the Total Environment 542, 955–964.

Baattrup-Pedersen, A., Göthe, E., Riis, T. and OHare, MT. (2016). Functional trait composition of aquatic plants can serve to disentangle multiple interacting stressors in lowland stream. Science of the Total Environment 543 230–238

Backhaus, T. and Karlsson, M. (2014). Screening level mixture risk assessment of pharmaceuticals in STP effluents, Water Research. Elsevier Ltd, 49(0), pp. 157–165.

Bahamonde, P.A., Tetreault, G.R., McMaster, M.E., Servos, M.R., Martyniuk, C.J., Munkittrick, K.R., 2014. Molecular signatures in rainbow darter (Etheostoma caeruleum) inhabiting an urbanized river reach receiving wastewater effluents. Aquat. Toxicol. 148, 211–220.

Baird DJ, Burton GA, Culp JM, Maltby M. 2007. Summary and recommendations from a SETAC Pellston Workshop on *in situ* measures of ecologcial effects. Integr Environ Assessm Mngmt 3: 275-278

Baird, D.J., P.J. Van den Brink, A.A. Chariton, K.A. Dafforn and E.L. Johnston (2016). New diagnostics for multiply stressed marine and freshwater ecosystems: integrating models, ecoinformatics and big data. Marine and Freshwater Research 67: 391–392.

Baldigo B, George SD, Phillips PJ, Hemming JDC, Denslow ND, Kroll KJ (2015) Potential estrogenic effects of wastewaters on gene expression in Pimephales promelas and fish assemblages in streams of southeastern New York, 34(12), 2803–2815.

Bammer, V., György, A., Pehlivanov, L., Schabuss, M., Szaloky, Z., Zornig, H. (2015). Fish In: Liška, I., Wagner, F., Sengl, M., Deutsch, K. and Slobodník, J. Joint Danube Survey 3, Final Scientific Report. International Comission for the Protection of the Danube River, Vienna, Austria. pp. 126-139

Barbout MT, Gerritsen J, Snyder BT, and Stubbing JB. (1999). Rapid bioassessment protocols for use in streams and rivers. 2nd edition. US EPA. www.epa.gov/OWOW/monitoring/techmom.html

Bauer, D. E., Conforti, V., Ruiz, L. and Gómez, N. (2012). An *in situ* test to explore the responses of Scenedesmus acutus and Lepocinclis acus as indicators of the changes in water quality in lowland streams, Ecotoxicology and Environmental Safety, 77, pp. 71–78. doi: 10.1016/j.ecoenv.2011.10.021.

Becker RA, Ankley GT, Edwards SW, Kennedy SW, Linkov I, Meek B, Sachana M, Segner H, van der Burg B, Villeneuve DL, Watanabe H, Barton-Maclaren TS (2015). Increasing scientific confidence in adverse outcome pathways: application of tailored Bradford-Hill considerations for evaluating weight of evidence. Regulatory Toxicology and Pharmacology 72: 514-537.

Beliaeff, B. and Burgeot, T. (2002). Integrated biomarker response: a useful tool for ecological risk assessment., Environmental toxicology and chemistry, 21(6), pp. 1316–1322.

Beiras, R. and Durán, I. (2014). Objective classification of ecological status in marine water bodies using ecotoxicological information and multivariate analysis. Environmental Science and Pollution Research 21: 13291–13301.

Besse, A. Deliverable No.270: Indicator value database for European freshwater diatoms. Euro-limpacs: Integrated Project to evaluate the Impacts of Global Change on European Freshwater Ecosystems, WP7: Indicators of ecosystem health. Alterra, Wageningen Available at http://www.refresh.ucl.ac.uk/eurolimpacs/deliverables&view=abstract&delid=270&download=1&doc=Deliverable_270.pdf [July 2017]

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., Van De Bund, W., Zampoukas, N. and Hering, D. (2012). Three hundred ways to assess Europes surface waters: An almost complete overview of biological methods to implement the Water Framework Directive, Ecological Indicators. Elsevier Ltd, 18, pp. 31–41. doi: 10.1016/j.ecolind.2011.10.009.

Birk, S., Strackbein, J. and Hering, D. (2010). WISER methods database. Version: May 2010. Available at http://www.wiser.eu/programme-and-results/dataand-guidelines/method-database/.

Birk, S., Korte, T. and Hering, D., (2006). Intercalibration of assessment methods for macrophytes in lowland streams: direct comparison and analysis of common metrics. Hydrobiologia 566, 417–430.

Blanck, H. (2002) A Critical Review of Procedures and Approaches Used for Assessing Pollution-Induced Community Tolerance (PICT) in Biotic Communities, Human and Ecological Risk Assessment: An International Journal, 8(5), pp. 1003–1034. doi: 10.1080/1080-700291905792.

Bonacci, S., I. Corsi, R. Chiea, F. Regoli, Focardi, S. (2003). Induction of EROD activity in European eel (*Anguilla anguilla*) experimentally exposed to benzo[a]pyrene and beta-naphthoflavone. Environment International 29, 467-473.

Bonet, B., Corcoll, N., Acuna, V., Sigg, L., Behra, R. and Guasch, H. (2013). Seasonal changes in antioxidant enzyme activities of freshwater biofilms in a metal polluted Mediterranean stream, Science of the Total Environment, 444, pp. 1–10. doi: 10.1016/j.scitotenv.2012.11.036.

Bonet, B., Corcoll, N., Tlili, A., Morin, S. and Guasch, H. (2014). Antioxidant enzyme activities in biofilms as biomarker of Zn pollution in a natural system: An active bio-monitoring study, Ecotoxicology and Environmental Safety. Elsevier, 103(1), pp. 82–90. doi: 10.1016/j.ecoenv.2013.11.007.

Bozcaarmutlu, A., Arinc, E. (2004). Inhibitory effects of divalent metal ions on liver microsomal 7-ethoxyresorufin O-deethylase (EROD) activity of leaping mullet. Marine Environmental Research 58, 521-524.

Brack, W. (2011) Effect-Directed Analysis of Complex Environmental Contamination. Handbook of Environmental Chemistry Vol 15, Springer, Heidelberg

Brack W, Ait-Aissa S, Burgess RM, Busch W, Creusot N, Di Paolo C, Escher BI, Hewitt ML, Hilscherova K, Hollender J, Hollert H, Jonker W, Kool J, Lamoree M, Muschket M, Neumann S, Rostkowski P, Ruttkies C, Schollee J, Schymanski EL, Schulze T, Seiler TB, Tindall AJ, De Aragão Umbuzeiro G, Vrana B, Krauss (2016) Effect-directed analysis supporting monitoring of aquatic environments — An in-depth overview, Science of the Total Environment, 544, 1073-1118.

Brack, W., Dulio, V., Ågerstrand, M., Allan, I., Altenburger, R., Brinkmann, M., Bunke, D., Burgess, R.M., Cousins, I., Escher, B.I. and Hernandez, F.J., (2017). Towards the review of the European Union Water Framework Directive: Recommendations for more efficient assessment and management of chemical contamination in European surface water resources. Science of the Total Environment, 576, pp.720-737.

Brammell, B.F., Price D.J., Birge W.J., Elskus, A.A. (2004). Apparent lack of CYP1A response to high PCB body burdens in fish from a chronically contaminated PCB site. Marine Environmental Research 58: 251-255.

Brauns, M., Offinger, W. (2002). Bemerkenswerte Nachweise von Wasserinsekten (Ephemeroptera, Coleoptera, Trichoptera) aus dem Nordharz, Sachsen-Anhalt. Erik Mauch Verlag - Lauterbomia 44.

Brüschweiler, B.J., Würgler, F.E., Fent, K. (1996). Inhibitory Effects of Heavy Metals on Cytochrome P4501A Induction in Permanent Fish Hepatoma Cells. Archives of Environmental Contamination and Toxicology 31, 475-482.

Buesing, N. and Gessner, M.O. (2003). Incorporation of radiolabelled leucine into protein to estimate bacterial production in plant litter, sediment, epiphytic biofilms and water samples. Microbial Ecology 45, 291-301.

Burki R, Vermeirssen ELM, Körner O, Joris C, Burkhardt-Holm P, Segner H (2006). Assessment of estrogenic exposure in brown trout (*Salmo trutta*) in a Swiss midland river: integrated analysis of passive samplers, wild and caged fish, and vitellogenin mRNA and protein. Environmental Toxicology and Chemistry 25:2077-2086.

Burton, G.A. Jr., Batley, G.E., Chapman, P.M., Forbes, V.E., Smith, E.P., Reynoldson, T., Schlekat, C.E., den Besten, P.J., Bailer, A.J., Green, A.S. and Dwyer, R.L. (2002). A weight-of-evidence framework for assessing sediment (or other) contamination: Improving certainty in the decision-making process. Human Ecological Risk Assessment 8, 1675–1696.

Busch, W., Schmidt, S., Kühne, R., Schulze, T., Krauss, M. and Altenburger, R., (2016). Micropollutants in European rivers: A mode of action survey to support the development of effect-based tools for water monitoring. Environmental toxicology and chemistry, 35(8), pp.1887-1899.

Cazenave J, Bacchetta C, Rossia A, Aleb A, Campana M, Parma MJ (2014) Deleterious effects of wastewater on the health status of fish: A field caging study. Ecological Indicators 38 : 104– 112

CEN - European Committee for Standardisation (2014). Guidance for the surveying of aquatic macrophytes in running waters. EN 14184:2014

Chapman, P.M. (1990). The Sediment Quality Triad approach to determining pollution-induced degradation. Science of the Total Environment 97(8), 815-825.

Chapman, P.M. (2002). Integrating toxicology and benthic ecology: Putting the eco back into ecotoxicology. Marine Pollution Bulletin 44, 7-15.

Chapman, P.M. and Anderson, J., (2005). A Decision-Making Framework for Sediment Contamination. Integrated Environmental Assessment and Management 1(3), 163-173.

Chapman PM, Hollert H (2006) Should the sediment quality triad become a tetrad, a pentad, or possibly even a hexad? J Soils Sediments 6: 4-8.

Colas, F., Baudoin, J. M., Chauvet, E., Clivot, H., Danger, M., Guérold, F. and Devin, S. (2016). Dam-associated multiple-stressor impacts on fungal biomass and richness reveal the initial signs of ecosystem functioning impairment. Ecological Indicators, *60*, 1077–1090.

Colin, N. C., Porte, D., Fernandes, C., Barata, F., Padrós, M., Carrassón, M., Monroy, O. Cano-Rocabayera, A., de Sostoa, B., Piña, A. and Maceda-Veiga (2016). Ecological relevance of biomarkers in monitoring studies of macro-invertebrates and fish in Mediterranean rivers. Science of The Total Environment 540: 307–323

Connon, R. E., Geist, J. and Werner, I. (2012). Effect-based tools for monitoring and predicting the ecotoxicological effects of chemicals in the aquatic environment, Sensors (Switzerland), pp. 12741–12771.

Corcoll, N., Bonet, B., Morin, S., Tlili, A., Leira, M. and Guasch, H. (2012). The effect of metals on photosynthesis processes and diatom metrics of biofilm from a metal-contaminated river: A translocation experiment. Ecological Indicators, 18, 620–631.

Cormier, SM. and Suter, GW II. (2013). A method for assessing causation of field exposure-response relationships. Environmental Toxicology and Chemistry 32: 272-276.

Cortes R, Hughes S, Coimbra A, Monteiro S, Pereira V, Lopes M, Pereira S, Pinto A, Sampaio A, Santos C, Carrola J, de Jesus J, Varandas S (2013) A multiple index integrating different levels of organization. Ecotoxicology and Environmental Safety 132, 270–278.

Coste, M., Boutry, S., Tison-Rosebery, J. and Delmas, F. (2009) 'Improvements of the Biological Diatom Index (BDI): Description and efficiency of the new version (BDI-2006)', Ecological Indicators, 9(4), pp. 621–650..

Culp, J., D.G. Armanini, M.J. Dunbar, J.M. Orlofske, N.L. Poff, A.I. Pollard, A.G. Yates and G.C. Hose (2011). Incorporating traits in aquatic biomonitoring to enhance causal diagnosis and prediction. Integrated Environmental Assessment and Management 7: 187–197.

Davis, A.P., Murphy, C.G., Johnson, R., Lay, J.M., Lennon-Hopkins, K., Saraceni-Richards, C., Sciaky, D., King, B.L., Rosentein, M.C., Wiegers, T.C. and Mattingly, C.J. (2013). The comparative Toxicogenomics database: update 2013. Nucleic Acids Res. 2013, D1104-D1114 (Database issue).

De La Broise, D. and Palenik, B., (2007). Immersed *in situ* microcosms: a tool for the assessment of pollution impact on phytoplankton. Journal of experimental marine biology and ecology, 341(2), pp.274-281.

Deutschmann, B., Kolarevic, S., Brack, W., Kaisarevic, S., Kostic, J., Kracun-Kolarevic, M., Liska, I., Paunovic, M., Seiler, TB., Shao, Y., Sipos, S., Slobodnik, J., Teodorovic, I., Vukovic-Gacic, B and, Hollert, H. (2016). Longitudinal profiles of the genotoxic potential of the River Danube on erythrocytes of wild fish assessed using the comet and micronucleus assay. Science of the Total Environment, 573, *1441-1449.*

Di Paolo, C., Ottermanns, R., Keiter, S., Ait-Aissa, S., Bluhm, K., Brack, W., Breitholtz, M., Buchinger, S., Carere, M., Chalon, C. and Cousin, X., (2016). Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts–towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring. Water research, 104, pp.473-484.

Dolbeth, M., Dolédec, S. and Pardal M.Â. (2015). Relationship between functional diversity and benthic secondary production in a disturbed estuary. Marine Ecology Progress Series 539: 33–46

Dorazio, R.M., Gotelli, N.J. and Ellison, A.M. (2011). Modern methods of estimating biodiversity from presence-absence surveys. In: Biodiversity loss in a changing planet. InTech Publishers.

Dorigo, U., Bérard, A., Bouchez, A., Rimet, F. and Montuelle, B. (2010). Transplantation of microbenthic algal assemblages to assess structural and functional recovery after diuron exposure. Archives of Environmental Contamination and Toxicology, *59*(4), 555–563.

Dorigo, U., Berard, A., Rimet, F., Bouchez, A. and Montuelle, B. (2010a). *In situ* assessment of periphyton recovery in a river contaminated by pesticides. Aquatic Toxicology, 98(4), 396–406.

dos Santos, M. M., Moreno-Garrido, I., Gonçalves, F., Soares, A. M. V. M. and Ribeiro, R. (2002). An *in situ* bioassay for estuarine environments using the microalga Phaeodactylum tricornutum. Environmental Toxicology and Chemistry / SETAC, 21(3), 567–574.

Dranguet, P., Cosio, C., Le Faucheur, S., Beauvais-Fluck, R., Freiburghaus, A., Worms, IAM., Petit, B., Civic, N., Docquier, M. and Slaveykova VI (2017). Transcriptomic approach for assessment of the impact on microalga and macrophyte of in-situ exposure in river sites contaminated by chlor-alkali plant effluents. Water Research, 121, 86-94.

Dussling U., Berg R., Klinger H., Wolter C. (2004): Assessing the ecological status of river systems using fish assemblages. Handbuch Angewandte Limnologie 12/04 (20.Erg.Lfg.): 1/84.

EC (2013) Commission decision of 20 September 2013 establishing, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, the values of the Member State monitoring system

classifications as a result of the intercalibration exercise and repeal, Offical Journal of the European Union, L266, pp. 1–47.

EFI+ CONSORTIUM (2009). Manual for the application of the new European Fish Index – EFI+. A fish-based method to assess the ecological status of European running waters in support of the Water Framework Directive. June 2009.

Eltzov, E. and Marks, R. S. (2011). Whole-cell aquatic biosensors. Analytical and Bioanalytical Chemistry, 400(4), 895–913.

Espinoza, H.M., Williams, C.R., Gallagher, E.P. (2012). Effect of cadmium on glutathione S-transferase and metallothionein gene expression in coho salmon liver, gill and olfactory tissues. Aquatic Toxicology 110-111, 37-44.

Escher BI, Allinson M, Altenburger R, Bain PA, Balaguer P, Busch W, CragoJ, Denslow ND, Dopp E, Hilscherova K, Humpage KR, Kumar A, Grimaldi M, Jayasinghe BS, Jarosava B, Jia A, Makarov S, Maruya KA, Medvedev A, Mehinto AC, Mendez JE, Poulsen A, Prochazka E, Richard J, Schifferli A, Schlenk D, Scholz S, Shiraishi F, Snyder S, Su G, Tang JY, van der Burg B, van der Linden SC, Werner I, Westerheide SD, Wong CK, Yang M, Yeung BH, Zhang X, Leusch FD. 2014. Benchmarking organic micropllutants in wastewater, recycled water and drinking water with *in vitro* bioassays. Environ Sci Technol 48: 19401956.

EU Guidance Document No. 10 (2003): Common Implementation Strategy for the Water Framework Directive (2000/60/EC), River and lakes – Typology, reference conditions and classification systems. Office for Official Publications of the European Communities, Luxemburg,

European Commission (2008). Commission Decision establishing, pursuant to Directive 2000/60/EC, the values of the Member State monitoring system classifications as a result of the intercalibration exercise. Official Journal of the European Union, Series L, 332: 20–44.

European Commission (2011). Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 27. Technical Guidance for Deriving Environmental Quality Standards (https://circabc.europa.eu/sd/a/0cc3581b-5f65-4b6f-91c6-433a1e947838/TGD-EQS%20CIS-WFD%2027%20EC%202011.pdf).

European Commission (2013). Commission Decision establishing, pursuant to Directive 2000/60/EC the values of the Member State monitoring system classifications as a result of the intercalibration exercise and repealing Decision 2008/915/ EC. Official Journal of the European Union, Series L, 266: 1–47.

European Chemicals Agency, (2015). Guidance on the Biocidal Products Regulation. Volume III Human Health – Part B Risk Assessment. Version 1.1.

Exponent (2003). Focused Literature Review of Weight-of-Evidence Approaches. Prepared for Science Advisory Board for Contaminated Sites in British Columbia, Victoria, British Columbia. Document No. 0906563.000 01F1 1109 SD23, Seattle, WA, USA.

Fame consortium (2004). Manual for the application of the European Fish Index – EFI. A fish/based method to assess the ecological status of European rivers in support of the Water Framework Directive. Version 1.1., January 2005.

Fechner, L. C., Gourlay-Francé, C., Bourgeault, A. and Tusseau-Vuillemin, M. H. (2012). Diffuse urban pollution increases metal tolerance of natural heterotrophic biofilms. Environmental Pollution, 162, 311–318.

Fechner, L. C., Versace, F., Gourlay-Francé, C. and Tusseau-Vuillemin, M.-H. (2012a). Adaptation of copper community tolerance levels after biofilm transplantation in an urban river. Aquatic Toxicology (Amsterdam, Netherlands), 106–107, 32–41.

Fechner, L. C., Gourlay-Francé, C. and Tusseau-Vuillemin, M. H. (2014). Linking community tolerance and structure with low metallic contamination: A field study on 13 biofilms sampled across the seine river basin. Water Research, *51*, 152–162.

Feckler, A., Kahlert, M. and Bundschuh, M. (2015). Impacts of contaminants on the ecological role of lotic biofilms. Bulletin of Environmental Contamination and Toxicology, 95(4), 421–427.

Feiler U., Ratte M., Arts G., Bazin C., Brauer F., Casado C., Dören L., Eklund B., Gilberg D., Grote M., Gonsior G., Hafner C., Kopf W., Lemnitzer B., Liedtke A., Matthias U., Okos E., Pandard P., Scheerbaum D., Schmitt-Jansen M., Stewart K., Teodorovic I., Wenzel A. and Pluta H.J. (2014). Inter-laboratory trial of a standardized sediment contact test with the aquatic plant *Myriophyllum aquaticum* (ISO 16191) Environmental Toxicology and Chemistry 33, 662-670.

Flinders, CA., Ragsdale, RL. and Hall, TJ. (2009). Patterns of Fish Community Structure in a Long-Term Watershed-Scale Study to Address the Aquatic Ecosystem Effects of Pulp and Paper Mill Discharges in Four US Receiving Streams. Integrated Environmental Assessment and Management 5 (2), 219–233.

Focks A, Rico A; van den Brink P; Velki M, Deutschmann B, Seiler TB, Hollert H; Segner H, Rehberger K; Sipos S, Jovanovic V, Novkovic M, Cvijanovic D, Radulovic S, Tenji D, Kaiserevic S, Teodorovic I (2015) ID T5.1 Report on WoE and trait-based results for JDS3 samples. Internal Deliverable, Solutions for present and future emerging pollutants in land and water resources management (SOLUTIONS), FP 7 No no. 603437

Forbes, V.E., Palmqvist A. and Bach, L. (2006). The use and misuse of biomarkers in ecotoxicology. Environmental Toxicology and Chemistry 25: 272–280.

Foulquier, A., Morin, S., Dabrin, A., Margoum, C., Mazzella, N. and Pesce, S. (2015). Effects of mixtures of dissolved and particulate contaminants on phototrophic biofilms: new insights from a PICT approach combining toxicity tests with passive samplers and model substances. Environmental Science and Pollution Research, *22*(6), 4025–4036.

Frimpong, E.A. and Angermeier, P.L. (2010). Trait-Based Approaches in the Analysis of Stream Fish Communities. American Fisheries Society Symposium 73:109–136, 2010. American Fisheries Society.

Garcia-Reyero, N., Lavelle, CM., Escalon, B.L., Martinovic, D., Kroll, K.J., Sorensen, P.W. and Denslow, N.D. (2011). Behavioral and genomic impacts of a wastewater effluent on the fathead minnow. Aquatic Toxicology, 101, 38-48.

Good, I.J. (1991). Weight of evidence and the Bayesian likelihood ratio. In: Aitken CGG, Stoney, D., editors. The Use of Statistics in Forensic Science. Boca Raton: CRC Press, 85-106.

Gosetti, F., Mazzucco, E., Gennaro, M.C. and Marengo, E. (2016). Contaminants in water: non-target UHPLC/MS analysis. Environmental chemistry letters, 14(1), pp.51-65.Gotelli, N.J. and Colwell, R.K., 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology letters, 4(4), pp.379-391.

Gotelli, N., Chao, A. (2013) Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. Encyclopedia of Biodiversity, Volume 5, p. 195-211.

Graf, W., Csányi, B., Leitner, P., Paunović, M., Huber, T., Szekeres, J., Nagy, C. and Borza, P (2015). Macroinvertebrates. In: Liška, I., Wagner, F., Sengl, M., Deutsch, K. and Slobodník, K. Joint Danube Survey 3, Final Scientific Report. International Commission for the Protection of the Danube River, Vienna, Austria. pp 81-99.

Guasch, H., Bonet, B., Bonnineau, C., Corcoll, N., López-Doval, J. C., Munoz, I., Ricart, M., Serra, A. and Clements, W. (2012). How to Link Field Observations with Causality? Field and Experimental Approaches Linking Chemical Pollution with Ecological Alterations, in Emerging and Priority Pollutants in Rivers., pp. 147–179. doi: 10.1007/978-3-642-25722-3_7.

Gustavsson, M., Kreuger, J., Bundschuh, M. and Backhaus, T. (2017). Pesticide mixtures in the Swedish streams: Environmental risks, contributions of individual compounds and consequences of single-substance oriented risk mitigation, Science of the Total Environment. Elsevier B.V., 598, pp. 973–983.

Hall, TJ., Fisher, R., Rodgers, Jr. J.H., Minshall, W., Landis, W.G., Kovacs, T., Firth, B.K., Dube, M.G., Flinders, C.A., Deardorff, T.L. and Borton, D.L. (2009). A Long-Term, Multitrophic Level Study to Assess Pulp and Paper Mill Effluent Effects on Aquatic Communities in Four US Receiving Waters: Lessons Learned. Integrated Environmental Assessment and Management 5 (2), 283–290

Hamers T, Legler J, Blaha L, Hylland K, Marigomez I, Schipper CA, Segner H, Vethaak AD, Witters H, de Zwart D, Leonards PEG. (2013). Expert opinion on toxicity profiling – report from a NORMAN expert group meeting. Integrated Environmental Assessment and Management 9: 185-191.

Haunschmid, R., Wolfram, G., Spindler, T., Honsig/Erlenburg, W., Wimmer, R., Jagsch, A., Kainz, E., Hehenwarter, K., Wagner, B., Konecny, R., Riedmüller, R., Ibel, G., Sasano, B. and Schotzko, N. (2006). Erstellung einer fischbasierten Typologie österreichischer Fließgewässer sowie einer Bewertungsmethode des fischökologischen Zustandes gemäß EU Wasserrahmenrichtlinie. Schriftenreihe des BAW Band 23, Wien, 104 pp.

Hassan, S. H. A., Van Ginkel, S. W., Hussein, M. A. M., Abskharon, R. and Oh, S. E. (2016). Toxicity assessment using different bioassays and microbial biosensors. *Environment International*, *92–93*, 106–118.

Havelkova, M., T. Randak, J. Blahova, I. Slatinska, Svobodova, Z. (2008). Biochemical markers for the assessment of aquatic environment contamination. Interdisciplinary Toxicology 1, 169-181

He, F., Dong, X., Sun, M. and Cai, Q. (2015). Altitudinal pattern of stream periphyton biomass in tributaries of the Lancang-Mekong River: An indicator of anthropogenic impact? *Quaternary International*, *380–381*, 282–287.

Herman, M.R. and Nejadhashemi A.P. (2015). A review of macroinvertebrate and fish-based stream health indices. Ecohydrology & Hydrobiology 15, 53–6754

Hernandez-Moreno, D., de la Casa-Resino, I., Maria Flores, J., Gonzalez-Gomez, M.J., Maria Neila, C., Soler, F., Perez-Lopez, M. (2014). Different enzymatic activities in carp (*cyprinus carpio L.)* as potential biomarkers of exposure to the pesticide methomyl. Arhiv za higijenu rada i toksikologiju 65, 311-318.

Higley E, Grund S, Jones PD, Schulze T, Seiler TB, Varel U Lv, Brack W, Wölz J, Zielke H, Giesy JP, Hollert H and Hecker M (2012) Endocrine disrupting, mutagenic, and teratogenic effects of upper Danube River sediments using effect-directed analysis. Environmental Toxicology and Chemistry, 31: 1053–1062.

Hilscherova, K,, Vrana, B,, Novak, J,, Smedes, F,, Rusina, T,, Okonski, K,, Ait-Aissa, S,, Brack, W,, Creusot, N,, Escher, B,, Grabic, R,, Krauss, M,, Neale, P,, Schulze, T, and Slobodnik, J. (2017). Chemical and toxicological profiling of large rivers using mobile passive sampling and comparison to large volume active sampling. SETAC Europe, 27 Annual Meeting, 7-11 May, Brussels, Belgium, Extended Abstract

Höss, S., Ahlf, W., Fahnenstich, C., Gilberg, D., Hollert, H., Melbye, K., Meller, M., Hammers-Wirtz, M., Heininger, P., Neumann-Hensel, H., Ottermanns, R., Ratte, H.T., Seiler, T.B., Spira, D., Weber, J. and Feiler, U. (2010). Variability of sediment-contact tests in freshwater sediments with low-level anthropogenic contamination: Determination of toxicity thresholds. Environmental Pollution 158, 2999-3010.

Hollert, H., Heise, S., Pudenz, S., Bruggemann, R., Ahlf, W., Braunbeck, T. (2002). Application of a sediment quality triad and different statistical approaches (Hasse diagrams and fuzzy logic) for the comparative evaluation of small streams. Ecotoxicology 11:, 311-321.

Hollert H, Keiter S, Boettcher M, Grund S, Seitz N, Otte J, Bluhm K, Wurm K, Hecker M, Higley E, Giesy J, Takner H, van Bavel B, Engwall M, Reifferscheid G, Manz W, Erdinger L, Schulze T, Luebcke-van Varel U, Kammann U, Schoeneberger R, Suter M, Brack W, Straehle U, Braunbeck T (2009): Assessing sediments and fish health using a weight-of-evidence approach - in search for the causes of fish decline in the Danube River. Umweltwissenschaften und Schadstoff-Forschung 21, 260-263

Hollert H, Durr M, Olsman H, Halldin K, Van Bavel B, Brack W, Tysklind M, Engwall M, Braunbeck T (2002): Biological and chemical determination of dioxin-like compounds in sediments by means of a sediment triad approach in the catchment area of the River Neckar. Ecotoxicology 11, 323-336

Hong, S., Giesy, J. P., Lee, J., Lee, J. and Khim, J. S. (2016). Effect-Directed Analysis: Current Status and Future Challenges, Ocean Science Journal, 51, pp. 413–433.

Hope, B.K. and Clarkson, J.R., (2013). A Strategy for Using Weight-of-Evidence Methods in Ecological Risk Assessment. Human and Ecological Risk Assessment: An International Journal 20(2), 290-315.

Huggett, R.J., Kimerle, R.A. and Mehrle Jr, P.M. (1992). Biomarkers: biochemical, physiological and histological makers of anthropogenic stress, dans: Proceedings of the 8th Pellston Workshop. Ann Harbor, Keystone, Colorado, July 1989.

Hull, R.N. and Swanson, S. (2006). Sequential Analysis of Lines of Evidence – An Advanced Weight-of-Evidence Approach for Ecological Risk Assessment. Integrated Environmental Assessment and Management 2(4), 302-311.

Hunting, E.R., Jong, S.D. and Vijver, M.G., (2017). Assessment of monitoring tools and strategies safeguarding aquatic ecosystems within the European water framework directive. CML reports.

Hutchinson T, Ankley GT, Segner H, Tyler CR (2006). Screening and testing for endocrine disruption in fish – biomarkers as „signposts" not „traffic lights" in risk assessment. Environmental Health Perspectives 114, supplement 1:106-114.

Inostroza, P.A., Iván, V.-E., Wicht, A.-J., Krauss, M., Brack, W., Norf, H. (2016a). Anthropogenic Stressors Shape Genetic Structure: Insights from a Model Freshwater Population along Land Use Gradient. Environmental Science and Technology 50, 11346-11356.

Inostroza, P. A., Wicht, A. J., Huber, T., Nagy, C., Brack, W. and Krauss, M. (2016b) Body burden of pesticides and wastewater-derived pollutants on freshwater invertebrates: Method development and application in the Danube River, Environmental Pollution. Elsevier Ltd, 214, pp. 77–85.

Inostroza, P.A. (2016). Organic micropollutants in freshwater ecosystems – Pollution dynamic and adverse effects at population genetic level in a model freshwater population. PhD thesis, RWTH Aachen University, Aachen, 35 pp.

Irons KS, Sass GG, McClelland MA, Satfford JD. 2007. Reduced condition factor of two native fish species coincident with invasion of non-native Asian carps in the Illinois River, USA. Is this evidence for competition and reduced fitness? Journal of Fish Biology 71 (Supplement D): 258-273.

ISO - International Organisation for Standardisation (2005): Water quality – Determination of the toxic effect of water constituents and waste water on duckweed (*Lemna minor*) – Duckweed growth inhibition test, ISO 20079:2005

Jeppesen E, Meerhoff M, Holmgren K, Gonzalez-Bergonzoni I, Teixeira-de Mello F, Declerck SAJ, De Meester L, Sondergaard M, Lauridsen TL, Bjerring R, Conde-Porcuna JM, Mazzeo N, Iglesias C, Reizenstein M, Malmquist HJ, Liu Z, Balaya D,Lazzaro X. 2010. Impacts of climate warming on lake fish community structure and potential ecosystem effects. Hydrobiologia 646: 73-90

Jasinska, E.J., Goss, G.G., Gillis, P.L., Van Der Kraak, G.J., Matsumoto J., de Souza Machado A.A., Giacomin, M., Moon, T.W., Massarsky, A., Gagne, F., Servos, M.R., Wilson, J., Sultana, T. and Metcalfe, C. (2015). Assessment of biomarkers for contaminants of emerging concern on aquatic organisms downstream of a municipal wastewater discharge. Science of the Total Environment 530–531, 140–153.

Jobling, S., Coey, S.,Whitmore, J.G., Kime, D.E., Van Look, K.J.W., McAllister, B.G., Beresford, N., Henshaw, A.C., Brighty, G., Tyler, C.R., Sumpter, J.P., 2002. Wild intersex roach (Rutilus rutilus) have reduced fertility. Biol. Reprod. 67 (2), 515–524.

Johnson AC, Chen Y (2017) Does exposure to domestic wastewater effluent (including steroid estrogens) harm fish populations in the UK? Science of the Total Environemnt, 589, 89-96.

Johnston, E.L., Mayer-Pinto, M. and Crowe, T.P. (2015). Chemical contaminant effects on marine ecosystem functioning. Journal of Applied Ecology 52: 140–149.

Jouanneau, S., Durand, M. J. and Thouand, G. (2015). Technological Design of Optical Bacterial Biosensors for the Online Environmental Monitoring of Water Pollutants, in Standford, P. (ed.) Bioluminescent Microbial Biosensors: Design, Construction, and Implementation. CRC Press.

Kammerad, B. (2011). Angler und Fischer in Sachsen-Anhalt. Landesanglerverband Sachsen-Anhalt e.V im DAV.

Karr J.R. (1981). Assessment of biological integrity using fish communities. Fisheries 6:, 21-27.

Karr J.R.(1991). Biological integrity: a long-neglected aspect of water resource management. Ecological Applications, l(l), 1991, pp. 66-84

Karr J.R. (1993). Defining and assessing ecological integrity: beyond water quality. Environmental Toxicology and Chemistry 12: 1521-1531.

Karr, J.R., and Dudley, D.R. (1981). Ecological perspective on water quality goals. Environmental Management 5:55-68.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P.R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. Illinois Natural History Survey, Champaigne, Illinois, Special Publication 5.

Kautza, A., Mazeika, S. and Sullivan, P. (2015). Spatially – dependent human alterations determine fish assemblage composition in a modified river system. River Systems 21 (2-3): 93-108

Keiter S, Rastall A, Kosmehl T, Wurm K, Erdinger L, Braunbeck T, Hollert H (2006) Ecotoxicological assessment of sediment, suspended matter and water samples in the Upper Danube River. A pilot study in search for the causes for the decline of fish catches. Environ Sci Pollut Res Int 13: 308-319.

Keiter S, Boettcher M, Grund S, Seitz N, Braunbeck T, Hollert H (2009): Decrease in fish populations in the upper Danube River. Umweltwissenschaften und Schadstoff-Forschung 21, 186-196

Kelly, M. (2013) 'Data rich, information poor? Phytobenthos assessment and the Water Framework Directive', European Journal of Phycology, 48(4), pp. 437–450.

Kolarević, S., Kračun-Kolarević, M., Kostić, J., Slobodnik, J., Liška, I., Gačić, Z., Paunović, M., Knežević-Vukčević, J. and Vuković-Gačić, B. (2016). Assessment of the genotoxic potential along the Danube River by application of the comet assay on haemocytes of freshwater mussels: The Joint Danube Survey 3. Science of The Total Environment, 540, 377-385.

Krimsky, S. (2005). The weight of scientific evidence in policy and law. American Journal of Public Health 95, 129-135.

Kuhn, M., Szklarczyk, D., Franceschini, A., von Mering, C., Jensen, L.J. and Bork, P. (2012). STITCH 3: zooming in on protein-chemical interactions. Nucleic Acids Res. 40, D876-D880 (Database issue).

Kuzmanovic, A.S., Dolédec, N., de Castro-Catala, A., Ginebreda, S., Sabater, I., Muñoz, D. and Barceló (2017). Environmental stressors as a driver of the trait composition of benthic macroinvertebrate assemblages in polluted Iberian rivers. Environmental Research 156 (2017) 485–493

König, M., Escher, B., Neale, P., Krauss, M., Hilscherová, K., Novák, J., Teodorović, I., Schulze, T., Seidensticker, S., Hashmi, M.A.K., Ahlheim, J. and Brack, W. (2017). Impact of untreated wastewater on a major European river evaluated with a combination of *in vitro* bioassays and chemical analysis, Environmental Pollution, 220, B, 1220-1230

Larras, F., Rimet, F., Gregorio, V., Bérard, A., Leboulanger, C., Montuelle, B. and Bouchez, A. (2016). Pollution-induced community tolerance (PICT) as a tool for monitoring Lake Geneva long-term *in situ* ecotoxic restoration from herbicide contamination, Environmental Science and Pollution Research, 23(5), pp. 4301–4311

Letelier, M.E., Martinez, M., Gonzalez-Lira, V., Faundez, M., Aracena-Parks, P. (2006). Inhibition of cytosolic glutathione S-transferase activity from rat liver by copper. Chemico-biological Interactions 164, 39-48.

Li S, Villeneuve DL, Berninger JP, Blackwell BR, Cavallin JE, Hughes MN, Jensen KM, Jorgenson Z, Kahl MD, Schroeder AL, Stevens KE, Thomas LM, Weberg MA, Ankley GT (2017) An integrated approach for identifying priority contaminant in the Great Lakes Basin – Investigations in the Lower Green Bay/Fox River and Milwaukee Estuary areas of concern, Science of the Total Environment, 579, 825-837.

Lindemeyer, D.B. and Likens, G.E. (2010). The science and application of ecological monitoring. Biological Conservation 143: 1317-1328.

Linkov, I., Loney, D., Cormier, S., Satterstrom, F.K. and Bridges, T. (2009). Weight-of-evidence evaluation in environmental assessment: Review of qualitative and quantitative approaches. Science of the Total Environment 407, 5199-5205.

Liška, I., Wagner, F., Sengl, M., Deutsch, K. and Slobodník, J. (2015). Joint Danube Survey 3 - A Comprehensive Analysis of Danube Water Quality, Final Scientific Report. International Commission for the Protection of the Danube River, Vienna, Austria; ISBN: 978-3-200-03795-3

Long, E.R. and Chapman, P.M. (1985). A sediment quality triad: measures of sediment contamination, toxicity and infaunal community composition in Puget Sound. Marine Pollution Bulletin 16, 405-415.

Lopez-Roldan, R., Kazlauskaite, L., Ribo, J., Riva, M.C., González, S. and Cortina, J.L., (2012). Evaluation of an automated luminescent bacteria assay for *in situ* aquatic toxicity determination. Science of the total environment, 440, pp.307-313.

Lowell, R.B., Culp, J.M. and Dube, M.G. (2000). A weight-of-evidence approach for northern river risk assessment: Integrating the effects of multiple stressors. Environmental Toxicology and Chemistry 19, 1182–1190.

Lyche-Solheim, A., Feld, C.K., Birk, S., Phillips, G., Carvalho, L., Morabito G., Mischke, U., Willby, N., Søndergaard, M., Hellsten, S., Kolada, A., Mjelde, M., Böhmer, J., Miler, O., Pusch, M.T., Argillier, C., Jeppesen, E., Lauridsen T.L. and Poikane, S. (2013). Ecological status assessment of European lakes: a comparison of metrics for phytoplankton, macrophytes, benthic invertebrates and fish. Hydrobiologia 704: 57–74.

Magurran, A. E. (2003) Measuring Biodiversity, Blackwell Science Ltd.

Malaj, E., von der Ohe, P.C., Grote, M., Kühne, R., Mondy, C.P., Usseglio-Polatera, P., Brack, W. and Schäfer R.B. (2014). Organic chemicals jeopardize the health of freshwater ecosystems on the continental scale. PNAS 111: 9549–9554.

Marques, C.R., Pereira, R., Antunes, S.C., Cachada, A., Duarte, A.C. and Gonçalves, F., (2011). *In situ* aquatic bioassessment of pesticides applied on rice fields using a microalga and daphnids. Science of the Total Environment, 409(18), pp.3375-3385.

Martinez-Haro, M., Beiras, R., Bellas, J., Capela, R., Coelho, JP., Lopes, I., Moreira-Santos, M., Reis-Henriques, AM., Ribeiro, R., Santos, M.M. and Marques, JC. (2015). A review on the ecological quality status assessment in aquatic systems using community-bases indicators and ecotoxicologcial tools: what might be the added value of their combination? Ecological Indicators 48: 8-16.

Masner, P., Javůrková, B. and Bláha, L., (2017). Rapid *in situ* toxicity testing with luminescent bacteria Photorhabdus luminescens and Vibrio fischeri adapted to a small portable luminometer. Environmental Science and Pollution Research, 24(4), pp.3748-3758.

Meador MR, Goldstein RM. 2003. Assessing water quality at large geographic scales: relations among land use, water physicochemistry, riparian condition and fish community structure. Environmental Management 31: 504-517.

Mengoni, A., Gonnelli, C., Galardi, F., Gabbrielli, R., Bazzicalupo, M. (2000). Genetic diversity and heavy metal tolerance in populations of Silene paradoxa L. (Caryophyllaceae): a random amplified polymorphic DNA analysis. Molecular Ecology 9(9), 1319-1324.

Montuelle, B., Dorigo, U., Bérard, A., Volat, B., Bouchez, A., Tlili, A., Gouy, V. and Pesce, S., (2010). The periphyton as a multimetric bioindicator for assessing the impact of land use on rivers: an overview of the Ardières-Morcille experimental watershed (France). Hydrobiologia, 657(1), pp.123-141.

Moschet, C., Lew, B.M., Hasenbein, S., Anumol, T. and Young, T.M. (2017). LC-and GC-QTOF-MS as Complementary Tools for a Comprehensive Micropollutant Analysis in Aquatic Systems. Environmental science & technology, 51(3), pp.1553-1561.

Morin, S., Gómez, N., Tornés, E. Licursi M., Rosebery, J. 2016. Benthic Diatom Monitoring and Assessment of Freshwater Environments: Standard Methods and Future Challenges. In: Aquatic Biofilms: Ecology, Water Quality and Wastewater Treatment (Edited by: Anna M. Romaní, Helena Guasch and M. Dolors Balaguer). Caister Academic Press, U.K.

Munoz I, Lopez-Doval JC, Ricart M, Villagrasa M, Brix R, Ggeiszinger A, Ginebreda A, Guasch H, Lopez de Alda MJ, Romani AM, Sabater S, Barcelo D. 2009. Bridiging levels of phaarmaceuticals in river water with biological community structure in the Llobregat river basin (Northeast Spain). Envviron Toxicol Chem 28: 2706-2714.

Nadig, S.G., Lee, K., Adams, S. (1998). Evaluating alterations of genetic diversity in sunfish populations exposed to contaminants using RAPD assay. Aquatic Toxicology 43(2), 163-178.

Neale, P. A., Altenburger, R., Aït-Aïssa, S., Brion, F., Busch, W., Umbuzeiro, G. de A., Denison, M. S., Du Pasquier, D., Hilscherová, K., Hollert, H., Morales, D. A., Novák, J., Schlichting, R., Seiler, T.-B., Serra, H., Shao, Y., Tindall, A. J., Tollefsen, K. E., Williams, T. D. and Escher, B. I. (2017) Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water, Water Research, 123.

Neale, P.A., Ait-Aissa, S., Brack, W., Creusot, N., Denison, M.S., Deutschmann, B., Hilscherova, K., Hollert, H., Krauss, M., Novaak, J. ,Schulze, T., Seiler, T.B., Serra, H., Shao, Y. and Escher, B.I. (2015). Linking *in vitro* Effects and Detected Organic Micropollutants in Surface Water Using Mixture-Toxicity Modeling. Environmental Science & Technology 49, 14614−14624

Neale, P.A., Altenburger, R., Aït-Aïssa, S., Brion, F., Busch, W., de Aragão Umbuzeiro, G., Denison, M.S., Du Pasquier, D., Hilscherová, K., Hollert, H. and Morales, D.A., (2017). Development of a bioanalytical test battery for water quality monitoring: Fingerprinting identified micropollutants and their contribution to effects in surface water. Water Research, 123, 734-750.

Noble RAA, Cowx IG, Goffaux D, Kestemont P (2007) Assessing the health of European rivers using functional ecological guilds of fish communities: standardising species classification and approaches to metric selection, Fisheries Management and Ecology, 14, 381–392.

Nõges, P., Argillier, C., Borja, A., Garmendia, J.M., Hanganu, J. Kodeš, V., Pletterbauer, F., Sagouis, A. and Birk, S. (2016). Quantified biotic and abiotic responses to multiple stress in freshwater, marine and ground waters. Science of the Total Environment 540, 43–52

Oberdorff, T., Pont, D., Hugueny, B. and Porcher J.P. (2002): Development and validation of a fish/based index for the assessment of »river health« in France. Freshwater Biology 47: 1720/1734.

Oikari, A. (2006). Caging techniques for field exposures of fish to chemical contaminants. Aquatic Toxicology 78:370-381

Ormerod, S.J., Dobson, M., Hildrew, A.G. and Townsend, C.R. (2010). Multiple stressors in freshwater ecosystems. Freshwater Biology 55, 1-4.

Palace VP, Doebel C, Baron CL, Evans RE, Wautier KG, Klaverkamp JF, Werner J, Kollar S. 2005. Caging small-bodied fish as an alternative method for environmental effects monitoring. Water Qual Res J Canada 40: 328-333

Peters K., Bundschuh, M. and Schäfer, R.B. (2013). Review on the effects of toxicants on freshwater ecosystem functions. Environmental Pollution 180: 324-329.

Petrie, B., Barden, R. and Kasprzyk-Hordern, B. (2014). A review on emerging contaminants in wastewaters and the environment: Current knowledge, understudied areas and recommendations for future monitoring, Water Research. Elsevier Ltd, 72(0), pp. 3–27. doi: 10.1016/j.watres.2014.08.053.

Poikane, S., Kelly, M. and Cantonati, M. (2016) 'Benthic algal assessment of ecological status in European lakes and rivers: Challenges and opportunities', Science of the Total Environment. British Geological Survey, NERC, 568, pp. 603–613. doi: 10.1016/j.scitotenv.2016.02.027.

Pont, D., Hugueny, B., Roset, N. and Rogers, C. (2007). Development of a fish/based index for the assessment of river health in Europe: the European Fish Index. Fisheries Management and Ecology 14: 427/439.

Pringault, O., Lafabrie, C., Avezac, M., Bancon-Montigny, C., Carre, C., Chalghaf, M., Delpoux, S., Duvivier, A., Elbaz-Poulichet, F., Gonzalez, C. and Got, P., (2016). Consequences of contaminant mixture on the dynamics and functional diversity of bacterioplankton in a southwestern Mediterranean coastal ecosystem. Chemosphere, 144, pp.1060-1073.

Ra, J.S., Kim, D.S., Chang, N.I. and An, G.K. (2007). Ecological health assessments based on whole effluent toxicity tests and the index of biological integrity in temperate streams influenced by wastewater treatment plant effluents. Environmental Toxicology and Chemistry, 26, 2010–2018

Rehberger K, Werner I, Hitzfeld, B, Segner H, Baumann L (2017). 20 years of fish immunotoxicology – what we know and where we are. Critical Reviews in Toxicology 47: 509-535.

Reuter, H., Krause, G., Möning, A., Wulkow, M., Horn, H. (2003). RIONET: A water quality management tool for river basins. Water Science and Technology 48(10), 47-53.

Richardson, S. D. and Kimura, S. Y. (2016). Water Analysis: Emerging Contaminants and Current Issues, Analytical Chemistry, 88(1), pp. 546–582..

Rico, A., Van den Brink, P.J., Leitner, P., Graf W. and Focks A. (2016). Relative influence of chemical and non-chemical stressors on invertebrate communities: a case study in the Danube River. Science of the Total Environment 571: 1370 – 1382.

Riddle, B.R., Ladle, R.J., Lourie, S.A. and Whittaker, R.J., 2011. Basic biogeography: estimating biodiversity and mapping nature. Conservation biogeography, pp.45-92.

Rosi-Marshall, E.J., Kincaid, D.W., Bechtold, H.A., Royer, T.V., Rojas, M. and Kelly, J.J., 2013. Pharmaceuticals suppress algal growth and microbial respiration and alter bacterial communities in stream biofilms. Ecological Applications, 23(3), pp.583-593.

Sanchez, W., Piccini, B., Maillot-Marechal, E. and Porcher, J.M. (2010). Comparison of two reference systems for biomarker data analysis in a freshwater biomonitoring context. Environment International 36: 377-382.

Sanchez, W., Burgeot, T. and Porcher, J. M. (2013). A novel “Integrated Biomarker Response” calculation based on reference deviation concept, Environmental Science and Pollution Research, 20(5), pp. 2721–2725.

Sánchez-Quiles, D., Marbà, N. and Tovar-Sánchez, A. (2017). Trace metal accumulation in marine macrophytes: Hotspots of coastal contamination worldwide. Science of the Total Environment, 576, 520-527.

Schleier III, J.J., Marshall, L.A., Davis, R.S. and Peterson, R.K.D. (2015). A quantitative approach for integrating multiple lines of evidence for the evaluation of environmental health risks. PeerJ 3: e730; DOI 10.7717/peerj.730

Schroeder, A.L., Ankley, G.T., HKouck, K.A. and Villeneuve, D.L. (2016). Environmental surveillance and monitoring - the next frontiers for high-throughput toxicology. Environmental Toxicology and Chemistry 35 (3), 513-525.

Schroeder, AL., Martinovic-Weigelt, D., Ankley, G.T., Lee, K.E., Garcia-Reyero, N., Perkins, E.J., Schoenfuss, H.L. and Villeneuve, D.L. (2017). Prior knowledge-based approach for associating contaminants with biological effects: A case study in the St. Croix River basin, MN, WI, USA. Environmental Pollution 221, 427-436.

Schlüter-Vorberg, L., Knopp, G., Cornel, P., Ternes, T. and Coors, A. (2017). Survival, reproduction, growth, and parasite resistance of aquatic organisms exposed on-site to wastewater treated by advanced treatment processes. Aquatic Toxicology, 186, 171-179

Schulze, T., Krauss, M., Novak, J., Hilscherova, K., Ait-Aissa, S., Creusot, N., Macova, M., Neale, P., Escher, B.I., Gomes, T., Tollefsen, K.E., Tarcai, Z., Shao, Y., Deutschmann, B., Seiler, T.B., Hollert, H., Tarabek, P., Tousova, Z., Slobodnik, J., Walz, K.H. and Brack, W. (2015). Large volume sampling and effect-based screening In: Liška I, Wagner F, Sengl M, Deutsch K, Slobodník J. Joint Danube Survey 3, Final Scientific Report. International Comission for the Protection of the Danube River, Vienna, Austria. pp. 284-295

Segner, H. (2011). Moving beyond a descriptive aquatic toxicology: the value of biological process and trait information. Aquatic Toxicology 105:50-55.

Segner, H., Schmitt-Jansen, M. and Sabater, S., 2014. Assessing the impact of multiple stressors on aquatic biota: the receptor's side matters. Environmetnal Science and Technology, 48 (14), 7690–7696

Shannon, C.E., 1949. The Mathematical Theory of Communication. Univ. Illinois Press, Urbana, p. 144.

Shinn C, Blanchet S, Loot G, Lek S, Grenouillet G. 2015. Phenotypic variation a an indicator of pesticide stress in gudgeon: accounting for confounding factors in the wild. Science of the total Environment 538: 733-742.

Smith, E.P., Lipkovich, I. and Ye, K. (2002). Weight-of-Evidence (WOE): Quantitative Estimation of Probability of Impairment for Individual and Multiple Lines of Evidence. Human and Ecological Risk Assessment 8(7), 1585-1596.

Stalter, D., Peters, L.I., O'Malley, E., Tang, J.Y.M., Revalor, M., Farré, M.J., Watson, K., von Gunten, U. and Escher, B.I. (2016). Sample Enrichment for Bioanalytical Assessment of Disinfected Drinking Water: Concentrating the Polar, the Volatiles, and the Unknowns. Environmental science & technology, 50(12), pp.6495-6505.

Stankovic, I., Pall, K. and Janauer, G. (2015). Macrophytes In: Liška I, Wagner F, Sengl M, Deutsch K, Slobodník J. Joint Danube Survey 3, Final Scientific Report. International Comission for the Protection of the Danube River, Vienna, Austria. pp. 110-118.

Suter, G.W. II. (1993a). Ecological risk assessment. Boca Raton (FL): Lewis. 538 pp.

Suter, G.W. II. (1993b). A critique of ecosystem health concepts and indices. Environ Toxicol Chem 12: 1533-1539.

Suter, G.W. II. and Cormier, S.M. (2011). Why and how to combine evidence in environmental assessments: Weighing evidence and building cases. Science of the Total Environment 409, 1406-1417.

Theodorakis, C.W., Shugart, L.R. (1997). Genetic ecotoxicology II: population genetic structure in mosquitofish exposed *in situ* to radionuclides. Ecotoxicology 6 (6), 335-354.

Theodorakis, C.W., Lee, K.-L., Adams, S.M., Law, C.B. (2006). Evidence of altered gene flow, mutation rate, and genetic diversity in redbreast sunfish from a pulp-mill-contaminated river. Environmental Science and Technology 40(1), 377-386.

Tlili, A., Marechal, M., Montuelle, B., Volat, B., Dorigo, U. and Bérard, A., (2011). Use of the MicroResp™ method to assess pollution-induced community tolerance to metals for lotic biofilms. Environmental pollution, 159(1), pp.18-24.

Tlili, A., Berard, A., Blanck, H., Bouchez, A., Cássio, F., Eriksson, K. M., Morin, S., Montuelle, B., Navarro, E., Pascoal, C., Pesce, S., Schmitt-Jansen, M. and Behra, R. (2016). Pollution-induced community tolerance (PICT): towards an ecologically relevant risk assessment of chemicals in aquatic systems, Freshwater Biology, 61(12), pp. 2141–2151.

Tlili, A., Dorigo, U., Montuelle, B., Margoum, C., Carluer, N., Gouy, V., Bouchez, A. and Berard, A. (2008). Responses of chronically contaminated biofilms to short pulses of diuron - An experimental study simulating flooding events in a small river. Aquatic Toxicology 87(4), 252-263.

Tlili, A., Hollender, J., Kienle, C. and Behra, R. (2017). Micropollutant-induced tolerance of *in situ* periphyton: Establishing causality in wastewater-impacted streams, Water Research. Elsevier Ltd, 111, pp. 185–194. doi: 10.1016/j.watres.2017.01.016.

Traven L, Mićović V, Vukić Lušić D, Smital T (2013) The responses of the hepatosomatic index (HSI), 7-ethoxyresorufin-O-deethylase (EROD) activity and glutathione-S-transferase (GST) activity in sea bass (Dicentrarchus labrax, Linnaeus 1758) caged at a polluted site: implications for their use in environmental risk assessment. Environmental Monitoring and Assessment, 185(11), 9009-9018.

Triebskorn, R., Adam, S., Behrens, A., Beier, S., Böhmer, J., Braunbeck, T., Casper, H., Dietze, U., Gernhöfer, M., Honnen, W., Köhler, H.R., Konradt, J., Lehmann, R., Luckenbach, T., Oberemm, A., Schwaiger, J., Segner, H., Strmac, M., Schüürmann, G., Siligato, S. and Traunspurger, W. (2003). Establishing causality between pollution and effects at different levels of biological organization: the VALIMAR project. Human and Ecological Risk Assessment 9:171-194.

Trute, M., Gallis, B., Doneanu, C., Shaffer, S., Goodlett, D., Gallagher, E. (2007). Characterization of hepatic glutathione S-transferases in coho salmon (*Oncorhynchus kisutch*). Aquatic Toxicology 81, 126-136.

Tyler, C.R., Spary, C., Gibson, R., Shears, J., Santos, E. and Sumpter, .JP. (2005). Accounting for differences in the vitellogenic responses of rainbow trout and roach exposed to oestrogenic effluents from wastewater treatment works. Environmental Science & Technology 39:2599-2607.

US Environmental Protection Agency (2000). Stressor identification guidance document. Washington DC: Office of Water, Office of Research and Development. EPA/822/B-00/025.

Van den Brink, P.J., Baird, D.J., Baveco, H.J.M. and Focks, A. (2013). The use of trait-based aproaches and eco(toxico)logical models to advance the ecological risk assessemnt framework for chemicals. Integrated Environmental Assessment and Management 9: 47-57.

Van den Brink, P.J., Blake, N., Brock, T. C.M. and Maltby L. (2006). Predictive value of Species Sensitivity Distributions for effects of herbicides in freshwater ecosystems. Journal of Human and Ecological Risk Assessment 12: 645-674.

Van den Brink, P.J., Rubach, M.N., Culp, J.M., Maund, S.J. and Baird, D.J. (2010). Traits-based ecological risk assessment (TERA): realizing the potential of ecoinformatics approaches in ecotoxicology. Integrated Environmental Assessment and Management 7, 169-171.

Van den Brink, P.J., Sibley, P.K., Ratte, H.T., Baird, D.J., Nabholz J.V. and Sanderson H. (2008). Extrapolation of effects measures across levels of biological organization in ecological risk assessment. In: Extrapolation practice for ecological effect and exposure characterization of chemicals, Solomon, K.R., T.C.M. Brock, D. De Zwart, S.D. Dyer, L. Posthuma, S.M. Richards, H. Sanderson, P.K. Sibley and P.J. Van den Brink (eds.) SETAC, CRC press. Boca Raton FL, USA.

Van der Oost, R., Beyer, J. and Vermeulen, N.P.E., (2003). Fish bioaccumulation and biomarkers in environmental risk assessment: a review. Environmental Toxicology and Pharmacology 13, 57-149.

van Straalen, N.M. and Timmermans, M. (2002). Genetic variation in toxicant-stressed populations: An evaluation of the “Genetic Erosion” hypothesis. Human and Ecological Risk Assessment 8(5), 983-1002.

Van Wijngaarden, R.P.A., T.C.M. Brock and P.J. Van den Brink. (2005). Threshold levels for effects of insecticides in freshwater ecosystems, a review. Ecotoxicology 14: 353-378.

Verbrugge, L.N.H., Schipper, A.M., Huijbregts, M.A.J., Van der Velde, G. and Leuven, R.S.E.W. (2012). Sensitivity of native and non-native mollusc species to changing river water temperature. Biological Invasions 14, 1187–1199.

Verschoor, A.M., Vos, M. and van der Stap, I. (2004). Inducible defences prevent strong population fluctuations in bi- and tritrophic food chains. Ecol. Lett. 7, 1143-1148.

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

Vos, P.J., Meelis, E. and ter Keus, W.K. (2000). A framework for the design of ecological monitoring programs as a tool for environmental and nature management. Environ Monitor Assessm 61: 317-344.

Vrana, B., Smedes, F., Rusina, T., Okonski, K., Allan, I., Grung, M., Hilscherova, K., Novak, J., Tarabek, P. and Slobodnik, J. (2015). Passive sampling: chemical analysis and toxicological profiling In: Liška I, Wagner F, Sengl M, Deutsch K, Slobodník J. Joint Danube Survey 3, Final Scientific Report. International Commission for the Protection of the Danube River, Vienna, Austria. pp. 304-318

Wang, C., Wang, X., Su, R., Liang, S. and Yang, S., (2011). No detected toxic concentrations in *in situ* algal growth inhibition tests—a convenient approach to aquatic ecotoxicology. Ecotoxicology and environmental safety, 74(3), pp.225-229.

Weed, D. (2005). Weight of evidence: a review of concepts and methods. Risk Analysis 25(6), 1545-1557.

Wiegleb, G., Gebler, D., van de Weyer, K. and Birk, S. (2016). Comparative test of ecological assessment methods of lowland streams based on long-term monitoring data of macrophytes. Science of the Total Environment 541, 1269–1281

Whitehead, A., Anderson, S.L., Kuivila, K.M.L., Roach, J., May, B. (2003). Genetiv variation among interconnected populations of Catostomus occidentalis: implications for distinguishing impacts of cantaminants from biogeographical structuring. Molecular Ecology 12(10), 2817-2833.

Whyte, J. J., Tillit, D.E. (2000). Ethoxyresorufin-O-deethylase (EROD) activity in fish as a biomarker of chemical exposure. Critical Reviews in Toxicology 30,347–570.

Wolfram, G., Höss, S., Orendt, C., Schmitt, C., Adamek, Z., Bandow, N., Großschartner, M., Kukkonen, J.V.K., Leloup, V., Lopez Doval, J.C., Munoz, I., Traunspurger, W., Tuikka, A., Van Liefferinge, C., von der Ohe, P.C. and de Deckere, E. (2012). Assessing the impact of chemical pollution on benthic invertebrates from three different European rivers using a weight-of-evidence approach. Scince of the Total Environment 438, 498-509.

Wu, N., Dong, X., Liu, Y., Wang, C., Baattrup-Pedersen, A. and Riis, T. (2017) 'Using river microalgae as indicators for freshwater biomonitoring: Review of published research and future directions', Ecological Indicators. Elsevier, 81(March), pp. 124–131. doi: 10.1016/j.ecolind.2017.05.066.

Zelinka, M., Marvan, P., 1961. Zur Prazisierung der biologischen Klassifikation der Reinheit flie Bender Gewasser. Arch. Hydrobiol. 57, 389–407.