Advancements in effect-based water quality assessment

de Baat, M.L.

CHAPTER 6

THIS CHAPTER IS BASED ON THE PAPER:

ADVANCEMENTS IN EFFECT-BASED SURFACE WATER QUALITY ASSESSMENT

Published in: Water Research (2020), 183: 116017
ML de Baat, R van der Oost, GH van der Lee, N Wieringa, T Hamers, PFM Verdonschot, P de Voogt, MHS Kraak

ABSTRACT

Legally-prescribed chemical monitoring is unfit for determining the pollution status of surface waters, and there is a need for improved assessment methods that consider the aggregated risk of all bioavailable micropollutants present in the aquatic environment. Therefore, the present study aimed to advance effect-based water quality assessment by implementing methodological improvements and to gain insight into contamination source-specific bioanalytical responses. Passive sampling of non-polar and polar organic compounds and metals was applied at 14 surface water locations that were characterized by two major anthropogenic contamination sources, agriculture and WWTP effluent, as well as reference locations with a low expected impact from micropollutants. Departing from the experience gained in previous studies, a battery of 20 *in vivo* and *in vitro* bioassays was composed and subsequently exposed to the passive sampler extracts. Next, the bioanalytical responses were compared to effect-based trigger values to identify ecotoxicological risks. The bioanalytical assessment of the joint risks of metals and (non-)polar organic compounds resulted in the successful identification of pollution source-specific ecotoxicological risk profiles. Cumulative ecotoxicological risks were lowest for reference locations, followed by agriculture locations and the highest for WWTP locations, and were mainly driven by polar organic contaminants. It is concluded that the presently employed advanced effect-based methods can readily be applied in surface water quality assessment and that the integration of chemical- and effect-based monitoring approaches will foster future-proof water quality assessment strategies on the road to a non-toxic environment.

INTRODUCTION

Surface waters are contaminated with an increasing diversity of anthropogenic compounds, giving rise to the presence of complex contaminant mixtures that can cause serious harm to aquatic ecosystems.[3,7,41] Legislations like the European Water Framework Directive (WFD)[36] and the United States Clean Water Act (CWA)[192] aim to protect surface waters from human impacts by the implementation of chemical and ecological water quality criteria. However, the separate interpretations of the chemical and ecological status of water bodies often yield divergent water quality management advice, which poses practical problems for the implementation of measures to protect surface waters from further degradation.[6] As a result, there is a growing consensus among scientists and authorities that the methods currently used for chemical and ecological water quality assessment require a revision to obtain a more coherent and future-proof approach.[79] Traditionally, chemical water quality is assessed by the monitoring of concentrations of a limited list of individual priority compounds. However, environmental concentrations of these compounds are decreasing, and consequently, currently identified risks to aquatic ecosystems are caused by complex mixtures of (un)known, unregulated and unmonitored compounds.[25,87] Hence, the legally-prescribed strategies are unfit for the monitoring of chemical pollution of surface waters, and there is thus a need for improved assessment methods that consider the aggregated risk of all bioavailable micropollutants present in the aquatic environment. Consequently, there is an increasing interest in the use of bioanalytical tools in environmental quality assessment.[30,46,87] Bioanalytical responses to environmental samples are caused by the combined action of all bioavailable mixtures of (un)known compounds and their metabolites present in the sample, thereby overcoming the limitations posed by chemical analysis of a limited number of target compounds.[86]

Effect-based strategies have been successful in the identification of ecotoxicological risks in surface waters and the ranking of locations based on these risks.[11,14,15,48,88] Nonetheless, clear suggestions for further improvements of the applied methods were also made, including the addition of environmentally relevant compound groups and toxicity endpoints.[88] Moreover, the complex and diluted pollution present at previously studied locations, often larger bodies of water like lakes and rivers, made it difficult to identify land use- and contamination source-specific bioanalytical response profiles. A better understanding of contamination source-specific response profiles can aid in the application of mitigation efforts following from effect-based water quality assessment. Hence, refinement of the current methods and an improved interpretation of bioanalytical responses is recommended for the implementation of effect-based methods in regulatory frameworks like the CWA and the WFD.[193] The present study aimed to advance effect-based water quality assessment by implementing methodological improvements and to gain insight into contamination source-specific bioanalytical responses. To this end, the presently applied monitoring strategy combined passive sampling, a battery of *in vivo* and *in vitro* bioassays and effect-based trigger values (EBTs) to screen for potential ecotoxicological risks in surface waters.

The methodological improvements explored here were the bioanalytical risk assessment of metals and the streamlining of previously used bioassay batteries to represent those endpoints most relevant to aquatic ecosystem health.[88] Due to a strong focus on emerging organic contaminants, metals have only rarely been included in the combination of passive sampling and bioanalytical assessment of chemical surface water quality,[82] despite their potential toxicity. Therefore, in the present study, bioanalytical risk assessment of metals was integrated with that of organic contaminants. Furthermore, to simultaneously investigate the increasing risk of polar compounds in aquatic ecosystems,[85] the *in vivo* bioassays were performed not only on non-polar organic extracts, as in previous studies, but also on polar organic and metal extracts. The streamlining of the bioassay battery followed from the experience gained in previous studies[88] and resulted in the exclusion of tests that were previously unresponsive in surface water quality assessment (GR CALUX, antibiotics waterSCAN and algal growth inhibition) and their replacement with relevant and responsive endpoints (*anti*-PR CALUX and algal photosynthetic inhibition).

Bioassay battery responses for the investigated locations were used to gain insight in contamination source-specific toxicity profiles and the potential risks they pose to aquatic ecosystems. Therefore, locations were selected that were characterized by two major anthropogenic contamination sources, agriculture and WWTP effluent, as well as reference locations with a low expected impact from micropollutants.

MATERIAL & METHODS

Sampling locations

Sampling locations were selected in collaboration with nine Dutch regional water authorities. This resulted in a set of 14 lowland streams and drainage ditches in The Netherlands within three location types (Figure S1 and Table S1), either surrounded by ornamental flower bulb horticulture (horticulture; n=5), directly receiving WWTP effluent (WWTP; n=4), or reference locations with no known contamination sources (reference; n=5). The locations were comparable in width, depth and flow velocity (Table S2). Sampling was conducted between August 20th and October 5th, 2018.

Passive sampler deployment, extraction and sampled volume estimation

Passive sampling devices

Silicone rubber (SR) sheets, with a weight of 20 g per set of six sheets, spiked with performance reference compounds (PRCs), were obtained from Deltares (Utrecht, The Netherlands) and applied for the sampling of non-polar compounds.[194] Polar organic chemical integrative samplers (POCIS) containing 0.2 g of Oasis hydrophilic-lipophilic balance sorbent (HLB; Waters, Etten-Leur, The Netherlands) were constructed in the laboratory at the University of Amsterdam (SI 2) and applied for the sampling of compounds in the more polar range.[92] Diffusive gradients in thin-films (DGT) containing a 0.15 mL mixed chelex and TiO_2 (Metsorb) binding layer were obtained from DGT Research (Lancaster, UK) and applied for the sampling of metals from the surface water.[195]

The samplers were transported to the study sites in airtight packaging at 4˚C. Unexposed blanks of all sampler types were included in all subsequent analyses. Additional information on passive sampler construction, extraction and sampled volume calculation is given in SI 2.

Field deployment of passive samplers

SR sheets and POCIS were deployed simultaneously at each sampling location in separate stainless steel cages. The mesh size of the cages allowed a largely unobstructed flow of water around the samplers. Cages with samplers were suspended in the middle of the water column to ensure permanent inundation of the samplers, while avoiding direct diffusion of compounds from the sediment to the samplers. Per location, six SR sheets and four POCIS were exposed for a period of six wk. After exposure, the samplers were cleaned in the field with local water and a scrubbing sponge to remove biofouling, transported to the laboratory on ice and stored at –20 °C until extraction.

Three DGTs per location were deployed for two wk, halfway through the POCIS and SR deployment period. DGTs were retained in polyacrylate holders in the middle of the water column. After exposure, DGTs were rinsed in the field with deionized water, transported to the laboratory on ice and stored at 4°C until extraction.

Extraction of SR

All equipment used in the SR extraction procedure was cleaned with acetone and LC grade acetonitrile (Biosolve, The Netherlands) before use. SR sheets were thawed and dried and the six sheets per location were folded and stacked in a harmonica shape to maximize the surface contact area with the extraction solvent and placed at the bottom of a 150 mL Erlenmeyer flask (Figure S2). After the addition of 75 mL LC grade acetonitrile, the flasks were closed and placed on a shaker for 2 d at 110 rpm. Extracts were stored at 4°C and the extraction procedure was repeated once more. Both extracts were combined in round bottom flasks and evaporated on a Büchi Rotavapor system (Flawil, Switzerland) at 45°C and 117 mbar to approximately 5 mL. The extracts were subsequently transferred to glass vials, filled up to exactly 10 mL with LC grade acetonitrile by weight and stored at -20°C until analyses.

Extraction of POCIS

Frozen POCIS were freeze-dried overnight at −53°C in a Scanvac CoolSafe freeze-dryer. All equipment used in the POCIS extraction procedure was cleaned with acetone and LC grade acetonitrile before use. Each POCIS was disassembled and the dry sorbent of the four POCIS that were exposed per location was pooled and transferred to a 6 mL glass Supelco SPE column with Teflon frit (Sigma-Aldrich, The Netherlands) using a glass funnel. The mass of the recovered sorbent per location was recorded with an analytical balance. The SPE columns were placed on an SPE manifold and eluted three times with 3 mL LC grade acetonitrile under vacuum. Finally, the extracts were topped up to exactly 10 mL with acetonitrile by weight and stored at -20°C until analyses.

Extraction of DGT

All equipment used in the DGT extraction procedure was acid cleaned with 0.1 M HNO_3 and ultrapure water. The three DGTs per location were disassembled and their binding layers combined in 3 mL of 1.0 M HNO_3, extracted for 24 h at room temperature, after which the extracts were stored at 4°C until analyses.

Estimation of sampled water volumes

SR

Sampling rates for SR were calculated based on the rate of PRC dissipation from the sheets during the field exposure.[50] PRC chemical analysis was performed at the laboratory of TNO (Utrecht, The Netherlands; analytical details provided in SI 2). Subsequently, 50% of the calculated sampling rate for each location was used as a provisional estimation of the average extracted water volume per day, based on the assumption that 50% of the organic contaminants present in the surface water reach equilibrium with the SR during field exposure, as described by Van der Oost *et al*. (2017).[14]

POCIS

The reported average sampling rate for POCIS of 0.18 L/d,[52] that was previously successfully applied in combination with effect-based water quality assessment,[88] was used to determine the concentration factor of the field deployed POCIS to compare bioassay responses between sites. A correction for the HLB sorbent recovery was applied to incorporate sorbent loss during the extraction procedure. To this end, the remaining sorbent mass after extraction was divided by the initial sorbent mass (0.8 g for four POCIS) and the total estimated volume per location (30.24 L for four POCIS) was multiplied by this fraction to obtain a final sampled volume and to ensure an impartial comparison between locations.

DGT

Since no general approach for the interpretation of bioassay results in combination with DGT extracts was available,[82] a novel approach to determine sampled volumes of DGT samplers was presently developed. By using sampled water volumes for toxicity interpretation, this new approach is now in line with that for organic extracts. The sampling rate for the DGT samplers was determined using a theoretical approach, as well as an approach based on the detected masses of metals that had accumulated in the samplers (calculations provided in SI 2). Both approaches rely on DGT theory, as outlined in numerous publications that confirm the usability of DGTs to obtain time-weighted average field concentrations of metals [e.g. Allan *et al.* (2007)[196]; Davison and Zhang (2012)[197]] Both approaches resulted in very similar outcomes and a mean sampling rate for three DGT samplers of 44.9 mL/d was used in the interpretation of the bioassay responses.

Bioassay battery

A battery of 20 bioassays (i.e. 20 unique bioassay x passive sampler extract combinations) was applied for the detection of ecotoxicological effects at the investigated locations (Table S5). The whole organism *Daphnia* and PAM tests were performed at the laboratory of the University of Amsterdam, and the *Aliivibrio fischeri* bioluminescence inhibition assay was performed at the laboratory of the Vrije Universiteit Amsterdam. The *in vitro* CALUX assays were performed at the BioDetection Systems laboratory (Amsterdam, The Netherlands).

Sample pre-treatment

Organic extracts were transferred to dimethyl sulfoxide (DMSO) before application in the bioassays. To this end, the extracts were evaporated to dryness under N_2 flow at room temperature and redissolved in DMSO. Bioassays with organic extracts were performed at a 0.1-1% DMSO concentration to improve compound solubility in the exposure media and a control was always included to confirm the non-toxicity of the solvent. Inorganic extracts were freeze-dried overnight at −53°C in a Scanvac CoolSafe freeze-dryer and redissolved in exposure medium before exposure in the bioassays, to eliminate the HNO_3 from the extracts. Full recovery of metal concentrations using this sample treatment method was confirmed in a separate experiment using internal standards (data not shown).

Whole organism bioassays

The whole organism *Aliivibrio fischeri* bioluminescence inhibition, *Daphnia* and PAM bioassays were performed on dilution series of the extracts of all three passive samplers, resulting in nine *in vivo* responses. The *Aliivibrio fischeri* bioluminescence inhibition assay (further referred to as bacterial bioluminescence assay) was performed according to Hamers *et al.* (2001).[95] Luminescence inhibition was measured after 15 minutes of exposure to the passive sampler extracts. The *Daphnia* test was performed with *D. magna* (<24 h) originating from an in house culture, according to OECD guideline 202 with reduced test volumes, as previously described.[14] Daphnid immobilization was recorded after 48 h of exposure. The PAM test was performed using the freshwater microalga *Raphidocelis subcapitata* originating from an in house culture, according to de Baat *et al.* (2018).[72] Photosynthetic inhibition was measured after 4.5 h of exposure.

CALUX assays

The passive sampler extracts were analysed by a panel of *in vitro* CALUX® bioassays. Specific CALUX assays were performed on either non-polar (SR) or polar (POCIS) organic extracts. SR extracts were subjected to DR, PAH, PPARγ, Nrf2, PXR and p53 (without S9 metabolism) assays and POCIS extracts were subjected to ERα, *anti*-AR and *anti*-PR assays, according to previously described protocols.[96] The DR CALUX assay was performed with a sulfuric acid clean-up step to eliminate degradable compounds (e.g. PAHs) and to isolate the persistent compounds (e.g. dioxins and dioxin-like polychlorinated biphenyls). Cytotoxicity of the CALUX cells was monitored in both POCIS and SR extracts to rule out confounding influences on test outcomes.

Data analysis

Bioanalytical effect expression

Toxicity in the *in vivo* and genotoxicity assays was expressed as toxic units (TU), wherein one TU represented the dilution at which the extract caused 50% effect for the respective endpoints (EC_{50}). EC_{50} values were determined by nonlinear regression analysis with the built-in log logistic model in GraphPad Prism® (GraphPad Software Inc., v. 5.00, San Diego, CA, USA). Responses in the *in vitro* assays were expressed as concentrations of bioanalytical equivalents (BEQ) of the reference compounds (Table S5). Next, the bioassay responses were expressed as TU (*in vivo*) or BEQ concentrations (*in vitro*) and corrected for the estimated sampled water volumes of the passive samplers to represent the TU and BEQ/L at the sampling locations.

Risk interpretation using EBTs

Bioanalytical responses were compared to EBTs for ecotoxicological risk interpretation. EBTs reported by Van der Oost *et al.* (2017)[21] were used, unless more recently derived EBTs were available, which was the case for the ERα[32] and *anti*-AR[30] CALUX assays. For the PAH and PXR CALUX assays, strongly divergent EBTs were reported by Van der Oost *et al.* (2017)[21] and Escher *et al.* (2018),[30] hampering consolidated conclusions on ecotoxicological risks for these endpoints.[96] Therefore, the influence of the EBTs on the risk interpretation for these tests was explored in the present study, and intermediate values were derived based on the methods outlined by Escher *et al.* (2018)[30] as described in SI 4 (PAH 62.1 ng BEQ/L; and PXR 5.4 µg NEQ/L). Additionally, a preliminary EBT was derived for the *anti*-PR CALUX assay (13 ng Ru486 eq./L) based on the value previously reported by Escher *et al.* (2018),[30] as the reported reference compound differed from the one used in the present study (SI 4).

Since no EBTs were previously defined for the application of DGT extracts in bioassays, a preliminary EBT of 0.05 TU was presently derived for all three *in vivo* bioassays based on the approach outlined by Van der Oost *et al.* (2017)[21] (SI 4). This allowed for the interpretation of the bioassay responses to the DGT extracts in line with the approach for the organic extracts.

The responses of all bioassays were divided by their respective EBTs to obtain an effect-based risk quotient, where a quotient ≥1 represents a potential ecotoxicological risk indicated by that particular bioassay. These effect-based risk quotients were used for two purposes: i) The sum of these values yielded a cumulative ecotoxicological risk (Σ effect-based risk quotient) for each location, and ii) the quotients were subjected to multivariate analysis to gain insight into location type-specific ecotoxicological response profiles. To this end, non-metric multidimensional scaling (nMDS) was performed in R (R Core Team, v. 3.6.1, Vienna, Austria) using the 'metaMDS' function in the 'vegan' package, based on dissimilarities calculated with the Bray–Curtis index. Statistical differences between the location types were investigated using an analysis of similarity (ANOSIM) using the 'anosim' function. The 'multipatt' function (with r.g association function, 9999 permutations, and α=0.05) in the 'indicspecies' package was then used to perform a multilevel pattern analysis to identify the bioassays that were significantly associated with the different location types.

RESULTS

Bioassay response frequencies

All 20 unique bioassay x passive sampler extract combinations were successfully performed and all assays met their respective validity criteria. Responses of all bioassays for all locations, converted to surface water concentrations, are given in SI 5. Next, it was determined how frequently the different extract x bioassay combinations resulted in the detection of potential ecotoxicological risks (Figure 6.1). Bioassay responses were categorized as no response, or a response below or above the EBT of that test. The response frequencies ranged from no response at all locations for the *Daphnia* assay exposed to metal extracts to EBT exceedance at >75% of locations for the ERα and anti-AR CALUX assay, which were exposed to polar extracts. Out of the battery of 20 bioassays, 11 showed responses above their EBTs. Hence, 55% of the applied bioassays indicated the presence of a potential ecotoxicological risk at one or multiple locations. The most responsive assays (EBT exceedance at >50% of locations) were the ERα, *anti*-AR, PXR and *anti*-PR CALUX assays. The least responsive assays (no response at >50% of locations) were the DR and PPARγ CALUX assays and the *Daphnia* assay exposed to non-polar and metal extracts, and the PAM algae bioassay in combination with all three extracts.

Bioassay battery response profiles

EBT exceedances were observed for all location types, including the reference locations (Figure 6.2). The cumulative effect-based risk quotients allowed the ranking of sites based on

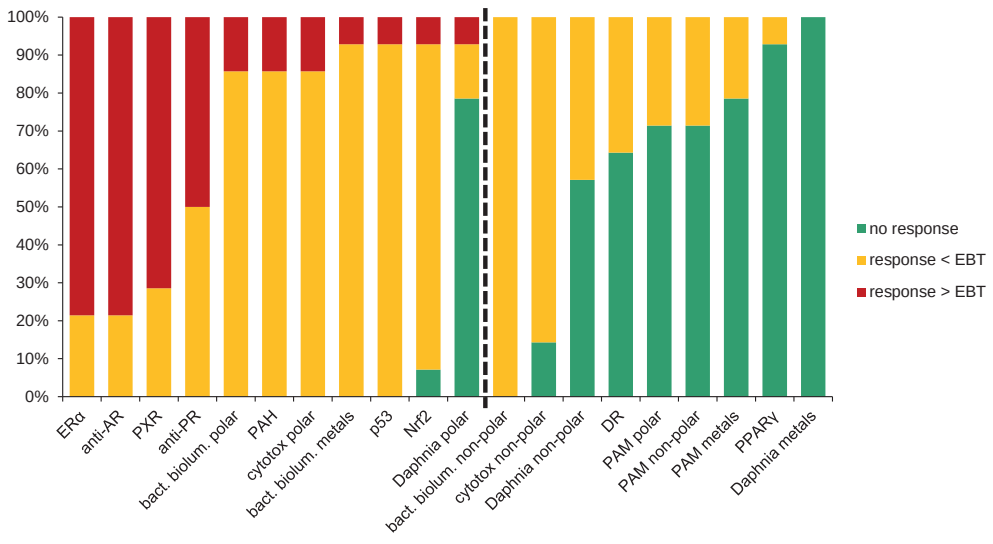

Figure 6.1. Frequency of responses of a panel of 20 bioassays to passive sampler extracts from 14 surface water locations. Colours indicate the bioassay responses and effect-based trigger value (EBT) exceedances at the percentage of study locations. The dashed line indicates the division between bioassays with and without EBT exceedance in the present study.

the potential ecotoxicological risks, and the specific bioassay battery response profiles gave insight into the compound groups responsible for the risks at each location. Reference locations exhibited the lowest cumulative effect-based risk quotients (4.3 – 10.9), followed by horticulture locations (11.3 – 27.2) and WWTP locations (12.8 – 47.7). On average, EBTs were most frequently exceeded at horticulture locations (22% of bioassays), followed by WWTP locations (18%) and least frequently at reference locations (13%).

The nMDS ordination showed that the locations could be grouped based on the location type (Figure 6.3; stress = 0.086), and the ANOSIM test confirmed that the bioassay battery response profiles differed significantly between location types (ANOSIM statistic R = 0.6414, p = 0.0001). The multilevel pattern analysis revealed that none of the bioassays were significantly associated with reference locations, nor were any bioassays significantly associated with multiple location types. Contrastingly, horticulture locations were significantly characterized by responses in the *anti*-PR (stat = 0.962, p = 0.0001), cytotoxicity (polar: stat = 0.811, p = 0.0014), and *anti*-AR (stat = 0.651, p = 0.0052) CALUX assays. WWTP locations on the other hand were significantly characterized by responses in the bacterial bioluminescence assay (polar: stat = 0.899, p = 0.0006; metals: stat = 0.548, p = 0.0036), ERα CALUX (stat = 0.845, p = 0.0006), *Daphnia* (non-polar: stat = 0.713, p = 0.0087; polar: stat = 0.674, p = 0.0106), and PAM algae (polar: stat = 0.663, p = 0.021) assays.

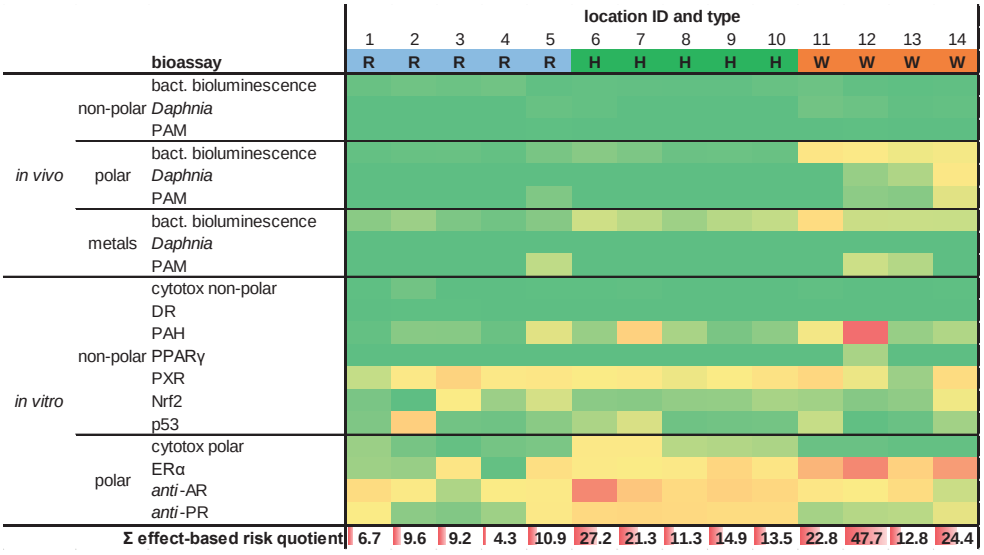

Figure 6.2. Heat map depicting the fold effect-based trigger value exceedance (effect-based risk quotient) for 20 bioassays and sum effect-based risk quotients at 14 surface water locations impacted by flower bulb horticulture (H) and wastewater treatment plant (W) effluent and for reference (R) locations. Effect-based risk quotients are depicted as follows: Green = 0, yellow =1, red = maximum value.

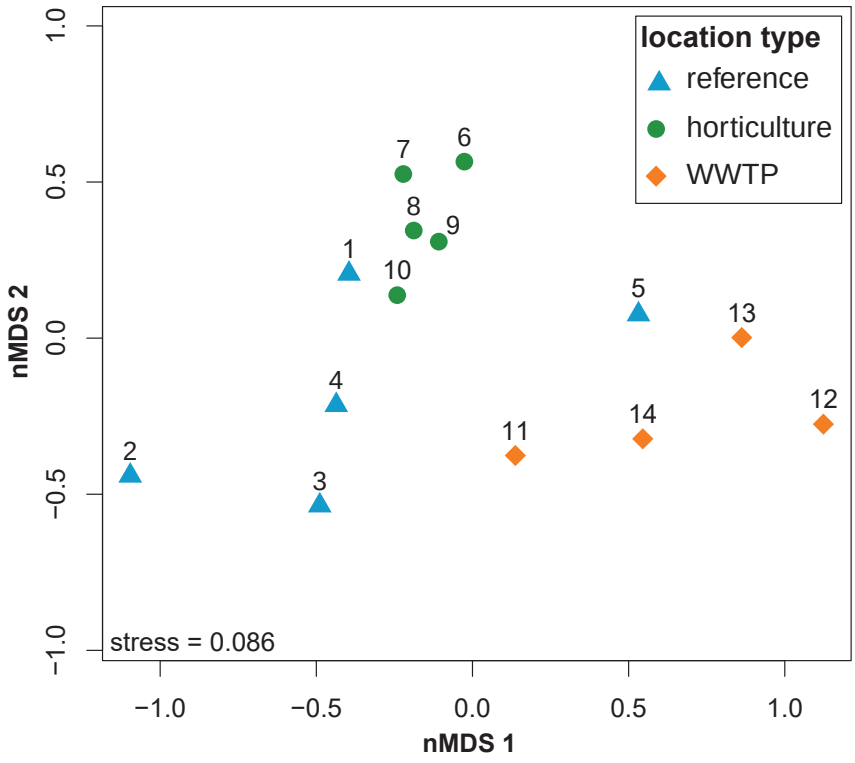

Figure 6.3. Non-metric multidimensional scaling (nMDS) plot depicting the difference in bioassay battery responses for 14 surface water locations impacted by flower bulb horticulture and wastewater treatment plant (WWTP) effluent and for reference locations, where points closer together represent a more similar bioassay response profile than those further apart.

DISCUSSION

Methodological improvements for a better ecotoxicological risk identification

Bioanalytical risk assessment of metals

The identification of ecotoxicological risks in effect-based surface water quality assessment depends strongly on the applied sampling methodology. Only compounds that are captured by the applied sampling methods, present at concentrations above bioanalytical detection limits, will elicit effects in the bioassays, highlighting the importance of effective sampling strategies that ensure the sequestration of a wide range of compounds.[89] Passive sampling is often used in combination with bioassays, as it allows for the sampling of a wide variety of bioavailable compounds and simultaneously concentrates the water, resulting in lower bioanalytical detection limits.[79] However, effect-based strategies often have a strong focus on organic contamination and only rarely have metals been included in the combination of passive sampling and bioanalytical assessment of chemical surface water quality.[82]

In the present study, passive sampling of metals was applied in combination with three *in vivo* bioassays, matching the approach used for the bioanalytical risk assessment of organic compounds. Toxic effects of the metal extracts were observed in the PAM algae and bacterial bioluminescence bioassays and comparison of the effects to the presently derived EBTs elucidated potential risks to bacteria by metals at WWTP locations, highlighting the relevance of effect-based risk assessment of metals in surface waters. As shown here, this novel approach can easily be merged with existing effect-based monitoring strategies to include the bioanalytical assessment of risks of bioavailable metal concentrations in aquatic systems.

Streamlining of previously used bioassay batteries to better represent endpoints relevant to aquatic ecosystem health

To encompass a wide range of responsive endpoints that are representative of micropollutant risks in surface waters, several adjustments to previously applied bioassay batteries were made. The revised battery allowed for the detection of potential ecotoxicological risks caused by the presence of metals and polar and non-polar organic compounds. The addition of the *anti*-PR CALUX assay resulted in the detection of potential ecotoxicological risks at 50% of the investigated locations and is thus a relevant addition to previously applied bioassay batteries.[14,88] Furthermore, performing the three *in vivo* assays not only on non-polar organic extracts but also on polar organic and metal extracts elucidated potential ecotoxicological risks of polar compounds and metals that would have otherwise gone undetected. This is in line with the study of Hamers *et al.* (2018),[15] who found generally higher *in vivo* responses to polar extracts than to non-polar extracts, and reflects the expected increased risk caused by the increasing presence of polar compounds in surface waters.[85] However, to meet the monitoring requirements that are related to future shifts in the chemical properties of contaminants of emerging concern (CECs), effect-based monitoring strategies should be open to further modifications and improvements. Improved (passive) sampling techniques for highly polar as well as ionizable organic compounds,[102,103] combined with bioassays responsive to such compounds, should result in future-proof solutions that allow for risk assessment of these CECs.

Considering assays that were not responsive in the currently applied bioassay battery, the presently observed lack of DR CALUX activity is in line with previous predictions that dioxins and dioxin-like compounds do not contribute substantially to the risks of organic micropollutants in surface waters.[88] Therefore, the inclusion of the DR CALUX assay in bioassay batteries for surface water quality monitoring appears to present little relevance. However, as the sediment is the ultimate sink for dioxins and as such also represents a repository for legacy contamination with dioxins, the use of the DR CALUX assay in sediment quality assessment remains relevant.

In the present study, the traditional algal growth inhibition test was substituted by the PAM algae bioassay, which was expected to better elucidate the frequent presence of herbicides in surface waters.[70,72] However, the assay never showed an EBT exceedance and was, in fact, one of the least responsive assays in the battery. Nonetheless, the PAM algae assay gave a response

at ~29% of the locations, which is a substantial increase compared to the previously observed response frequency of only 4% in the standard 72 h algae growth inhibition test.[88] The lack of responses that exceed the EBT may be attributable to an actual low risk caused by herbicides in surface waters in The Netherlands,[198] at least at the sites presently sampled in late summer. In many other intensive agricultural areas, however, the presence of hazardous concentrations of herbicides has been reported[198] and hence, an even more sensitive algal bioassay may better elucidate the risks of herbicides in surface waters in effect-based monitoring strategies.[199]

The presently applied bioassay battery represents endpoints at all organizational levels that are relevant to aquatic ecosystem health, as was proposed for holistic effect-based water quality assessment by Neale *et al*. (2017).[25] Yet although it spans a wide variety of relevant endpoints, some gaps remain in terms of the identification of groups of compounds that are contamination source-specific and are expected to potentially cause serious harm to aquatic ecosystems, most notably pesticides[200,201] and antibiotics.[77,202] Pesticides, in general, do elicit toxic responses in *anti*-AR and *anti*-PR assays, amongst others.[16,104,105] However, other endocrine-disrupting compounds, like pharmaceuticals and flame retardants, can also elicit responses in such assays.[16] Hence, attributing the observed effects to specific compounds requires confirmation either by highly specific bioassays or by chemical analysis. For example, the PAM algae and *Daphnia* bioassays can help tease out the effects of herbicides and insecticides, respectively. However, specific effects of fungicides are as of yet not covered in the bioassay battery, and expansion of the battery with fungal bioassays should allow for the isolation of fungicide toxicity. Similarly, the bacterial bioluminescence assay responds to toxicity caused by certain antimicrobials, but will also respond to a multitude of other compounds with specific and narcotic modes of action, and is not able to isolate the effects of antibiotics. Highly specific bacterial reporter assays that can elucidate the activity of specific groups of antibiotics are currently being developed,[203] yet the lack of available EBTs presently stands in the way of their application in bioassay batteries.

Bioassays that allow for the identification of compound group-specific effects will strongly aid in the identification of the responsible compounds in subsequent chemical analysis by narrowing down the list of suspects. Promising setups have been developed in recent years that combine ecologically relevant *in vivo* bioassays with liquid chromatography to obtain high-throughput setups for effect-directed analysis of ecologically relevant contaminants. This is an approach with a high applicability in effect-based water quality monitoring strategies like the present. For all three *in vivo* bioassays that were applied in the present study, effect-directed analysis approaches were recently developed that can thus be readily implemented into effect-based monitoring strategies to aid in the identification of causative compounds.[199,204,205] In the future, bioassay battery compositions can be tailored to include relevant assays depending on research or monitoring aims and to anticipate the ever-changing nature of chemical pollution of surface waters.

The influence of effect-based trigger values on the outcome of bioanalytical surface water quality assessment

EBTs are critical in the determination of the significance of effects observed in bioassay batteries. Similar to what environmental quality standards represent for single compounds, EBTs indicate predicted no-risk levels for mixtures of compounds that are present in environmental samples. This highlights the importance of the establishment of reliable EBTs, a field of research that is gaining traction in recent years.[21,30–32] Although there is consensus on the EBTs for many bioassays, for several, strongly divergent EBTs are reported, hindering consolidated conclusions on ecotoxicological risks for those endpoints.[96] This is most strikingly the case for the PAH and PXR CALUX assays, for which the EBTs derived by Van der Oost *et al.* (2017)[14] and Escher *et al.* (2018)[30] differ substantially (PAH 150 vs. 6.2 ng BEQ/L; and PXR 3 vs. 54 µg NEQ/L, respectively). Therefore, in the present study, the influence of the EBTs of these two assays on the ecotoxicological risk assessment was investigated by comparing the resulting number of EBT exceedances and effect-based risk quotients for all investigated locations (Table 6.1). Additionally, to merge the divergent EBTs, preliminary empirical intermediate EBTs for both assays are presently proposed and used in the final effect-based risk assessment (PAH 62.1 ng BEQ/L; and PXR 5.4 µg NEQ/L). For these two assays, it appears that the activity, except for two locations where the PAH CALUX assay exhibited very high responses, is uniformly present at all the investigated locations. The application of the different EBTs clearly illustrates their large and divergent impact on the resulting risk interpretation. The Van der Oost *et al.* (2017)[14] values would result in almost no EBT exceedance for the PAH CALUX and exceedance at almost all locations for the PXR CALUX. Contrastingly, the Escher *et al.* (2018)[30] values would result in EBT exceedances at almost all locations for the PAH CALUX and no exceedance at all for the PXR CALUX. Whether the presently proposed intermediate EBTs are, in fact, more representative of the risks of non-specific chemical stress and PAHs in surface waters is to be determined in future research.

The present exploration of the influence of EBTs on the outcome of effect-based risk assessments highlights the need for a consensus on EBTs for a unified application in environmental monitoring frameworks. The continuation of empirical research, that links bioassay responses with adverse effects on the ecological status of water bodies, is expected to further develop the scientific basis that is necessary for the reliable derivation of environmentally relevant EBTs. Nonetheless, bioanalytical responses are absolute and can be compared and ranked between locations and between studies, regardless of the availability of EBTs for risk interpretation. Moreover, for spatiotemporal monitoring of ecotoxicological risks, currently obtained bioanalytical responses can retroactively be compared to refined EBTs that may be developed in the future. Hence, the current lack of a consensus on EBTs for a few bioassays is no practical limitation to the wide application of effect-based tools in surface water quality assessment.

Table 6.1. Side-by-side comparison of effect-based risk quotients for PAH and PXR CALUX assays at 14 surface water locations for effect-based trigger (EBT) values reported by Van der Oost *et al.* (2017)[21] and Escher *et al.* (2018)[30] and preliminary EBT values derived in the present study. EBT exceedances are indicated with a grey cell fill. WWTP = wastewater treatment plant.

		PAH effect-based risk quotient			PXR effect-based risk quotient		
location ID	location type	a	b	c	a	b	c
1	reference	0.0	0.5	0.1	1.3	0.1	0.7
2	reference	0.1	3.2	0.3	2.7	0.1	1.5
3	reference	0.1	3.1	0.3	8.0	0.4	4.5
4	reference	0.0	1.0	0.1	2.4	0.1	1.4
5	reference	0.4	8.6	0.9	3.1	0.2	1.7
6	horticulture	0.2	4.2	0.4	1.8	0.1	1.0
7	horticulture	2.0	47.7	4.8	2.7	0.1	1.5
8	horticulture	0.2	5.3	0.5	1.7	0.1	0.9
9	horticulture	0.1	2.2	0.2	2.0	0.1	1.1
10	horticulture	0.1	3.6	0.4	4.2	0.2	2.4
11	WWTP	0.4	9.6	1.0	6.7	0.4	3.7
12	WWTP	9.5	230.2	23.0	1.7	0.1	0.9
13	WWTP	0.2	4.1	0.4	0.8	0.0	0.4
14	WWTP	0.2	5.8	0.6	5.2	0.3	2.9
	EBT	150	6.21	62.1	3	54	5.4
	response < EBT	12	2	12	1	14	4
	response > EBT	2	12	2	13	0	10
	% > EBT	14	86	14	93	0	71

EBT
a Van der Oost *et al*. 2017
b Escher *et al*. 2018
c Present study

Location type-specific bioanalytical response profiles

The cumulative effect-based risk quotients obtained in the present study indicated that ecotoxicological risks are potentially present even at reference locations. This illustrates that micropollutants are ubiquitous and pervasive in densely populated river deltas like The Netherlands, which is corroborated by the general presence of non-specific chemical stress at all locations as indicated by the 'promiscuous' PXR CALUX assay. Nonetheless, horticulture and WWTP locations always exhibited higher cumulative effect-based risk quotients than the reference locations.

Ecotoxicological profiles at horticulture locations were characterised by responses to polar extracts in the *anti*-AR, *anti*-PR, and cytotoxicity CALUX assays. Apart from toxicity to target organisms, pesticides and their metabolites can have endocrine-disrupting activities, and the presently observed characteristic response profile for horticulture locations is likely a result of agricultural activity and the resulting use of pesticides on the surrounding fields.[105]

The WWTP locations, contrastingly, were characterised by responses to polar extracts in the ERα CALUX assay and the three *in vivo* bioassays, and for the *Daphnia* bioassay to non-polar extracts and the bacterial bioluminescence assay to metal extracts. These responses were partly previously reported for WWTP effluent-impacted surface waters, in which they were related to the presence of complex mixtures of CECs, like pharmaceuticals, personal care products, pesticides and industrial chemicals.[96,109] Hence, the two main anthropogenic contamination sources investigated in the present study give rise to unique ecotoxicological response profiles. This is important because characteristic bioassay responses that are related to specific sources of pollution can aid the identification of potential causative contamination sources at impacted surface water locations for which the origin of pollution is not known. Furthermore, this will allow the targeted implementation of mitigation measures that reduce the risks of chemical contamination in surface waters.

Interestingly, the majority of the potential ecotoxicological risks in the present study were caused by polar organic contaminants, in both *in vivo* and *in vitro* assays, underlining the urgency of the increasing risks caused by polar CECs in surface waters.[85] These risks were especially pronounced in WWTP effluent impacted surface waters, which highlights the critical need for the use of safer compounds, input prevention, and the implementation of advanced wastewater treatment technologies.[206]

Conclusions

Passive sampling combined with effect-based methods allows the detection of ecotoxicological risks of mixtures of a much wider range of bioavailable compounds than traditional chemical-based methods prescribed by the WFD and CWA. Thus, effect-based methods are highly effective and superior to traditional chemical analytical methods in the screening of surface waters for potential ecotoxicological risks. An elaborate bioanalytical toolbox is now available that allows the identification of contamination source-specific ecotoxicological response profiles, paving the way for the identification of causative (groups of) compounds. The advancement of effect-based monitoring methods, and their implementation in regulatory frameworks like the WFD and CWA, will empower scientists and authorities to work together on the way forward to protect water resources. Nonetheless, chemical analyses, that transcend *a priori* selected target compound lists, are still fundamental to the identification of specific compounds that drive the observed risks and, as such, allow mitigation efforts for risk abatement. Ultimately, the integration of chemical- and effect-based monitoring approaches will foster future-proof water quality assessment strategies on the road to a non-toxic environment.

ACKNOWLEDGEMENTS

This study was part of the Smart Monitoring project (443.324), funded by the Foundation for applied water research (STOWA), The Netherlands. The authors want to thank Eline Reus for her assistance with field- and laboratory work. Jasperien de Weert and Henry Beeltje are acknowledged for their advice on passive sampler extraction and sampled volume estimations, and Peter Cenijn for his guidance during the bacterial bioluminescence assay procedure.

SUPPORTING INFORMATION

SI 1 – Sampling locations

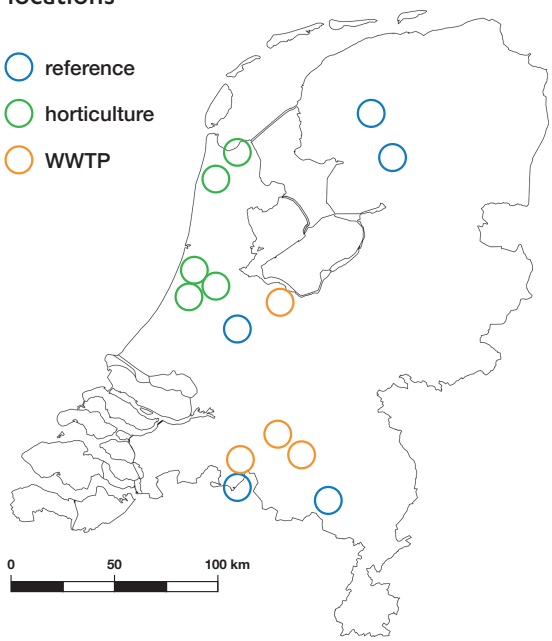

Figure S1. Surface water sampling locations in The Netherlands. WWTP = wastewater treatment plant.

SI 2 – Passive sampler construction, extraction and sampled volume calculation

POCIS construction

Polar organic chemical integrative samplers (POCIS) were applied for sampling of polar compounds from the surface water (Alvarez *et al.*, 2004).[92] POCIS were constructed using two stainless steel rings, with an inner diameter of 5.4 cm, to retain the sorbent between two polyether sulfone (PES) membranes, leaving approximately 46 cm^2 of surface area exposed to the surrounding water. Stainless steel rings (Exposmeter, Sweden), nuts and bolts, as well as all tools were cleaned in acetone before assembly of the samplers. PES diffusion limiting membrane filters (Pall Corporation, NY, USA; 0.1 µm pore size, 90 mm diameter) were cleaned before POCIS assembly in LC grade methanol:ultra-pure water (50:50, v:v) followed by rinsing in ultra-pure water. As a receiving phase, 0.2 g of Oasis hydrophilic-lipophilic balance (HLB) sorbent (Waters, Etten-Leur, The Netherlands) was enclosed between the PES membranes. The HLB was conditioned in its original column by sequentially eluting with 40 mL acetone, 40 mL dichloromethane and 40 mL methanol (Biosolve, The Netherlands; all chromatography grade) and dried under vacuum, followed by the final assembly of the POCIS. POCIS were stored at 4°C in food-grade Mylar zip lock bags until deployment.

Table S1. GPS coordinates for the surface water sampling locations in The Netherlands. WWTP = wastewater treatment plant.

location ID	location type	latitude	longitude
1	reference	52°49'22.7"N	5°54'26.5"E
2	reference	53°00'22.3"N	5°48'43.4"E
3	reference	52°08'08.2"N	4°48'37.6"E
4	reference	51°25'40.9"N	4°46'46.8"E
5	reference	51°18'09.7"N	5°29'09.6"E
6	horticulture	52°53'29.0"N	4°49'34.8"E
7	horticulture	52°45'51.4"N	4°40'52.0"E
8	horticulture	52°17'07.2"N	4°32'34.6"E
9	horticulture	52°17'23.2"N	4°30'37.7"E
10	horticulture	52°17'05.3"N	4°29'54.7"E
11	WWTP	52°12'43.4"N	4°53'10.6"E
12	WWTP	51°30'46.1"N	4°50'57.2"E
13	WWTP	51°36'08.3"N	5°04'32.9"E
14	WWTP	51°30'15.0"N	5°10'19.9"E

Table S2. General field parameters (mean ± SD) of drainage ditches and streams at reference (n = 5), horticulture (n = 5) and WWTP (n = 4) surface water locations. Measurements were taken once during the sampling period along 25 m stretches, except for mean temperature which was measured every 10 minutes for six weeks with the HOBO® Temperature/Light Logger UA-002-64 (Onset Computer Corporation, Bourne, MA, USA). Different letters in superscript indicate significant differences between sites (ANOVA, Post-hoc Tukey's test, $p < 0.05$).

		location type				
parameter	unit	reference	horticulture	WWTP	F	*p*
width	m	4.5±2.4	5.6±1.6	3±1.8	1.9	n.s.
depth	m	0.9±0.2	0.9±0.3	0.5±0.2	3.4	n.s.
flow velocity	m/s	4.7±7.3	3.7±4.5	12.3±10.4	1.7	n.s.
temperature	°C	15.7±1[a]	16.9±0.4[ab]	18.2±1.2[b]	8	0.007

SR extraction

Figure S2. Schematic depiction of SR sheets stacked in harmonica shape at the bottom of an Erlenmeyer flask for the extraction of the organic compounds.

SR PRC chemical analysis

SR sheets were spiked with PRCs with a wide hydrophobicity range (biphenyl D10 and the polychlorinated biphenyl (PCB) congeners 1, 2, 3, 10, 14, 21, 30, 50, 55, 78, 104, 145 and 204) that do not occur in Dutch surface waters. PRC chemical analysis was performed at the laboratory of TNO (Utrecht, The Netherlands). SR extracts were transferred to hexane by adding 0.5 mL extract to 100 mL hexane and concentrated with rotary film evaporator at 45˚C. After the solvent change the extract was cleaned up with 3% deactivated florisil column chromatography. The cleaned extract was evaporated to exactly 0.5 mL and analysed with an Agilent 7890 gas chromatograph connected to an Agilent 7000 Triple Quadrupole mass spectrometer (GC-MS/MS) equipped with Edwards pump. Quantification of PRCs was performed using the relative response factors to an external calibration standard.

DGT sampled volume calculation

The sampled volume for the DGT samplers was determined using a theoretical approach as well as an approach based on the detected masses of metals that had accumulated in the samplers. Both approaches rely on DGT theory as outlined in numerous publications that confirm the usability of DGTs to obtain time weighted average field concentrations of metals [e.g. Allan *et al.* (2007)[196]; Davison and Zhang (2012)[197]]. As no approach for the derivation of sampling rates for DGTs was previously reported, a formula was derived by combining two equations reported by Allan *et al.* (2007)[196] (definitions of constants are given in Table S3):

$$\textit{Eq. 1: } \mathbf{C_{water} = M / R_s t}$$

$$\textit{Eq. 2: } \mathbf{C_{water} = M\Delta g / DAt}$$

Combining the two equations resulted in equation 3:

Eq. 3: $\mathbf{R_s = DA / \Delta g}$

By using the values of the constants given in Table S3 and assuming a mean value for D of $5.0*10^{-6}$ cm^2/s, a daily (86400 s) sampling rate per 3 DGT samplers was derived: R_s = 44.2 mL/d.

This theoretically derived sampling rate was subsequently confirmed using C_{water} values calculated from metal concentrations detected in the DGT extracts (C_e). To allow for these calculations, concentrations of Cd, Cu, Fe, Pb and Zn in the DGT extracts were determined using an inductively coupled plasma mass spectrometer (ICP-OES OPTIMA 8300; Perkin Elmer, Groningen, The Netherlands). Only Cu, Fe and Zn were detected (Table S2) and the calculations were therefore based on these concentrations. C_{water} values were calculated as follows, using the variables and constants listed in Table S3. First, the mass of metal accumulated in the resin gel layer (M) was calculated for each metal using equation 4:

Eq. 4: $\mathbf{M = C_e (V_{HNO3} + V_{gel}) / f_e}$

Secondly, the labile metal concentration in the water (C_{water}) was calculated using equation 2. Lastly, the sampling rates of the DGT samplers was calculated using equation 5:

Eq. 5: $\mathbf{R_s = (M / C_{DGT}) / 14}$

These calculations resulted in an experimentally derived mean sampling rate per 3 DGT samplers of 44.9 mL/d, which is very close to the theoretically derived sampling rate (44.2 mL/d). The small difference between the theoretical and experimental sampling rate is likely attributable to the variation of D with temperature, which was accounted for in the experimentally derived sampling rate calculation. Based on these calculations, a mean sampled volume of 44.9 mL/d for 3 DGT samplers was used in the subsequent data interpretation.

Table S3. Detected metal concentrations in pooled extracts of three DGT samplers exposed to surface water for 14 d at locations with a variety of contaminations sources. WWTP = wastewater treatment plant. Definitions for variables are listed in Table S4.

location ID	location type	T (mean; °C)	Ce (µg/L)			M (µg)			D ($cm^2/s*10^{-6}$)			C_{DGT} (µg/L)			R_s (mL/d for 3 DGTs)		
			Zn	Fe	Cu	Zn	Fe	Cu	Zn	Fe	Cu	Zn	Fe	Cu	Zn	Fe	Cu
1	reference	18	110	700	20	0.47	3.02	0.09	5.0	5.0	5.1	0.77	4.86	0.14	44.1	44.4	45.3
2	reference	17	130	3890	10	0.56	16.78	0.04	4.9	4.9	5.0	0.93	27.81	0.07	42.9	43.1	44.0
3	reference	18	440	2330	10	1.90	10.05	0.04	5.0	5.0	5.1	3.07	16.16	0.07	44.1	44.4	45.3
4	reference	16	130	910	20	0.56	3.92	0.09	4.7	4.7	4.8	0.96	6.70	0.14	41.6	41.8	42.6
5	reference	20	1110	940	20	4.79	4.05	0.09	5.3	5.3	5.4	7.31	6.15	0.13	46.8	47.1	47.9
6	horticulture	17	100	1260	20	0.43	5.43	0.09	4.9	4.9	5.0	0.72	9.01	0.14	42.9	43.1	44.0
7	horticulture	17	40	880	10	0.17	3.80	0.04	4.9	4.9	5.0	0.29	6.29	0.07	42.9	43.1	44.0
8	horticulture	17	60	500	10	0.26	2.16	0.04	4.9	4.9	5.0	0.43	3.57	0.07	42.9	43.1	44.0
9	horticulture	18	80	600	-	0.35	2.59	-	5.0	5.0	-	0.56	4.16	-	44.1	44.4	-
10	horticulture	18	50	700	-	0.22	3.02	-	5.0	5.0	-	0.35	4.86	-	44.1	44.4	-
11	WWTP	21	430	39580	30	1.85	170.69	0.13	5.4	5.5	5.6	2.75	251.95	0.19	48.1	48.4	49.4
12	WWTP	19	1070	1280	30	4.61	5.52	0.13	5.1	5.2	5.3	7.25	8.62	0.20	45.5	45.7	46.6
13	WWTP	20	1440	1890	30	6.21	8.15	0.13	5.3	5.3	5.4	9.48	12.37	0.19	46.8	47.1	47.9
14	WWTP	19	640	4150	30	2.76	17.90	0.13	5.1	5.2	5.3	4.34	27.95	0.20	45.5	45.7	46.6
														mean	**44.9**		
														st. err.	0.3		

Table S4. Constants and variables used in the calculation of the DGT sampling rate.

variable	definition	unit	
C_e	concentration in HNO_3 extract	µg/L	
M	metal mass in HNO_3 extract	µg	
D	diffusion coefficient of metal in gel	$cm^2/s*10^{-6}$	
C_{water}	labile metal concentration in water	µg/L	
R_s	exchange rate	L/d	
constant	**definition**	**unit**	**value**
V_{HNO3}	HNO_3 volume used in extraction	mL	3
V_{gel}	volume of resin gel (3 samplers)	mL	0.45
Δg	diffusive layer thickness	cm	0.092
t	deployment time	s	1209600
A	exposure area (3 samplers)	cm^2	9.42
f_e	elution factor	-	0.8

SI 3 – Bioassay battery

Table S5. Bioassay battery applied to assess the toxicity of surface water from 14 locations in The Netherlands. Effect-based trigger (EBT) values were previously defined by Escher *et al.*, 2018 (*anti*-AR), Brion *et al.*, 2019 (ERα), and Van der Oost *et al.*, 2017b. EBT values for *anti*-PR as well as for the *in vivo* bioassays performed with inorganic extracts were defined in the present study. Previously reported EBT values for the PAH and PXR assays by Van der Oost *et al.*, 2017b and Escher *et al.*, 2018 were strongly divergent and intermediate EBT values were presently proposed. TU = toxic unit, ...EQ/L = equivalent concentration of the reference compound.

	bioassay	endpoint	reference compound	EBT	unit
in vivo	Daphnia	Mortality	n/a	0.05	TU
all extracts	PAM algae	Photosynthetic inhibition	n/a	0.05	TU
	Bacterial bioluminescence inhibition	Luminescence inhibition	n/a	0.05	TU
in vitro	cytotox nonpolar	Cytotoxicity	n/a	0.05	TU
CALUX	DR	Dioxin(-like) activity	2,3,7,8-TCDD	50	pg TEQ/L
organic	PAH	PAH activity	benzo(a)pyrene	62.1	ng BEQ/L
non-polar	PPARγ	Lipid metabolism inhibition	rosiglitazone	10	ng REQ/L
	Nrf2	Oxidative stress	curcumin	10	µg CEQ/L
	PXR	Toxic compound metabolism	nicardipine	5.4	µg NEQ/L
	p53	Genotoxicity	n/a	0.005	TU
in vitro	cytotox polar	Cytotoxicity	n/a	0.05	TU
CALUX	ERα	Estrogenic activity	17ß-estradiol	0.28	ng EEQ/L
organic	*anti*-AR	Antiandrogenic activity	flutamide	14.4	µg FEQ/L
polar	*anti*-PR	Antiprogestagenic activity	Ru486	13	ng REQ/L

SI 4 – Effect-based trigger value derivation

Anti-PR CALUX

A preliminary EBT for the *anti*-PR CALUX assay was derived by Escher *et al.* (2018),[30] expressed as 1967 ng endosulfan equivalents/L. Since the reference compound for the *anti*-PR CALUX assay used in the present study, Ru486 (mifepristone), differed from the reference compound of the EBT, relative effect potencies were used to translate the EBT to Ru486 equivalents (REQ). The activity of 1 ng Ru486 matches that of 1500 ng endosulfan in the *anti*-PR CALUX assay [SI of Escher *et al.* (2018)[30]], resulting in an EBT of 1.3 ng REQ/L. However, the use of this EBT value resulted in substantial exceedance of the EBT at all locations, including relatively unpolluted reference sites, suggesting that this value is too low for the diagnosis of surface water quality. Interestingly, Escher *et al.* (2018)[30] identified this EBT value as "too preliminary to derive a final effect threshold", and advised that it should be treated "with caution". Therefore, for use in the present study, the mixture factor of 100 used by Escher *et al.* (2018)[30] was increased to 1000, resulting in a revised EBT value of 13 ng REQ/L. This value indeed allowed for a clearer distinction between sites and contamination sources. It must be noted, however, that this value

is still preliminary and requires further research before it can be adopted for wider use in surface water quality assessment.

PAH and PXR CALUX

Since the EBT values for the PAH and PXR CALUX assays that were previously reported by Van der Oost *et al.* (2017)[21] and Escher *et al.* (2018)[30] were strongly divergent, preliminary intermediate EBT values assays were derived in the present study to explore their applicability in surface water quality assessment. To do so, the mixture factors for these tests as proposed by Escher *et al.* (2018)[30] were adjusted in the same way as was done for the *anti*-PR CALUX assay EBT value in the present study. For the PAH CALUX assay, the mixture factor was increased from 100 to 1000, resulting in the preliminary EBT value of 62.1 ng BEQ/L. For the PXR CALUX assay EBT value, Escher *et al.* (2018)[30] stated that "… it is necessary to invoke a mixture factor of at least 100 to account for mixture effects.". However, empirical data of the present paper indicated that adjusting the mixture factor from 100 to 10 allows for more diagnostic power to differentiate between locations. This exploration resulted in the preliminary EBT value of 5.4 µg NEQ/L for the PXR CALUX assay.

In vivo assays with DGT extracts

An EBT for the *in vivo* assays, when used in combination with DGT extracts, was derived based on the method outlined by Van der Oost *et al.* (2017).[21] In this method, the acute-to-chronic ratio (ACR) of compound groups and the extraction efficiency of passive samplers for surface water are used to derive an indication of chronic effects in surface water from acute effects in a concentrated water sample. An ACR for the toxicity of metals of 15.31 was previously reported.[207] The average elution factor (f_e) of metals from surface water using DGT samplers is 0.8 (~80% extraction efficiency).[208] Using these values, the derivation of the EBT for the *in vivo* assays was performed as follows:

$$EBT = f_e / ACR = 0.8 / 15.31 = 0.05 \text{ TU}$$

Therefore, the preliminary EBT value of 0.05 TU was used for the interpretation of ecotoxicological risk of metals in the *in vivo* bioassays in the present study. Interestingly, this value is identical to the EBT value used for the interpretation of ecotoxicological risk of organic compounds.

SI 5 – Responses of all bioassays at all investigated locations

Table S6. Responses of three in vivo bioassays to extracts of three passive sampler types (converted to water concentrations) exposed at 14 surface water locations. WWTP = wastewater treatment plant, EBT = effect-based trigger value, TU = toxic unit. Grey cell fills indicate EBT exceedances.

species		*A. fischeri*			*D. magna*			*R. subcapitata*		
bioassay name		Bacterial bioluminescence inhibition			Daphnia			PAM algae		
extract		polar	non-polar	metal	polar	non-polar	metal	polar	non-polar	metal
location ID	location type	TU	TU	TU	TU	TU	TU	TU	TU	TU
1	reference	0.003	0.004	0.017	0	0	0	0	0	0
2	reference	0.004	0.007	0.023	0	0	0	0	0	0
3	reference	0.004	0.006	0.012	0	0	0	0	0	0
4	reference	0.003	0.007	0.008	0	0	0	0	0	0
5	reference	0.010	0.001	0.015	0	0.004	0	0.013	0.001	0.034
6	horticulture	0.016	0.002	0.037	0	0.003	0	0	0	0
7	horticulture	0.012	0.002	0.032	0	0	0	0	0	0
8	horticulture	0.005	0.002	0.023	0	0	0	0	0	0
9	horticulture	0.006	0.001	0.031	0	0	0	0	0	0
10	horticulture	0.004	0.004	0.034	0	0	0	0	0	0
11	WWTP	0.101	0.006	0.153	0	0.008	0	0	0	0
12	WWTP	0.060	0.002	0.036	0.021	0.006	0	0.018	0.000	0.037
13	WWTP	0.047	0.000	0.036	0.028	0.002	0	0.016	0.000	0.030

Table s6. (continued)

species		*A. fischeri*			*D. magna*			*R. subcapitata*		
bioassay name		Bacterial bioluminescence inhibition			Daphnia			PAM algae		
extract		polar	non-polar	metal	polar	non-polar	metal	polar	non-polar	metal
location ID	location type	TU	TU	TU	TU	TU	TU	TU	TU	TU
14	WWTP	0.048	0.001	0.036	0.077	0.003	0	0.043	0.000	0
EBT		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
no response		0	0	0	11	8	14	10	10	11
response < EBT		12	14	13	2	6	0	4	4	3
response > EBT		2	0	1	1	0	0	0	0	0
% > EBT		14	0	7	7	0	0	0	0	0
% > EBT	reference	0	0	0	0	0	0	0	0	0
	horticulture	0	0	0	0	0	0	0	0	0
	WWTP	50	0	25	25	0	0	0	0	0

Table S7. Responses of 11 in vitro CALUX bioassays to extracts of polar and non-polar passive samplers (converted to water concentrations) exposed at 14 surface water locations. WWTP = wastewater treatment plant, EBT = effect-based trigger value, TU = toxic unit. Grey cell fills indicate EBT exceedances.

extract		nonpolar							polar			
bioassay name		cytotox n-p	DR	PAH	PPARγ	PXR	Nrf2	p53	cytotox p	ERα	*anti*-AR	*anti*-PR
location ID	location type	TU	pg TEQ/L	ng BEQ/L	ng REQ/L	µg NEQ/L	µg CEQ/L	TU	TU	ng EEQ/L	µg FEQ/L	ng REQ/L
1	reference	0.000	0	3.2	0	3.8	2.224	0.001	0.022	0.128	42.8	14.1
2	reference	0.008	0.482	19.6	0	8.0	0	0.025	0.010	0.117	14.3	4.6
3	reference	0	0.374	19.2	0	24.1	10.076	0.001	0.003	0.522	8.1	3.7
4	reference	0.000	0	5.9	0	7.3	4.451	0.001	0.009	0.014	16.8	5.9
5	reference	0.000	0.118	53.3	0	9.4	7.853	0.002	0.011	0.743	17.3	19.5
6	horticulture	0.001	0	25.9	0	5.4	3.374	0.003	0.073	0.345	252.4	45.6
7	horticulture	0.001	0	296.3	0	8.0	3.185	0.004	0.074	0.322	94.3	49.1
8	horticulture	0.000	0.079	33.0	0	5.0	3.825	0.001	0.031	0.385	46.5	45.3
9	horticulture	0.000	0	13.6	0	5.9	3.934	0.001	0.029	1.110	70.1	37.4
10	horticulture	0.000	0	22.5	0	12.7	5.247	0.001	0.027	0.526	54.8	38.4
11	WWTP	0.001	0	59.3	0	20.1	4.757	0.003	0.005	2.574	25.7	6.6
12	WWTP	0	0	1429.6	5.3	5.0	3.013	0.000	0.005	4.922	15.1	7.8
13	WWTP	0.000	0.030	25.7	0	2.4	3.659	0.001	0.003	1.329	41.9	8.2
14	WWTP	0.001	0	35.7	0	15.5	9.464	0.002	0.003	3.832	10.5	11.5
EBT		0.05	50	62.1	10	5.4	10	0.005	0.05	0.28	14.4	13

Table S7. (continued)

extract		nonpolar							polar			
bioassay name		cytotox n-p	DR	PAH	PPARγ	PXR	Nrf2	p53	cytotox p	ERα	*anti*-AR	*anti*-PR
location ID	location type	TU	pg TEQ/L	ng BEQ/L	ng REQ/L	µg NEQ/L	µg CEQ/L	TU	TU	ng EEQ/L	µg FEQ/L	ng REQ/L
no response		2	9	0	13	0	1	0	0	0	0	0
response < EBT		12	5	12	1	4	12	13	12	3	3	7
response > EBT		0	0	2	0	10	1	1	2	11	11	7
% > EBT		0	0	14	0	71	7	7	14	79	79	50
% > EBT	reference	0	0	0	0	80	20	20	0	40	60	40
	horticulture	0	0	20	0	80	0	0	40	100	100	100
	WWTP	0	0	25	0	50	0	0	0	100	75	0