SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part II. Field feasibility survey

Ron van der Oost[†*], Giulia Sileno[†,‡], Theo Janse[†], Mai Thao Nguyen[§], Harrie Besselink[¶] & Abraham Brouwer[¶,#]

[†]: Waternet Institute for the Urban Water Cycle, Department of Technology, Research and Engineering, Amsterdam, The Netherlands
[‡]: University of Amsterdam, Institute for Biodiversity and Ecosystem Dynamics (IBED), Amsterdam, The Netherlands
[§]: Waterproef Laboratory, Research & validation, Edam, The Netherlands
[¶]: Biodetection Systems BV, Amsterdam, The Netherlands
[#]: VU University, Department of Ecological Sciences, Faculty of Earth & Life Sciences, Amsterdam, The Netherlands

[*]: corresponding author:
Dr. Ron van der Oost
Waternet Institute for the Urban Water Cycle
Department of Technology, Research and Engineering
PO Box 94370,
1090 GJ Amsterdam,
The Netherlands
Email: ron.van.der.oost@waternet.nl
Phone: +31-652534378

Published in Environmental Toxicology & Chemistry 36: 2400-2416 (2017)

ABSTRACT

Since it is impossible to chemically analyze all relevant micropollutants, the implementation of bioanalytical tools is essential to estimate ecological risks of chemical mixtures in regular water monitoring programs. The first tier of the Smart Integrated Monitoring (SIMONI) strategy, which was described in part I, is based on the combination of passive sampling and bioanalytical measurements. Bioassay responses are compared to effect-based trigger values (EBT) and an overall SIMONI score on all bioassay data was designed to indicate environmental risks. The present paper is focused on analyzing the feasibility of the hazard identification tier by evaluating results of 45 field campaigns at sites with different pollution profiles near the city of Amsterdam. A Daphnia assay was performed *in situ*, while silicon rubber or POCIS passive sampler extracts were tested with four non-specific (daphnids, algae, bacteria and cell culture) and ten specific bioassays (nine CALUX assays and antibiotics scan).

Sensitivity analyses demonstrated the relevance of two classification variables in the SIMONI score formula on all bioanalytical data. The model indicated increased risks for the ecosystem at surface waters in greenhouse areas and undiluted WWTP effluents. The choice of testing specific bioassays on either polar or non-polar passive sampling extracts is cost-effective and still provided meaningful insights on micropollutant risks. Statistical analyses revealed that the model provides a relevant overall impact assessment based on bioassay responses. Data analyses on the chemically determined mixture toxic pressure and bioanalytical methods provided similar insights in relative risk ranking of water bodies. The SIMONI combination of passive sampling and bioanalytical testing appears to be a feasible strategy to identify chemical hazards.

Key words: Micropollutants, Environmental risk assessment, Bioanalytical tools, Passive sampling

1. INTRODUCTION

1.1 The SIMONI strategy

It is virtually impossible to chemically analyze all relevant micropollutants and their transformation products in water. Moreover, the ecological effects of micropollutant mixtures are practically all unknown. Therefore, it is of paramount importance to apply bioanalytical tools in regular water quality monitoring programs, to integrate the effects of all micropollutants activating certain receptors that may trigger adverse outcome pathways. The design of the Smart Integrated Monitoring (SIMONI) strategy and the derivation of effect-based trigger values (EBT) have been described in detail in a previous paper [1]. This two-tiered strategy will be briefly summarized in this paragraph. The first tier of the strategy is hazard identification of micropollutants, based upon the combination of field-exposed passive samplers, one *in situ* bioassay and fourteen laboratory bioassays (Figure 1). The second tier is a customized risk assessment, based upon the results of tier 1 and additional information on various aspects of the water system (influences of other ecological key factors). Tier 2 assessments should preferably be performed on concentrated large-volume time-integrated water samples instead of passive samplers, in order to more accurately quantify the chemical and bioanalytical results (see discussion).

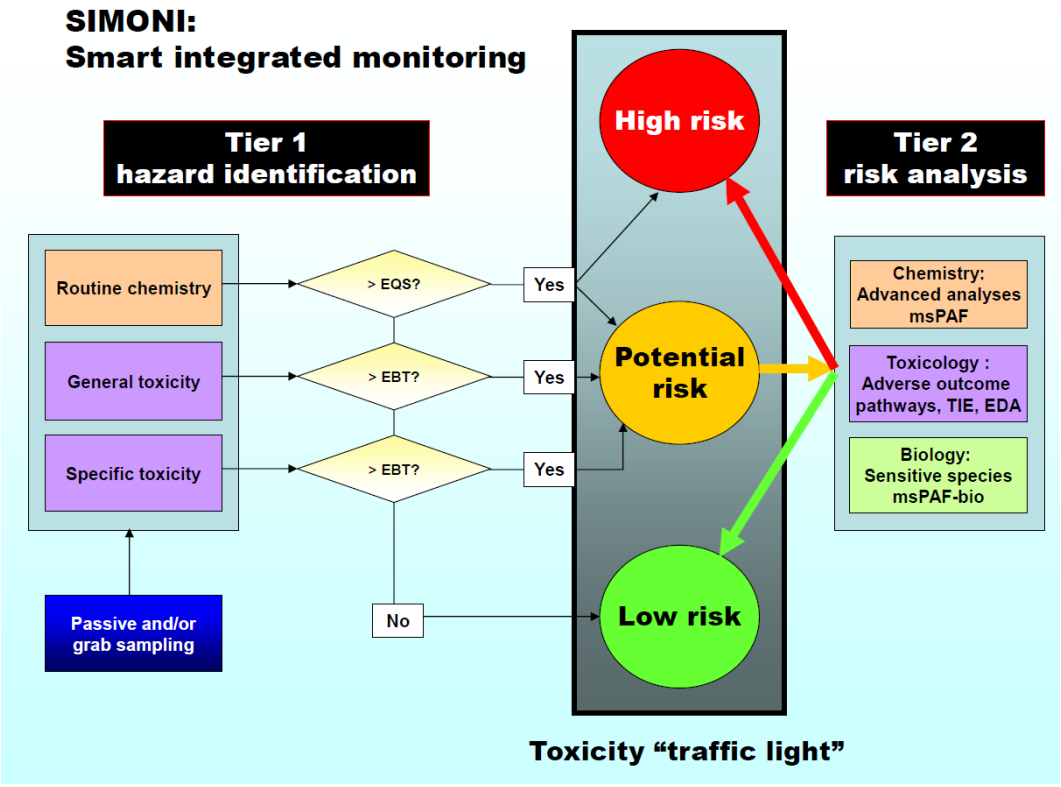

Figure 1: Schematic presentation of the SIMONI effect-based monitoring strategy; EQS = environmental quality standard; EBT = effect-based trigger value; msPAF = multiple-substance potentially affected fraction; TIE = toxicity identification & evaluation; EDA = effect directed analysis.

The first tier of the strategy, hazard identification, is applied to assess the potential risks of a broad spectrum mixture of chemical micropollutants. The main objective of this screening phase is to identify the 'hot spots' of chemical water pollution. Hazards of organic micropollutants are characterized by evaluating the responses of a suite of validated bioassays, using effect based trigger values (EBT) as criteria for potential risks. In this tier, chemical analyses are only performed on relevant inorganic chemicals, such as metals and ammonium. A model has been designed to calculate an overall SIMONI score that should be indicative for the ecological risks, based on general and specific modes of action [1]. Only a limited number of sites, where bioassay responses indicate ecological risks, should be examined by a more expensive tier 2 for the actual risk assessment. In this way, the more advanced and expensive chemical analyses (e.g. WFD priority pollutants) and bioanalytical methods (e.g. fish biomarkers) are only carried out at sites were they are most relevant. Results of the risk assessment should be verified with ecological observations, such as the reduced occurrence of species that are sensitive to certain micropollutants.

1.2 Bioanalytical hazard assessment

A selection of bioanalytical endpoints has been established in the first paper of this series [1], based upon literature data [2, 3, 4, 5] and own research (section 3.1 of the present paper). The initial selection is summarized in Table 1, together with the bioassays that were applied in the present study to measure these endpoints. The bioanalytical assessment can be performed on alternative bioassays that assess similar endpoints, but adjustment of EBT is required if relative effect potencies of key toxicants are different. The endpoint selection aims to cover a broad range of micropollutants, multiple modes of action (MoA: non-specific, specific & reactive) and multiple biological levels (*in vitro* and *in vivo*). Non-specific *in vivo* assays were included in the panel since they are responsive to the broadest range of micropollutants. Responses of these assays are expressed as relative enrichment factor (REF) of a sample for the EC50 measurement (concentration causing a 50% response). The REF can be converted to toxic units (TU=1/REF). Specific *in vitro* responses were selected since these are generally much more sensitive for targeted MoA than *in vivo* responses, and are able to detect specific activities caused by unknown mixtures of compounds with the same MoA, such as estrogenicity [6]. *In vitro* responses are expressed as bioanalytical equivalents (BEQ), i.e. a measure to express the effect of a mixture of unknown and potentially unidentified chemicals as the concentration of a known reference compound eliciting the same effect [7].

Table 1: Selected toxicological endpoints and applied bioassays for tier 1 SIMONI strategy

Category	Endpoint (mode of action)	Bioassay
Non-specific (*in situ*)	Non-specific toxicity Daphnia magna	Daphnia magna survival
Non-specific (*in vivo*)	Non-specific toxicity bacteria	Microtox
	Non-specific toxicity phytoplankton	Algaltoxkit
	Non-specific toxicity zooplankton	Daphniatoxkit
	Non-specific toxicity cytotoxicity	Cytotox CALUX
Specific (*in vitro*)	Estrogenic activity	ER CALUX
	Anti-androgenic activity	Anti-AR CALUX
	Glucocorticoid activity	GR CALUX
	Pregnane X receptor	PXR CALUX
	Aryl hydrocarbon receptor (POP)	DR CALUX
	Aryl hydrocarbon receptor (degradable)	PAH CALUX
	Peroxisome proliferation	PPAR CALUX
	Antibiotic activity (tetracyclines, quinolones, β-lactams and macrolides, aminoglycosides, sulphonamides)	RIKILT WATERSCAN
Reactive (*in vitro*)	Genotoxicity	P53 CALUX
	Oxidative stress	Nrf2 CALUX

1.3 Combining passive sampling and bioanalytical measurements

Water systems are generally sampled by snapshot grab sampling and concentrated with solid-phase extraction (SPE). This method has large disadvantages, since environmental concentrations of many micropollutants vary significantly over time. An alternative method is to apply time-integrated sampling with passive samplers that are able to concentrate bioavailable micropollutants on site, and may be a good reflection of the micropollutants that accumulate in tissues of water organisms [8, 9], thus reflecting the actual exposure conditions of a site. Chemicals passively diffuse along a gradient in chemical activity toward the samplers. The initial fast uptake occurs in a linear way (kinetic phase), slows down (intermediate phase) and reaches a plateau (equilibrium phase). Adsorption-based samplers, mainly for polar and ionic organic chemicals, are generally operated in the kinetic mode. An advantage of passive samplers is that they can enrich a broad mixture of chemicals, while largely leaving the matrix (e.g., lipids) and confounding factors (e.g., salinity and pH) behind. There are, however, certain pitfalls when combining passive sampling and bioassays [10]. First, the composition of the mixture extracted from the passive samplers is not the same as the one the organisms are exposed to in the field. Second, since the compounds causing significant responses in the bioassay are unknown, it is impossible to derive exact concentrations of toxic units or bioanalytical equivalents in the water phase. These pitfalls will be discussed in the present paper, and suggestions will be given for a provisional tier 1 interpretation of bioassay responses in passive sampling

extracts. Since the assumptions made for this quantification will not be suited for an exact tier 2 risk assessment, large-volume sampling (preferably time-integrated) has to be performed in follow-up studies.

1.4 Objectives of the present study

The present study is focused on analyzing the practical feasibility and interpretation of the tier 1 SIMONI hazard identification, as described in a previous paper [1]. The SIMONI strategy is a pragmatic model to prioritize sites with highest chemical risks, using passive sampling and bioanalytical tools. The primary objective of the present study is to determine whether the SIMONI strategy provides a relevant assessment for prioritizing sites with increased ecological risks due to micropollutants at relatively low costs. The assumptions made for the model and their uncertainties are evaluated on the basis of field studies that were performed over the last five years. Choices were made on the most relevant passive sampler extracts (polar or non-polar) to be tested for the different endpoints of Table 1. An estimation of extracted water volumes by passive sampling will be proposed in order to roughly estimate water levels of bioanalytical equivalents. The assumptions that were made for the SIMONI score for potential ecological risks (bioassay weight factors and threshold of 50% EBT exceedance as risk indicator) were evaluated by sensitivity analyses. The SIMONI Tier 1 bioanalytical hazard assessment will be compared to a mixture toxic pressure classification [11], based upon chemical analyses of a range of organic micropollutants. Finally, the repeatability of the SIMONI score over time was investigated, and statistical analyses were performed on relationships between SIMONI results and (assumed and measured) chemical pollution levels.

2. MATERIALS & METHODS

2.1 Sampling sites

Several monitoring campaigns have been performed over the last five years, in order to analyse and optimize the SIMONI model. The sites that were sampled are listed in Table 2. Most sites were within the Waternet management area near the City of Amsterdam, but presumed 'unpolluted' reference sites were also selected in other parts of the Netherlands. Unpolluted refers to sites without known pollution sources and a good ecological status, while moderately and highly polluted sites were discriminated based on known pollution sources and historic chemical and ecological data.

Table 2: Sampling sites used for the SIMONI evaluation, with known pollutant sources.

Sites	Code	Potential sources of micropollutants
Unpolluted		
Lake Waterleidingplas	LWP	no source identified
Lake Naardermeer	LNM	no source identified
Lake Botshol	LBH	no source identified
Lake Reeuwijk	LRW	no source identified
Peelkanaal	PKN	no source identified
Lake Geestmerambacht	LGA	no source identified
Lake Kennemerland	LKL	no source identified
Moderately polluted		
Maarsseveense Zodden	MZD	agriculture
Strook Lake Loosdrecht	SLL	recreational shipping
River Vecht Maarssen	RVM	agriculture
Waterleiding canal	WLC	agriculture, shipping, wwtp effluent
River Amstel before Uithoorn	ABU	agriculture, shipping, wwtp effluent
River Amstel after Uithoorn	AAU	agriculture, shipping, wwtp effluent
River Vecht Utrecht	RVU	agriculture, shipping, wwtp effluent
River Vecht Loenen	RVL	agriculture, shipping, wwtp effluent
Smal Weesp at Solvay	SWS	pharmaceutical industry
River Vecht Horstermeer	RVH	agriculture, shipping, wwtp effluent
River Amstel Uithoorn	RAU	agriculture, shipping, wwtp effluent
River Amstel Ronde Venen	RAR	agriculture, shipping, wwtp effluent
River Amstel Amstelveen	RAA	agriculture, shipping, wwtp effluent
Lake Eemmeer	LEM	agriculture, shipping, wwtp effluent
Heavily polluted		
Zuider Legmeerpolder 1	ZL1	greenhouses
Zuider Legmeerpolder 2	ZL2	greenhouses
Zuider Legmeerpolder 3	ZL3	greenhouses
Zuider Legmeerpolder 4	ZL4	greenhouses
Zuider Legmeerpolder 5	ZL5	greenhouses
Zuider Legmeerpolder 6	ZL6	greenhouses
Noorder Legmeerpolder 1	NL1	greenhouses
Noorder Legmeerpolder 2	NL2	greenhouses
Noorder Legmeerpolder 3	NL3	greenhouses
Ditch Zevenhoven	DZH	greenhouses
Gooiergracht Hilversum	GHI	undilluted wwtp effluent
Gooiergracht Blaricum	GBL	undilluted wwtp effluent

2.2 Passive sampling

2.2.1 Field deployment of passive samples. Silicone rubber passive samplers with and without performance reference compounds (PRCs) were obtained from Deltares, Netherlands. Polar Organic Chemical Integrative Samplers (POCIS) were obtained from Exposmeter AB, Sweden. Silicone rubbers (six blades on a holder, 20 gram total weight, 627 cm^2 total surface) with and/or without PRCs as well as four POCIS samplers (Oasis HLB sorbent, 41 cm^2 total surface) were deployed at each sampling site, and exposed for six weeks. After exposure, the samplers were cleaned with water from the sampling site to remove attached particulates and biofilm. Cleaned samplers were transported to the lab in plastic (SR) or metal foil (POCIS) containers, and stored at -20°C until extraction.

2.2.2 Extraction of silicone rubbers. The six silicone rubber blades were cut into small pieces and put in pre-cleaned thimbles for the Tecator® Soxtec Avanti 2050 extraction system. The extraction was performed with 80 mL of methanol:acetonitrile (1:2 V/V) mixture with boiling stones. The extraction program was 120 minutes boiling at 180°C, 30 minutes rinsing, 5 minutes recovery and 1 minute drying. Cooled extracts were filtered over glass fibre filters and collected in 250 mL glass bottles. Extraction jars were rinsed twice with 10 mL of extraction mixture. Extracts were evaporated by TurboVap® II Zymark at 45°C to approximately 5 mL. Extracts were transferred quantitatively (two times rinsing with 5 mL extraction mixture) to 15 mL conical tubes and evaporated under nitrogen, volume was filled up to exactly 10 mL.

2.2.3 Extraction of POCIS. Sorbent between the POCIS membranes (0.2 g of Oasis HLB powder per sampler) was transferred quantitatively into an empty SPE column with polyethylene frit. Columns were dried under vacuum extraction, followed by centrifugation (2000 rpm, 15 minutes), and nitrogen flow. Dry columns were eluted three times with 3 mL of acetone, with 5 minutes equilibration time between elutions. Eluates were collected in 10 mL conical tubes, and the end volumes were filled up to exactly 10 mL.

2.3 Estimations of passive samplers extracted water volumes

Silicone rubbers are partitioning-based samplers that are spiked with performance reference compounds (PRC) with a wide hydrophobicity range (biphenyl D10 and PCB IUPAC nrs. 1, 2, 3, 10, 14, 21, 30, 50, 55, 78, 104, 145 and 204), that do not occur in Dutch surface waters. PRC analysis is described under section 2.4.1. The rate of PRC dissipation was used to calculate the exchange rates (R_S values in L/day) of the samplers [12]. The R_S values can only be calculated for individual substances, preferably with known partitioning coefficients between sampler and water (K_{SW}), or alternatively with octanol-

water partitioning coefficients (K_{OW}). However, since the compounds causing an effect in the bioassays are unknown, a provisional estimation has to be made for the sampled water volumes. The calculated R_S values describe the water volume of compounds with a molecular weight of 300 Da that do not reach equilibrium (kinetic mode). Assuming that approximately 50% of the compounds reach equilibrium during exposure, 50% of this calculated R_S has been used as a provisional estimation of the average extracted water volume per day.

For the project 'Time-Integrative Passive sampling combined with Toxicity Profiling' (TIPTOP), estimations for extracted volumes of water with silicone rubbers were made by using the concentration-weighted average sampling volumes (Vs-cwa) [13]. Sampling volumes were calculated for all individual substances and then weighted for the concentration of the substance relative to the total concentration of all the analyzed substances. The Vs-cwa was calculated by summation of all weighted concentrations (Equation 1):

$$\mathrm{Vs_{cwa}} = \sum_{i=1}^{n} \left(\frac{\mathrm{concentration_i \ x \ sampled \ volume_i}}{\mathrm{total \ chemical \ concentration}} \right) \quad (1)$$

POCIS samplers for more polar compounds are adsorption-based samplers that are not spiked with PRCs. R_S rates for different polar compounds may vary from 30 to 300 mL per day [14, 15, 16]. As a provisional estimate of the average extraction volume we propose to use 100 mL water per day per sampler for both chemical and bioanalytical analyses.

Obviously, these provisional estimations for both types of samplers are not suited for exact calculations, but they can be used for the tier 1 screening phase to indicate differences between low risk and potential risk situations (green and orange in the SIMONI strategy).

2.4 Chemical analyses and sample clean-up

Chemical analyses were performed at the Waterproef laboratory (Edam, Netherlands), unless stated otherwise.

2.4.1 Polychlorinated biphenyls (PCBs), organo chlorine pesticides (OCPs) and Polycyclic aromatic hydrocarbons (PAHs). SR extracts were transferred to petroleum ether (PE) by adding 2 mL extract to 40 mL PE, and concentrated with Kuderna Danish at 80ºC. The PE extract was cleaned-up with aluminium oxide and silica gel column chromatography. The cleaned extract was evaporated to exactly 2 mL and analysed with Agilent 7890 Triple Quadrupole GC-MS/MS, Edwards pump for PCBs (including performance reference

compounds), OCPs and PAHs. Quantification was performed using external calibration series of six concentrations. The analyses were performed according to protocols of the Dutch Accreditation Council and the Dutch Standardization Institute. Detection limits using silicone rubber passive sampling were approximately 0.01 ng/L water.

2.4.2 Polar pesticides (PPs). POCIS extracts were prepared for the analyses of positive ion mode and negative ion mode by carefully evaporating 1 mL extract to dryness. Methanol:HPLC-water (1:9 V/V) was added for positive mode, methanol:HPLC-water (1:1 V/V) was added for negative mode. The final extracts were analysed with Thermo TSQ Quantum Discovery LC-MS/MS, ESI interface, Surveyor LC pump. Quantification was performed using an external calibration series of six concentrations. The method was validated by calculating the recovery and standard deviation in four surface water samples spiked with polar pesticides (average recovery was 85 ± 8%). Detection limits using POCIS passive sampling were approximately 0.1 ng/L water.

2.4.3 Nitrogen/phosphorus pesticides (NPPs). One mL of SR extract was transferred to an OASIS HLB (Waters) SPE column that is conditioned with methanol and dichloromethane (DCM). The SPE column was extracted with 3 times 3 mL DCM, with 10 minutes equilibration time between separate elutions. Eluates collected in 10 mL conical tubes were evaporated to exactly 1 mL and analysed with Hewlett Packard 6890 GC connected with Agilent 5973N mass selective detector. Quantification was performed using an external calibration series of six concentrations. The analysis was based on the Dutch Standardization Institute protocol, with some validated modifications regarding SPE extraction. Detection limits using silicone rubber passive sampling were approximately 1 ng/L water.

2.4.4 Pharmaceuticals. One mL portions of the POCIS acetone extracts were evaporated to dryness, and residues were dissolved in 100 µL of methanol (ultra-LC/MS grade, Biosolve) plus 1 mL of MilliQ water. Pharmaceuticals were analysed using ultra-HPLC (Waters Acquity), equipped with a quaternary pump, combined with a Quattro Xevo triple-quadrupole mass selective detector (Waters Micromass) with electrospray ionization. The average recovery was 91 ± 14%. A detailed description of the method and its validation is given in [17]. Detection limits using POCIS passive sampling were approximately 0.01 ng/L water. Analyses of pharmaceuticals were performed at Het Waterlaboratorium (HWL, Haarlem, Netherlands).

2.5 Bioassay analyses and sample treatment

Bioassays were performed at the Waterproef laboratory (Edam, Netherlands), unless stated otherwise. Silicone rubbers without PRC spikes were used from 2011 till 2013 for the bioassay analyses. From 2014 silicone rubbers with PRC were used for both chemical and bioassay analyses since blank effects were negligible (see 3.1). Passive sampling extracts were converted to other solvents before exposure in bioassays for specific and reactive toxicity. Details on solvent transfer are given below.

2.5.1 Daphnia in situ assay. For the Daphnia *in situ* assay daphnids were exposed to the water of all sites. Two cohorts of 10 daphnids (eight days old) were exposed to the water phase in 250 mL glass jars with a 300 µm mesh gauze cover. Experiments were carried out at the first week of the passive sampling deployment. Two jars were connected to the cages that contained the passive samplers, below the water surface, so that fresh water, small algae and bacteria (food for the daphnids) could enter the jars. The gauze cover kept the daphnids inside the jars and kept predators outside. Percentage survival of the *in situ* exposed Daphnia magna was monitored after one week of exposure. An observed mortality of 20% was used as trigger for potential ecological effects, since this percentage is used as blank validity criterion for the chronic Daphnia assay [18].

2.5.2 Preparation of extracts for non-specific toxicity with laboratory bioassays. For the laboratory assays of non-specific toxicity, SR extracts were evaporated to dryness under nitrogen and residues were dissolved in 60 mL of 'Dutch Standard Water' (DSW). DSW (200 mg of $CaCl_2.2H_2O$, 180 mg of $MgSO_4.7H_2O$, 100 mg of $NaHCO_3$ and 20 mg of $KHCO_3$ per liter MilliQ water, final pH 8.2) was freshly prepared from concentrated stock solutions and aerated for two hours. The DSW extracts were used for three acute bioassays. Bioassays for non-specific toxicity in 2011 and 2012 were performed at IMARES Institute for Marine Resources & Ecosystem Studies (IJmuiden, Netherlands). From 2013 they were performed at the Waterproef laboratory, according to the same procedures.

2.5.3 Daphniatox bioassay. The Daphnia immobilisation assay was performed according to [19] and [20], with reduced test volumes. Tests were performed in quadruplicate, wherein each concentration had a volume of 1 mL. In each test system five juvenile daphnids (<24 hours old) were exposed for 48 hours to a concentration range of the DSW extracts. After 24 and 48 hours, immobile daphnids were counted. EC50 values (volume-%) were determined by non-linear regression analysis with a log-logistic model by the statistical program SPSS (SPSS Inc, Chicago). The quality of the bioassay was determined by exposure to potassium dichromate (Boom BV, Netherlands). Quality of the daphnids was

checked with the requirement that >90% of the daphnids in the blank exposures was still mobile at the end of the test.

2.5.4 Algaltox bioassay. The inhibition of growth of algae was determined by the Algaltox assay, according to [21] and [22] procedures, with reduced test volumes, based on [23]. A known amount of algae from an exponentially growing culture was exposed to a dilution series of the DSW extracts. The assay was performed in 96-well microtiter plates, with eight wells for each concentration with a total volume of 250 µL. Algae are added at a starting concentration of about 10,000 cells/mL and measured with fluorescence (emission 670 nm, excitation 460 nm). After 24, 48 and 72 hours, the algal growth was determined. Exponential algal growth curves were determined to assess the percentage of growth inhibition, as compared to controls. Quality assurance of the algae was performed by exposure to potassium dichromate (Boom B.V, Netherlands). The algal growth in the controls should reach a rate of 0.92/day, according to [22]. EC50s were calculated using sigmoidal dose response curves with variable slopes [24].

2.5.5 Microtox assay. The bacterial luminescence inhibition assay is also known as Microtox® test. The test was performed by exposing the bioluminescent marine bacterium *Vibrio fischeri* to the water extracts. The degree of acute toxicity, expressed as EC50, was determined by the inhibition of the luminescence produced by *Vibrio fischeri* exposed to a concentration range of the DSW extracts. The procedure of the manufacturer [25], with reconstituted freeze-dried bacteria, was applied. Light emission was measured after 5, 15 and 30 minutes of exposure. The quality of the used batch of bacteria was monitored by testing of phenol (Acros Organics, USA). Microtox Omni software (version 1.18) was used for determination of the EC50 values and 95% confidence intervals.

2.5.6 RIKILT WaterSCAN (SCreeening ANtibiotics assay). Activities of five groups of antibiotics were determined with the WaterSCAN assay, obtained at RIKILT (Netherlands). The test system comprises five plates with different composition and specific bacteria (details outlined in [26]): the T-plate for tetracyclines, the Q-plate for quinolones, the B&M-plate for β-lactams and macrolides, the A-plate for aminoglycosides and the S-plate for sulphonamides. Inoculated agar was poured as a 2.5-3 mm thick layer and nine holes (14 mm diameter) were punched in each plate. Plates were stored for less than one week (4°C).

POCIS acetone extracts (2 mL) were evaporated to dryness under nitrogen and dissolved in 3 mL of methanol:water (1:1). The methanol:water extracts (250 µL) were pipetted into punch holes of each of the 5 plates, supplemented with a plate specific buffer (one drop)

and incubated for 16-18 hours at 30°C (T- and Q-plates) or 37°C (other plates). A plate-specific positive control solution (250 µL) was added to the centre punch hole of each plate. Positive controls consisted of 100 µg/L oxytetracycline (T-plate), 200 µg/L flumequine (Q-plate), 15 µg/L penicillin G (B&M-plate), 100 µg/L sulphamethoxazole (S-plate) and 200 µg/L neomycine (A-plate). After incubation of the test plates, antibiotics activities were estimated by measuring the diameters (d) of bacterial inhibition zones. The effect is proportional to the surface areas of cleaned zones ($=0.25*\pi*d^2$) minus the areas of the punch holes (154 mm^2). Estimations of the antibiotics-equivalents in the samples were made by comparing the inhibition zones of samples and positive controls. Antibiotic activities are expressed as bioanalytical equivalent concentrations (BEQ) of the reference antibiotics.

2.5.7 Specific and reactive CALUX reporter gene bioassays. Polar and non-polar passive sampling extracts were analysed by a panel of *in vitro* CALUX® (Chemical Activated LUciferase gene eXpression) bioassays. The non-polar SR extracts (2 mL in methanol:acetonitrile) were evaporated to dryness and taken up in 5 mL of hexane. Half of this hexane extract was evaporated to dryness and taken up in 50 µL of DMSO. This fraction was used for determinations of ER*a*, anti-AR, GR, and p53 CALUX activities. The remaining half of the hexane extract was cleaned on an acidic silica-column after which it was evaporated to dryness, and taken up in 25 µL of DMSO. This fraction was only used for DR CALUX activity, in order to specifically measure the effects of persistent dioxin-like compounds on the aryl-hydrocarbon receptor (AhR). The polar POCIS extracts (2 mL in acetone) were evaporated to dryness and taken up in 50 µL of DMSO. This extract was used for determinations of ER*a*, anti-AR, GR and p53 CALUX activities. All CALUX analyses were performed at BioDetection Systems BV (Amsterdam, Netherlands).

A CALUX bioassay panel for multiple modes of action was carried out, using previously described protocols [3, 6, 27, 28]. In short, dilution series were made of all DMSO extracts, after which the activity was determined in various CALUX bioassays. Specific CALUX cells were plated in 96-well microplates and after 24 hours of pre-incubation (37°C, at 7.5% CO2) exposed to the DMSO extracts (0.1 to 1.0% DMSO, triplicate measurements). After 24 hours of exposure (six hours for PAH CALUX), cells were lysed, and the luciferase activity was determined after addition of luciferin, using a multiwell luminometer (Lucy 2, Anthos, Austria). To rule out confounding influences, cells were monitored for cytotoxicity. The effects of water extracts were expressed as BEQ of the reference compounds. Dose-response curves of the reference compounds were included on each 96-well plate:

- DR CALUX: Dioxin-like effects, expressed as 2,3,7,8-TCDD EQ (AhR-agonist)

- PAH CALUX: PAH-like effects, expressed as benzo(a)pyrene EQ (AhR-agonist)
- ERα CALUX: Estrogenic activity, expressed as 17ß-estradiol EQ (ER-agonist)
- Anti-AR CALUX: Androgenic inhibition, expressed as flutamide EQ (AR-antagonist)
- GR CALUX: Glucocorticoid activity, expressed as dexamethasone EQ (GR-agonist)
- PPARγ CALUX: Peroxisome proliferation, expressed as rosiglitazone EQ (PPAR agonist)
- Nrf2 CALUX: Oxidative stress, expressed as curcumine EQ (Nrf2 inducing compound)
- PXR CALUX: Xenobiotics metabolism, expresses as nicardipine EQ (PXR agonist)
- p53 CALUX: Genotoxicity, expressed as relative enrichment factor (REF) for significant genotoxic effect

2.6 SIMONI modelling as bioanalytical indication for environmental risks

The SIMONI model uses a simple formula that aims to quantify the combined ecological hazards due to micropollutants, by integrating all individual bioassay responses [1]. All bioassays have been given a weight factor, i.e. 2 for apical toxicity endpoints (*in vivo*) and 1 for specific and reactive toxicity endpoints (*in vitro*), in order to get an equal weight for 5 non-specific and 10 specific endpoints. The SIMONI model divides all bioassay responses (toxic units [TU] or bioanalytical equivalents [BEQ]) by their associated EBT and multiplies them with this weight factor. One average relative response was calculated for the five antibiotics assays and two genotoxicity assays (with and without S9 metabolic activation). Results are then summed for all applied bioassays and divided by a proposed percentage (50%) of the total weight of the bioassays:

$$\text{SIMONI score} = \frac{\sum_{i=1}^{n} \left(\frac{\text{bioassay response}_i}{\text{EBT}_i} \right) \text{x weight}_i}{0.5 \text{ x total bioassay weight}} \qquad (2)$$

As a requirement for a reliable result it is assumed that the total weight of the applied bioassays must be at least 10 (weight of the entire bioassay battery is 20, i.e. a 50:50 distribution between *in vivo* and *in vitro* bioassays). A total SIMONI score above 1 is a provisional indication for ecological risks due to elevated concentrations of micropollutants in the water phase. In this SIMONI score it is assumed that an increased hazard for the ecosystem occurs when the responses of all bioassays are, on average, more than 50% of the proposed trigger values (i.e., total weight factor in Equation 2). The choices embedded in this scoring method were based on the experiences over the last years with sets of raw scores, weighted scores and ecological effects information (section 3.4).

2.7 Mixture toxic pressure modelling for chemical risk assessment

The mixture toxic pressure of all known substances in a sample is determined by quantifying the Potentially Affected Fraction (PAF) per compound, using species sensitivity distribution (SSD) modelling. This is followed by aggregation to the total mixture toxic pressure based upon mixture modelling [11]. The model uses all known concentrations of micropollutants in the water phase as input for compound-specific SSDs, in order to derive the fraction of species probably affected at the level of acute EC50. The SSD-method based on NOECs is often used in the derivation of water quality standards [29]. The protective water quality criterion for a chemical substance is determined for various aquatic organisms. For the so-called multiple-substances potentially affected fraction (msPAF), the method is used in the opposite direction to derive an impact metric based on concentrations. The impact metric is expressed as the percentage of aquatic organisms that may have adverse effects after exposure to the detected micropollutants, the msPAF [11]. The determination of mixture toxic pressures for the sampling sites was executed with a specifically designed software program [30]. The output of the program provides the msPAF as well as a ranking of substances probably contributing most to the potential ecological effects. The msPAF results are based on acute toxicity. While a generally accepted threshold for chronic effects is 5% of affected species, the provisional threshold level for acute effects on environmental health is an $msPAF_{EC50\text{-}acute}$ of 0.5%.

2.8 Statistical analyses

The bioanalytical dataset, limited to 11 bioassays at 39 sites without missing values, was analyzed with StatGraphic Centurion XVI software. A Factor Analysis (a type of Principal Component Analysis) was performed in order to obtain a small number of linear combinations of all variables which account for most of the variability in the dataset. The purpose of the analysis is to obtain a small number of factors which account for most of the variability in the 11 variables of the dataset. Initial communality estimates have been constructed from the squared multiple correlations of each variable with all of the other variables. Equations are constructed that estimate the common factors. A Varimax rotation is performed on the original equations, in order to simplify the explanation of the factors. The values of the variables in these equations are standardized by subtracting their means and dividing by their standard deviations.

Euclidean cluster analysis (Ward's method) was performed on the same dataset in order to create 1 cluster from 39 observations supplied. The clusters represent groups of observations (sampling sites) with similar characteristics. In Ward's method the distance between two clusters is the sum of squares between two clusters summed over all

variables (relative bioassay responses). To form the clusters, the procedure began with each observation in a separate group. It then combined the two observations which were closest together to form a new group. After recomputing the distance between the groups, the two groups then closest together were combined. This process was repeated until only 1 group remained. This statistical analysis is visualized by a dendrogram (section 3.6).

3. RESULTS

3.1 Bioassay responses in polar and non-polar passive sampling extracts

In 2013, a bioassay blank comparison of silicone rubbers with and without the performance reference compounds (PRC) was performed. Both blanks did not show significantly different effects in the bioassays, except from a slightly increased DR CALUX response in the PRC blank. This blank DR CALUX response was much lower than responses that were found at clean reference sites. Therefore, from 2014 the extracts of silicone rubbers with PRC were used for both chemical and bioassay analyses.

Results on the percentages of detectable responses in these bioassays used in field surveys with passive sampling from 2011 until 2015 are presented in Figure 2 (polar concentrates in POCIS) and Figure 3 (non-polar concentrates in silicone rubbers [SR]). The actual motivation behind the selection of SIMONI endpoints has been described earlier [1], so this paragraph only deals with the choice of analyzing polar or non-polar passive sampling (PS) extracts for the different endpoints.

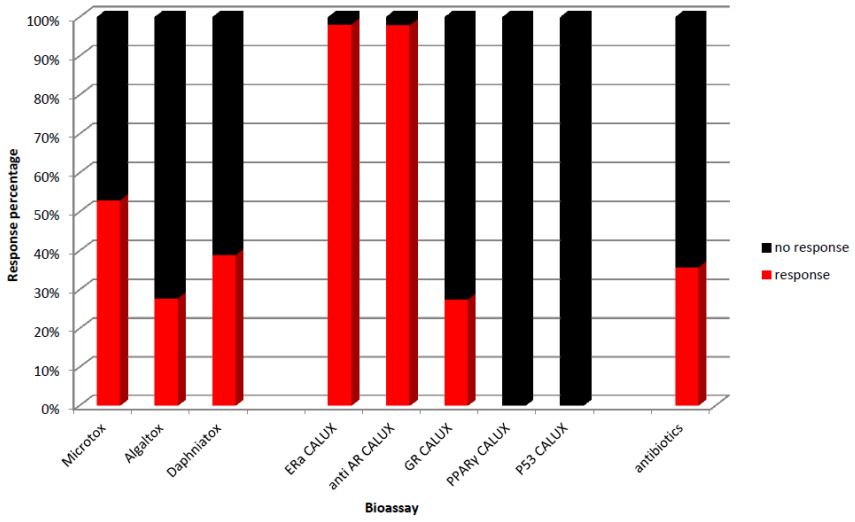

Figure 2: Overview of the percentages of detectable bioassay responses in all experiments performed on polar passive sampling extracts (POCIS) from Dutch freshwater sites (2010-2015). Number of assays performed varies from 7 (PPARγ CALUX) to 56 (antibiotics).

Although it would be relevant to test both polar and non-polar PS extracts on the entire bioassay battery, this would double the bioanalytical costs. For a cost-effective strategy, therefore, choices were made for the most relevant PS extracts to be tested on the selected endpoints.

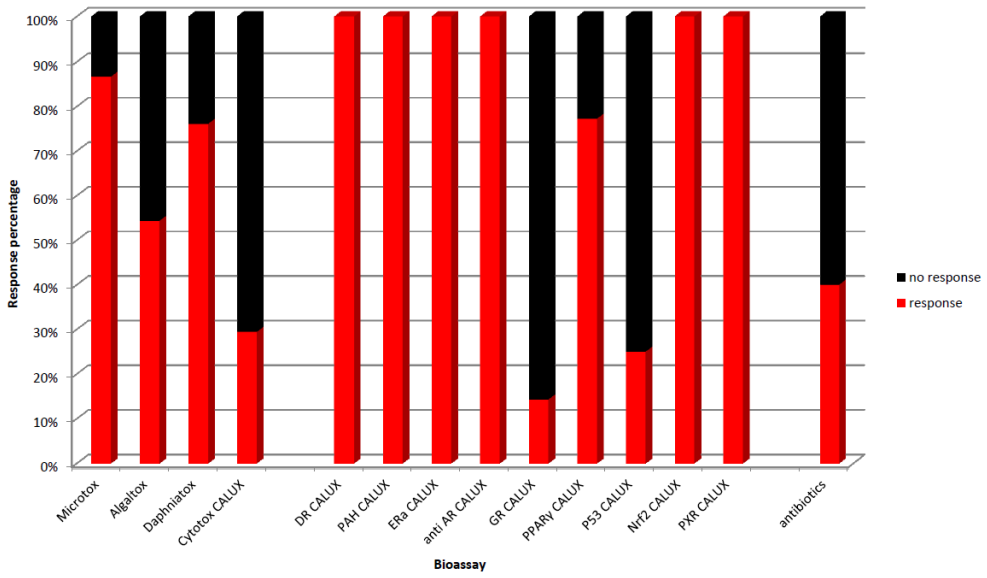

Figure 3: Overview of the percentages of detectable bioassay responses in all experiments performed on non-polar passive sampling extracts (silicone rubber) from Dutch freshwater sites (2010-2015). Number of assays performed varies from 7 (GR CALUX) to 59 (DR CALUX).

The non-specific *in vivo* assays appeared to be most responsive to the non-polar PS extracts, probably due to the higher water volumes that can be extracted (estimates of more than 100L in six weeks). Dioxin- and PAH-like effects were analyzed by DR and PAH CALUX in non-polar extracts that all showed detectable responses, even at 'unpolluted' reference sites. The situation is less clear for endocrine disrupting chemicals (EDC), since significant responses were found in both polar and non-polar extracts. However, since both EDC and antibiotics activities showed clearly increased activities in polar extracts of water affected by WWTP effluents and the fact that many well-known EDCs and antibiotics have polar properties, it was decided to measure these endpoints in polar extracts. Genotoxicity seems to be relevant for both polar and non-polar substances, but a choice was made for the non-polar PS extracts that extract the largest water volumes. For the same reason the 'promiscuous' bioassays for oxidative stress response (Nrf2 CALUX) and xenobiotic metabolism (PXR CALUX) are preferably tested in the non-polar extracts, although many polar compounds are also responsive in these bioassays.

3.2 Extracted water volumes with silicone rubbers

Assumptions on extracted water volumes of passive samplers have to be made in order to interpret bioassay responses of unknown compounds. The SIMONI assumption made for $Vs = 0.5*R_S$ was compared with the Equation for concentration-weighted average sampling volumes (Vs-cwa), derived by Hamers et al. [13], and the average volume without weighing concentrations (Vs-mean). Sampling volumes were calculated for individual compounds analyzed at eight sampling campaigns performed in 2012. Vs-cwa

and Vs-mean were determined using the data on 16 PAHs, 7 PCBs and 33 OCPs. The results of the estimated extraction volumes vs. the calculated Rs of the sampler (section 2.3) are presented in Figure 4.

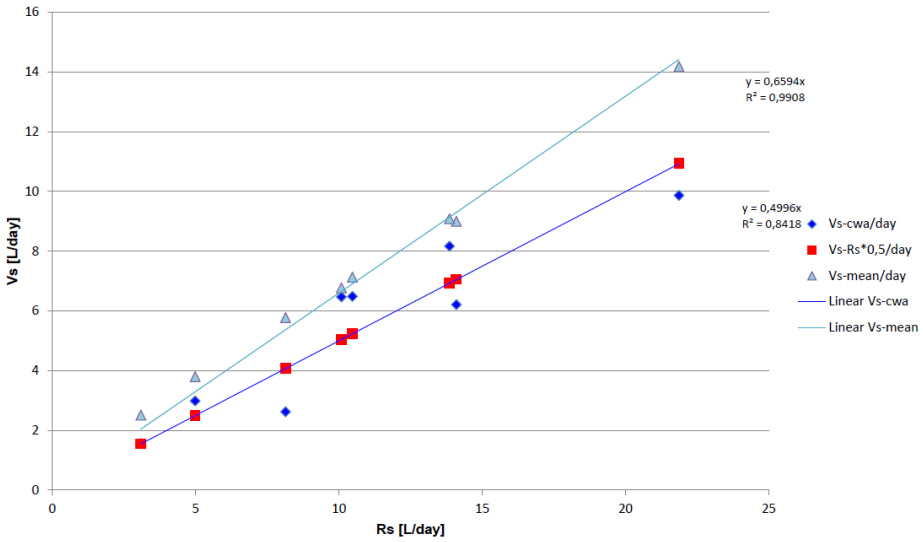

Figure 4: Estimated sampling volumes of silicone rubbers determined as mean sampling volumes (Vs-mean) and as concentration-weighted average sampling volumes (Vs-cwa) determined with individual PAHs, PCBs and OCPs, and as function of the exchange rate (0.5*R_S) of the samplers.

The Vs-cwa correlated well with R_S (R^2 = 0.84) and has the exact same trend line for R_S correlation as the SIMONI assumption Vs-R_S*0.5. The Vs-mean correlated extremely well with R_S (R^2 = 0.99), and mean volumes were 32% higher than those estimated with the SIMONI assumption.

3.3 Results of SIMONI analyses 2011-2015

A SIMONI hazard identification, according to the procedures described in the first paper of this series [1], was performed on all collected data over the last five years. Bioassay responses (BEQ or TU), measured in PS extracts, were converted to estimated water levels, using the Vs-R_S*0.5 estimation proposed in the present paper. These values were compared with specific effect-based trigger values (EBT) for each bioassay, in order to identify potential ecological hazards [1]. Results of all relative bioassay responses that were determined by dividing the water-based response by their respective EBT, are listed in Tables S1 (non-specific toxicity), S2 (specific toxicity polar extracts) and S3 (specific/reactive toxicity non-polar extracts) of the Supplemental Data. All relative bioassay responses were used to calculate the SIMONI score, as an indicator for ecological risks. A heat map of the individual relative bioassay responses (white = 'not measured', green = 'no response', yellow = 'response <EBT' and orange = 'response >EBT') is shown in Table 3, together with the SIMONI scores of all sites. Calculated SIMONI scores >1

(red) indicate a level of exposure to the chemical mixture that causes potential risks for the ecosystem.

Table 3: Heat map of relative bioassay responses divided by their effect-based trigger values and overall SIMONI scores (red: >1 = increased risk) of 45 campaigns at sites that are assumed to be unpolluted, moderately polluted and heavily polluted; names of abbreviated bioassays are listed in Table 1 and site codes are explained in Table 2; 2013 campaigns were performed in July (J) and September (S).

Sites	year	General toxicity					Specific toxicity										Antibiotics						TOTAL
		field	bact	algae	daphnid	cytotox	ER	anti-AR	GR	DR	PPARg	PAH	Nrf2	PXR	p53-	p53+	p53	amino	macro	sulfon	tetra	quino	antibio	SIMONI 1.2
EBT		20	0,05	0,05	0,05	0,05	0,5	25	100	50	10	150	10	3	0,005	0,005		500	50	100	250	100		
units *		%M	TU	TU	TU	TU	EEQ	FluEQ	DexEQ	TEQ	RosEQ	BaPEQ	CurEQ	NicEQ	TU	TU	TU	NeoEQ	PenEQ	SulEQ	OxyEQ	FlqEQ	AEQ	
Unpolluted																								
LWP	2012																							0,2
LNM	2015																							0,3
LWP	2015																							0,3
LBH	2015																							0,4
LRW	2015																							0,3
PKN	2015																							0,7
LGA	2015																							0,4
LKL	2015																							0,1
Moderately polluted																								
MZD	2011																							0,2
SLL	2011																							0,3
RVM	2011																							0,3
WLC	2011																							0,3
ABU	2012																							0,8
AAU	2012																							0,6
RVU	2012																							0,4
RVL	2012																							0,3
SWS	2012																							0,2
RVH	2014																							0,6
RAU	2014																							0,4
RAA	2014																							0,7
RAR	2014																							0,4
RAA	2015																							0,5
RAR	2015																							0,5
LEM	2015																							0,9
Heavily polluted																								
ZL1	2012																							2,7
NL1	2012																							1,5
ZL2	2013-J																							1,4
ZL3	2013-J																							1,4
ZL4	2013-J																							1,4
ZL5	2013-J																							1,8
ZL6	2013-J																							0,3
NL2	2013-J																							1,1
NL3	2013-J																							0,6
DZH	2013-J																							1,1
ZL2	2013-S																							0,8
ZL3	2013-S																							1,2
ZL4	2013-S																							1,3
ZL5	2013-S																							2,2
ZL6	2013-S																							0,9
NL2	2013-S																							1,3
NL3	2013-S																							0,5
DZH	2013-S																							1,0
GHI	2014																							1,7
GHI	2015																							1,7
GBL	2015																							1,7

not measured
no response
response < EBT
response ≥ EBT

*: % mortality [%M], Toxic Units [TU], or expressed as equivalents of the reference compounds: EEQ = estradiol; FluEQ = flutamide; TEQ = 2378-TCDD; DexEQ = dexamethasone; RosEQ = rosiglitazone; BaPEQ = benzo[a]pyrene; CurEQ = curcumine; NicEQ = nicardipine; NeoEQ = neomycine; PenEQ = penicillin; SulEQ = sulfamethoxazole; OxyEQ = oxytetracyclin; FlqEQ = flumequine.

The results of the bioassay responses at the seven reference sites with a good ecological status were also used as 'background BEQ' for the derivation of effect-based trigger values

[1]. EBT of the individual bioassays were exceeded at many sites, even at some of the sites that are presumed to be unpolluted references. The threshold of 1 for the overall SIMONI score that should indicate increased chemical risks, however, is exceeded only at surface waters in greenhouse areas (ZL and NL sites & DZH, potentially polluted with pesticides) and in surface water at GHI and GBL that consists of undiluted effluent of waste water treatment plants (WWTP).

3.4 Impact of varying assumptions for the SIMONI score

Several assumptions were made for the calculation of the SIMONI score, which was designed to serve as bioanalytical indication for the overall environmental risks due to micropollutants. First, the individual bioanalytical endpoints were given a weight factor of 2 for apical *in vivo* endpoints or 1 for *in vitro* endpoints for specific or reactive toxicity [1]. The impact of the weight factors on the overall SIMONI score is demonstrated in Table 4, in which all sites are arranged by increasing SIMONI score.

SIMONI scores in the first column were calculated with the original SIMONI formula (weight factors 1&2), while data for the second column were calculated without adding weight factors to the endpoints. The comparison shows that weight factors only cause minor differences on the overall SIMONI scores: without weight factors SIMONI scores of the polluted NL2-J and ZLP3-S sites (1.09 and 1.17) are below the threshold of 1, while the threshold is just exceeded at the LEM site (1.03) that is considered moderately polluted. When the two scores are compared, the outcomes show a relatively robust sequence of sites, arranged from low to high predicted pollution and SIMONI scores.

The second assumption in the original SIMONI formula is the cut-off percentage of 50%. This means that an assumed environmental risk due to micropollutants occurs when the responses of all bioassays are, on average, more than 50% of the proposed effect-based trigger values (EBT). The results of calculations in Table 4 show that this cut-off percentage has a significant impact on the number of sites with SIMONI values exceeding the threshold of 1. With a low cut-off percentage of 20% average EBT exceedance the threshold is exceeded at almost 70% of the investigated sites, while 16% exceedance is observed with a 75% cut-off. Cut-off percentages between 40% and 60% seem to be most realistic, since no thresholds are exceeded at unpolluted sites (white) and not many thresholds are exceeded at sites that are considered moderately polluted (grey).

Table 4: Impact of assumptions on weight factors and EBT exceedance cut-off percentages on the SIMONI scores >1 (red), that indicate environmental risks due to organic micropollutants.

Sites	pollution	SIMONI 1.2 50%-EBT weight1/2	no weight 50%-EBT weight1	lower cut-off percentage 20%-EBT weight1/2	30%-EBT weight1/2	40%-EBT weight1/2	higher cut-off percentage 60%-EBT weight1/2	75%-EBT weight1/2
LKL		0,14	0,16	0,36	0,24	0,18	0,12	0,10
WLP 2012		0,19	0,22	0,46	0,31	0,23	0,15	0,12
MZD		0,20	0,19	0,50	0,33	0,25	0,17	0,14
SWS		0,23	0,28	0,57	0,38	0,28	0,19	0,15
WLC		0,26	0,28	0,64	0,43	0,32	0,21	0,18
LRW		0,27	0,30	0,66	0,44	0,33	0,22	0,18
LNM		0,29	0,45	0,73	0,48	0,36	0,24	0,19
WLP 2015		0,31	0,30	0,77	0,51	0,39	0,26	0,21
RVM		0,32	0,36	0,79	0,53	0,40	0,26	0,23
SLL		0,32	0,38	0,81	0,54	0,40	0,27	0,23
RVL		0,34	0,41	0,86	0,57	0,43	0,29	0,23
ZL6 J		0,34	0,41	0,86	0,57	0,43	0,29	0,23
RAR 2014		0,35	0,49	0,88	0,58	0,44	0,29	0,23
RVU		0,39	0,49	0,97	0,65	0,49	0,32	0,26
LGA		0,41	0,49	1,03	0,69	0,52	0,34	0,27
LBH		0,42	0,45	1,04	0,69	0,52	0,35	0,28
RAU		0,42	0,59	1,05	0,70	0,52	0,35	0,28
RAR 2015		0,48	0,54	1,20	0,80	0,60	0,40	0,32
NL3 S		0,49	0,51	1,23	0,82	0,61	0,41	0,33
RAA 2015		0,50	0,55	1,24	0,83	0,62	0,41	0,33
NL3 J		0,56	0,47	1,40	0,94	0,70	0,47	0,37
RVH		0,62	0,62	1,55	1,03	0,77	0,52	0,41
AAU		0,65	0,80	1,62	1,08	0,81	0,54	0,43
RAA 2014		0,68	0,82	1,71	1,14	0,86	0,57	0,46
PKN		0,74	0,68	1,84	1,23	0,92	0,61	0,49
ABU		0,78	0,75	1,95	1,30	0,97	0,65	0,52
ZL2 S		0,83	0,77	2,07	1,38	1,04	0,69	0,55
ZL6 S		0,86	0,95	2,16	1,44	1,08	0,72	0,58
LEM		0,88	1,03	2,20	1,47	1,10	0,73	0,59
DZH S		0,96	0,93	2,40	1,60	1,20	0,80	0,64
DZH J		1,06	1,39	2,64	1,76	1,32	0,88	0,70
NL2 J		1,09	0,89	2,71	1,81	1,36	0,90	0,72
ZL3 S		1,17	0,99	2,92	1,94	1,46	0,97	0,78
NL2 S		1,30	1,10	3,24	2,16	1,62	1,08	0,86
ZL4 S		1,33	1,34	3,31	2,21	1,66	1,10	0,88
ZL3 J		1,40	1,40	3,51	2,34	1,76	1,17	0,94
ZL4 J		1,44	1,16	3,60	2,40	1,80	1,20	0,96
ZL2 J		1,45	1,13	3,62	2,41	1,81	1,21	0,97
NL1		1,53	1,31	3,82	2,55	1,91	1,27	1,02
GBL		1,65	1,89	4,13	2,75	2,06	1,38	1,10
GHI 2015		1,69	1,99	4,23	2,82	2,11	1,41	1,13
GHI 2014		1,70	1,95	4,24	2,83	2,12	1,41	1,13
ZL5 J		1,85	1,89	4,62	3,08	2,31	1,54	1,23
ZL5 S		2,21	1,78	5,52	3,68	2,76	1,84	1,47
ZL1		2,74	3,86	6,86	4,57	3,43	2,29	1,83
SIMONI >1 % exceedance		15 33%	14 31%	31 69%	24 53%	19 42%	12 27%	7 16%

unpolluted
moderately pollluted
heavily polluted

3.5 Comparison of risk analysis with chemical and bioanalytical measurements

In the campaign that was carried out at eight sites in 2012, the passive sampling extracts were analyzed with both extensive chemical and bioanalytical methods, so that a cross-comparison between SIMONI results and other risk assessment methods can be made.

3.5.1 Chemical analyses and water quality standards. Total water concentrations of six groups of organic micropollutants are shown in Table 5. The results of Table 5 are the

summed concentrations of individual compounds that are listed in the Supplemental Data: 16 polycyclic aromatic hydrocarbons (PAHs: Table S4), 7 polychlorinated biphenyls (PCBs: Table S5), 33 organochlorine pesticides (OCPs: Table S6), 58 nitrogen/phosphorous pesticides (NPPs: Table S7), 38 polar pesticides (PPs: Table S8) and 44 pharmaceuticals (Pharma: Table S9). Water levels of PAHs, PCBs, OCPs and NP pesticides were based upon calculations with silicone rubber data and exchange rates (R_S) of the samplers that were determined with performance reference compounds. Water concentrations of polar pesticides and pharmaceuticals were provisional estimations obtained with POCIS data (section 2.3).

Table 5: Concentrations of micropollutants in water, calculated with levels in silicone rubber (SR) or POCIS extracts; PAHs: polycyclic aromatic hydrocarbons; PCBs: polychlorinated biphenyls; OCPs: organochlorine pesticides, NPPs: nitrogen/phosphorous pesticides; PPs: Polar pesticides; Pharma: pharmaceuticals.

	PAHs	PCBs	OCPs	NPPs	PPs	Pharma
	total	total	total	total	total	total
PS extract >	SR	SR	SR	SR	POCIS	POCIS
Site	ng/L	ng/L	ng/L	ng/L	ng/L	ng/L
ZL1	86	0,6	8,2	**1799**	18	7
NL1	97	0,4	3,5	**1830**	**240**	31
ABU	197	1,5	**8,1**	**496**	38	18
AAU	**704**	0,7	**8,7**	**582**	73	37
RVU	170	2,7	**8,6**	**630**	34	420
RVL	166	1,3	**3,8**	**698**	10	59
SWS	75	0,5	3,9	**147**	17	22
WLP	36	1,8	2,7	**315**	4	4
Blank	3	0,3	0,4	0	1	6

= levels above MAC-EQS or maximum tolerable risk (MTR) values detected
= levels above provisional guideline values or AA-EQS detected

The measured concentrations were compared to regulatory water quality standards. Based upon these guidelines for chemical analytical data, a potential acute chemical risk for the aquatic ecosystem was observed at all sites (individual EQS exceedances are indicated in Tables S4 to S9 of the Supplemental Data). Most exceedances of the chemical guidelines, however, were observed for the Dutch regulatory Maximum Tolerable Risk (MTR) values (DDT and derivatives, DEET, pirimicarb and carbendazim). The ad hoc MTR for DEET (110 ng/L) was exceeded at all eight sites. The EU Water Framework Directive environmental quality standards for the maximum acceptable concentrations (MAC-EQS) were only exceeded for pirimiphos-methyl at ZL1 and fluoranthene at AAU. No MAC-EQS of DDTs and DEET are available while MAC-EQS of pirimicarb and carbendazim are higher than the MTR guidelines. Annual average EQS (AA-EQS) of heptachlor epoxides were exceeded at all sites, and several PAH AA-EQS were exceeded at all sites except for the WLP reference. Although estimated time-weighted average concentrations over several weeks are

obtained, these passive sampling campaigns cannot be directly compared with annual average concentrations of twelve monthly grab samples.

3.5.2 Bioanalytical analyses and SIMONI modelling. EBT exceedances of the 2012 bioassay responses are shown in different compartments of Table 3 (site codes indicated in light yellow). The numerical values of bioassay/EBT values are listed in Tables S1, S2 and S3 of the Supplemental Data. Individual bioassay analyses revealed various toxicity profiles. High percentages of *in situ* Daphnia mortality were observed at the greenhouse area sites ZL1 en NL1 (100% and 70%, respectively). This bioassay responds to acute toxicity of all substances present in the water phase, including inorganic substances like metals, salts and nutrients. A moderate mortality (30%) was observed at the ABU site of river Amstel that partly receives its water from this greenhouse area, while no mortality was observed at the other 2012 sites. All other bioassays were performed on passive sampler extracts. The proposed EBT [1] were exceeded for Daphniatox at the ZL1 and NL1 sites, for anti-androgenic activity at ZL1, NL1 and AAU sites, for genotoxicity at the ZL1 site, and for PAH activities at ABU, AAU (both river Amstel sites) and RVL (river Vecht). The increased PAH CALUX activity at the river Amstel sites was in agreement with the observed AA-EQS exceedance of fluoranthene. No EBT exceedances were observed at the WLP reference site (both in 2012 and 2015). Regarding the overall bioanalytical results that indicate environmental risks, SIMONI scores >1 were found at the ZL1 and NL1 sites that receive run-off water from greenhouses, which indicates a potentially increased risk of micropollutants for the ecosystem. SIMONI scores of both river Amstel sites (ABU and AAU), both river Vecht sites (RVL and RVU), Smal canal Weesp (SWS) and WLP reference lake were all below the SIMONI threshold that signals an increased chemical risk for the ecosystem. According to this toxicity screening, environmental risks due to agriculture seem higher than those at rivers Amstel and Vecht that are, amongst others, influenced by WWTP effluent discharges. These results do not seem to be consistent with the assessment based on individual chemical guidelines.

3.5.3 Chemical analyses and mixture toxic pressure modelling. An alternative approach to assess the environmental risks with chemically measured pollutant concentrations is the determination of the overall 'mixture toxic pressure'. The toxic pressure is determined by calculating the potential ecological effects of the measured concentrations of chemicals. It is expressed as the percentage of aquatic organisms that may exhibit adverse effects after exposure to the detected micropollutants, aggregated to the so-called multiple-substances potentially affected fraction (msPAF, [11]). The msPAF values of the eight sites investigated in 2012 were calculated with the extensive dataset of compounds that were analysed in the passive sampler extracts (Supplemental Data, Tables SI-4 to SI-9). Figure

5 shows the relationship between the acute msPAF values based upon chemical concentrations and the SIMONI score based upon bioanalytical measurements. A good correlation (R^2 = 0.80) was observed between these two parameters, which implies that both methods provide similar insights in relative risk ranking for the studied water bodies. The provisional threshold levels for increased environmental risks (1.0 for SIMONI score and 0.5% for msPAF) are both exceeded at the ZL1 and NL1 sites (red bullets), while risk estimations for the other 2012 sites (blue bullets) are all below both SIMONI and msPAF thresholds.

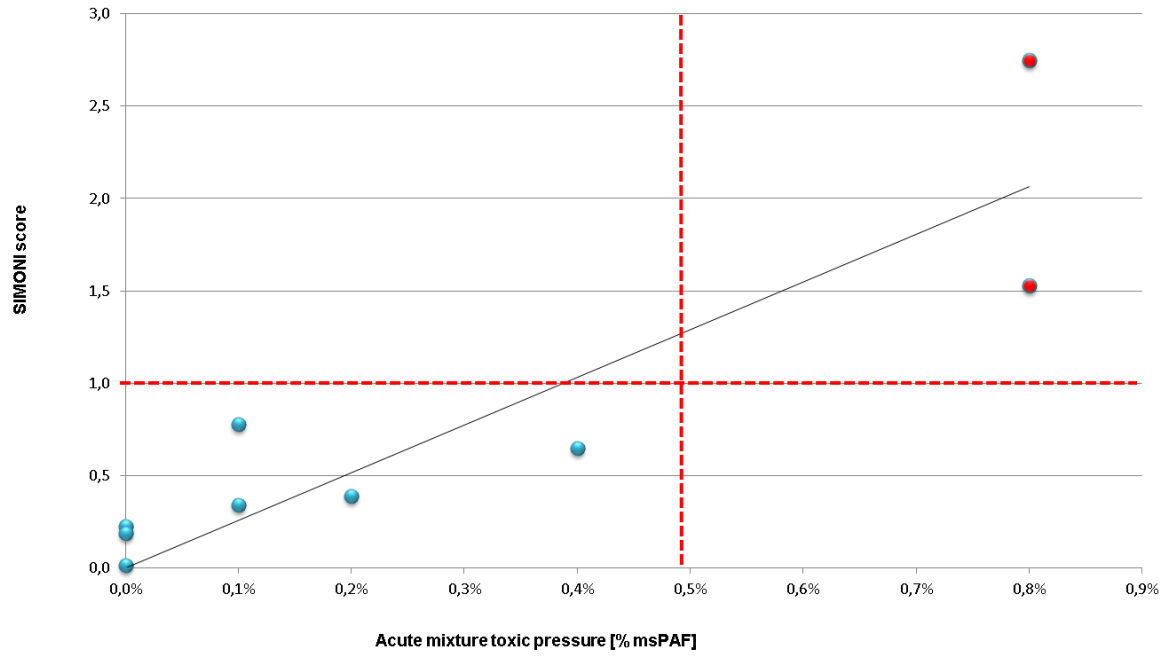

Figure 5: Comparison of environmental risk estimations with bioanalytical SIMONI score and mixture toxic pressure (chemical msPAF); provisional threshold levels are indicated as a dashed red lines; red bullets indicate ZL1 and NL1 sites with increased environmental risks.

A disaggregation analysis of the mixture toxic pressure data identified the compounds contributing most to the msPAF and provided a ranking of 'most important chemicals within a mixture' [31]. This analysis indicated that the highest chemicals risks at ZL1 were due to pirimicarb, diazinon and pirimiphos-methyl and highest risks at NL1 were due to chlorpyrifos, pirimicarb and imidacloprid. These are logical outcomes given the greenhouse-area in which the samples were taken. The MTR guideline for pirimicarb (90 ng/L) was exceeded at the ZL1 and NL1 sites, but water concentrations at both sites were below the MAC-EQS of 1800 ng/L. The MTR and AA-EQS for diazinon (37 ng/L) were not exceeded, but the MAC-EQS for pirimiphos-methyl (1.6 ng/L) was exceeded at ZL1. The MAC-EQS of chlorpyrifos (100 ng/L) was not exceeded at NL1. The MAC-EQS guideline for imidacloprid (200 ng/L) was not exceeded, but estimated concentrations at NL1 were above the AA-EQS (8 ng/L).

3.6 Statistical identification of major bioanalytical driving factors for risk

A limited dataset of relative bioassay responses (Table 3, except missing values) was used for statistical analyses. The limited data set was constructed as follows. Data on CALUX assays for cytotoxicity, PXR and PAHs with only a limited number of data were removed and average responses were used for p53 CALUX with and without S9 addition and for the 5 antibiotic activity assays. The sites with missing bioassay values (MZD, SLL, RVM, WLC, DZH-J and LGA) were then removed. The remaining dataset with 11 bioassays and 39 sites is presented in Table S10 of the Supplemental Data. This dataset was subjected to Factor analysis, in order to obtain a small number of factors which account for most of the variability in the 11 variables. This analysis showed that 2 factors could be extracted with eigenvalues greater than or equal to 1.0. Together these factors accounted for 83.3% of the variability in the original data, which was much higher than the variability obtained by the first two components of Principal Component Analysis (53.6%). The table and graph of the equations of factors 1 and 2 loading are presented in Figure 6.

Bioassays	factor 1	factor 2
D-field	-0.0987605	0.633046
bact	-0.364492	0.37025
algae	-0.174881	0.528211
daphnid	-0.0902826	0.844664
ER	0.797673	0.0234768
anti-AR	0.0922907	0.203581
GR	0.976467	0.0399557
DR	0.313052	-0.0931241
PPARg	-0.260925	0.226291
p53	-0.0202287	0.892514
antibio	0.914328	0.0118654

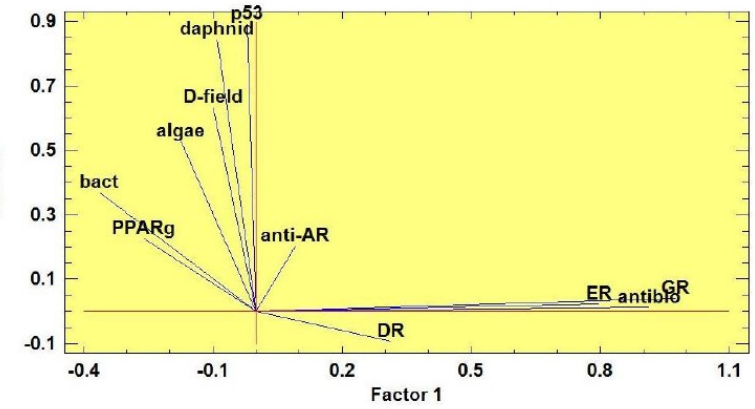

Figure 6: Factor loading matrix and graph after Varimax rotation; factor loadings are presented on a colour gradient from low (green) to high (red); full bioassay names are mentioned in Table 1.

Factor 1 is mainly determined by the endocrine disruptive responses on ER and GR CALUX and the antibiotic activities. Factor 2 is mainly determined by effects on genotoxicity (p53 CALUX) and non-specific toxicity on daphnids (both field and lab assays) and algae. If the factor equations are applied on the relative bioassay responses per site, all 39 sited can be presented in the scatterplot of Figure 7. The majority of sites are clustered together, due to low values of both factors. The GHI and GBL sites that are heavily impacted by undiluted WWTP effluents, however, had clearly elevated factor 1 values. The ZL and NL sites that were influenced by emissions of greenhouse agriculture, on the other hand, had elevated factor 2 values. This implies that sites that were impacted by two different sources of micropollutants are clearly separated.

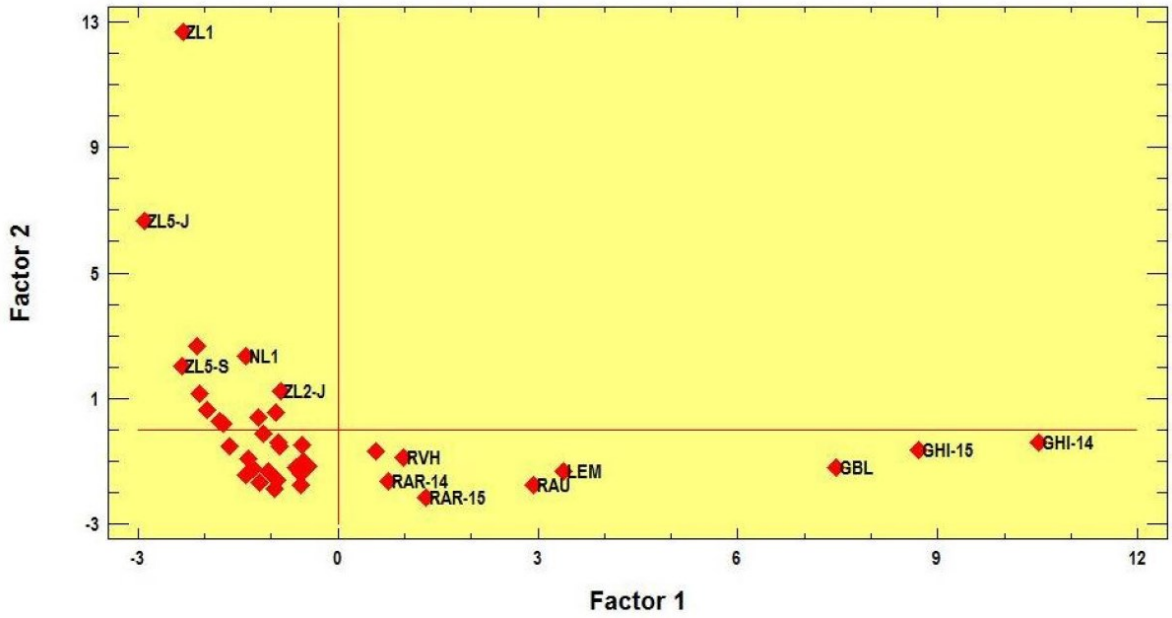

Figure 7: Scatterplot of 39 investigated sites after factor analysis. Site codes are explained in Table 2.

The dendrogram of the Euclidean cluster analysis on the same dataset used for the Factor analysis is presented in Figure 8, together with SIMONI scores of all sites and a heat map of relative bioassay responses. This analysis showed that the clustering of groups of observations with similar characteristics correlates quite well with the overall SIMONI scores of the sites. Most of the unpolluted sites with SIMONI scores <0.5 are in the left hand cluster, most of the moderately polluted sites with SIMONI scores from 0.5 to 1.0 are in the middle cluster, and most of the heavily polluted sites with SIMONI scores >1 can be found in the right hand cluster. The three sites that consisted of undiluted WWTP effluents were clustered separately on the far right.

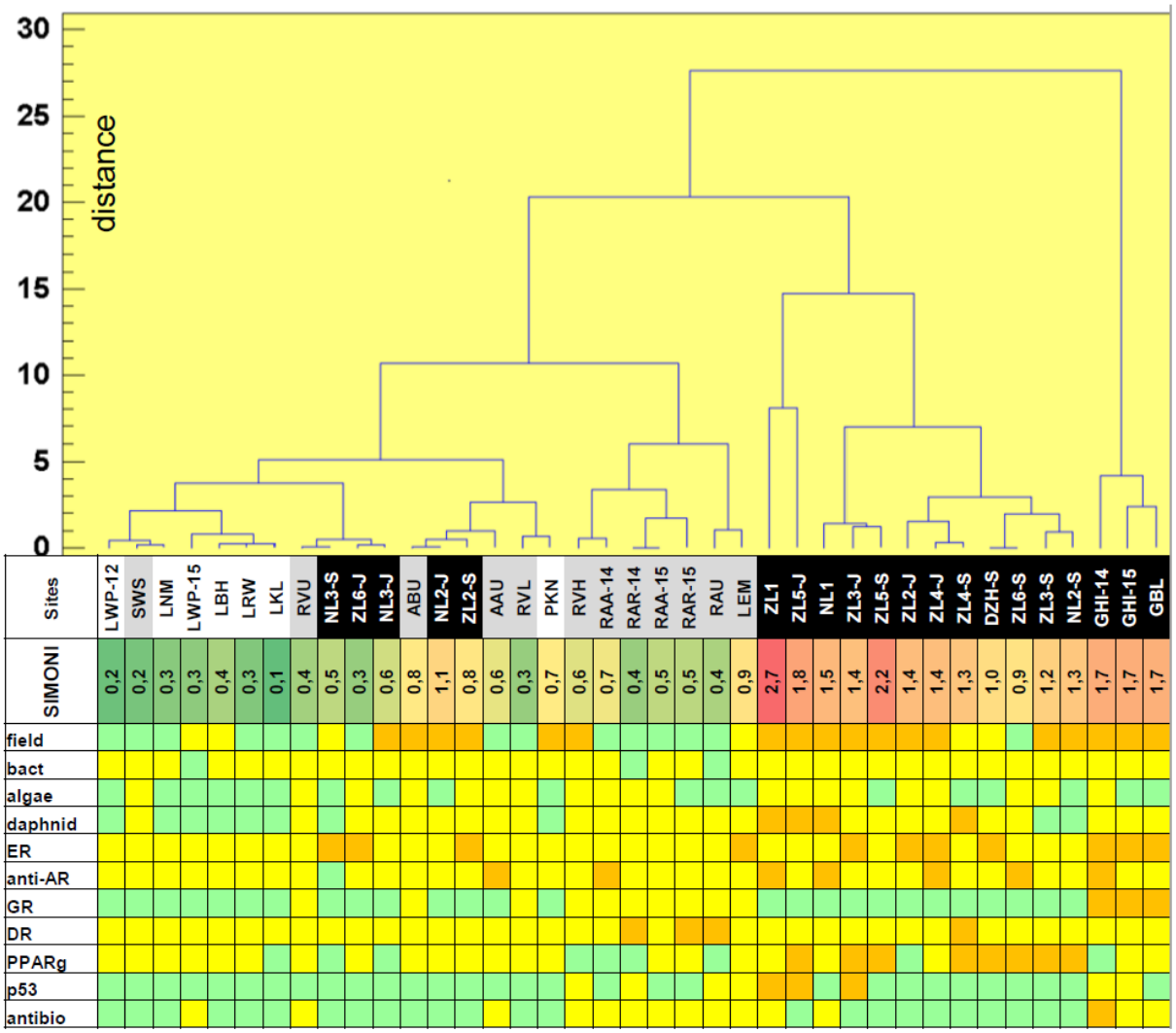

Figure 8: Dendrogram of a statistical cluster analysis of 39 sites that were investigated with 11 bioassays; site abbreviations are explained Table 2; white, grey and black background colours are assumed unpolluted, moderately polluted and heavily polluted status, respectively; SIMONI scores are presented on a colour gradient from low (green) to high (red); bioassay colours indicate no response (green), response below EBT (yellow) and response above EBT (orange).

4. DISCUSSION

The Smart Integrated Monitoring (SIMONI) strategy has been described in detail in an earlier paper by Van der Oost and coworkers [1]. That paper describes a selection of bioanalytical endpoints that are most relevant for water quality assessment, adaptation and derivation of effect-based trigger values (EBT) for a battery of *in vivo* and *in vitro* bioassays and the design of a simple model that converts all bioassay results into a single SIMONI score. The SIMONI score is proposed to be an overall indicator for environmental risks due to micropollutants. The present paper is focused on the practical feasibility and interpretation of this method, with a comparison to other risk assessment methods. Since only organic substances concentrate in the passive sampling devices, this hazard identification is based upon levels of organic micropollutants. For a more complete tier 1 hazard identification additional chemical analyses have to be performed only on inorganic compounds, such as metals and ammonium.

4.1 Combining passive sampling and bioanalytical measurements

The application of passive samplers as a pre-concentration for environmental toxicants applied in bioanalytical testing strengthens toxicity assessment by focusing on cumulative effects of the organic toxicants present [32]. Although the analyses for the SIMONI strategy can also be performed on concentrated grab samples of water, the preferred method is to concentrate the micropollutants by passive sampling. Despite the existence of several knowledge gaps, passive sampling presently is the best available technology for chemical monitoring of nonpolar organic compounds [33]. The advantages of passive sampling are the collection of time-weighted average concentrations instead of snapshots, the concentration of the bioavailable fraction of micropollutants and the eliminations of confounding factors in the matrix (e.g. nutrients, pH and salinity). Bioconcentration of the freely dissolved fraction, sorbed by the passive samplers, is the principal uptake route of compounds with log K_{OW} from 1 to 6 for many aquatic species [34]. In several studies it was demonstrated that the coupling of passive sampling techniques with *in vivo* and *in vitro* bioassays is feasible and offers a cost-effective early warning signal on water quality deterioration, e.g. [32, 35].

A challenge in combining passive sampling and bioassays is to expose the test organisms or cells to comparable concentration ratios of substances as those in the surface waters. Due to varying sampling rates for different substances this is impossible with passive sampling and solvent spike in bioassays [10]. The concentration ratio in POCIS samplers will also be disturbed due to an increased uptake rate with increasing logD_{OW}, the pH-

dependent n-octanol–water distribution ratio [36]. To a lesser extent, however, this disadvantage also applies for other concentration techniques of micropollutants, due to differences in recovery. The concentration ratios can be disturbed both by sampling and dosing. During the solvent spike an additional effect may occur due to differences in sorption to the matrix of the bioassay, e.g. specific sorption to proteins or plastic walls. Passive dosing may improve comparability of concentration ratios in water and bioassay for substances that reach equilibrium during silicone rubber sampling (approximately log K_{OW} < 5.5), but underestimates the risk of more hydrophobic compounds (K. Booij, PaSOC Netherlands, personal communication). It appears to be possible to apply passive dosing in miniaturized 24-well plate bioassays [37]. A way to concentrate substances in the bioassay medium using passive dosing is to spike the extracts to smaller pieces of sampler material, e.g. 20 times concentration will be achieved when extracts of the 20 gram field exposed silicone rubbers are spiked to 1 gram of silicone rubber (K. Booij, PaSOC Netherlands, personal communication). Since passive dosing techniques are more time-consuming than sorbent spikes, however, this would reduce the cost-effectiveness of the strategy.

A major disadvantage of passive sampling is the unknown amount of extracted water. Extracted water volumes for compounds that reach equilibrium during sampling with silicone rubbers are defined by sampler weight multiplied by the sampler-water partitioning coefficient (K_{PW}). For compounds that are in the kinetic phase, the sampled water volume can be estimated by multiplying the sampling rate (R_S) with the deployment time. Since the identity of the substances causing the bioanalytical responses is unknown it is impossible to make an exact calculation of the extracted water volumes for the effect measurements. Two assumptions were made to estimate extracted water volumes of partitioning samplers for non-polar compounds and adsorption samplers for polar compounds (section 2.3, Materials and Methods). The partitioning samplers (silicone rubbers) are spiked with performance reference compounds (PRC), in order to calculate the exchange rate (R_S in L/day) of the samplers, which is affected by factors such as flow rate of the water, water temperature and biofouling. Based on PRC dissipation of water exposed passive samplers, >75% equilibrium is typically observed for compounds with log K_{OW} values up to 5.5, with silicone samplers that are exposed to water for 12 to 50 d [38, 39]. In the present paper it is assumed that 50% equilibrium was reached between non-polar compounds in water en silicone rubbers. The actual equilibrium percentages for compounds with log K_{OW} range 3-8 will probably vary between 25% and 100% after six weeks of exposure. An assumed average 50% of the R_S is proposed as the daily extracted water volume. This assumption correlated very well with the estimation of the concentration-weighted average sampling volume [13]. The actual water volumes for

different substances probably vary within one order of magnitude of the volume estimated with the SIMONI assumption.

A second assumption had to be made for POCIS samplers that concentrate the more polar compounds. Extracted water volumes are less affected by turbidity and temperature, but depend upon water agitation and substance properties [40]. For POCIS samplers, R_S variations between 30 and 300 mL/day were found for different types of substances, e.g., pharmaceuticals, pesticides, (alkylated) phenols, hormones, UV blockers [14, 15, 16]. A linear uptake for up to 56 days was observed for herbicides and pharmaceuticals with log K_{ow} <4.0 [40]. For the adsorption samplers an average extraction volume is assumed of 100 mL per day per sampler. The actual water volumes and the estimated average volume for different chemicals will vary within one order of magnitude.

It is the opinion of the authors that differences between actual and estimated water volumes are acceptable for a tier 1 screening (hazard identification), because the results are sensitive for potential impact signals such that the test outcomes yield useful SIMONI scores (slight chance of false negatives) even despite this uncertainty. A tier 2 risk assessment, however, should preferably be performed with large-volume filtrated grab samples (if possible with time-integrated steps).

For a cost-effective monitoring strategy a choice had to be made which extracts (polar or non-polar) to use for the bioanalyses. It would be possible to combine both PS extracts for an overall bioanalytical screening, but then a provisional indication of the total extracted water volumes would be impossible. The difference between estimated water volumes of the two types of samplers are around one order of magnitude. Therefore, a selection has been made, based upon percentages of detectable responses and theoretical or practical interpretations of the responses (section 3.1 of the present paper). The non-polar extracts are analyzed for non-specific *in vivo* endpoints, AhR receptor (dioxin and PAH like effects), pregnane X receptor, oxidative stress, genotoxicity and peroxisome proliferation. The polar extracts are analyzed for endocrine disrupting effects (estrogenic, anti-androgenic and glucucorticoid effects) and antibiotic activities. The SIMONI classifications of the present paper were all based upon a significant part of this endpoint/extract selection, and the assumptions made on extracted water volumes mentioned above. If additional budget is available, it would be relevant to analyze apical *in vivo* endpoints, genotoxicity, oxidative stress and PXR responses in both PS extracts. Alternatively, additional bioanalytical test on both PS extracts can be performed in a tier 2 follow-up study.

4.2 Risk analysis with chemical and bioanalytical measurements

A cross-comparison between SIMONI-results and other methods to characterize risks or ecological impacts may further support the interpretation of SIMONI-scores, in comparison to the endpoint of e.g. good chemical and ecological status. An initial classification of the investigated sites (unpolluted, moderately polluted and heavily polluted) was made, based upon inventory of potential sources of micropollutants and historical data on chemistry and biology. Although individual EBT exceedances were found at all classes, the provisional threshold of the overall SIMONI score was only exceeded at the sites that were considered to be heavily polluted. The individual scores that signal potential risks to ecosystems may be further elaborated when deemed necessary, e.g., when a specific threat of endocrine disrupting activity hypothesized from known emissions is verified by an associated EBT exceedance. Looking at the overall picture, the frequency of individual EBT exceedances increases from reference sites (4 scores >EBT at 8 sites), via the moderately polluted sites (18 scores >EBT at 16 sites), to the polluted sites (53 scores >EBT at 21 sites), according to site-selecting criteria and expectations. The difference in frequency of scoring bioassay impacts amongst the selected subgroups suggests that the 'raw data' scores of the bioassay battery associate with increasing chemical exposure. Overall SIMONI scores below 1 were found for samples in which three or less individual bioassays exceeded the EBT signal. The bioassay battery thus seems informative for spatio-temporal comparisons across sites, ranking net mixture risks for ecosystems.

The initial assumptions on bioassay weight factors of 2 (*in vivo*) and 1 (*in vitro*) for the SIMONI score model did not have a significant impact on the classification results. The weight factors, however, provide a balance between the impact of *in vivo* and *in vitro* bioassays on the SIMONI score [1]. The provisional threshold level of the SIMONI score is defined in a way that an increased environmental risk due to micropollutants is indicated when the responses of all bioassays are, on average, more than 50% of the proposed effect-based trigger values (EBT). Based upon the initial classification of the sites, a cut-off percentage between 40% and 60% seems to give the most realistic indication of the environmental risks. Therefore, the 50% cut-off value and the bioassay weight factors will be maintained in the calculation of the SIMONI score, unless future research indicates that adjustments are needed. The overall SIMONI-score helps to easily discern groups of sites in a summary overview, while the underlying quantitative scores provide detailed risk-ranking information when needed.

Data of extensive measurements with both chemical and bioanalytical methods were only obtained for eight passive sampling campaigns in 2012. Varying classifications were found if chemical water concentrations are compared to different water quality guidelines (MTR,

MAC-EQS and AA-EQS). Based upon exceedances of the ad hoc MTR for DEET, all sites should be classified as posing a potential chemical risk for the ecology. At some sites the MTR guidelines for DDT & derivatives, pirimicarb and carbendazim were exceeded. Based upon the Water Framework Directive MAC-EQS guideline, only the ZL1 and ABU sites were classified as an acute chemical risk for the ecology, due to exceedances of pirimiphos-methyl and fluoranthene, respectively. The AA-EQS levels for chronic chemical risks were exceeded at all sites for heptachlorepoxide-cis, at seven sites for one or more PAHs and at three sites for imidacloprid. A major disadvantage of the risk analyses using individual chemical guidelines is that due to the one-out-all-out principle, as applied in the Water Framework Directive, no distinction is made between classification of sites with a slight exceedance of a single guideline and sites with large exceedances of several guidelines. Opposed to this, an overall mixture risk may be present when all measured variables are below their criterion value. Implicitly, this may result in a biased interpretation of acute chemical risks: many sites receive the status of non-conformity due to exceedance of one or more criteria. At first sight, the SIMONI results are not consistent with the assessment based on chemical guidelines described above, which signalled ecological impacts at all sites. However, the exceedance of regulatory guidelines implies that the concentrations of the compounds at the sites are not sufficiently safe, in terms of certainty that the concentration would not cause impacts of any kind (direct ecotoxicity, food-chain mediated effects or human-toxicological effects) under all ambient conditions. Thus, exceedances of generic water quality standards do not necessarily imply an ecological impact, while ecological risks per chemical are signalled more frequently than using bioassay scores.

The determination of the mixture toxic pressure (msPAF) uses an underlying SSD-modeling tool that is based upon a large aquatic toxicity database for almost 5000 substances, collected in addition to the data used for the different guideline systems. An advantage of this model is that its output, like the SIMONI score, provides quantitative mixture risk estimation: higher concentrations of chemicals will contribute to the indication of higher ecological risks (increased percentage of potentially affected species). In addition, like the SIMONI score, a single number for the mixture toxic pressure is easier to interpret for regulators than a long list of substances that are compared to varying guidelines. The classifications that were made with the msPAF and SIMONI models correlated well and resulted in a similar risk ranking and classification of the respective water bodies. This can be interpreted as mutual reinforcement of the interpretation that both strategies yield a meaningful relative risk ranking. Turning to optional absolute judgement criteria (guidelines) for SIMONI and msPAF, a preliminary proposal defined the values indicating increased environmental risks as 1.0 for the SIMONI score (50% average exceedance of bioanalytical EBT) and 0.5 for msPAF (0.5% of species potentially affected

after acute exposure). Micropollutant levels at the two sites with extensive greenhouse agriculture (ZL1 and NL1) were similarly classified as an increased micropollutants risk for the ecosystem. Mixture toxic pressure analyses revealed that these risks were mainly due to NP-pesticides, some of which also exceeded chemical guidelines (pirimicarb MTR and pirimiphos-methyl MAC-EQS). Classifications based upon SIMONI and msPAF results are not similar to the assessment based on chemical concentration guidelines, either because the guidelines do not cover the full mixture composition, or because the route by which (direct toxicity, food chain toxicity or human toxicity) or due to the safety factors that are used to derive the protective chemical guidelines. Therefore regulatory guidelines frequently signal the absence of full protection of all endpoints, without implying a certain ecotoxicological impact.

4.3 Statistics

A statistical Factor analysis was performed on a restricted database of relative bioassay responses (i.e. response divided by the effect-based trigger value). Factor 1 was mainly determined by endocrine disruption and antibiotic activities, while Factor 2 was mainly determined by genotoxicity and apical endpoints on daphnids and algae. Together these factors made a clear distinction between the sites that were classified as exhibiting increased chemical risks for the ecosystem. Euclidean cluster analysis on the relative bioassay responses revealed four main groups of sites: unpolluted sites with SIMONI scores <0.5, moderately polluted sites with SIMONI scores 0.5-1.0 and heavily polluted sites impacted by either greenhouse emissions or WWTP effluents, both with SIMONI scores >1.0. The formula that was used to generate the SIMONI scores thus seems to provide statistically sound and ecologically meaningful results on overall bioanalytical assessment.

The repeatability of SIMONI analysis over time, observed at sites that were analysed by more PS campaigns, relates to the sources of pollution. At the WLP reference and the WWTP-polluted GHI, RAA and RAR sites, a good agreement was observed for the overall SIMONI score in different years (Table 3), despite variations in individual bioassay responses. These similarities are also demonstrated with the cluster analysis, which shows a close clustering of these sites over different years (Figure 7). Some of the ZL and NL sites, polluted by greenhouse residual water discharges, that were monitored in July and September (ZL3, ZL4, ZL5, NL3) are clustered close together. Other greenhouse sites (ZL2, ZL6 and NL2), however, show large differences between the two campaigns (Figure 7). These seasonal differences are most probably due to varying pesticide spraying and emission regimes, whereas the consistency at the WWTP sites signals the time-constancy of this emission source.

5. CONCLUSIONS

The SIMONI model provides a good and reproducible indication of the overall ecological risks due to micropollutants in a tier-1 application, despite uncertainties in the technical execution of the underlying approaches. Certain exceedances of individual effect-based trigger values (EBT) were observed at sites that are considered to be moderately polluted. The overall SIMONI score, however, only indicated increased risks for the ecosystem at sites that were assumed to have elevated concentrations of micropollutants from known sources. Statistical cluster analysis on the dataset revealed that sites with comparable SIMONI scores clustered well together. The SIMONI formula that was used to generate these scores thus seems to provide a meaningful overall assessment of the ecological consequences associated to a set of bioassay responses. The weight factors for the bioassays and the cut-off percentage of average EBT exceedance to indicate a provisional threshold level seem to be relevant choices.

The combination of passive sampling and bioanalytical testing appears to be a relevant combination for an initial tier 1 hazard assessment of organic micropollutants in water systems. A tier 2 risk assessment should preferably be performed with large-volume filtered grab samples for a more exact quantification of the observed bioanalytical and chemical results. Although it would be better to investigate both polar and non-polar PS extracts with the entire bioassay battery, this would double the bioanalytical costs. The selection of testing selected bioassays on polar or non-polar extracts is more cost-effective and appears to provide meaningful results on the identification of potential hazards due to organic micropollutants.

Extensive chemical and bioanalytical data were collected only at a limited number of campaigns. A comparison between classification of the sites with regular water quality guidelines and the SIMONI strategy was inconsistent. Certain chemical guidelines were exceeded at all eight investigated sites (even a clean reference), while the SIMONI score only identified two of the sites with a potential chemical hazard. These differences are most probably due to the process choice made in the derivation of the chemical guidelines and the one-out-all-out principle applied to the chemical classification. An alternative approach to assess the mixture toxic pressure on the ecosystem with results of chemical analyses is the msPAF determination (potentially affected fraction of water organisms due to multiple substances). A similar classification of elevated micropollutant risks at two sites affected by greenhouse emissions was obtained with the SIMONI score and the mixture

toxic pressure (msPAF) values. This implies that these chemical and bioanalytical methods provide similar insights in relative risk ranking for water bodies.

Based upon the results obtained thus far, the tier 1 hazard identification of the SIMONI strategy seems to provide an accurate picture of the potential ecological risks of a wide array of organic micropollutants. The design of this practical monitoring approach will be further developed, optimized and validated over time. The University of Amsterdam just started a follow-up study to optimize the strategy, which is funded by the Dutch water authorities. Part of this investigation is a nationwide feasibility and validation study with the tier 1 strategy that is described in part I of this paper by Van der Oost et al. (2017). Due to its low costs and high relevance, the SIMONI model has the potential to become the first bioanalytical strategy to be applied in regular monitoring of surface water quality.

6. ACKNOWLEDGMENTS

The authors thank the Waternet Research & Innovation group for funding the main part of the SIMONI project. Laura Moria, Leon Kors and Cor Verkerk (Waternet Institute for the Urban Water Cycle, Netherlands) are acknowledged for useful input on practical applications and funding of the monitoring campaigns. We thank the STOWA Foundation for Applied Water Research for funding most monitoring campaigns at the reference sites. The authors like to thank Dick de Zwart and Leo Posthuma (both National Institute for Public Health and the Environment, RIVM, Netherlands) for input on the mixture toxic pressure (msPAF) analyses, and Kees Booij (PaSOC, Netherlands) for useful discussions on passive sampling. Tim Williams (University of Birmingham, UK) Leo Posthuma (RIVM, Netherlands) and Beate Escher (Centre for Environmental Research, UFZ, University of Tübingen, Germany and University of Queensland, Australia) are highly acknowledged for critically reading and improving quality of the manuscript.

7. SUPPLEMENTAL DATA

Table S1: General toxicity

Table S2: Specific toxicity and antibiotics in polar PS extracts

Table S3: Specific toxicity in non-polar PS extracts

Table S4: PAHs in water (SR calculations)

Table S5: PCBs in water (SR calculations)

Table S6: OCPs in water (SR calculations)

Table S7: NP pesticides in water (SR calculations)

Table S8: Polar pesticides in water (POCIS estimations)

Table S9: Pharmaceuticals in water (POCIS estimations)

Table S10: Dataset relative bioassay responses used for statistical analyses

8. REFERENCES

1: Van der Oost R, Sileno G, Suarez Muños M, Besselink H, and Brouwer A, 2017. SIMONI (Smart Integrated Monitoring) as a novel bioanalytical strategy for water quality assessment: Part I. model design and effect-based trigger values. Environ. Toxicol. Chem. 36: 2385-2399.

2: Willemsen A, Vaal MA, De Zwart D, 1995. Microbiotests as tools for environmental monitoring. RIVM report 607042005.

3: Van der Linden S, Heringa M, Man H-Y, Sonneveld E, Puijker LM, Brouwer A and Van der Burg B, 2008. Detection of Multiple Hormonal Activities in Wastewater Effluents and Surface Water, Using a Panel of Steroid Receptor CALUX Bioassays. Environ. Sci. Technol. 42: 5814–5820.

4: Macova M, Toze S, Hodgers L, Mueller JF, Bartkow M, Escher BI, 2011. Bioanalytical tools for the evaluation of organic micropollutants during sewage treatment, water recycling and drinking water generation. Water Res. 45: 4238-4247.

5: Escher BI, Allinson M, Altenburger R, Bain PA, Balaguer P, Busch W, Crago J, Denslow NA, Dopp E, Hilscherova K, Humpage AR, Kumar A, Grimaldi M, Jayasinghe BS, Jarosova 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 JYM, Van der Burg B, Van der Linden SC, Werner I, Westerheide SD, Wong CKC, Yang M, Yeung BHY, Zhang X and Leusch FDL, 2014. Benchmarking Organic Micropollutants in Wastewater, Recycled Water and Drinking Water with In Vitro Bioassays. Environ. Sci. Technol. 48: 1940–1956.

6: Sonneveld E, Jansen HJ, Riteco JAC, Brouwer A, Van der Burg B, 2005. Development of androgen- and estrogen-responsive bioassays, members of a panel of human cell line-based highly selective steroid responsive bioassays. Toxicol. Sci. 83: 136–148.

7: Escher BI and Leusch FDL, 2012. Bioanalytical Tools in Water Quality Assessment. IWA publishing, London (UK).

8: Smedes, F, 2007. Monitoring of Chlorinated Biphenyls and Polycyclic Aromatic Hydrocarbons by Passive Sampling in Concert with Deployed Mussels. In: Comprehensive Analytical Chemistry, Eds. R. Greenwood, G. Mills and B. Vrana, Elsevier, Volume 48, pp 407-448.

9: Li J-Y, Tang JYM, Jin L and Escher BI, 2013. Understanding bioavailability and toxicity of sediment-associated contaminants by combining passive sampling with in vitro bioassays in an urban river catchment. Environm. Toxicol. Chem 32: 2888-2896.

10: Jahnke A, Mayer P, Schäfer S, Witt G, Haase N and Escher BI, 2016. Strategies for Transferring Mixtures of Organic Contaminants from Aquatic Environments into Bioassays. Environ. Sci. Technol. 50: 5424-5431.

11: De Zwart D and Posthuma L, 2005 Complex mixture toxicity for single and multiple species: Proposed methodologies. Environ. Toxicol. Chem 24: 2665-2676.

12: Rusina TP, Smedes F, Koblizkova M, and Klanova J, 2010. Calibration of Silicone Rubber Passive Samplers: Experimental and Modeled Relations between Sampling Rate and Compound Properties. Environ. Sci. Technol. 44: 362–367.

13: Hamers T, Legradi J, Zwart N, Smedes F, De Weert J, Van den Brandhof EJ, Van de Meent D, De Zwart D, 2016. Time-Integrative Passive sampling combined with Toxicity Profiling (TIPTOP): an effect-based strategy for cost-effective chemical water quality assessment. Final report of the LRI-ECO23 project.

14: Morin N, Camilleri J, Cren-Olivé C, Coquery M, Miège C, 2013. Determination of uptake kinetics and sampling rates for 56 organic micropollutants using ''pharmaceutical'' POCIS. Talanta 109: 61–73.

15: Roberts PH, Balaam JL, 2006. Offline extraction and passive sampling. EU Modelkey progress report SSPI-CT-2003-511237-2.

16: Vrana B, Vermeirssen ELM, Allen IA, Kohoutek J, Kennedy K, Mills GA, Greenwood R, 2009. Passive sampling of emerging pollutants in the aquatic environment: state of the art and perspectives, Position paper, Norman Association report Nr. W604002510.

17: Houtman CJ, Ten Broek R, De Jong K, Pieters B and Kroesbergen J, 2013. A multi-component 'snapshot' of pharmaceuticals and pesticides in the river Meuse basin. Environ. Toxicol. Chem. 32: 2449–2459.

18: OECD, 2008. Test No. 211: Daphnia magna Reproduction Test, OECD Publishing, Paris. URL: http://dx.doi.org/10.1787/9789264070127-en

19: ISO, 1998. International Standard. Water quality – Determination of the inhibition of the mobility of *Daphnia magna* Straus (*Cladocera, Crustacea*) – Acute toxicity test. Update ref. no. ISO 6341:1996/Cor.1:1998(E).

20: OECD, 2004. OECD Guideline for testing of Chemicals. Daphnia sp., Acute Immobilisation Test. Adopted on 13 April 2004. OECD Test Guideline 202.

21: ISO, 2004. International Standard. Water quality - Freshwater algal growth inhibition test with unicellular green algae. Second edition 2004-10-01. ISO 8692:2004(E).

22: OECD, 2006. OECD Guideline for testing of Chemicals. Freshwater Alga and Cyanobacteria, Growth Inhibition Test. Adopted on 23 March 2006. OECD Test Guideline 201.

23: Peterson HG, Nyholm N and Ruecker N, 2005. Algal microplate toxicity test suitable for heavy metals. In: C. Blaise & J.-F. Férard (eds.), Small-scale Freshwater Toxicity Investigations. Vol. 1. pp. 243-270.

24: ISO, 2006. International Standard. Water quality - Guidance on statistical interpretation of ecotoxicity data. First edition 2006-04-01. ISO/TS 20281:2006(E).

25: AZUR Environmental, 1998. Microtox® Basic test procedure.

26: Pikkemaat MG, Dijk SO, Schouten J, Rapallini M, van Egmond HJ, 2008. A new microbial screening method for the detection of antimicrobial residues in slaughter animals: The Nouws antibiotic test (NAT-screening). Food Control 19:781–789.

27: Murk AJ, Legler J, Denison MS, Giesy JP, Van der Guchte C, and Brouwer A, 1996. Chemical-Activated Luciferase Gene Expression (CALUX): A Novel in Vitro Bioassay for Ah Receptor Active Compounds in Sediments and Pore Water. Fundam. Appl. Toxicol. 33: 149-160.

28: Hamers T, Kamstra JH, Sonneveld E, Murk AJ, Kester MH, Andersson PL, Legler J, and Brouwer A, 2006. In vitro profiling of the endocrine-disrupting potency of brominated flame retardants. Toxicol. Sci. 92: 157–173.

29: Posthuma L, Traas TP and Suter II, GW, (editors) 2002. Species sensitivity distributions in ecotoxicology. Lewis Publishers, Boca Raton, FL.

30: STOWA, 2016. Posthuma L, De Zwart D, Osté L, Van der Oost R, Postma J Ecologische Sleutelfactor Toxiciteit. Deel 1: Methode voor het in beeld brengen van de toxiciteit. STOWA report 2016-15A (in Dutch), Amersfoort, the Netherlands, 2016. URL: http://www.stowa.nl/projecten/ecologische_sleutelfactor_8_toxiciteit__ontwikkeling_instrument_voor_ecologische_effectanalyse_toxiciteit_

31: De Zwart D, Posthuma L, Gevrey M, Von Der Ohe PC, De Deckere E, 2009. Diagnosis of ecosystem impairment in a multiple-stress context - How to formulate effective river basin management plans. Integrat. Environ. Assess. Managem. 5: 38-49.

32: Shaw M, Negri A, Fabricius K, Mueller JF, 2009. Predicting water toxicity: Pairing passive sampling with bioassays on the Great Barrier Reef. Aquatic Toxicology 95: 108–116.

33: Booij K, Robinson CD, Burgess RM, Mayer P, Roberts CA, Ahrens L, Allan IJ, Brant J, Jones L, Kraus UR, Larsen MM, Lepom P, Petersen J, Pröfrock D, Roose P, Schäfer S, Smedes F, Tixier C, Vorkamp K, and Whitehouse P, 2016. Passive sampling in regulatory chemical monitoring of nonpolar organic compounds in the aquatic environment. Environ. Sci. Technol. 50: 3−17.

34: Rastall AC, Getting D, Goddard J, Roberts DR, Erdinger L, 2006. A biomimetic approach to the detection and identification of estrogen receptor agonists in surface waters using semipermeable membrane devices (SPMDs) and bioassay directed chemical analysis. Environ. Sci. Pollut. Res. 13: 256–267.

35: Emelogu ES, Pollard P, Robinson CD, Smedes F, Webster L, Oliver IW, McKenzie C, Seiler T, Hollert H, Moffat CF, 2013. Investigating the significance of dissolved organic contaminants in aquatic environments: Coupling passive sampling with in vitro bioassays. Chemosphere 90: 210-219.

36: Li H, Helm, PA, Paterson G, and Metcalfe CD, 2011. The effects of dissolved organic matter and pH on sampling rates for polar organic chemical integrative samplers (POCIS). Chemospher 83: 271-280.

37: Smith KEC, Oostingh GJ, and Mayer P, 2010. Passive dosing for producing defined and constant exposure of hydrophobic organic compounds during in vitro toxicity tests. Chem. Res. Toxicol. 23: 55-65.

38: Allan IJ, Harman C, Ranneklev SB, Thomas KV, Grung M, 2013. Passive sampling for target and nontarget analyses of moderately polar and nonpolar substances in water. Environ. Toxicol. Chem. 32: 1718-1726.

39: Booij K and Smedes F, 2010. An improved method for estimating in situ sampling rates of nonpolar passive samplers. Environ. Sci. Technol. 44: 6789-6794.

40: Alvarez DA, Petty JD, Huckins JN, Jones-Lepp TL, Getting DT, Goddard JP, Manahan SE, 2004. Development of a passive, in situ, integrative sampler for hydrophilic organic contaminants in aquatic environments. Environ. Toxicol. Chem. 23: 1640–1648.