Effect-based and chemical analytical methods to monitor estrogens under the European Water Framework Directive

Sarah Könemann‡ [a,b,*], Robert Kase‡ [b], Eszter Simon [b], Kees Swart [c], Sebastian Buchinger [d], Michael Schlüsener [d], Henner Hollert [a], Beate I. Escher [e,f], Inge Werner [b], Selim Aït-Aïssa [g], Etienne Vermeirssen [b], Valeria Dulio [g], Sara Valsecchi [h], Stefano Polesello [h], Peter Behnisch [c], Barbora Javurkova [i], Olivier Perceval [k], Carolina Di Paolo [a], Daniel Olbrich [b], Eliska Sychrova [i], Rita Schlichting[e], Lomig Leborgne [l], Manfred Clara [m], Christoph Scheffknecht [n], Yves Marneffe [o], Carole Chalon [o], Petr Tušil [p], Přemysl Soldàn [p], Brigitte von Danwitz [q], Julia Schwaiger [r], Maria Isabel San Martín Becares [s], Francesca Bersani [t], Klara Hilscherová [i], Georg Reifferscheid [d], Thomas Ternes [d], Mario Carere [u]

[a] *Institute for Environmental Research, RWTH Aachen University, Worringerweg 1, 52074 Aachen, DE;*

[b] *Swiss Centre for Applied Ecotoxicology Eawag-EPFL, Überlandstrasse 131, 8600 Dübendorf, CH;*

[c] *BioDetection Systems b.v., Science Park 406, 1098 XH Amsterdam, NL;*

[d] *Bundesanstalt für Gewässerkunde, Am Mainzer Tor 1, 56068 Koblenz, DE;*

[e] *Helmholtz Centre for Environmental Research - UFZ, Permoserstrasse 15, 04318 Leipzig, DE;*

[f] *Eberhard Karls University Tübingen, Environmental Toxicology, Center for Applied Geosciences, 72074 Tübingen, DE;*

[g] *INERIS, Rue Jaques Taffanel, Parc Technologique ALATA, 60550 Verneuil-en-Halatte, FR;*

[h] *Istituto di Ricerca sulle Acque, Via del Mulino 19, 20861 Brugherio (MB), IT;*

[i] *Masaryk University, Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, 625 00 Brno, CZ;*

[k] *French National Agency for Water and Aquatic Environments, 5 Square Felix Nadar, 94300 Vincennes, FR;*

[l] *Agence de l'eau Adour-Garonne, 90 rue de Fétéra, CS 87801, 31078 Toulouse Cedex 4, FR;*

[m] *Umweltbundesamt, Spittelauer Lände 5, 1090 Wien, AT;*

[n] *Umweltinstitut, Institut für Umwelt und Lebensmittelsicherheit des Landes Vorarlberg, Montfortstraße 4,6901 Bregenz, AT;*

[o] *Institut Scientifique de Service Public (ISSeP), Rue Chéra 200, 4000 Liège, BE;*

[p] *T.G. Masaryk Water Research Institute, Podbabská 2582/30, Praha 6, 16000, CZ;*

[q] *Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV), Auf dem Draap 25, 40221 Düsseldorf, DE;*

[r] *Bayerisches Landesamt für Umwelt, Demollstrasse 32, 82407 Wielenbach, DE;*

[s] *Instituto de Recursos Naturales, Universidad de León, Avenida de Portugal 42, 24071 León, ES;*

[t] *Centro Ricerche (SMAT), Società Metropolitana Acque Torino S.p.A.C. so Unità d'Italia 235/3, 10127 Torino, IT;*

[u,*] *National Institute of Health, Department Environment and Health, Roma, IT;*

‡Authors contributed equally to this manuscript.

**Corresponding author:*
Sarah Könemann
Institute for Environmental Research, RWTH Aachen University
Worringerweg 1, 52074 Aachen, Germany
Tel.:+41587655254
E-mail: sarah.koenemann@rwth-aachen.de

Detailed description of the authors contributions can be found in the Supplementary Information (SI, Table S1).

Abstract

The European Decision EU 2015/495 included three steroidal estrogens, estrone, 17β-estradiol and 17α-ethinyl estradiol, in the "watch-list" of the Water Framework Directive (WFD). As consequence, these substances have to be chemically monitored at the level of their environmental quality standards, which can be challenging. This project aimed to identify reliable effect-based methods (EBMs) for screening of endocrine disrupting compounds, to harmonise monitoring and data interpretation methods, and to contribute to the current WFD review process. Water and wastewater samples were collected across Europe and analysed using chemical analyses and EBMs. The results showed that 17β-estradiol equivalents were comparable among methods, while results can vary between methods based on the relative potencies for individual substances. Further, derived 17β-estradiol equivalents were highly correlated with LC-MS/MS analyses. This study shows that the inclusion of effect-based screening methods into monitoring programmes for estrogens in surface waterbodies would be a valuable complement to chemical analysis.

Keywords

Science-policy interface	Estrogen screening	Endocrine disruption
Surface and waste water assessment	Emerging pollutants	EU watch-list
Steroid analyses	*In vitro* bioassays	Integrated effects of mixtures

1 State of the Art

Over the past two decades, numerous scientific studies have demonstrated that endocrine disrupting chemicals (EDCs) elicit adverse effects on sensitive aquatic species, such as fish [1-7]. Steroidal estrogens, like the natural hormones estrone (E1) and 17β-estradiol (E2), as well as the synthetic hormone 17α-ethinyl estradiol (EE2), are of particular environmental concern [8-11]. Due to their steady release via waste water effluents into surface waters [12, 13] and their high biological activity, even very low concentrations of E2 and EE2 have been shown to cause reproductive toxicity with negative effects at the population level [14-16]. As a consequence, E1, E2, and EE2 were included in a European Union (EU) Water Framework Directive (WFD) "watch-list" [17-20]. The WFD watch-list mechanism aims to collect high-quality monitoring data on concentrations of emerging pollutants and potentially hazardous substances, whose currently available monitoring information shows either quantitative or qualitative deficiencies [21]. To collect more high-quality data, listed substances have to be monitored at representative EU sampling sites for a period of at least 12 and up to 48 months. The watch-list mechanism is expected to support future substance prioritisation processes, enable the implementation of measures, and facilitate environmental risk assessment across the EU.

Chemical monitoring of estrogens for the watch-list mechanism is challenging, because the European Commission set maximum acceptable method detection limits (MDLs) at EQS levels of 400 pg/L for E1 and E2, and 35 pg/L for EE2 [18, 22]. Most routine analytical methods used by the Member States cannot meet these requirements, especially for EE2, based on [23, 24]. Hence, the quality assessment of water bodies based on current methods is a challenge for the detection/quantification limits that are too high to detect if EQS are being exceeded or not. Effect-based methods are able to detect estrogenic substances at sub-ng or even pg levels and have the potential to be used as a complementary screening tool [12, 25-27]. In addition, they do not require *a priori* knowledge of the substances to be monitored, as they are able to determine the biological response caused by complex mixtures of unknown compounds. Thus, effect-based methods may be suitable to serve as a valuable link between chemical analytical and ecological quality assessments, since the effects can rarely be linked to individual compounds.

As described in an EU technical report, which was elaborated in the context of the Chemical Monitoring and Emerging Pollutants (CMEP) expert group under the Common Implementation Strategy (CIS) of the WFD, effect-based tools can be categorised into three main groups: Bioassays (*in vitro*, *in vivo*), biomarkers, and ecological methods [28]. With regard to steroidal estrogens and other EDCs, *in vitro* reporter gene assays have been used predominantly to determine the total estrogen receptor (ER) mediated estrogenicity of an environmental sample [29]. Among the most commonly applied assays are *in vitro* methods such as estrogen receptor transactivation assays (ER-TAs), which use various cell types including yeast, human and other mammalian cell lines that were transfected with a human estrogen receptor coupled to a reporter gene [30]. Activation of the ER leads to the expression of the reporter

gene product, usually an enzyme that modifies another chemical, causing a quantifiable response. The resulting estrogenic potential of a sample is expressed as an E2 equivalent concentration (EEQ), indicating the estrogenic activity of the sample or sample dilution in terms of equivalency to the estrogenic activity of the corresponding E2 reference concentration [31].

Although ER-TAs are highly advantageous methods for the detection of ER activation and quantification of very low estrogen concentrations in surface waters [23], these methods are not included within current WFD monitoring programmes [20]. One reason for this is the lack of data that demonstrate their applicability as a monitoring and screening tool in combination with chemical analytical methods (see e.g. [14]). Such information would greatly increase their regulatory acceptance. As a response to this need, an EU-wide project involving 24 research organisations and environmental agencies from 12 countries was carried out to evaluate the usefulness of specific *in vitro* methods for identifying the presence of the watch-list substances, E1, E2, and EE2, in surface and waste waters. The project aimed to compare the chemical and effect-based data resulting from the analysis of 16 surface and 17 waste water treatment plant effluent samples. Analyses were conducted in seven participating laboratories using different LC/MS- (three laboratories) and effect-based methods (five laboratories). The objectives of the study were (i) the demonstration of reliable effect-based screening methods for the monitoring of estrogenic EDCs in waste water and surface water, (ii) the harmonisation of data interpretation methods, and (iii) providing recommendations for the implementation of cost-effective and reliable effect-based methods in WFD monitoring programmes.

2 The Project

2.1 Sampling

A total number of 16 surface water (SW) and 17 waste water (WW) samples were collected according to a protocol developed by the participants (SI, Part A). Selected sampling sites were located in seven European countries in Central and Southern Europe (Figure 1): Austria (1 SW/ 3 WW), Belgium (2/2), Czech Republic (2/2), France (1/1), Germany (4/4), Italy (5/3), and Spain (1/2). Sample collection was carried out from September to November 2015 by ten participating institutions. The samples were taken based on prior knowledge on their contamination with estrogens and represented a gradient of contamination from high to moderate.

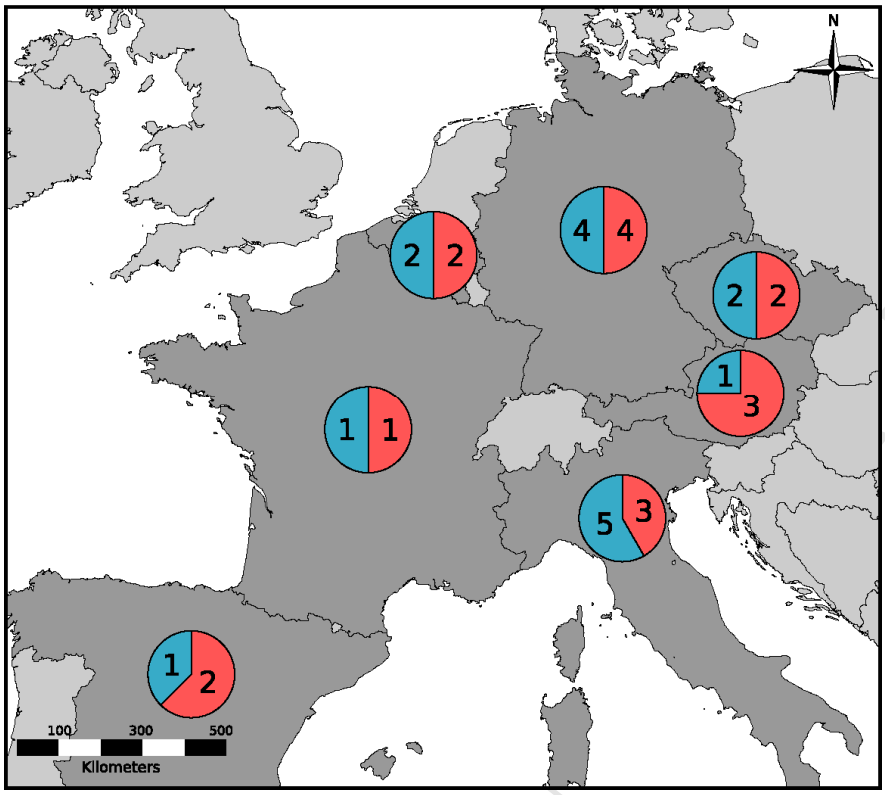


Figure 1: Samples taken in various European States (dark grey). The circles indicate the number of surface water (blue) and waste water samples (red) taken in each country.

2.2 Sample preparation

The sample preparation included the filtering of a part of the SW (see SI, Part A) and all WW samples over glass fibre filters (Millipore, type 4, retention 2.7 µm, circle size 4.7 cm). Since a filtration step can have an impact on the composition of a sample and its estrogenic activity [32], the filtration step was investigated during a feasibility study prior to the main study presented here. The results of the pre study did neither show a significant reduction in estrogenicity in the control nor in tested environmental samples (data not shown). Subsequently, all samples were enriched by means of solid-phase extraction (SPE; 11 L sample to 11 mL extract) and extracts were passed over silica gel (SiOH) columns (methods focusing on E1, E2 and EE2). While for surface water each extract was split into eleven 1 mL aliquots that were each passed over a single SiOH column, for waste water a single column was inadvertently used to treat the whole extract (11 mL). For LC-MS/MS analysis this means that matrix was less efficiently removed from WW extracts (relative to SW extracts) and higher matrix loads would have impeded low LOQs in WW LC-MS/MS analysis. For bioassay analysis this means that, should additional ER-agonists (i.e. other than E1, E2 and EE2) have been present in the extracts, a reduced clean-up efficiency would have reduced ER-agonist removal which in turn would have caused enhanced effects in bioassays. Full details of sample preparation are provided in SI, Part A.

2.3 Chemical and effect-based analyses

Participating laboratories received spiked reference samples, blanks and encoded water extracts. The chemical analyses were conducted in three different labs, which applied an LC-MS/MS with negative electron spray ionisation (detailed information in SI, Part D Table S2). The effect-based methods were conducted in five different labs: Estrogen Receptor Chemical Activated LUciferase gene eXpression (ER-CALUX) at Biodetection Systems (BDS), luciferase-transfected human breast cancer cell line (MELN) gene-reporter assay at INERIS [33], ER-GeneBLAzer assay at the Helmholtz Centre for Environmental Research (UFZ) [34], the stably transfected human estrogen receptor-alpha transcriptional activation Assay using hERα-HeLa-9903 cells (HeLa-9903 assay) at RECETOX [35], and planar Yeast Estrogen Screen (pYES) at the German Federal Institute of Hydrology (BfG) [36, 37]. The pYES is a method, which combines a chromatographic separation of the sample by thin layer chromatography (TLC) with a subsequent performance of the YES on the planar surface of the TLC-plate [38-40]. Like the common assays which are performed in micro-well-plates, this approach allows the quantification of the overall estrogenic activity present in the sample by means of E2-equivalence concentrations. Furthermore, like methods based on LC/MS, it also allows the estimation of concentrations of individual estrogenic compounds, e.g. E1, E2 and EE2, due to the chromatographic separation of the sample. For this purpose the respective standard compounds are used for a calibration on the same TLC plate – in the present study E1, E2, EE2, and estriol (E3) were applied in a mixture at three different levels. Due to the limited separation power of the thin layer chromatography compared to HPLC and GC in particular, a co-migration of estrogenic compounds cannot be excluded. Therefore, under the assumption of effect addition, the estimated individual concentrations represent the possible maximal concentration of the respective compound. This approach can be used to identify and quantify substance groups causing ER-activation.

2.4 Blanks and positive controls

Ultrapure water (11 L) was used as extraction blank. An extraction blank was included with each extraction run of 10 samples, subjected to clean-up and distributed the same as the sample extracts. Further, each analysis using effect-based methods included a negative control. To avoid solvent effects on cell viability, its concentrations did not exceed a defined value (see SI, Part D Table S3). As positive controls for ensuring the validity and enabling a comparison of the methods, surface water samples (11 L each) from the Netherlands were spiked with E2 and EE2 at two concentrations by the central lab (BDS). The “low spike” (600 pg/L) represented a concentration slightly above the proposed EQS for E2 (400 pg/L). The “high spike” (6000 pg/L) represented a concentration that is quantifiable with high certainty by both effect-based and chemical methods.

2.5 Data evaluation – effect-based methods

Raw data and information on relative enrichment factors (REF) of the extracts were collected from participating laboratories. The REF expresses the combination of: 1) sample enrichment using SPE and 2) extract dilution steps in each of the applied effect-based methods. Estrogenic activity of the extracts was expressed as E2-equivalence concentration (pg EEQ/L water) (described in detail in SI, Part B). Briefly, dose-response curves of the reference compound, E2, and the dilution series of the water extracts and blanks were fitted using a five-parametric non-linear regression with normalised data. The concentration of the positive control (E2) needed to induce 10 % effect of the maximum E2-induction (PC_{10}), was calculated. Subsequently, the relative REF of the sample, that stimulates the assay at PC_{10} level was determined by interpolation. The PC_{10} reference concentration was divided by the corresponding sample dilution (REF) to obtain the EEQ of the sample. EEQs derived by the PC_{10} method are presented in the results section.

2.6 Data evaluation – chemical analysis

Internal standard calibration and interpolation using a linear regression model were performed to determine concentrations (pg/L) of the individual steroidal estrogens in sample extracts. Identification of selected analytes was performed based on two to three Multiple Reaction Monitoring (MRM) transitions between the precursor ion and two or three most abundant product ions, depending on the laboratory where analyses were done. The first transition was used for quantification purposes whereas the second and third transitions were used to confirm the presence of the target compound in the sample. Quantified analytes were identified by comparing the retention time (RT) of the corresponding standard and the ratio between two ion transitions recorded (±20 %) in the standard and water samples.

2.7 Calculation of sample-dependent LOD and LOQ

The Limits of quantification (LOQ) for effect-based methods the LOQs were calculated as 3-fold the standard deviation (SD) of the averaged response of the negative control on each assay plate. The effect level of 3-fold the SD was interpolated from the E2 reference curve and divided by the REF of the sample to derive the LOQ. The actual reporting for effect-based methods occurred at the 10% effect level which was always above LOQ (typically at 2-5 % effect levels).

In case of the chemical analysis the limits of detection (LOD) were determined for each compound in each sample based on the signal intensity of the internal standards or the analyte peak by a signal-to-noise (S/N) ratio of 3:1 and LOQ by a S/N ratio of 10:1.

When comparing LOQs of effect-based methods with those of chemical analyses the various key differences between the two approaches need to be taken into account (for further background see SI, Part C).

2.8 Comparison of chemical and biological analysis

The EEQ_{bio} is the ratio of the effect concentration of the reference compound estradiol EC_{50}(E2) (pg/L) and the sample EC_{50}(sample) (Equation 1) and was derived in this study using the PC_{10} approach (see above). The EEQ_{chem} was calculated from the sum of the relative effect potencies REP_i times the detected concentration of estrogenic chemical i, c_i [41]. The REP, in turn, is the ratio of the effect concentration of the reference compound estradiol EC_{50}(E2) and the chemical i's EC_{50}(i) (Equation 2).

$$EEQ_{bio} = \frac{EC_{50}(E2)}{EC_{50}(sample)} \quad (1)$$

$$EEQ_{chem} = \sum_{i=1}^{n} REP_i \cdot c_i = \sum_{i=1}^{n} \frac{EC50(E2)}{EC50(i)} \cdot c_i \quad (2)$$

Due to the analytical method detection limits of E2 and EE2, we evaluated the potential contribution of non-detected estrogens to the overall $EEQ_{chem,LOD/2}$ using Equation 3, where values below the LOD ("non-detects") were included as LOD/2. If the analytical lab reported data as <LOQ, we used LOQ/2 in Equation 3 instead of LOD/2. In Equation 3, n refers to the total number of chemicals included in the analysis, m refers to the number of chemicals below LOD. Ci is the average value of three analytical measurements,

$$EEQ_{chem,\,LOD/2} = \sum_{i=1}^{n-m} REP_i \cdot c_i + \sum_{j=1}^{m} REP_j \cdot LOD_j/2 \quad (3)$$

2.9 Correlation analysis

The correlation analysis among effect-based methods ($EEQ_{bio)}$ was performed with GraphPad Prism, using the Pearson correlation (r). [42].

3 Results and discussion

3.1 Reference chemicals and validation

All essential criteria for method performance were fulfilled in this study (described in more detail in the SI, Part E). As shown in Table S4 (SI, Part E), the chemical analytical as well as effect-based methods showed good recovery in the spiked samples. No estrogenic activity or quantifiable concentrations of E1, E2, and EE2 were measured in the blank samples (i.e. procedure-, extraction- and solvent blanks). As the derived effect concentrations in the effect-based methods and chemically measured EE2

concentrations matched with the nominal concentrations of the spiked samples, the observed effects can be ascribed to the samples themselves.

3.2 Results of chemical analysis

Measured concentrations of the three estrogens E1, E2 and EE2 differed widely between sampling sites as well as between surface and waste water samples. Differences among SW samples can be explained by varying river characteristics, e.g. flow (dilution factor), or temperature, as well as differences in estrogenicity of treated WW, that are released into the SW. The results of the analyses, which are summarised in Figure 2, show a 3.2 to 3.6 times higher mean concentration for E1 and E2 in WW (Figure 2B) compared to SW (Figure 2A). Due to the highly contaminated WW sample M(23), possibly influenced by an industrial discharge of EE2, the mean concentration of EE2 across all WW samples was approximately 20 times higher compared to SW (Figure 2). Estrone (E1) was quantified in all samples. For E1 maximum concentrations of 5.6 ng/L (sample P(7)) and 20.5 ng/L (sample Q(20)) in SW and WW were measured, respectively. E2 was the second most frequently quantified estrogen and measured above LOQ in nine of 16 SW and six of 17 WW samples. Measured concentrations ranged from 0.4 ng/L (sample N(33)) to 1.1 ng/L (sample Q(20)) in WW, and from 0.06 ng/L (sample J(10)) to 0.5 ng/L (sample N(15)) in SW. The synthetic EE2 was least frequently quantified and measured above LOQ in four of 16 SW and four of 17 WW samples with a maximum concentration of 0.3 ng/L in SW sample O(3) and 7.5 ng/L in WW sample M(23). These concentration ranges and patterns are in accordance with recent review studies [43, 44].

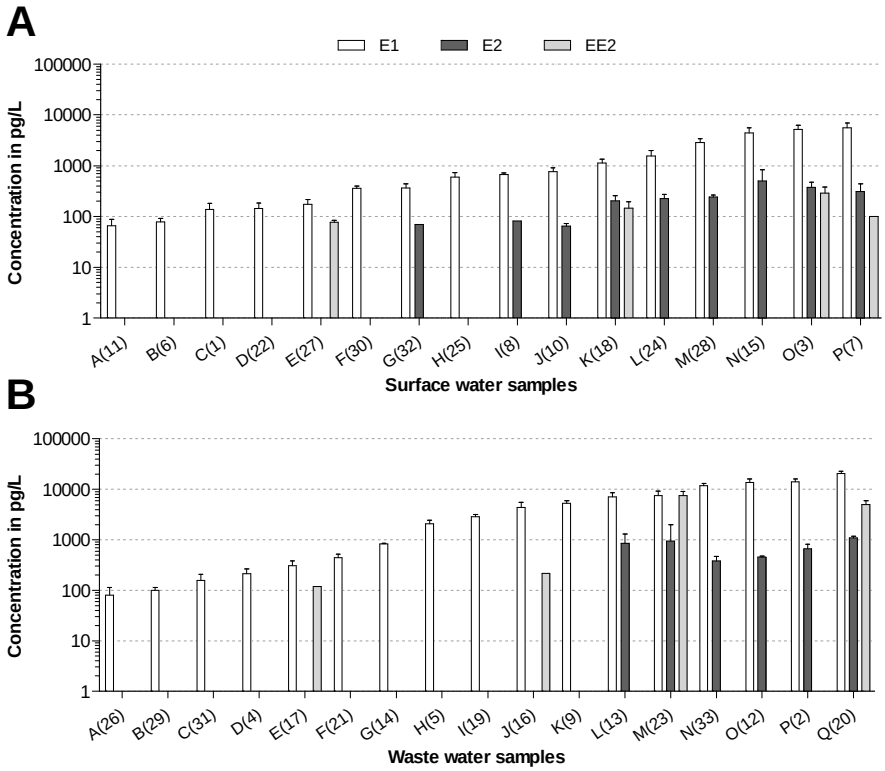


Figure 2: Chemical analytically measured concentrations for SW (A) and WW extracts (B) above LOQ for E1, E2 and EE2. The bars show the mean concentration of all three applied methods for each analyte showing results > LOQ, the standard deviation is shown when two or three methods reported results. The sample-dependent LOQs are listed in the supplementary information together with the measurement data of analytical methods (SI, Part F, Table S6 and S7).

Our results underline the analytical difficulties that have recently been highlighted for E2 and EE2 by several studies and workshops [16, 45], stressing the challenges that emerge for routine methods used in national monitoring programmes. Despite the use of quite advanced chemical analytical techniques (status 2015), the detection and quantification of E2 and EE2 in SW and WW samples was problematic in some cases. While it was possible to quantify E1 in almost all samples, the percentage of quantifications was significantly reduced for E2 and even more for EE2 (Figure 3). This was partially due to the fact that insufficient silica gel was used to reduce the matrix effects in WW. WW is considered as worst-case regarding matrix effects [46, 47].

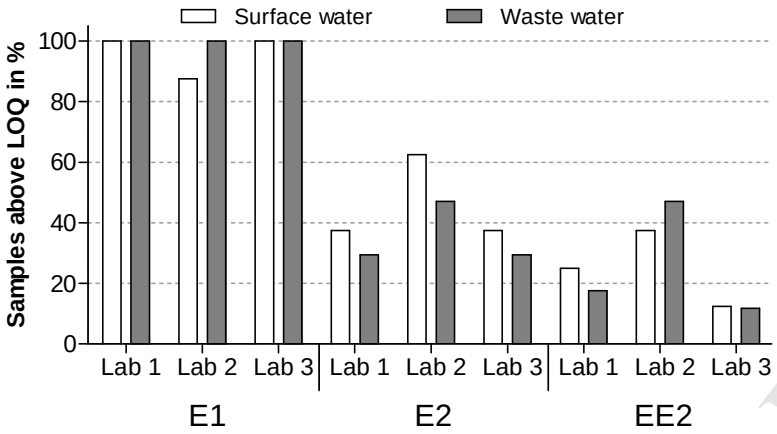


Figure 3: Mean percentage of quantified (>LOQ) samples for each substance in SW and WW. The sample-dependent LOQs are listed in the supplementary information together with the measurement data of the analytical methods (SI Part F, Table S7).

However, the quantification of substances itself is not the only challenge faced by those routinely applying analytical methods for watch-list monitoring. According to the EU Commission Decision 2015/495, which established the first watch-list, the indicative methods applied by Member States have to meet the minimum requirement for method detection limits (MDL) equal to the proposed EQSs of E1 at 3.6 ng/L, E2 at 0.4 ng/L and EE2 at 0.035 ng/L [18]. To take into consideration the matrix effects of different waters, LODs and LOQs had to be calculated for each sample (SI Part F, Table S7). The three techniques used in the current study were able to meet MDL requirements for E1 in all SW and WW samples. Also for E2, in 96 % of surface water samples and 94 % of waste water samples detection was possible at the level of the proposed EQS. In the case of EE2, the minimum criteria were not met, since only 56 % and 16 % of SW and WW samples, respectively, could be monitored at the EQS level. These findings are in accordance with a recent report from 2015, which showed that the lowest LOQ found in literature at that time was sufficient for compliance monitoring of E1 and E2 in inland surface waters, while the criteria were not met for EE2 by several Member States [24]. It has to be pointed out that, in this project, the silica clean-up step for the sample extracts differed between WW and SW samples (see methods section) favouring the presence of polar compounds in extracts of WW samples. This difference likely reduced the sensitivity of the analytical method for the target compounds in WW samples. Furthermore, sample extraction was performed at pH 3 possibly increasing concentrations of humic acids and thus lowering sensitivity of LC/MS-based methods applied. Under ideal conditions, we estimate that analytical methods can achieve LODs and LOQs of a factor 2 to 3 lower in WW samples. It has to be recognised that the LODs of chemical analytical methods used exclusively for steroidal estrogens already significantly decreased from 2013 (LOD E2 and EE2 of 100 pg/L) to 2015 (E2: 60 pg/L, EE2: 85 pg/L) and will certainly decrease further [16, 23].

Nevertheless, if steroidal estrogens were to be included in the EU priority list for monitoring, very strict minimum performance criteria would apply. As stated in the Commission Directive 2009/90/EC, an analytical method used for monitoring of priority substances needs a LOQ equal or below a value of

30 % of the EQS [48]. These requirements can presently be met only for E1, but not for E2 or EE2 in all SW. Regarding the quantification of E2, and EE2, existent routine analytical techniques still lag behind the requirements. This result is supported by two recent reviews on the performance of current analytical methods that have shown that 35 % of reviewed methods complied with the EQS for E2, while only one method complied with the EQS for EE2 [49, 50]. In order to not only detect but also quantify at such low concentrations as required for regulatory monitoring application, a further decrease of LOQs is necessary, which is difficult to achieve for routinely used non-tailored analytical methods in the short-term.

3.3 Quantification limits of chemical-analytical and *in vitro* effect-based methods

The LOQs for all methods applied in this study are summarised in Figure 4. Since E2 is used as the reference compound for all effect-based methods, the LOQ of E2 is shown for the chemical-analytical methods as an example. When comparing LOQs across the different methods it has to be taken into account that LOQs were derived along different approaches (see method section and SI, Part C for further details). The effect-based *in vitro* methods were generally able to quantify effects at one to two orders of magnitude lower concentrations than the analytical methods used. For effect-based methods, LOQs ranged between 0.002 ng/L and 0.2 ng/L for SW as well as WW, while for chemical-analytical methods LOQs for E2 were 0.04 ng/L to 1.5 ng/L in SW and 0.05 ng/L to 3 ng/L in WW. This increase in LOQs for chemical-analytical methods in WW samples (Figure 4B) compared to surface water (Figure 4A) can be ascribed to the higher complexity of the waste water matrix [46, 47] as well as the less efficient clean-up used for WW samples.

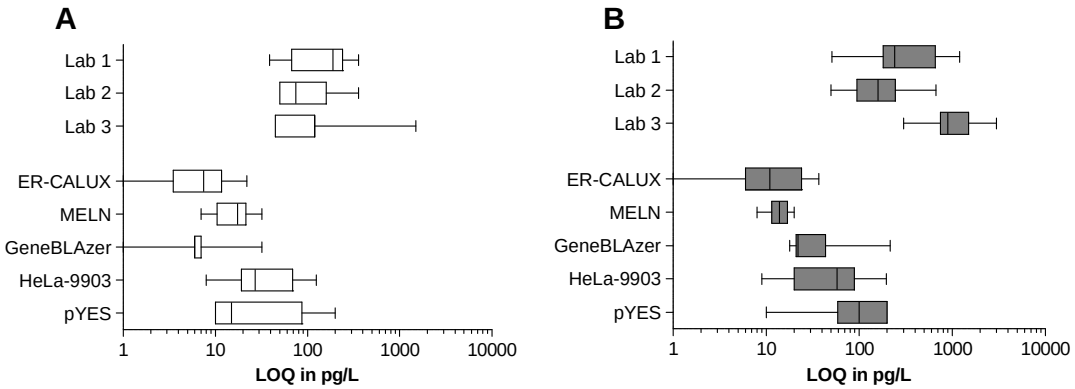


Figure 4: Sample-dependent LOQs in surface water (A) and waste water (B) extracts. For the chemical analytical method the LOQ of E2 is shown as an example and for the effect-based methods the LOQ of the integrated effects is represented. Plots indicate the distribution of data, thereby the bottom and the top of the box are the first and third quartiles, while the line inside the box is the median. The whiskers show the minimum and maximum of all data.

3.4 Measured estrogenic effects

As a result of these low effect-based quantification limits, estrogenic activities were detected in all tested samples. As expected, highest EEQs were measured in WW samples (Figure 5A and B) . In SW, EEQ_{bio}

ranged from 0.16 ng/L measured with HeLa-9903 in sample B(6) to up to 5.4 ng/L measured with pYES in sample O(3). In WW, the lowest EEQ_{bio} of 0.03 ng/L was measured in sample A(26) with ER-GeneBLAzer, while the highest EEQ_{bio} of 24 ng/L was measured in sample M(23) with HeLa-9903. Further, it is evident that EEQ_{bio} for SW samples determined with the MELN, as well as the pYES, were higher (> 50 %) than the EEQ_{bio} measured with the other effect-based methods. A possible reason for this pattern, which was less pronounced in WW, could be a higher sensitivity of the MELN and pYES towards E1 (see SI Part F, Table S8), combined with a larger proportion of E1 in surface water. Additionally, alterations in the method's performance occur due to differences between the test systems, which was already mentioned in previous studies [23, 44, 51] and is further discussed for this project in an associated publication [52].

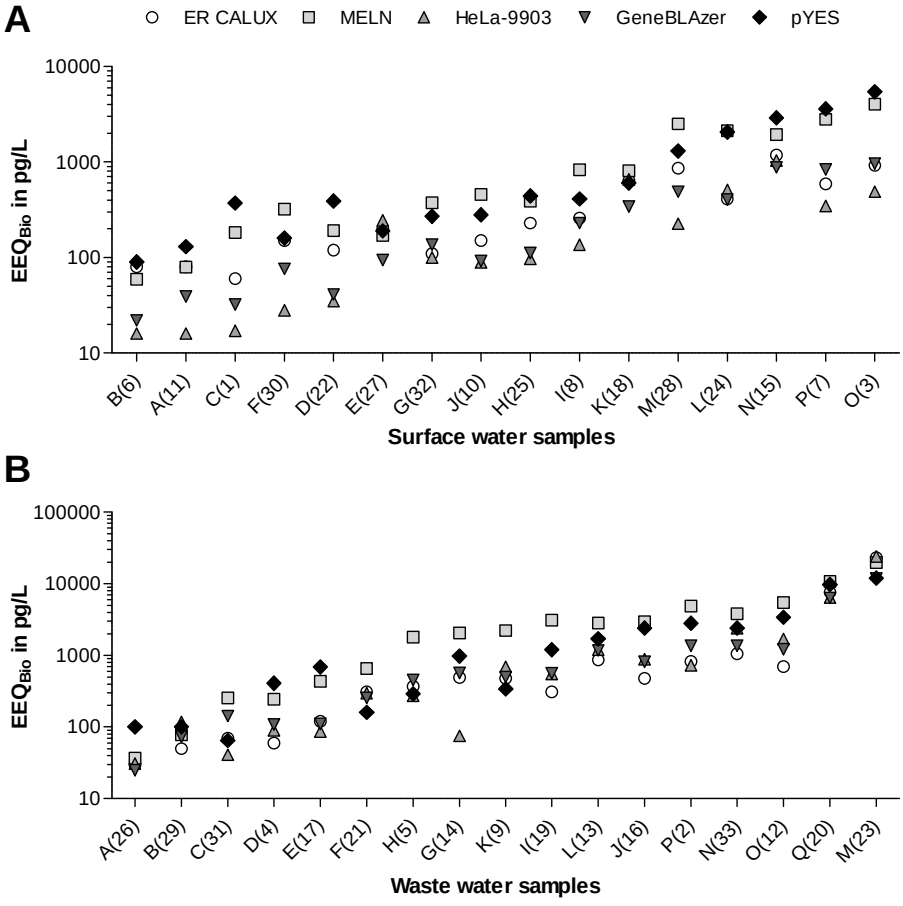


Figure 5: Measured E2-equivalents for all SW (A) and WW (B) extracts. The symbols show the EEQs for each bioassay, which were calculated according to the method described in section 2.5. The sample-dependent LOQs are mentioned in the supplementary information, together with the measurement data of effect-based methods (SI Part F, Table S8 and S9).

3.5 Comparison of chemical analysis and *in vitro* effect-based methods

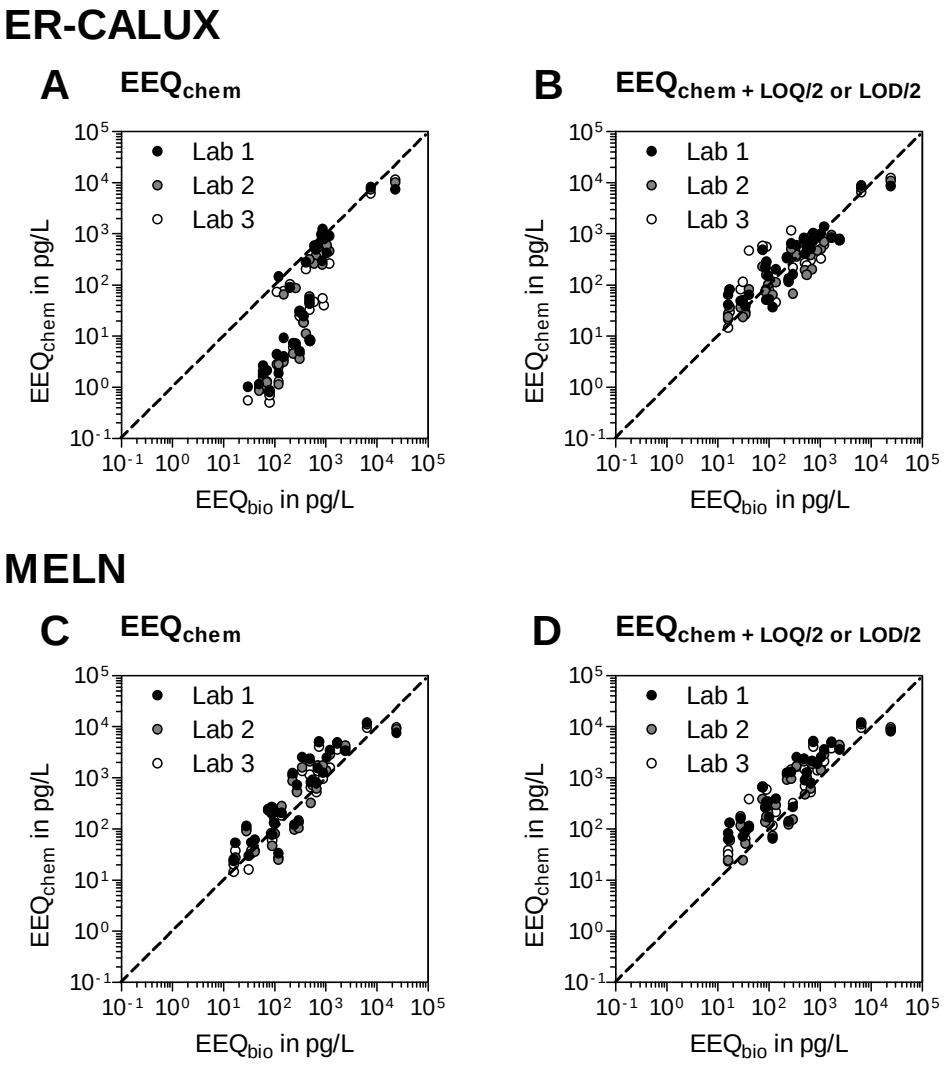
We cannot a priori expect consistency between EEQ_{chem} calculated from E1, E2, and EE2 concentrations and EEQ_{bio}. Although the extraction and clean-up method focused on E1, E2, and EE2, other natural estrogens and xenoestrogens (both agonists and antagonists) might still be present in the extracts and

contribute to the mixture effects detected by effect-based methods. Thus, there can be situations where EEQ_{chem} is lower than EEQ_{bio} because: 1) agonists other than E1, E2, and EE2 were present in the sample but not quantified by LC-MS/MS analyses or 2) some target compounds were present but below LOQ or LOD, thus they were not included in EEQ_{chem} but still contributed to EEQ_{bio}. Alternatively, EEQ_{chem} can be higher than EEQ_{bio} when antagonists supress the response of the assay.

For ER-CALUX, the comparison of EEQ_{bio} with EEQ_{chem} (Figure 6A) indicated an underestimation of EEQ_{bio} by EEQ_{chem} at low concentrations of steroidal estrogens. When E1 concentrations are low, typically E2 and EE2 concentrations are below LOQ (Figure 2). However, as stated above, also below their LOD/LOQ, these chemicals may be present and contribute to the biological mixture effect (i.e. EEQ_{bio}). We therefore also calculated the $EEQ_{chem,LOD/2}$ that uses the LOD/2 or LOQ/2 for those E2 and EE2 concentrations below the LOD or LOQ. The increase in EEQ_{chem}, due to the inclusion of LOQ/2 and LOD/2 data (SI, Part F, Table S10-14), shifts the EEQ_{chem} - EEQ_{bio} data cluster towards the one-to-one line (Figure 6B). In fact, there is now a slight overestimation of the biological effect in the range where EEQ concentrations are low (up to ca.100 pg/L). The fact that the agreement between EEQ_{chem} and EEQ_{bio} has become much better (going from Figure 6A to 6B) is a good indication that E2 and EE2 are indeed present and were captured by effect-based methods.

The situation for MELN is markedly different from that of ER-CALUX. For MELN the direct comparison between EEQ_{chem} and EEQ_{bio} is already very good (Figure 6C). In fact, EEQ_{chem} tends to be above EEQ_{bio} already before adding the additional EEQ_{chem} component using LOD/2 or LOQ/2 for E2 and EE2. The inclusion of LOD/2 or LOQ/2 in the EEQ_{chem} calculation caused a notable overestimation of EEQ_{chem} for almost all samples (>90 % of data above the 1 to 1 line in Figure 6C). The other three bioassays show results that are intermediate between ER-CALUX and MELN, with a general trend towards a slight underestimation of EEQ_{chem} for samples with low EEQ_{bio} and an overestimation after adding LOD/2 or LOQ/2 (see Figure S1).

The marked differences between ER-CALUX and MELN are not unexpected. MELN has the highest relative E1 effect potency of all tested bioassays (0.29 compared to 0.01 for ER-CALUX; Table S5). Thus, EEQ_{chem} results for MELN are strongly based on E1 concentrations – a compound that was always measured (except for a few samples by Lab 2, Figure 3). Consequently, for MELN the relative contribution of E2 and EE2 at LOD/2 or LOQ/2 on top of measured E1 concentrations is relatively small though still noticeable for samples with low EEQ concentrations (compare Figure 6C to 6D).



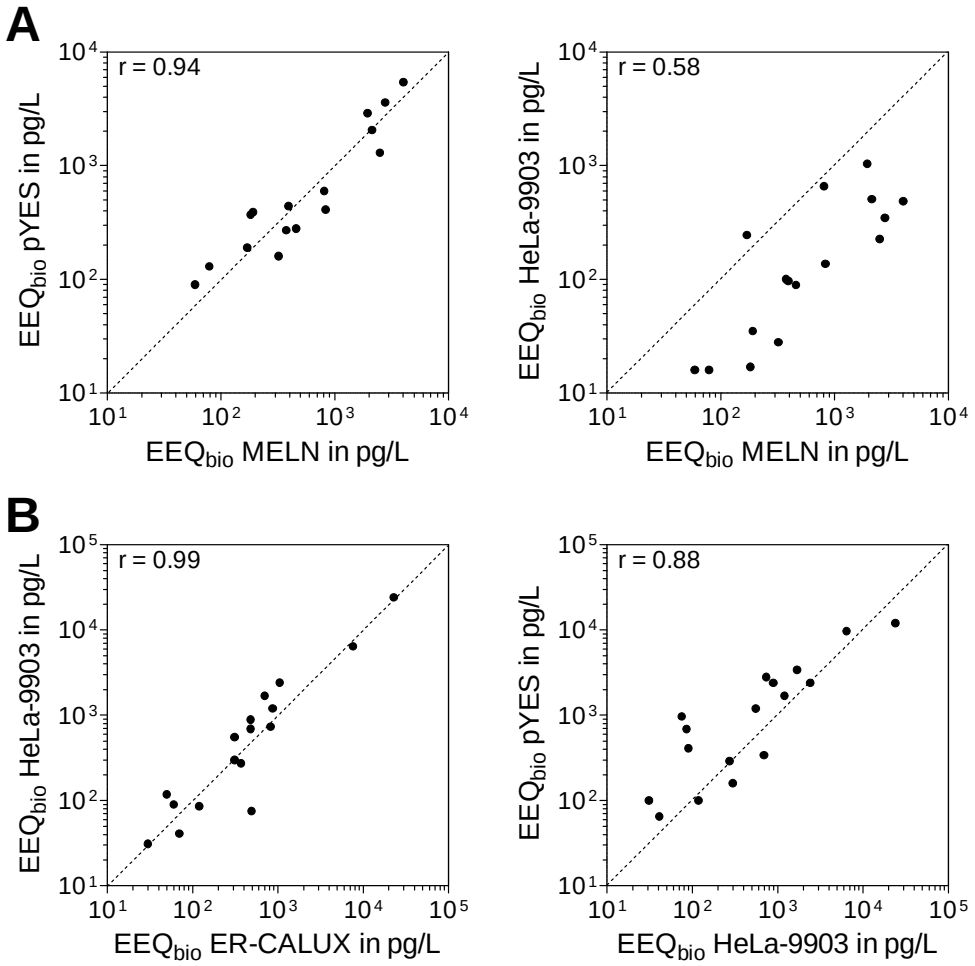
381

382 **Figure 6: Comparison of EEQ_{chem} with EEQ_{bio}.** Exemplary graphs are shown for the ER-CALUX (A, B) and MELN assay
383 (C, D) (further figures in the SI, Part G). Graphs on the left show the EEQ_{chem} derived from values >LOQ, while the graphs on
384 the right show the $EEQ_{chem + LOD/2 or LOQ/2}$ calculated by including LOD/2 or LOQ/2. The dashed line indicates perfect agreement
385 of EEQ_{chem} with EEQ_{bio}.

386 ### 3.6 Comparison of effect-based methods

387 To compare the five effect-based methods amongst each other, a correlation analysis was conducted by
388 plotting the EEQs of one method against the EEQs of all other methods for SW samples and WW
389 samples, respectively (Figure 7).

390



391

392 **Figure 7: Exemplary graphs of correlation analysis of effect-based methods for SW (A) and WW (B) showing the**
393 **strongest and weakest correlations.** The correlation analysis was based on the method described in section 2.9. The dashed
394 line indicates perfect agreement of the compared effect-based methods. All correlations were significant with a p value <0.0001
395 except for MELN and HeLa-9903 (top right panel) which had a p value ≈ 0.01. Further graphs are shown in SI, Part H, Figures
396 S2 and S3.

397 The results of this analysis are summarised in Table 1 and Table 2 and show a strong correlation and
398 thus good comparability of pYES, MELN and ER-CALUX. For SW samples, the strongest correlations
399 were seen for pYES/MELN (r°= 0.94) and pYES/ER-GeneBLAzer (r°= 0.94), while the weakest
400 correlation was determined for MELN/HeLa-9903 (r°= 0.58). For WW samples, test results correlated
401 strongly among all methods (Table), and the strongest correlation (r°= 0.99) was observed for ER-
402 CALUX/HeLa-9903. It is known that effect-based methods differ in their REPs for individual ER-
403 agonists [53-55] which can explain that results obtained by the HeLa-9903 assay correlated less strongly
404 with other test results . Based on these differences effect-based methods can be split into two groups:
405 pYES and MELN with high E1 REP and ER-CALUX, HeLa-9903 and ER-GeneBLAzer with lower E1
406 REP.

Table 1: Pearson correlation coefficients of all bioassays for SW. The values were calculated according to the method mentioned in section 2.9. All correlations were significant with a p value <0.0001 (***) and a p value ≈ 0.01 (*).

	MELN	ER-GeneBLAzer	HeLa-9903	pYES
ER-CALUX	0.81 ***	0.91 ***	0.86 ***	0.76 ***
MELN		0.93 ***	0.58 *	0.94 ***
ER-GeneBLAzer			0.77 ***	0.94 ***
HeLa-9903				0.61 *

Table 2: Pearson correlation coefficients of all bioassays for WW. The values were calculated according to the method mentioned in section 2.9. All correlations were significant with a p value <0.0001 (***).

	MELN	ER-GeneBLAzer	HeLA-9903	pYES
ER-CALUX	0.94 ***	0.98 ***	0.99 ***	0.89 ***
MELN		0.98 ***	0.94 ***	0.97 ***
ER-GeneBLAzer			0.97 ***	0.96 ***
HeLa-9903				0.88 ***

4 Conclusions and trends

By including E1, E2, and EE2 in the watch-list of the WFD, the European Commission recognised the need to assess environmental occurrence and impact of these endocrine disrupting substances. However, the current WFD monitoring approach, which is based on chemical analytical measurements and compliance with specific EQSs, has been shown to be limited with regard to the ability to detect these substances at required concentrations [18, 51]. As demonstrated in this study, chemical analytical methods (status 2015) were unable to quantify the steroidal estrogens E2 and EE2 at EQS concentrations in all samples although E1 was measured effectively. Using effect-based methods, EEQ concentrations could be determined in all samples. As these EEQ concentrations are the responses to mixtures of known as well as unknown substances, effect-based methods have the potential to be highly valuable tools complementing routine monitoring and water quality assessment for estrogenic compounds. Effect-based methods are of particular regulatory interest as tools to screen and prioritise samples for further analysis by chemical analytical methods. Furthermore, DIN/EN/ISO standards to determine the estrogenic potential of water samples – covering human cell lines (e.g. ER-CALUX) and yeast based assays – will be available in early 2018 under ISO/DIS19040. The availability of such standards will facilitate the integration of effect-based methods into regulatory schemes.

Our study showed that EEQ results obtained from all effect-based methods applied were comparable – especially at higher concentrations found in WW – but results can vary between methods based on the

relative effect potencies for individual substances. This has to be considered for the interpretation of data and determination of threshold values. As stated above: 1) *in vitro* effect-based methods cannot deliver single substance based measurements, but are suitable to assess overall estrogenicity in water samples and 2) results of these methods need to be confirmed by advanced chemical analysis. Along these lines, the inclusion of effect-based methods into monitoring programmes as a screening tool (detailed description in Kase et al., [52]) for estrogenic substances in surface water bodies would be a valuable complement to chemical analysis currently foreseen by the Directive 2013/39/EU and WFD [28, 56, 57].

Acknowledgements

The major part of the project was funded by in-kind contribution from numerous project partners showing high interest in this project. Essential parts of the project, e.g. extraction and reporting were externally co-funded by the Swiss Centre of Applied Ecotoxicology EAWAG-EPFL, NORMAN Network, INERIS and Pharmaceutical Associations (Pfizer, Johnson & Johnson, TEVA which joined the project later and only supported the extraction at central Lab BDS). In total, 70 colleagues associated to 24 institutes and agencies in 12 nations contributed to this project. The authors would like to thank: Joint Research Centre (EC), The French Agency for Biodiversity (previously ONEMA) (FR), INERIS (FR), Bio Detection Systems (NL), Swiss Centre for Applied Ecotoxicology (CH), Federal Institute of Hydrology (DE), Federal Environment Agency (DE), RWTH Aachen (DE), RECETOX (CZ), NORMAN-Network, Helmholtz Centre for Environmental Research-UFZ (DE), IRSA-CNR (IT), Italian Institute of Health (IT), University of Leon (ES), Water Research Institute T.G.Masaryk (CZ), Bavarian State Office for Environment (DE), LANUV (DE), Environment Agency Austria (AT), ISSeP (Scientific Institute of Public Service) Wallonia (BE), SMAT (IT), BrianzAcque (IT), Vettabbia scarl, MilanoDepur (IT), Agence de l'Eau Adour-Garonne (FR), Ontario Ministry of the Environment and Climate Change (CAN), McGill University (CAN), Environmental Institute (SK) and DG Environment of the European Commission (EC). The project partners from UFZ, RWTH Aachen, INERIS and RECETOX were supported by the SOLUTIONS project, that is funded by the European Union Seventh Framework Programme (FP7-ENV-2013-two-stage Collaborative project) under grant agreement number 603437. Additionally, we would like to thank Christin Kuehnert (UFZ), Maria König (UFZ) and Michael Gundlach (RWTH) for experimental assistance. Special thanks for stewardship of this study we would like to address to Helen Clayton and Stéphanie Schaan from European Commission, Environment DG.

The Federal Institute of Hydrology got financial support by the Federal Ministry for the Environment, Nature Conservation, and Building (BMUB).

Conflict of interests

The Federal Institute of Hydrology did not receive any kind of financial support from the Pharmaceutical Associations. Other authors declare no conflict of interest.

Bibliography

1. John P Sumpter and Andrew C Johnson, *10th anniversary perspective: Reflections on endocrine disruption in the aquatic environment: From known knowns to unknown unknowns (and many things in between).* Journal of Environmental Monitoring, 2008. **10**(12): p. 1476-1485.
2. Georg Streck, *Chemical and biological analysis of estrogenic, progestagenic and androgenic steroids in the environment.* TrAC Trends in Analytical Chemistry, 2009. **28**(6): p. 635-652.
3. Daniel J Caldwell, Frank Mastrocco, Paul D Anderson, Reinhard Länge, and John P Sumpter, *Predicted-no-effect concentrations for the steroid estrogens estrone, 17β-estradiol, estriol, and 17α-ethinylestradiol.* Environmental Toxicology and Chemistry, 2012. **31**(6): p. 1396-1406.
4. Helmut Segner, Ayako Casanova-Nakayama, Robert Kase, and Charles R Tyler, *Impact of environmental estrogens on yfish considering the diversity of estrogen signaling.* General and comparative endocrinology, 2013. **191**: p. 190-201.
5. Jinmiao Zha, Liwei Sun, Yiqi Zhou, Philip A Spear, Mei Ma, and Zijian Wang, *Assessment of 17α-ethinylestradiol effects and underlying mechanisms in a continuous, multigeneration exposure of the chinese rare minnow (gobiocypris rarus).* Toxicology and applied pharmacology, 2008. **226**(3): p. 298-308.
6. Karen A Kidd, Paul J Blanchfield, Kenneth H Mills, Vince P Palace, Robert E Evans, James M Lazorchak, and Robert W Flick, *Collapse of a fish population after exposure to a synthetic estrogen*, in *Proceedings of the National Academy of Sciences*. 2007. p. 8897-8901.
7. Stefanie Grund, Eric Higley, René Schönenberger, Marc JF Suter, John P Giesy, Thomas Braunbeck, Markus Hecker, and Henner Hollert, *The endocrine disrupting potential of sediments from the upper danube river (germany) as revealed by in vitro bioassays and chemical analysis.* Environmental Science and Pollution Research, 2011. **18**(3): p. 446-460.
8. Barbara V Rutishauser, Maija Pesonen, Beate I Escher, Gabriele E Ackermann, Hans-Rudolf Aerni, Marc J-F Suter, and Rik IL Eggen, *Comparative analysis of estrogenic activity in sewage treatment plant effluents involving three in vitro assays and chemical analysis of steroids.* Environmental Toxicology and Chemistry, 2004. **23**(4): p. 857-864.
9. Andrew C Johnson, Egon Dumont, Richard J Williams, Rik Oldenkamp, Iwona Cisowska, and John P Sumpter, *Do concentrations of ethinylestradiol, estradiol, and diclofenac in european rivers exceed proposed eu environmental quality standards?* Environmental science & technology, 2013. **47**(21): p. 12297-12304.
10. Adam R Schwindt, Dana L Winkelman, Kristen Keteles, Mark Murphy, and Alan M Vajda, *An environmental oestrogen disrupts fish population dynamics through direct and transgenerational effects on survival and fecundity.* Journal of applied ecology, 2014. **51**(3): p. 582-591.
11. Eric Higley, Stefanie Grund, Paul D Jones, Tobias Schulze, Thomas-B Seiler, Urte Lübcke-von Varel, Werner Brack, Jan Wölz, Hanno Zielke, and John P Giesy, *Endocrine disrupting, mutagenic, and teratogenic effects of upper danube river sediments using effect-directed analysis.* Environmental Toxicology and Chemistry, 2012. **31**(5): p. 1053-1062.
12. CEJR Desbrow, EJ Routledge, GC Brighty, JP Sumpter, and M Waldock, *Identification of estrogenic chemicals in stw effluent. 1. Chemical fractionation and in vitro biological screening.* Environmental science & technology, 1998. **32**(11): p. 1549-1558.
13. Patricia Burkhardt-Holm, Walter Giger, Herbert GUttinger, Ueli Ochsenbein, Armin Peter, Karin Scheurer, Helmut Segner, Erich Staub, and Marc J-F Suter, *Where have all the fish gone?* Environmental science & technology, 2005. **39**(21): p. 441A-447A.
14. Frederic DL Leusch, Christiaan De Jager, Yves Levi, Richard Lim, Leo Puijker, Frank Sacher, Louis A Tremblay, Vickie S Wilson, and Heather F Chapman, *Comparison of five in vitro bioassays to measure estrogenic activity in environmental waters.* Environmental science & technology, 2010. **44**(10): p. 3853-3860.
15. Jon P Nash, David E Kime, Leo TM Van der Ven, Piet W Wester, François Brion, Gerd Maack, Petra Stahlschmidt-Allner, and Charles R Tyler, *Long-term exposure to environmental concentrations of the pharmaceutical ethynylestradiol causes reproductive failure in fish.* Environmental health perspectives, 2004: p. 1725-1733.
16. Christiane Heiss. *Recommendation for a monitoring strategy for estrogens in coastal and continental surface waters*. 2013 [cited 2018 February 5th]; Available from: http://www.bafg.de/DE/05_Wissen/02_Veranst/2013/2013_02_27_votum_en.pdf?__blob=publicationFile.
17. European Union, *Directive 2013/39/eu of the european parliament and of the council of 12 august 2013 amending directives 2000/60/ec and 2008/105/ec as regards priority substances in the field of water policy.* Official Journal of the European Union, 2013.
18. EU Decision, *495/2015, commission implementing decision (eu) 2015/495 of 20 march 2015 establishing a watch list of substances for union-wide monitoring in the field of water policy pursuant to directive 2008/105/ec of the european parliament and of the council.* Off. J. Eur. Union L, 2015. **78**: p. 40-42.
19. Mario Carere, Stefano Polesello, Robert Kase, and Bernd Manfred Gawlik, *The emerging contaminants in the context of the eu water framework directive*, in *Emerging contaminants in river ecosystems*. 2015, Springer. p. 197-215.

20. EU WFD, *Directive 2000/60/ec of the european parliament and of the council establishing a framework for the community action in the field of water policy.* The European Parliament and the Council of the European Union: Brussels, Belgium, 2000.
21. RN Carvalho, L Ceriani, and A Ippolito, *Development of the first watch list under the environmental quality standards directive water policy*. 2015.
22. AS Wernersson, C Maggi, and M Carere, *Technical report on aquatic effect-based monitoring tools.* Office for Official Publications of the European Communities, 2014.
23. Petra Y Kunz, Cornelia Kienle, Mario Carere, Nadzeya Homazava, and Robert Kase, *In vitro bioassays to screen for endocrine active pharmaceuticals in surface and waste waters.* Journal of pharmaceutical and biomedical analysis, 2015. **106**: p. 107-115.
24. Robert Loos. *Analytical methods for possible wfd 1st watch list substances*. 2015 [cited 2018 February 5th]; Report EUR 27046 EN:[Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.829.1423&rep=rep1&type=pdf.
25. Mario Carere, Stefano Polesello, Caterina Sollazzo, and Bernd Manfred Gawlik, *Chemical monitoring and emerging pollutants in the common implementation strategy of the water framework directive.* TrAC Trends in Analytical Chemistry, 2012. **36**: p. 12-14.
26. Markus Hecker and Henner Hollert, *Endocrine disruptor screening: Regulatory perspectives and needs.* Environmental Sciences Europe, 2011. **23**(1): p. 1.
27. AD Vethaak and AC Belfroid, *Estrogens and xeno-estrogens in the aquatic environment of the netherlands: Occurence, potency and biological effects*. 2002: Rijkswaterstaat, RIKZ.
28. Ann-Sofie Wernersson, Mario Carere, Chiara Maggi, Petr Tusil, Premysl Soldan, Alice James, Wilfried Sanchez, Valeria Dulio, Katja Broeg, and Georg Reifferscheid, *The european technical report on aquatic effect-based monitoring tools under the water framework directive.* Environmental Sciences Europe, 2015. **27**(1): p. 1-11.
29. Karin Kinnberg, *Evaluation of in vitro assays for determination of estrogenic activity in the environment.* 2003, Danish Ministry of the Environment, Danish Environmental Protection Agency.
30. Diane M Klotz, Barbara S Beckman, Steven M Hill, John A McLachlan, Marian R Walters, and Steven F Arnold, *Identification of environmental chemicals with estrogenic activity using a combination of in vitro assays.* Environmental health perspectives, 1996. **104**(10): p. 1084.
31. Beate I Escher, Nadine Bramaz, Jochen F Mueller, Pamela Quayle, Sibylle Rutishauser, and Etiënne LM Vermeirssen, *Toxic equivalent concentrations (teqs) for baseline toxicity and specific modes of action as a tool to improve interpretation of ecotoxicity testing of environmental samples.* Journal of Environmental Monitoring, 2008. **10**(5): p. 612-621.
32. Charles W Walker and John E Watson, *Adsorption of estrogens on laboratory materials and filters during sample preparation all rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.* Journal of environmental quality, 2010. **39**(2): p. 744-748.
33. Patrick Balaguer, Fabienne François, Franck Comunale, Hélène Fenet, Anne-Marie Boussioux, Michel Pons, Jean-Claude Nicolas, and Claude Casellas, *Reporter cell lines to study the estrogenic effects of xenoestrogens.* Science of the Total Environment, 1999. **233**(1): p. 47-56.
34. Beate I Escher, Mayumi Allinson, Rolf Altenburger, Peter A Bain, Patrick Balaguer, Wibke Busch, Jordan Crago, Nancy D Denslow, Elke Dopp, and Klara Hilscherova, *Benchmarking organic micropollutants in wastewater, recycled water and drinking water with in vitro bioassays.* Environmental science & technology, 2013. **48**(3): p. 1940-1956.
35. OECD, *Test no. 455: Performance-based test guideline for stably transfected transactivation in vitro assays to detect estrogen receptor agonists*. OECD Publishing.
36. Denise Spira, Georg Reifferscheid, and Sebastian Buchinger, *Combination of high-performance thin-layer chromatography with a specific bioassay-a tool for effect-directed analysis.* JPC-Journal of Planar Chromatography-Modern TLC, 2013. **26**(5): p. 395-401.
37. M. B. Müller, C. Dausend, C. Weins, and F. H. Frimmel, *A new bioautographic screening method for the detection of estrogenic compounds.* Chromatographia, 2004. **60**(3): p. 207-211.
38. Sebastian Buchinger, Denise Spira, Kathrin Bröder, Michael Schlüsener, Thomas Ternes, and Georg Reifferscheid, *Direct coupling of thin-layer chromatography with a bioassay for the detection of estrogenic compounds: Applications for effect-directed analysis.* Analytical chemistry, 2013. **85**(15): p. 7248-7256.
39. Ines Klingelhöfer and Gertrud E Morlock, *Bioprofiling of surface/wastewater and bioquantitation of discovered endocrine-active compounds by streamlined direct bioautography.* Analytical chemistry, 2015. **87**(21): p. 11098-11104.
40. Andreas Schönborn and Andrea Grimmer, *Coupling sample preparation with effect-directed analysis of estrogenic activity-proposal for a new rapid screening concept for water samples.* JPC-Journal of Planar Chromatography-Modern TLC, 2013. **26**(5): p. 402-408.

41. Maria König, Beate I Escher, Peta A Neale, Martin Krauss, Klára Hilscherová, Jiří Novák, Ivana Teodorović, Tobias Schulze, Sven Seidensticker, and Muhammad Arslan Kamal Hashmi, *Impact of untreated wastewater on a major european river evaluated with a combination of in vitro bioassays and chemical analysis.* Environmental Pollution, 2017. **220**: p. 1220-1230.
42. Lothar Sachs, *Angewandte statistik: Anwendung statistischer methoden*. 2013: Springer-Verlag.
43. Ze-hua Liu, Yoshinori Kanjo, and Satoshi Mizutani, *Removal mechanisms for endocrine disrupting compounds (edcs) in wastewater treatment—physical means, biodegradation, and chemical advanced oxidation: A review.* Science of the Total Environment, 2009. **407**(2): p. 731-748.
44. Barbora Jarošová, Luděk Bláha, John P Giesy, and Klára Hilscherová, *What level of estrogenic activity determined by in vitro assays in municipal waste waters can be considered as safe?* Environment international, 2014. **64**: p. 98-109.
45. R Loos, *Analytical methods relevant to the european commission's 2012 proposal on priority substances under the water framework directive.* JRC scientific and policy report, 2012.
46. Tom Benijts, Riet Dams, Willy Lambert, and André De Leenheer, *Countering matrix effects in environmental liquid chromatography–electrospray ionization tandem mass spectrometry water analysis for endocrine disrupting chemicals.* Journal of chromatography A, 2004. **1029**(1): p. 153-159.
47. YKK Koh, TY Chiu, A Boobis, E Cartmell, JN Lester, and MD Scrimshaw, *Determination of steroid estrogens in wastewater by high performance liquid chromatography–tandem mass spectrometry.* Journal of chromatography A, 2007. **1173**(1): p. 81-87.
48. EC Directive, *Commission directive 2009/90/ec of 31 july 2009 laying down, pursuant to directive 2000/60/ec of the european parliament and of the council, technical specifications for chemical analysis and monitoring of water status.* Official Journal of the European Union L, 2009. **201**: p. 36.
49. Sándor Görög, *Advances in the analysis of steroid hormone drugs in pharmaceuticals and environmental samples (2004–2010).* Journal of pharmaceutical and biomedical analysis, 2011. **55**(4): p. 728-743.
50. Helena Tomšíková, Jana Aufartová, Petr Solich, Lucie Nováková, Zoraida Sosa-Ferrera, and José Juan Santana-Rodríguez, *High-sensitivity analysis of female-steroid hormones in environmental samples.* TrAC Trends in Analytical Chemistry, 2012. **34**: p. 35-58.
51. Carolina Di Paolo, Richard Ottermanns, Steffen Keiter, Selim Ait-Aissa, Kerstin Bluhm, Werner Brack, Magnus Breitholtz, Sebastian Buchinger, Mario Carere, and Carole Chalon, *Bioassay battery interlaboratory investigation of emerging contaminants in spiked water extracts–towards the implementation of bioanalytical monitoring tools in water quality assessment and monitoring.* Water research, 2016. **104**: p. 473-484.
52. Robert Kase, Barbora Javurkova, Eszter Simon, Kees Swart, Sebastian Buchinger, Sarah Könemann, Beate Escher, Mario Carere, Valeria Dulio, Selim Ait-Aissa, Henner Hollert, Sara Valsecchi, Stefano Polesello, Peter Behnisch, Carolina Di Paolo, Daniel Olbrich, Eliska Sychrova, Michael Gundlach, Rita Schlichting, Lomig Leborgne, Manfred Clara, Christoph Scheffknecht, Yves Marneffe, Carole Chalon, Petr Tusil, Premysl Soldan, Brigitte von Danwitz, Julia Schwaiger, Antonio Moran, Francesca Bersani, Olivier Perceval, Cornelia Kienle, Etienne Vermeirssen, Klara Hilscherova, Georg Reifferscheid, and Inge Werner, *Screening and risk management solutions for steroidal oestrogens in surface and wastewater.* Trends in Analytical Chemistry, 2018. Under review.
53. Juliette Legler, Martine Dennekamp, A Dick Vethaak, Abraham Brouwer, Jan H Koeman, Bart van der Burg, and Albertinka J Murk, *Detection of estrogenic activity in sediment-associated compounds using in vitro reporter gene assays.* Science of the Total Environment, 2002. **293**(1): p. 69-83.
54. Albertinka J Murk, Juliette Legler, Marola MH Van Lipzig, John HN Meerman, Angelique C Belfroid, Albertus Spenkelink, Bart Van Der Burg, Gerard BJ Rijs, and Dick Vethaak, *Detection of estrogenic potency in wastewater and surface water with three in vitro bioassays.* Environmental Toxicology and Chemistry, 2002. **21**(1): p. 16-23.
55. Frederic DL Leusch. *Tools to detect estrogenic activity in environmental waters.* 2008 [cited 2018 February 5th]; Available from: http://www.waterrf.org/resources/Lists/SpecialReports/Attachments/2/GWRC_EDC_ToolsToDetect.pdf.
56. Werner Brack, Valeria Dulio, Marlene Ågerstrand, Ian Allan, Rolf Altenburger, Markus Brinkmann, Dirk Bunke, Robert M Burgess, Ian Cousins, and Beate I Escher, *Towards the review of the european union water framework directive: Recommendations for more efficient assessment and management of chemical contamination in european surface water resources.* Science of the Total Environment, 2017. **576**: p. 720-737.
57. Ian J Allan, Branislav Vrana, Richard Greenwood, Graham A Mills, Benoit Roig, and Catherine Gonzalez, *A "toolbox" for biological and chemical monitoring requirements for the european union's water framework directive.* Talanta, 2006. **69**(2): p. 302-322.

Highlights

In vitro effect-based methods integrate effects of mixtures of chemical compounds with the same mode of action

E2 equivalents are highly correlated with LC-MS/MS

E2 equivalents are highly correlated among effect-based methods

Implementation of effect-based methods in the water framework directive is highly recommended