2024

Ongoing Laboratory Performance Study on Chemical Analysis of Hydrophobic and Hydrophilic Compounds in Three Aquatic Passive Samplers

Kees Booij

Steven Crum

Branislav Vrana

Roman Grabic

Nicolas A.O. Morin

See next page for additional authors

Citation/Publisher Attribution

Booij, K. *et al.* (2024) 'Ongoing Laboratory Performance Study on Chemical Analysis of Hydrophobic and Hydrophilic Compounds in Three Aquatic Passive Samplers', *Environmental Science & Technology*, *58*(15), 6772–6780. https://doi.org/10.1021/acs.est.3c10272
Available at: https://doi.org/10.1021/acs.est.3c10272

Ongoing Laboratory Performance Study on Chemical Analysis of Hydrophobic and Hydrophilic Compounds in Three Aquatic Passive Samplers

Authors

Kees Booij, Steven Crum, Branislav Vrana, Roman Grabic, Nicolas A.O. Morin, Koen Parmentier, Cécile Kech, Petra Krystek, Kazushi Noro, Benjamin Becker, Rainer Lohmann, Laure Malleret, Sarit L. Kaserzon, Cécile Miège, Fabrice Alliot, Fabienne Pfeiffer, Denis Crowley, Magdalena Rakowska, Tomas Ocelka, Gi Beum Kim, and Laura Röhler

Ongoing laboratory performance study on chemical analysis of hydrophobic and hydrophilic compounds in three aquatic passive samplers

Kees Booij[a,*], Steven Crum[b], Branislav Vrana[c], Roman Grabic[d], Nicolas A.O. Morin[e], Koen Parmentier[f], Cécile Kech[g], Petra Krystek[h], Kazushi Noro[i], Benjamin Becker[j], Rainer Lohmann[k], Laure Malleret[l], Sarit L. Kaserzon[m], Cécile Miège[n], Fabrice Alliot[o], Fabienne Pfeiffer[p], Denis Crowley[q], Magdalena Rakowska[r], Tomas Ocelka[s], Gi B. Kim[t], Laura Röhler[u]

[a] PaSOC, 8821 LV Kimswerd, The Netherlands

[b] Wageningen Environmental Research, 6708 PB Wageningen, The Netherlands

[c] RECETOX, Faculty of Science, Masaryk University, 61137 Brno, Czech Republic

[d] University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, 38925 Vodňany, Czech Republic

[e] Laboratoire de l'Environnement et de l'Alimentation de la Vendée, 85021 La Roche sur Yon, France

[f] Royal Belgian Institute of Natural Sciences (RBINS), 8400 Oostende, Belgium

[g] Scientific Institute of Public Service (ISSeP), 4000 Liège, Belgium

[h] Deltares, 3584 BK Utrecht, The Netherlands

[i] University of Shizuoka, 422-8526 Shizuoka, Japan, and Research Institute of Environment, Agriculture, and Fisheries, Osaka Prefecture, 583-0862 Habikino, Osaka, Japan

[j] Federal Institute of Hydrology, 56068 Koblenz, Rheinland-Pfalz, Germany

[k] University of Rhode Island, Graduate School of Oceanography, Narragansett, RI 02882, USA

[l] Laboratoire Chimie Environnement, Aix Marseille University, CNRS, 13545 Aix-en-Provence, France

[m] Queensland Alliance for Environmental Health Sciences (QAEHS), The University of Queensland, Woolloongabba, Queensland 4102, Australia

[n] INRAE, UR RiverLy, 69625 Villeurbanne, France

[o] METIS, Sorbonne Université, CNRS, EPHE, PSL University, UMR 7619, 75005 Paris, France

[p] University of Lausanne, School of Criminal Justice, 1015 Lausanne, Switzerland

[q] Marine Institute, H91 R673 Galway, Ireland

[r] Civil, Environmental, and Construction Engineering, Texas Tech University, Lubbock, TX 79409-1023, United States and Envirostatus LLC, Lubbock, TX 79415, United States

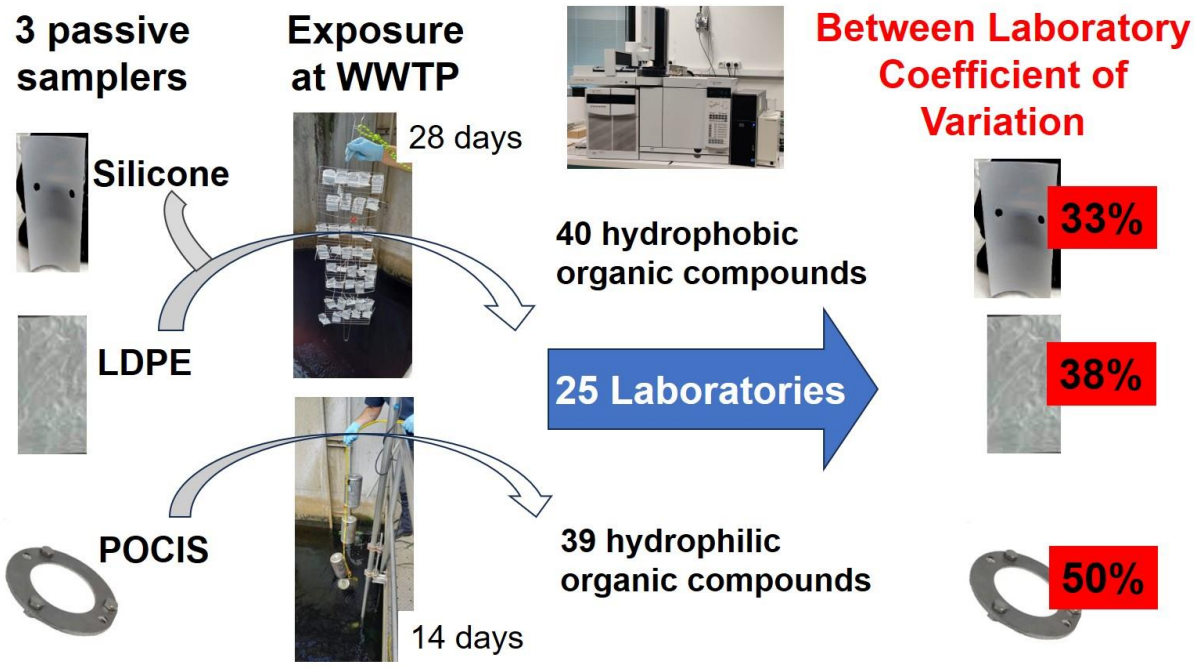
[s] E&H services Inc., DioxinLab, 739 51 Dobrá, Czech Republic

[t] Gyeongsang National University, Marine Environmental Engineering, 660-701 Jinju, Korea

[u] NIVA - Norwegian Institute for Water Research, 0579 Oslo, Norway

* Email: keesbooij@pasoc.eu

TOC/Abstract Graphic



Abstract

The quality of chemical analysis is an important aspect of passive sampling based environmental assessments. The present study reports on a proficiency testing program for the chemical analysis of hydrophobic organic compounds in silicone and low-density polyethylene (LDPE) passive samplers and hydrophilic compounds in polar organic chemical integrative samplers. The median between-laboratory coefficients of variation (CVs) of hydrophobic compound concentrations in the polymer phase were 33% (silicone) and 38% (LDPE), similar to CVs obtained in four earlier rounds of this program. The median CV over all rounds was 32%. Much higher variabilities were observed for hydrophilic compound concentrations in the sorbent: 50% for the untransformed data, and a factor of 1.6 after log transformation. Limiting the data to the best performing laboratories did not result in less variability. Data quality for hydrophilic compounds was only weakly related to the use of structurally identical internal standards and was unrelated to the choice of extraction solvent and extraction time. Standard deviations of the aqueous concentration estimates for hydrophobic compound sampling by the best performing laboratories were 0.21 log units for silicone and 0.27 log units for LDPE (factors 1.6 to 1.9). The implications are that proficiency testing programs may give more realistic estimates of uncertainties in chemical analysis than within-laboratory quality control programs, and that these high uncertainties should be taken into account in environmental assessments.

Synopsis: Chemical analysis of aquatic passive samplers results in an uncertainty of aqueous concentrations of at least 0.21 log units (factor 1.6).

Keywords: passive sampling, proficiency testing, interlaboratory comparison, silicone, polyethylene, POCIS, quality control

1. Introduction

The quality of chemical analysis in environmental monitoring programs is of crucial importance for assessing the risk of anthropogenic organic compounds in aqueous environments. Lower uncertainties in reported concentration levels facilitate the detection of temporal and spatial trends, and reduce the risk of making incorrect environmental management decisions. Data comparability between laboratories should be particularly high and well defined when multiple laboratories submit data to the same monitoring program. The data quality policy of the United States Environmental Protection Agency (US EPA) stipulates that the quality of chemical analysis should be appropriate for the intended use of the data within particular studies, and that measurement uncertainty should be weighed against sampling uncertainty.[1,2] US EPA requires that quality management plans (laboratory level) and quality assurance project plans (project level) are in place for each study.[3] Laboratories that analyze priority substances within the EU Water Framework Directive should be accredited according to ISO 17025 (or equivalent), participate in proficiency testing (PT) programs, and ensure that measurement uncertainty is less than 50% (coverage factor k=2) at the level of the environmental quality standards.[4,5] This corresponds to a relative standard error of 25%.

Regular participation in PT programs can inform laboratories about the adequacy of their internal quality assurance and control measures, and the comparability of their data with data from other laboratories. In addition, these programs inform clients and stakeholders of these laboratories about the achievable accuracy of the chemical analysis, both with respect to the individual laboratory and within the group of participants. Guidance on the selection of PT programs and the interpretation of results is given by Eurachem.[6] Participation in such studies is also important for laboratories that use passive sampling methods for assessing the global distribution of contaminants, evaluating contaminated sites, and monitoring the release of potentially harmful substances into the aquatic environment, among others.[7–9]

PT programs for the analysis of organic contaminants in the aquatic environment have been mainly focused on polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), and chlorinated

89 pesticides in sediments and biota. Results of these studies indicated median between-laboratory
90 coefficients of variation (CVs) in the range 23 to 65%.[10–13] A compilation of unpublished one-off
91 interlaboratory comparison studies on the analysis of nonpolar and slightly polar organic compounds
92 ($\log K_{ow}$ = 3.4 - 7.6) in surface water samples suggests CVs of 20 to 70%.[4] A PT study for the analysis of
93 pharmaceuticals in spiked clean water and spiked waste water reported CVs of 15 to 61%.[14] Four
94 interlaboratory studies for passive sampling in sediments and water revealed substantial between-
95 laboratory variability of aqueous concentrations of organic compounds based on the same sampler type
96 (silicone, LDPE, POCIS). Variability was either expressed as CVs with median values of 87% and 62%,[15,16]
97 or as ratios of high/low aqueous concentrations (C_w) of 4.4 and 2.[17,18] Two studies identified chemical
98 analysis of the sampler sorbent as a major source of variability.[16,17] This justifies a separate focus on
99 analytical quality.

100 A PT development exercise for passive sampling of hydrophobic organic compounds was initiated in
101 2014 by WEPAL-QUASIMEME, which is an ISO 17043 accredited not-for-profit PT organization under the
102 umbrella of Wageningen University, The Netherlands. Results for the first two rounds of this exercise
103 (2014 and 2015) indicated CVs for the analysis of hydrophobic compounds in silicone samplers between
104 3 and 97%, with a median value of 28%.[19] Additional rounds for silicone samplers were organized in
105 2016 and 2018.

106 The purpose of present study was to evaluate and improve the design of the PT scheme for passive
107 sampling of hydrophobic compounds, extend the program to LDPE samplers, and explore the challenges
108 of a PT scheme for passive sampling of hydrophilic compounds. The primary focus was on the variability
109 of the chemical analysis of passive samplers, rather than on variability associated with sampler
110 preparation, deployment, and modelling, but the consequences for the uncertainty in C_w estimates of
111 hydrophobic compounds are evaluated.

2. Materials and methods

2.1. Sampler preparation.

Silicone sheets with 0.25 mm thickness (Specialty Silicone Products, SSP-M823) and low-density polyethylene (LDPE) sheets with 0.080 mm thickness (Alte-Rego Corporation) were cut to a size of 55 × 90 mm and pre-extracted with ethyl acetate (silicone) or acetone/hexane 1:1 v/v (LDPE) using a 100 h Soxhlet extraction. Sheets were spiked with performance reference compounds (PRCs), following published guidelines.[20] PRCs were PCB congeners 1, 2, 3, 10, 14, 21, 50, 55, 78, 104, 145, 204. Sheets were exposed for 28 d (starting 25 August 2021) in the effluent of the municipal wastewater treatment plant serving Brno city, Czech Republic. Collected sheets were transferred to an amber glass jar (one jar per polymer type) and packed into 40 mL amber glass vials (3 tightly wrapped sheets per vial, Supporting Information Figure S1-1). Nonexposed sheets were transferred to 40 mL amber glass vials (3 per vial), immediately after spiking. All vials were stored in the dark at −20 °C.

Granular Oasis Hydrophilic-Lipophilic Balance (HLB) sorbent was washed three times with acetone, followed by solvent removal under mild vacuum. The sorbent was not spiked with PRCs. The sorbent was exposed in multiple Polar Organic Chemical Integrative Samplers (POCIS), equipped with a polyethersulfone membrane (Pall Supor 100, nominal pore size 0.1 μm, membrane thickness approximately 130 μm) sandwiched between two stainless steel washers (54 mm internal diameter, 101 mm external diameter). A higher sorbent mass per POCIS (400 mg vs. 220 mg for standard POCIS) was chosen to reduce the number of samplers. This was considered to be permissible because the focus of the present study was on the chemical analysis of hydrophilic compounds in the sorbent, rather than the determination of aqueous concentrations of these compounds. Exposure time was 14 d (starting 25 August 2021), at the same site as for the silicone and LDPE exposures. Sorbents of all POCIS were pooled after exposure, freeze-dried, mixed, distributed over 4 mL amber glass vials (approximately 600 mg per vial), and stored at −20°C. The choice for a granular sorbent instead of extraction disks optimized sample homogeneity, which eliminates any difference in the accumulation by individual POCIS. Sampler preparation and deployment was done by RECETOX.

QUASIMEME sent samplers to participants by courier. Participants were instructed to check samples for damage and to store them dark and frozen, preferably at −20 °C.

Homogeneity tests were carried out to ensure sufficiently similar concentrations in the samplers. These tests were done by RECETOX for silicone and LDPE and by the University of South Bohemia for HLB. Eight randomly selected vials (containing silicone sheets, LDPE sheets, or HLB sorbent) were processed in one analysis batch. For silicone, CVs (relative to the within-batch mean) were 1-6% for organochlorine compounds, 2-12% for polybrominated biphenyl ethers (PBDEs), and 3-9% for PAHs, with the exception of naphthalene (75%), acenaphthene (11%) and acenaphthylene (16%). Standard deviations of retained PRC fractions (amount ratios in exposed/nonexposed samplers) in silicone were 0.0001- 0.06 (dimensionless). Similar results were obtained for LDPE. CVs for hydrophilic compounds in HLB were between 8 and 17%. Samples were therefore considered to be homogeneous. Further details are given in section S2 and in the supplementary data file.

2.2. Data collection.

Participants from 25 laboratories reported 1545 concentration values (669 for silicone, 474 for LDPE, and 402 for HLB) for 78 hydrophobic compounds and 75 hydrophilic compounds. In addition, 245 retained fractions for 12 PRCs were reported. All compounds are listed in section S3, and all reported concentrations are available from the supplementary data file. Laboratories were based in North America (3), Europe (19), Asia (2), and Australia (1), and were from academia (12), public sector (11), and private sector (2). Participants used their own methods of analysis. Data collection and storage was done by QUASIMEME.

To remove ambiguity in reported concentrations of chrysene and triphenylene due to co-elution, participants had to report the separate compounds and/or their sum. Participants also had to report benzofluoranthenes as individual compounds (b, j, k congeners) and/or their sums, for the same reason. A data evaluation could be made for 44 non-PRCs in silicone, 36 non-PRCs in LDPE, 11 PRCs in silicone, 12 PRCs in LDPE, and 39 hydrophilic compounds in Oasis HLB (Tables S3-1 and S3-2).

For comparison, results from previous rounds with silicone samplers in 2014, 2015, 2016, and 2018 were included in the data analysis. Sampler deployments for these rounds were made at four sites in The Netherlands: Western Scheldt near Hansweert (2014, 91 d), River Rhine near Lobith (2015, 49 d), River Meuse near Eijsden (2016, 42 d), and Rotterdam Harbor near Maassluis (2018, 43 d). The design of these studies was the same as for silicone in 2021. The number of participants in all rounds was 21 (2014), 13 (2015), 10 (2016), 13 (2018), 19 (silicone 2021), 13 (LDPE 2021), and 15 (HLB 2021).

During the initial data evaluation we noted particularly large variability (50% or more) for hydrophilic compounds and for 2- and 3-ring PAHs. Participants were therefore asked to provide additional information about their certification status (ISO 17025 or similar), use of certified or internal reference materials (hydrophobic compounds only), internal standards, and extraction methods. This information was obtained from 22 out of 25 laboratories; one laboratory was shut down after participating, and two laboratories did not respond.

2.3. Data analysis.

Robust means and standard deviations were obtained using the Cofino model.[21–23] This model does not require separate outlier tests and can incorporate left-censored values (below a reporting limit). A normal distribution of errors was assumed, unless indicated otherwise. Considerations for the assumed error distribution are given in the Results and Discussion section. Robust means are further referred to as “consensus values”.

Laboratory performance was evaluated in terms of Z’-scores for all analytes. These are defined as the difference between reported and consensus value, normalized by a total error (all in ng/g).

$$Z' = \frac{\text{Reported Value} - \text{Consensus Value}}{\text{Total Error}} \quad (1)$$

The total error was obtained from the target error (the error that was assumed to be achievable) and the uncertainty (u_x) in the consensus value.

$$\text{Total Error} = \sqrt{\left(\text{Target Error}\right)^2 + u_x^2} \quad (2)$$

$$u_x = 1.25\frac{s}{\sqrt{n}} \quad (3)$$

where s is the standard deviation, and n is the number of observations. The u_x is 1.25 times the standard error of the mean. The factor 1.25 is prescribed by ISO 13528 to account for possible deviations from normality in the data.

The target error was expressed in terms of a constant error and a proportional error (%).

$$\text{Target Error} = \frac{\text{Proportional Error} \times \text{Consensus Value}}{100\%} + 0.5 \times \text{Constant Error} \quad (4)$$

This approach allows for a higher relative error at low concentrations. The proportional error was set at 25%, based on previous experience with QUASIMEME PT studies for hydrophobic compound analysis in sediments and biota, and on legal EU requirements.[5] The constant error was set at 0.03 ng/g for organochlorine compounds, 0.1 ng/g for PAHs, and 0.01 ng/g for PBDEs, also based on previous experience. These errors were set at 10% and 0.005 for retained PRC fractions (dimensionless) and at 25% and 0.1 ng/g for hydrophilic compounds.

Consensus values were set when data fell into either of two categories. Category 1: ≥ 7 numerical values, ≥ 50% of the reported values with $|Z'| < 2$, and ≥ 5 values with $|Z'| < 3$. Category 2: 4 to 6 numerical values, and ≥ 4 values with $|Z'| < 2$, and ≥ 70% values with $|Z'| < 3$.

Water sampling rates (R_s, L/d) for sampling of hydrophobic compounds by silicone and LDPE were evaluated from the retained PRC fractions (f), assuming rate control by the water boundary layer.

$$f = \exp\left(-\frac{R_s t}{m K_{sw}}\right) = \exp\left(-\frac{\beta_M t}{m K_{sw} M_r^{0.47}}\right) \quad (5)$$

where K_{sw} (L/kg) is the sampler-water partition coefficient, m (kg) is the sampler mass, t (d) is time, M_r is the relative molecular mass (dimensionless), and β_M is a proportionality constant that reflects the flow effect on the exchange kinetics. The power 0.47 accounts for the weak decrease of aqueous diffusion coefficients with molecular weight.[24] Literature values for K_{sw} were adopted (Table S4−1). C_w (ng/L) was calculated from

$$C_w = \frac{C_s}{K_{sw}\left[1-\exp\left(-\frac{\beta_M t}{m K_{sw} M_r^{0.47}}\right)\right]} \quad (6)$$

Data analysis was done by QUASIMEME and PaSOC.

3. Results and Discussion

3.1. Hydrophobic compounds analysis in silicone and LDPE

CVs for the analysis of hydrophobic compounds ranged between 10 and 150% and were similar for LDPE and silicone samplers (Figure 1, Figures S5-1 and S5-3). The CVs of bromodiphenylethers span a similar range as observed in the NORMAN Interlaboratory Study.[16,25] These CVs were 13-77% in the NORMAN study versus 19-150% in the present study, without an obvious correlation between the CVs of compounds reported in both studies (Table S6-1, Figure S6-1). Other hydrophobic compounds were not included in the NORMAN study.

Appreciable differences occur among laboratories with respect to the percentage of |Z'|<2. Data for silicone 2021 show that 5 laboratories have more than 90% |Z'|<2, another 5 have 80-90% |Z'|<2, and 9 laboratories have less than 80% |Z'|<2 (Figure 2). Nearly all laboratories reported extreme values

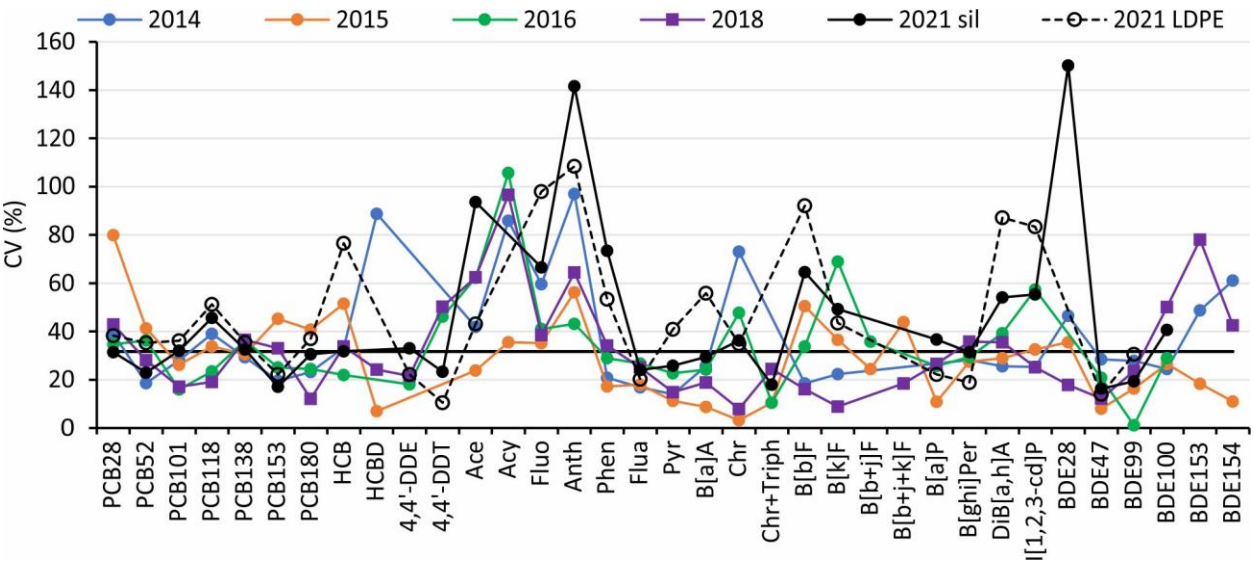


Figure 1. CVs for the analysis of hydrophobic compounds in silicone (2014, 2015, 2016, 2018, 2021) and LDPE (2021). The solid black line represents the median CV over all rounds and compounds (32%). Lines between the data points are shown as a visual guide. CVs of compounds only reported in 2021 are not shown (see Figure S5-3). Shorthand names are defined in Table S3-1.

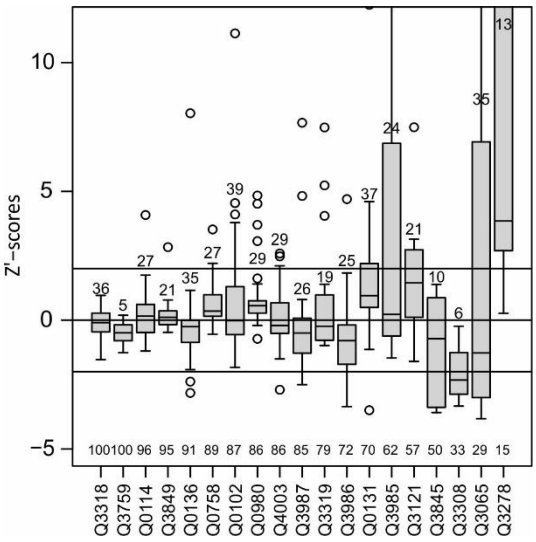


Figure 2. Boxplot of Z'-scores for the analysis of hydrophobic compounds in silicone for the 2021 study. Laboratories are sorted on descending percentage |Z'|<2 (numbers near the horizontal axis). The number of reported values is shown above the boxes. Horizontal lines represent Z'=±2 and Z'=0.

(|Z'|>3) for one or more compounds. Boxplots for LDPE are similar as for silicone (Figures S7-1 and S7-2). Percentages of favorable Z'-scores (|Z'|<2) were rather similar for laboratories with and without certification under ISO 17025 (Figure S8-1). Medians were 89% (range 15-100%) and 79% (range 26-96%), respectively. Laboratories that regularly analyze reference materials (certified and internal reference materials, including matrix spikes) had a similar percentage of favorable Z'-scores as laboratories that used none of these measures (Figure S8-1), with medians of 80% (range 15 to 99%) and 81% (range 26-100%), respectively. We found no relationship between %|Z'|<2 and methods used for extraction, cleanup, and instrumental analysis (section S8). This means that it is more important how methods are applied than which method is selected. Further details are given in section S8.

Comparison with four previous rounds (2014, 2015, 2016 and 2018), and taking silicone and LDPE for 2021 as separate datasets, shows that high and low CVs occur for individual compounds in separate rounds, but that the long-term average is fairly constant (Figure 1). Several very small CVs can be labelled as unrealistic. For example, the CVs of BDE99 in 2016 (1%, *n*=5) and chrysene in 2015 (3%, *n*=7) are exceptionally small. The CVs over all rounds are more realistic (23% for BDE99 and 49% for chrysene).

No relationship between CV and concentration level was observed (Figure S9-1), which means that the constant error (equation 4) cannot be determined until more data are available. Standard deviations of

retained PRCs fractions were smaller for LDPE than for silicone by a factor of approximately 2, for unknown reasons (Figure S5-2).

To further explore differences and similarities of CVs between compounds, we hypothesized that CVs can be characterized by a common value, obtained from the median CV over all rounds and compounds. Subsequently, we evaluated whether the 95% confidence intervals of the CVs (section S10) for individual compounds overlapped with this common value. This was the case for most compounds, except the 2- and 3-ring PAHs (Figure S11-1). Between-round differences in CVs are therefore likely caused by the small number of participants per round (typically 10 to 20). Median CVs per compound, taken over all rounds, ranged between 14% (chrysene + triphenylene) and 91% (acenaphthylene), with an overall median of 32%. This analysis indicates that between-laboratory CVs are best estimated from the median CV over multiple rounds and multiple compounds. The median CV within a single round may be adopted as an optimal initial estimate when multi-round data are not available, for example in the case of one-off studies or when compounds are included for the first time. For the present data, these medians amount to 28, 28, 29, 26, 33, and 38 % (silicone 2014, 2015, 2016, 2018, 2021, and LDPE in 2021, respectively), which is close to the multi-round median of 32% (Figure S11-2).

Median CVs of acenaphthene, acenaphthylene, fluorene, and anthracene over multiple rounds were substantially higher (53, 91, 50, 81%, respectively, Figure S11-1). The CV of naphthalene (available for silicone in 2021 only) was also high (120%). Evaporation losses during sample processing cannot explain the high CVs of these relatively volatile compounds because laboratories using structurally identical internal standards did not obtain lower |Z'|-scores (section S12). Evaporation before sample processing is also unlikely because samplers were tightly wrapped in a small vial, and sampler-air partition coefficients are >36000 mL/g (Section S1). We speculate therefore that these high variabilities originate from background contamination during sample processing and inadequate blank subtraction. Chromatographic separation is unlikely to explain the higher CV of anthracene as compared with phenanthrene (section S12).

3.2. Hydrophilic compounds in HLB

CVs for the analysis of hydrophilic compounds in HLB ranged from 10 to 100%, with a median value of 50% (Figure 3). The highest CVs were observed for compounds with concentrations > 1000 ng/g (Figure S9-1, right panel), possibly as a result of detector linearity issues. Six out of 39 compounds had CVs that differ significantly from the median (two-tailed Chi-squared test, $\alpha = 0.05$). The CVs are similar to the range of 16-119% and 39-111% that was found for hydrophilic compound analysis in POCIS in other studies (Table S6-1).[15,16,25] The lack of correlation between the CVs among studies suggests that high or low CVs are not compound-specific (Figure S6-1, right panel).

Appreciable differences in laboratory performance occur, as evidenced by the Z'-score boxplots and percentage |Z'|<2 (Figure 4, left panel). Seven laboratories obtained more than 90% |Z'|<2. The percentage was between 80 and 90% for three laboratories and less than 80% for five laboratories.

Laboratory certification status (ISO 17025) for hydrophilic compound analysis in surface waters was not strongly associated with the percentage |Z'|<2, and the same was observed for the choice of extraction solvent (methanol, acetone, mixed solvents), extraction time (3 to >60 min), and the use of matrix-matching standards methods (Table S13-1, Figures S13-1 to S13-4). Quantification with structurally identical internal standards generally yielded Z'-scores between −2 and 2, but the range of Z'-scores was not particularly narrow, and several Z'-scores >10 were observed (Figure S13-5). These observations suggest that sorbent extraction methods are not critical and that the use of structurally identical internal standards is no guarantee for high data comparability among laboratories.

To evaluate if the presently adopted proportional error of 25% is achievable by the best performing laboratories, the CVs were determined on three separate datasets: all participants, participants with > 80% |Z'|<2, and participants with > 90% |Z'|<2. These sub-datasets are further referred to as 40+, 80+, and 90+. A comparison could be made for 23 out of 39 compounds. For the other 16 compounds the number of observations in the 80+ and 90+ sub-datasets was too small to set a consensus value. Median

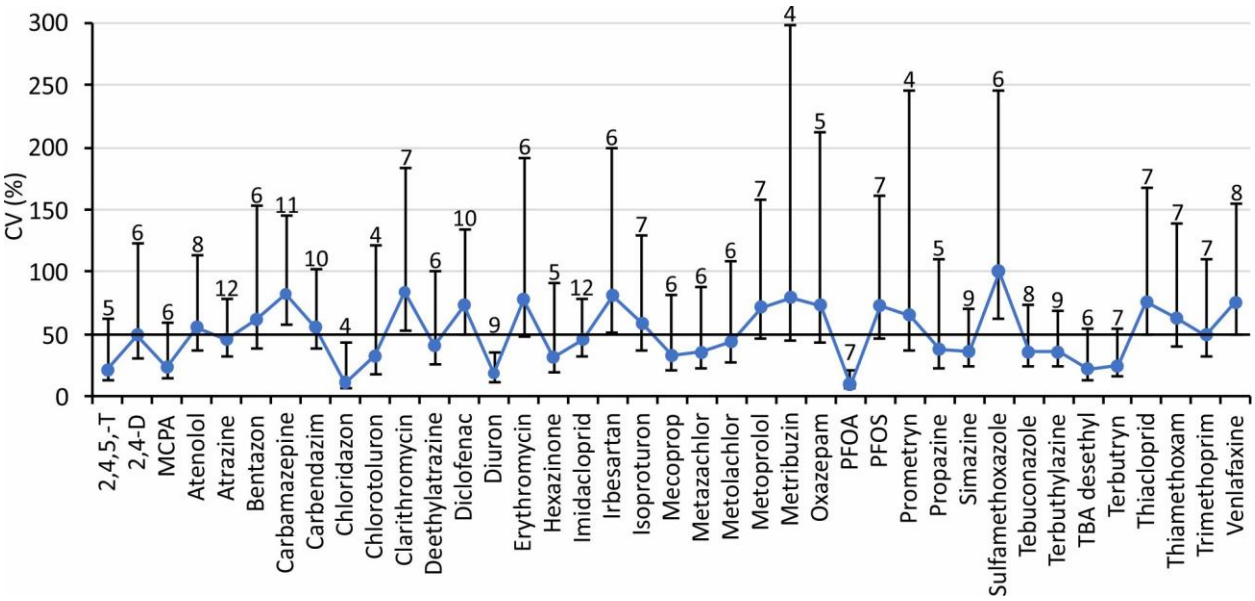


Figure 3. Coefficients of variation (CV) for the analysis of hydrophilic compounds in HLB. Error bars span the 95% confidence range. The number of observations is shown above the error bars. The solid line represents the median CV. Shorthand names are defined in Table S3-2.

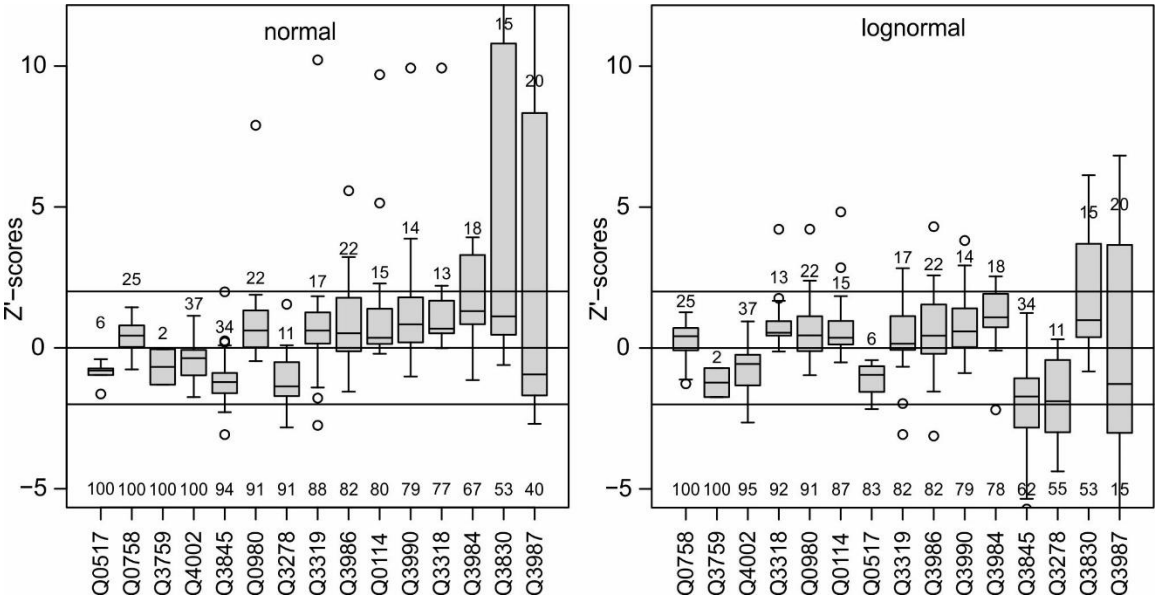


Figure 4. Boxplots of Z'-scores for the analysis of hydrophilic compounds in HLB for the 2021 study, assuming a normal (left) and lognormal (right) error distribution. Laboratories are sorted on descending percentage $|Z'|<2$ (numbers near the horizontal axis). The number of reported values is shown above the boxes. Horizontal lines represent $Z'=\pm 2$ and $Z'=0$.

CVs for the 90+, 80+, and 40+ sub-datasets were 30, 38, and 44%, respectively. These differences were only significant for the 40+ versus 90+ comparison (one-tailed *F*-test, $p=0.04$), which can be attributed to a large difference in the CVs for PFOS (73 vs 19%) and venlafaxine (76 vs 25%) (Figure S14-1).

311 The similarity of CVs for the respective sub-datasets illustrates first that the Cofino model is fairly
312 robust. Second, the adopted proportional error of 25% is presently not achievable, even by the best-
313 performing laboratories. Third, the high CVs shed doubt on the applicability of assuming a normal
314 distribution of errors, as CVs > 50% predict a significant probability of negative concentrations. The
315 latter occurred for 19 out of 39 compounds (>5% probability of negative C_s).

316 Moderate differences between assuming a normal and a lognormal distribution of errors were observed
317 for the consensus values. Data were $\log_{10}$-transformed and modeled in the same way as the
318 untransformed data. The ratios of consensus values (normal/lognormal) were 0.97 ± 0.05 (median $\pm$
319 standard deviation based on the median of absolute deviations). The highest differences occurred for
320 sulfamethoxazole (ratio = 1.26) and venlafaxine (ratio = 0.70). The probability density function for the
321 untransformed data showed significant overlap with negative concentrations for these compounds (the
322 modeled probability of C_s<0 was 0.18 and 0.11, respectively; Figure S15-1). The high ratio for
323 sulfamethoxazole can be understood by considering that geometric means (log-transformed data) are
324 always smaller than arithmetic means (untransformed data). The low ratio for venlafaxine is caused by
325 the fact that two high values appear as outliers for the untransformed data, and one low value appears
326 as an outlier for the log-transformed data, shifting the consensus concentration to a higher value for the
327 log-transformed data. The effect of the assumed error distribution on the consensus value is therefore
328 typically 5%, but can be as large as 30% for individual compounds.

329 The results of a Shapiro-Wilk normality test on the log-transformed and untransformed data were
330 inconclusive. Significant deviations from normality (p <0.05) occurred for 12 out of 39 compounds for
331 the untransformed data and 2 compounds after log transformation. However, the raw data contained
332 extreme concentrations that by their very nature result in low p-values in this test. Applying the test to
333 data after excluding values with $|Z'| > 2$ resulted in $0.04 < p < 1.00$ for data with or without log
334 transformation. This result is not surprising, as removing extreme values promotes normality of the
335 remaining data.

The assumed error distribution has a substantial effect on the standard deviations and resulting Z'-scores. Standard deviations (*s*) for the log-transformed data ranged between 0.04 and 0.54, with a median value of 0.21, taken over all compounds (Figure S14-2). These standard deviations are more easily interpreted as uncertainty factors after back-transformation (10^s). The median standard deviation of 0.21 corresponds to an uncertainty factor of 1.6 (68% of the data between 0.6 and 1.6 times the consensus value). This uncertainty is similar to the CV of 50% that was obtained for the untransformed data (68% of the data between 0.5 and 1.5 times the consensus value). Data analysis without log transformation becomes somewhat problematic with such high CVs. First, positive probabilities of negative concentrations that occur under the assumption of a normal error distribution are unrealistic. Second, assuming a normal distribution of errors is more permissive towards laboratories reporting very low values because Z'-values below −2 are impossible when CVs are >50%. For example, with a consensus value of 100 ng/g and a CV of 50%, a reported value of 1 ng/g is considered acceptable (Z'= −1.98), even though it is lower by two orders of magnitude. An advantage of adopting lognormal error distributions is that they are more generally applicable because the normal distribution is a limiting case of lognormal distributions when the errors are small. A disadvantage may be that CVs are more easily interpreted by participants of PT studies than uncertainty factors, but the primary indicators of laboratory performance are the Z'-scores, which can be reported for either error distribution.

Assuming a lognormal or a normal error distribution can make a large difference in the Z'-scores of individual laboratories. The target error for the log-transformed data was set to 0.10 log units to optimize comparability with the 25% target error used for the untransformed data (0.10 log units back-transform to a factor 1.26, which is approximately 25% higher or lower than the consensus value). The general appearance of Z'-score boxplot is similar for the analysis with and without log transformation, but some notable differences occur (Figure 4, right panel). The percentage of $|Z'|<2$ is higher for three laboratories, smaller for six laboratories, and the same for six laboratories. Some large differences in the $\%|Z'|<2$ were observed for some laboratories. A >10% increase was observed for two participants (77 to 92% and 67 to 78%), and a >10% decrease occurred for four participants (100 to 83%, 94 to 62%, 91 to 55%, and 40 to 15%). These differences can greatly impact laboratories when documenting the

analytical quality of their analyses for customers and funding agencies. For the moment it seems best that PT organizers analyze the data using both assumed error distributions and discuss with their clients which distribution yields the most realistic and useful results.

Between-laboratory variability for hydrophilic compounds is high compared with the requirements of environmental monitoring programs, whether expressed as a CV of 50% or an uncertainty factor of 1.6, and efforts by laboratories are needed to reduce this variability. For future rounds of the present PT program a first focus on the variability associated with instrumental analysis can be considered. Distribution of spiked and unspiked sampler extracts would allow the PT organizer to recalculate concentrations that are reported by participants, using the standard addition method, which is a powerful approach for dealing with matrix effects.

3.3. Uncertainties in aqueous concentrations

Uncertainties in sampling kinetics and chemical analysis can both have an impact on the C_w estimate in passive sampler based monitoring. An appreciable scatter of a factor two or more exists in reported sampling rates of individual hydrophilic compounds by POCIS and Chemcatchers, and the effects of flow, temperature and biofouling on R_s are only partly understood.[26–29] Organizers of monitoring programs for hydrophilic compounds should therefore separately evaluate the uncertainties in R_s, and take an additional factor of 1.6 (standard deviation) or 1.6^2 (95% confidence range) into account for the uncertainty of the chemical analysis.

By contrast, modelling of aqueous passive sampling of hydrophobic compounds with silicone and LDPE is well established, and C_w can be reliably estimated, using equations 5 and 6.[24,30,31] Applying the model using the consensus values showed that sampling by silicone and LDPE yielded similar results for R_s (21 ± 2 L/d vs. 24 ± 3 L/d at M_r=300, Figure S16-1) and C_w (ratio of $C_{w,LDPE}/C_{w,silione} = 0.94 \pm 0.26$, Figures S16-2 and S16-3). This indicates that equations 5 and 6 are adequate and that the consensus values are good approximations of the true values.

C_w values were also calculated using the C_s and retained PRC fractions reported by each individual laboratory. These C_w values were then normalized by the C_w values that were based on the consensus

values of C_s and retained PRC fractions ($C_{w,consensus}$). The ratio $C_w/C_{w,consensus}$ can be taken as a measure of accuracy (Figure 5). The ratio of the 2.5% and 97.5% percentiles was taken as an approximation of the width of the 95% confidence interval, which amounts to 2.7 log units for silicone and 3.6 log units for LDPE. This corresponds to a standard deviation of 0.68 log units (silicone, factor 5) and 0.91 log units (LDPE, factor 8), considering that the 95% confidence interval spans four standard deviations. Limiting the dataset to the best performing laboratories (more than 80% |Z'|<2, both for PRCs and native compounds) yields uncertainty factors of 1.6 for silicone (6 laboratories) and 1.9 for LDPE (3 laboratories). Outlying values originate mainly from data for 2- and 3-ring PAHs, which apparently are difficult to analyze. Eliminating these analytes yields uncertainty factors of 1.3 (silicone) and 1.7 (LDPE) for the best performing laboratories. This indicates that uncertainty factors of 1.3 to 1.7 (standard deviation of 0.11 to 0.23 log units) are achievable, but only for a limited number of the laboratories (6/15 for silicone and 3/9 for LDPE). These values are quite similar to the factor of 1.4 that can be attributed to chemical analysis for the study by Jonker et al. on passive sampling in sediments (Section S6).[17]

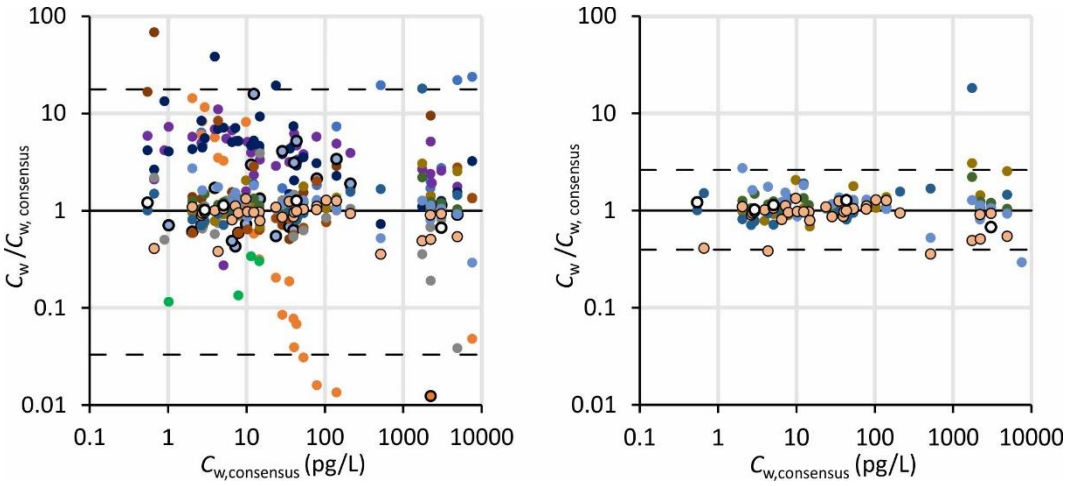


Figure 5. Ratio of silicone-based C_w estimates from data of individual laboratories and C_w estimates from the consensus values ($C_{w,consensus}$) for all laboratories (left panel) and laboratories with >80% |Z'|<2 (right panel). Different symbols represent different laboratories. Horizontal lines represent the reference value of 1, and the 2.5 and 97.5 percentiles. Seven data points for one laboratory are off scale in the left panel. Corresponding plots for individual laboratories, plots for LDPE, and C_w ratios vs. compound name are shown in Section S16.

4. Implications

Realistic estimates of achievable accuracy are best obtained from multi-round PT studies. Results from five rounds for the analysis of hydrophobic compounds in silicone and LDPE show that the consistency of between-laboratory CVs can be improved by taking the median among compounds and (if available) among study rounds, which amounts to 32%. The presently used proportional error of 25% is reasonably close to the multi-round average, and generally gives a fair assessment of laboratory performance for silicone and LDPE samplers. A higher proportional error is defensible for the 2- and 3-ring PAHs. Continuous monitoring of CVs for individual compounds over successive rounds strengthens the robustness of the laboratory performance assessment. Z'-scores for compounds that are included for the first time can be explicitly labeled as tentative. Adopting a normal error distribution is defensible for the data analysis of hydrophobic compounds in silicone and LDPE because the CVs are relatively small, although adopting a log-normal distribution would yield a more realistic assessment for the more volatile PAHs, while giving similar results for the other compounds.

Results for the analysis of hydrophilic compounds in Oasis HLB are characterized by a median CV of 50%. Evidence that lower CVs can be achieved by the best performing laboratories is weak, indicating that the presently adopted target error of 25% is presently not achievable for these compounds. The high CVs for hydrophilic compounds limit the usefulness of the assigned Z'-scores because any reported concentration close to zero results in an acceptable Z'-score. Log transformation prior to statistical analysis can be considered until CVs decrease to values of approximately 30%. Alternatively, organizers of laboratory performance studies can decide to not assign consensus values for compounds with CVs >30%, but this would result in loss of information, which is not in the best interest of participants and their stakeholders. Repeated participation in proficiency testing programs, amended by evaluation of analytical methods, is needed to bring between-laboratory variability in line with scientific and legal requirements.

Supporting information: raw data for round 2021 and summary data for previous rounds (xlsx). Sampler shipment vials, homogeneity test results, compound list, LDPE-water and silicone-water partition

coefficients, CVs of hydrophobic compounds, comparison with other studies, Z'-score boxplots for all compounds and sampler types, laboratory performance versus certification status and use of reference materials, concentration dependence of CVs, methods for confidence intervals of variances, hydrophobic compound CVs (all rounds), variability for 2- and 3-ring PAHs, laboratory performance versus certification and analytical methods for hydrophilic compounds, CV comparison of hydrophilic compounds among laboratory groups with different general performance, effect of assuming a normal versus lognormal distribution of errors, aqueous concentration estimates for hydrophobic compounds (pdf). This information is available free of charge via the Internet at http://pubs.acs.org.

Acknowledgments

We gratefully acknowledge the contribution of the following people to the present study: Henry Beeltje (Netherlands Organisation for Applied Scientific Research / TNO, Netherlands), Gerlinde Roskam (Deltares, Netherlands), Leen Vandekerckhove (Flanders Environment Agency / VMM, Belgium), Tomáš Mičaník (T. G. Masaryk Water Research Institute / VUV, Czech Republic), Mai Thao Nguyen (Waterproef, Netherlands), Tom Gallé (Luxembourg Institute of Science and Technology / LIST), Valérie Bouchart and Céline Passignat (LABÉO, France). The project results were created with the financial support of the Czech Science Foundation within the project 'Holistic exposure and effect potential assessment of complex chemical mixtures in the aquatic environment' No. 20-04676X. The authors thank the RECETOX Research Infrastructure (No. LM2023069) financed by the Ministry of Education, Youth and Sport, and the Operational Programme Research, Development and Education (the CETOCOEN EXCELLENCE project No. CZ.02.1.01/0.0/0.0/17_043/0009632) for supportive background. We thank four anonymous reviewers for their constructive comments.

5. References

(1) US EPA. Guidance on Systematic Planning Using the Data Quality Objectives Process, EPA QA/G-4., 2006. https://www.epa.gov/quality/guidance-systematic-planning-using-data-quality-objectives-process-epa-qag-4 (accessed 2024-03-18).

(2) US EPA. EPA Requirements for Quality Assurance Project Plans, EPA QA/R-5., 2001. https://www.epa.gov/quality/epa-qar-5-epa-requirements-quality-assurance-project-plans (accessed 2024-03-18).

(3) US EPA. Environmental Information Quality Procedure. CIO 2105-P-01.3., 2013. https://www.epa.gov/irmpoli8/environmental-information-quality-procedure (accessed 2024-03-18).

(4) EU. *Guidance on Surface Water Chemical Monitoring under the Water Framework Directive. Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 19.*; Publications Office: Luxembourg, 2009; pp 1–132. http://dx.doi.org/10.2779/72701.

(5) EU. 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. *Off. J. Eur. Union* **2009**, *L210*, 36–38. https://eur-lex.europa.eu/eli/dir/2009/90/oj (accessed 2024-03-18).

(6) Brookman, B.; Mann, I. *Selection, Use and Interpretation of Proficiency Testing (PT) Schemes (3rd Ed. 2021).*; 2021. https://eurachem.org/index.php/publications/guides/usingpt (accessed 2024-03-18).

(7) Lohmann, R.; Muir, D.; Zeng, E. Y.; Bao, L.-J.; Allan, I. J.; Arinaitwe, K.; Booij, K.; Helm, P.; Kaserzon, S.; Mueller, J. F.; Shibata, Y.; Smedes, F.; Tsapakis, M.; Wong, C. S.; You, J. Aquatic Global Passive Sampling (AQUA-GAPS) Revisited: First Steps toward a Network of Networks for Monitoring Organic Contaminants in the Aquatic Environment. *Environ. Sci. Technol.* **2017**, *51* (3), 1060–1067. https://doi.org/10.1021/acs.est.6b05159.

(8) McKay, S.; Tscharke, B.; Hawker, D.; Thompson, K.; O'Brien, J.; Mueller, J. F.; Kaserzon, S. Calibration and Validation of a Microporous Polyethylene Passive Sampler for Quantitative Estimation of Illicit Drug and Pharmaceutical and Personal Care Product (PPCP) Concentrations in Wastewater Influent. *Sci. Total Environ.* **2020**, *704*, 135891. https://doi.org/10.1016/j.scitotenv.2019.135891.

(9) U.S. EPA/SERDP/ESTP. *Laboratory, Field, and Analytical Procedures for Using Passive Sampling in the Evaluation of Contaminated Sediments: User's Manual*; EPA/600/R-16/357; Office of Research and Development: Washington, DC 20460, 2017; p 167. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100R14L.txt (accessed 2024-03-18).

(10) de Boer, J.; Wells, D. E. Chlorobiphenyls and Organochlorine Pesticides in Fish and Sediments - Three Years of QUASIMEME Laboratory Performance Studies. *Mar. Pollut. Bull.* **1997**, *35* (1–6), 52–63. https://doi.org/10.1016/S0025-326X(97)00123-9.

(11) IAEA. *Reference Products for Environment and Trade.* https://analytical-reference-materials.iaea.org/organic (accessed 2024-03-18).

(12) Law, R. J.; Kelly, C. A.; Nicholson, M. D. The QUASIMEME Laboratory Performance Study of Polycyclic Aromatic Hydrocarbons (PAHs): Assessment for the Period 1996-1999. *J. Environ. Monit.* **2000**, *2* (5), 517–523. https://doi.org/10.1039/B004290K.

(13) Van Leeuwen, S. P. J.; van Bavel, B.; de Boer, J. First Worldwide UNEP Interlaboratory Study on Persistent Organic Pollutants (POPs), with Data on Polychlorinated Biphenyls and Organochlorine Pesticides. *TRAC-Trends Anal. Chem.* **2013**, *46*, 110–117. https://doi.org/10.1016/j.trac.2012.12.020.

(14) Ziegler, E.; Boubetra, A.; Tirard, A. Drug Residues in Water. *AWE Int.* **2021**, *October 2021*, 45–53. https://www.bipea.org/scientific-papers/ (accessed 2024-03-18).

(15) Miege, C.; Mazzella, N.; Schiavone, S.; Dabrin, A.; Berho, C.; Ghestem, J. P.; Gonzalez, C.; Gonzalez, J. L.; Lalere, B.; Lardy-Fontan, S.; Lepot, B.; Munaron, D.; Tixier, C.; Togola, A.; Coquery, M. An in Situ Intercomparison Exercise on Passive Samplers for Monitoring Metals, Polycyclic Aromatic

Hydrocarbons and Pesticides in Surface Waters. *TRAC-Trends Anal. Chem.* **2012**, *36*, 128–143. https://doi.org/10.1016/j.trac.2012.01.009.

(16) Vrana, B.; Smedes, F.; Prokeš, R.; Loos, R.; Mazzella, N.; Miege, C.; Budzinski, H.; Vermeirssen, E.; Ocelka, T.; Gravell, A.; Kaserzon, S. An Interlaboratory Study on Passive Sampling of Emerging Water Pollutants. *TrAC Trends Anal. Chem.* **2016**, *76*, 153–165. https://doi.org/10.1016/j.trac.2015.10.013.

(17) Jonker, M. T. O.; van der Heijden, S. A.; Adelman, D.; Apell, J. N.; Burgess, R. M.; Choi, Y.; Fernandez, L. A.; Flavetta, G. M.; Ghosh, U.; Gschwend, P. M.; Hale, S. E.; Jalalizadeh, M.; Khairy, M.; Lampi, M. A.; Lao, W.; Lohmann, R.; Lydy, M. J.; Maruya, K. A.; Nutile, S. A.; Oen, A. M. P.; Rakowska, M. I.; Reible, D.; Rusina, T. P.; Smedes, F.; Wu, Y. Advancing the Use of Passive Sampling in Risk Assessment and Management of Sediments Contaminated with Hydrophobic Organic Chemicals: Results of an International Ex Situ Passive Sampling Interlaboratory Comparison. *Environ. Sci. Technol.* **2018**, *52* (6), 3574–3582. https://doi.org/10.1021/acs.est.7b05752.

(18) Lotufo, G. R.; Michalsen, M. M.; Reible, D. D.; Gschwend, P. M.; Ghosh, U.; Kennedy, A. J.; Kerns, K. M.; Rakowska, M. I.; Odetayo, A.; MacFarlane, J. K.; Yan, S.; Bokare, M. Interlaboratory Study of Polyethylene and Polydimethylsiloxane Polymeric Samplers for Ex Situ Measurement of Freely Dissolved Hydrophobic Organic Compounds in Sediment Porewater. *Environ. Toxicol. Chem.* **2022**, *41* (8), 1885–1902. https://doi.org/10.1002/etc.5356.

(19) Booij, K.; Smedes, F.; Crum, S. Laboratory Performance Study for Passive Sampling of Nonpolar Chemicals in Water. *Environ. Toxicol. Chem.* **2017**, *36* (5), 1156–1161. https://doi.org/10.1002/etc.3657.

(20) Smedes, F.; Booij, K. *Guidelines for Passive Sampling of Hydrophobic Contaminants in Water Using Silicone Rubber Samplers.*; ICES Techniques in Marine Environmental Sciences 52.; International Council for the Exploration of the Sea: Copenhagen, 2012; pp 1–20. https://dx.doi.org/10.17895/ices.pub.5077.

(21) Cofino, W. P.; van Stokkum, I. H. M.; van Steenwijk, J.; Wells, D. E. A New Model for the Inference of Population Characteristics from Experimental Data Using Uncertainties: Part II. Application to

Censored Datasets. *Anal. Chim. Acta* **2005**, *533* (1), 31–39. https://doi.org/10.1016/j.aca.2004.11.008.

(22) Cofino, W. P.; van Stokkum, I. H. M.; Wells, D. E.; Ariese, F.; Wegener, J. W. M.; Peerboom, R. A. L. A New Model for the Inference of Population Characteristics from Experimental Data Using Uncertainties. Application to Interlaboratory Studies. *Chemom. Intell. Lab. Syst.* **2000**, *53* (1–2), 37–55. https://doi.org/10.1016/S0169-7439(00)00093-9.

(23) Molenaar, J.; Cofino, W. P.; Torfs, P. J. J. F. Efficient and Robust Analysis of Interlaboratory Studies. *Chemom. Intell. Lab. Syst.* **2018**, *175*, 65–73. https://doi.org/10.1016/j.chemolab.2018.01.003.

(24) Rusina, T. P.; Smedes, F.; Koblizkova, M.; Klanova, J. Calibration of Silicone Rubber Passive Samplers: Experimental and Modeled Relations between Sampling Rate and Compound Properties. *Environ. Sci. Technol.* **2010**, *44* (1), 362–367. https://doi.org/10.1021/es900938r.

(25) Vrana, B.; Smedes, F.; Prokes, R.; Loos, R.; Mazzella, N.; Miege, C.; Budzinski, H.; Vermeirssen, E.; Ocelka, T.; Gravell, A.; Kaserzon, S. *NORMAN Interlaboratory Study (ILS) on Passive Sampling of Emerging Pollutants: A Chemical Monitoring on Site (CM Onsite) Organised by NORMAN Association and JRC in Support of the Water Framework Directive.*; EUR 27655 EN.; European Union Publications Office, 2016; p 239. https://dx.doi.org/10.2788/6757.

(26) Charriau, A.; Lissalde, S.; Poulier, G.; Mazzella, N.; Buzier, R.; Guibaud, G. Overview of the Chemcatcher® for the Passive Sampling of Various Pollutants in Aquatic Environments Part A: Principles, Calibration, Preparation and Analysis of the Sampler. *Talanta* **2016**, *148*, 556–571. https://doi.org/10.1016/j.talanta.2015.06.064.

(27) Glanzmann, V.; Reymond, N.; Weyermann, C.; Estoppey, N. An Improved Chemcatcher-Based Method for the Integrative Passive Sampling of 44 Hydrophilic Micropollutants in Surface Water – Part A: Calibration under Four Controlled Hydrodynamic Conditions. *Sci. Total Environ.* **2023**, *871*, 162037. https://doi.org/10.1016/j.scitotenv.2023.162037.

(28) Guibal, R.; Lissalde, S.; Guibaud, G. Experimental Estimation of 44 Pharmaceutical Polar Organic Chemical Integrative Sampler Sampling Rates in an Artificial River under Various Flow Conditions. *Environ. Toxicol. Chem.* **2020**, *39* (6), 1186–1195. https://doi.org/10.1002/etc.4717.

(29) Harman, C.; Allan, I. J.; Vermeirssen, E. L. M. Calibration and Use of the Polar Organic Chemical Integrative Sampler - a Critical Review. *Environ. Toxicol. Chem.* **2012**, *31* (12), 2724–2738. https://doi.org/10.1002/etc.2011.

(30) Booij, K. Passive Sampler Exchange Kinetics in Large and Small Water Volumes under Mixed Rate Control by Sorbent and Water Boundary Layer. *Environ. Toxicol. Chem.* **2021**, *40* (5), 1241–1254. https://doi.org/10.1002/etc.4989.

(31) Booij, K.; Smedes, F. An Improved Method for Estimating in Situ Sampling Rates of Nonpolar Passive Samplers. *Environ. Sci. Technol.* **2010**, *44* (17), 6789–6794. https://doi.org/10.1021/es101321v.