2018

Advancing the use of passive sampling in risk assessment and management of contaminated sediments: Results of an international passive sampling inter-laboratory comparison

Michael T. O. Jonker

Stephan A. van der Heijden

See next page for additional authors

Follow this and additional works at: https://digitalcommons.uri.edu/gsofacpubs

Authors

Michael T. O. Jonker, Stephan A. van der Heijden, Dave Adelman, Jennifer N. Apell, Robert M. Burgess, Yongju Choi, Loretta A. Fernandez, Geanna M. Flavetta, Upal Ghosh, Philip M. Gschwend, Sarah E. Hale, Mehregan Jalalizadeh, Mohammed Khairy, Mark A. Lampi, Wenjian Lao, Rainer Lohmann, Michael J. Lydy, Keith A. Maruya, Samuel A. Nutile, Amy M. P. Oen, Magdalena I. Rakowska, Danny Reible, Tatsiana P. Rusina, Foppe Smedes, and Yanwen Wu

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 inter-laboratory comparison

Michiel T.O. Jonker[1*], Stephan A. van der Heijden[1], Dave Adelman[2], Jennifer N. Apell[3], Robert M. Burgess[4], Yongju Choi[5,6], Loretta A. Fernandez[7], Geanna M. Flavetta[7], Upal Ghosh[8], Philip M. Gschwend[3], Sarah E. Hale[9], Mehregan Jalalizadeh[8], Mohammed Khairy[2,10], Mark A. Lampi[11] Wenjian Lao[12], Rainer Lohmann[2], Michael J. Lydy[13], Keith A. Maruya[12], Samuel A. Nutile[13], Amy M.P. Oen[9], Magdalena I. Rakowska[14], Danny Reible[14], Tatsiana P. Rusina[15], Foppe Smedes[15,16], and Yanwen Wu[5]

[1]Institute for Risk Assessment Sciences, Utrecht University; Yalelaan 104, 3584 CM Utrecht, the Netherlands
[2]Graduate School of Oceanography, University of Rhode Island, South Ferry Road, URI Bay Campus, Narragansett, RI, 02882, USA
[3]RM Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[4]Atlantic Ecology Division, Office of Research and Development, U.S. Environmental Protection Agency, Narragansett, RI, 02882, USA
[5]Department of Civil and Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA, 94305, USA
[6]Department of Civil and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
[7]Department of Civil and Environmental Engineering, Northeastern University, 360 Huntington Avenue, Boston, MA, 02115, USA
[8]Department of Chemical, Biochemical, and Environmental Engineering, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD, 21250, USA
[9]Norwegian Geotechnical Institute. Environmental Technology. Sognsveien 72, 0806 Oslo, Norway
[10]Department of Environmental Sciences, Faculty of Science, Alexandria University, 21511 Moharam Bek, Alexandria, Egypt
[11]ExxonMobil Biomedical Sciences, Inc., 1545 US 22 East, Annandale, NJ, 08822, USA
[12]Southern California Coastal Water Research Project Authority. 3535 Harbor Blvd. Suite 110, Costa Mesa, CA, 92626, USA
[13]Center for Fisheries, Aquaculture and Aquatic Sciences, and Department of Zoology, Southern Illinois University, Carbondale, IL, 62901, USA
[14]Civil, Environmental and Construction Engineering, Texas Tech University, Box 41023, Lubbock, TX, 79409-1023, USA
[15] Masaryk University, Faculty of Science, Research Centre for Toxic Compounds in the Environment (RECETOX), Kamenice 753/5, 62500, Brno, Czech Republic
[16] Deltares, P.O. Box 85467, 3508 AL Utrecht, the Netherlands

* Corresponding author e-mail: m.t.o.jonker@uu.nl; phone: +31 30 2535338.

48   **Abstract**

49   This paper presents the results of an international inter-laboratory comparison on ex situ

50   passive sampling in sediments. The main objectives were to map the state of the science in

51   passively sampling sediments, identify sources of variability, provide recommendations and

52   practical guidance for standardized passive sampling, and advance the use of passive

53   sampling in regulatory decision making by increasing confidence in the use of technique.

54   The study was performed by a consortium of 11 laboratories and included experiments with

55   14 passive sampling formats on 3 sediments for 25 target chemicals (PAHs and PCBs). The

56   resulting overall inter-laboratory variability was large (a factor of ~10), but standardization

57   of methods halved this variability. The remaining variability was primarily due to factors not

58   related to passive sampling itself, i.e., sediment heterogeneity and analytical chemistry.

59   Excluding the latter source of variability, by performing all analyses in one laboratory,

60   showed that passive sampling results can have a high precision and a very low inter-method

61   variability (< factor of 1.7). It is concluded that passive sampling, irrespective of the specific

62   method used, is fit for implementation in risk assessment and management of contaminated

63   sediments, provided that method setup and performance, as well as chemical analyses are

64   quality-controlled.

Introduction

Traditional methods for assessing risks and managing contaminated sediments are based on total, solvent-extractable concentrations of sediment-associated chemicals, following normalization to the sediment organic carbon content.[1] Within the environmental scientific community it is generally accepted that this approach does not lead to a realistic assessment of actual risks at field-contaminated sites.[2] Therefore, several methods for estimating the 'bioavailable' concentration or fraction of a chemical have been developed during the past decades. These methods aim to determine the concentration or fraction that is available for causing ecotoxicological effects and more closely reflects actual or potential exposure. Among these methods, partitioning-based, non-depletive extractions with polymers (colloquially referred to as "passive sampling methods", even though often active mixing of the polymer and the sediment is applied) are considered the best developed and have the most solid scientific basis.[3] Through passive sampling, the freely dissolved concentration (C_{free}) of a chemical in sediment pore water is determined, which is a good metric of the driving force behind accumulation and toxicological effects in organisms.[4] The technique involves direct exposure of a polymer phase to sediment, either *in situ* or *ex situ*. Hydrophobic organic chemicals present in the sediment system partition into the polymer and the resulting polymer-sorbed equilibrium concentration is used to calculate C_{free}. Several different polymers have been applied as a sampling phase, including polydimethylsiloxane (PDMS), polyethylene (PE), polyoxymethylene (POM), polyacrylate (PAc), and silicone rubber (SR), with the polymers being available in different formats.[5]

Despite the multitude of sampler formats and application possibilities, passive sampling is currently primarily used for scientific purposes and as an indicator of sediment remediation performance, rather than to design sediment management approaches. Acceptance in the risk assessment and regulatory community has been slow, among other reasons because so many

90   different types of passive samplers are applied and the lack of standardized methods. There

91   is a perception outside the scientific community that no scientific consensus exists on which

92   is the best method to use.[2] Although guidelines for selection of specific polymers have been

93   proposed,[5] and the application of different passive samplers and (calculation and analysis)

94   methods should theoretically yield identical C_{free} values, it is currently unknown if this

95   actually holds true and the inter-method variability has not yet been adequately quantified.

96   This information is crucial, however, when implementing passive sampling in risk

97   assessment practices for contaminated sediments, as recently suggested by the US EPA.[6]

98   In November 2012, a SETAC workshop on passive sampling in sediments was held in Costa

99   Mesa (CA, USA), with the goal of advancing the application of passive sampling in the risk

100 assessment and management of contaminated sediments.[2] During the workshop, several

101 research needs and bottlenecks for implementation were identified, including the above-

102 mentioned issue and the necessity for a round-robin inter-laboratory study, standardization

103 of methods, and characterization of sources of uncertainty.[2,5] In response, an international

104 inter-laboratory comparison study was initiated, with the main objectives to: (i) map the

105 state of the science in *ex situ* passive sampling in sediments, and the inter-laboratory and

106 inter-method variability in C_{free} determinations; (ii) identify the sources of variability in C_{free}

107 as determined with passive sampling; (iii) propose measures to reduce variability and to

108 provide practical guidance (standardized methods); and, (iv) increase the overall confidence

109 in passive sampling to advance its use outside the scientific domain, i.e., in regulatory

110 decision making. The results of the inter-laboratory comparison are presented in this paper.

111 Practical guidance (i.e., a proposed standard method) and polymer-water partition

112 coefficients (K_{pw}) needed to calculate C_{free} will be presented in separate, forthcoming papers.

113

114 **Study Design**

115   Eleven research laboratories from four different countries (USA, The Netherlands, Norway,

116   and the Czech Republic) participated in the study. The Utrecht University laboratory acted

117   as coordinating laboratory. Each participating laboratory had a proven track record in

118   passive sampling in sediments and contributed to the study by applying their own passive

119   sampling procedures (i.e., format, experimental setup), previously published in the peer-

120   reviewed literature. In total, 14 passive sampling formats were included, which differed in

121   polymer material, source, form (i.e., polymer sheet vs. coating on a glass (SPME) fiber), or

122   thickness. Five of the 11 laboratories applied multiple formats. Passive sampling

123   experiments were performed with three sediments, including two field-contaminated

124   sediments and one unpolluted sediment that was spiked in the coordinating laboratory.

125   Target chemicals included 12 polychlorinated biphenyls (PCBs) and 13 polycyclic aromatic

126   hydrocarbons (PAHs). C_{free} values of these chemicals were determined in five-fold for each

127   sediment in the following set of tiered experiments. In the first experiment, each laboratory

128   followed its own procedure(s). The resulting C_{free} values were reported to the coordinating

129   laboratory, along with the K_{pw} values used in the calculations and a description of the

130   methods applied. This experiment mapped the overall variability in passive sampling

131   methods. In the second experiment, participants were asked to redo the measurements, but to

132   strictly apply a 'standard' protocol that was dictated by the coordinating laboratory. This

133   experiment was performed in duplicate: one set of sample extracts was analyzed by the

134   respective participant, to quantify the contribution of employing different protocols to the

135   overall variability; the other set was shipped to, and analyzed by the coordinating laboratory,

136   in order to evaluate the contribution of analytical chemistry to the overall variability. All

137   participants were also provided with a standard solution of the target chemicals, of which the

138   reported concentrations yielded a direct measure of the analytical (instrumental calibration)

139   variability. In the third experiment, the coordinating laboratory applied the 'standard'

140   protocol to all 14 passive sampling formats (as shared by the participants) in order to

141   identify the inter-method variability. Finally, supplementary tests were performed to map

142   any additional sources of variation in C_{free}, including polymer mass determination, sediment

143   heterogeneity, and sediment storage time.

144

145   ## Materials and Methods

146   **Passive Samplers**

147   An overview of the applied passive samplers (polymer types, thicknesses, suppliers) is given

148   in the Supporting Information (Table S1).

149

150   **Target Chemicals**

151   Target chemicals were the PAHs phenanthrene, anthracene, fluoranthene, pyrene,

152   benz[*a*]anthracene, chrysene, benzo[*e*]pyrene, benzo[*b*]fluoranthene, benzo[*k*]fluoranthene,

153   benzo[*a*]pyrene, benzo[*ghi*]perylene, dibenz[*ah*]anthracene, and indeno[123,*cd*]pyrene; and

154   PCB congeners 18, 28, 52, 66, 77, 101, 118, 138, 153, 170, 180, and 187.

155

156   **Analytical Standard Solution**

157   A standard solution was prepared for each participant, by adding 50 µL of an acetone spike

158   containing the 25 target chemicals to 950 µL of the participant-specific injection solvent

159   applied during chemical analyses by the respective laboratory (either *n*-hexane, *n*-heptane, *n*-

160   hexane/acetone (1:1), dichloromethane, or acetonitrile). Nominal concentrations (not shared

161   with the participants) were about 50 µg/L for PCBs and 100 µg/L for PAHs.

162

163   **Sediments**

164   The three testing sediments differed in degree of complexity by passive sampling

165   application. The 'least complex' sediment ("SP sediment") was an unpolluted, sandy

166   sediment, sampled from the small river 'Kromme Rijn', near Werkhoven, the Netherlands. It

167   was sieved through a 1 mm sieve, yielding a 20-kg dry weight (dw) sample, which was

168   intensively mixed for 30 min with a mechanical mixer. Ten 2 kg (dw) portions of the

169   sediment were successively spiked in 5 L glass beakers with relatively high levels of the

170   target chemicals, by adding drop-wise 4 mL of an acetone solution containing the target

171   chemicals (PAHs at ~250 mg/L each; PCBs at ~150 mg/L each), while intensively

172   mechanically stirring (30 min). All portions were finally pooled in a 110 L concrete mixer,

173   which subsequently mixed this spiked (SP) sediment continuously for 4.5 weeks. The

174   sediment of 'intermediate complexity' ("BB sediment") originated from the "Biesbosch", a

175   Dutch sedimentation area. This sediment has been used in a previous study in outdoor

176   ditches,[7] and the sediment used in the present study was sampled from that site. It contained

177   relatively low native concentrations of the target chemicals, but was known to be

178   homogeneous. Therefore, it was mixed in a concrete mixer for a shorter period of time, i.e.,

179   1.5 week. The most complex sediment ("FD sediment") was a sediment composed by

180   combining (2:1) a French and a Dutch sediment. The French sediment was sampled from the

181   river Tillet (Aix les Bains, Savoie), was very sandy, and contained hardly any PAHs. PCBs

182   were however present at high concentrations, and originated from a former electric

183   transformer manufacturing facility 2 km upstream. The Dutch sediment was sampled from

184   the river Hollandsche IJssel and has been previously studied.[8] It contained no detectable

185   PCBs, but PAHs were present at intermediate concentrations, mostly originating from an

186   upstream diesel-powered water pumping station. This sediment also contained non-aqueous

187   phase liquids (NAPLs). The composited sediment was mixed in a concrete mixer for 4

188   weeks nonstop. Before mixing, a quantity of the biocide sodium azide (NaN_3) was added to

189 all sediments, producing a concentration of 100 mg/L water. After mixing, the sediments

190 were divided among amber-colored glass jars in portions sufficient to meet each

191 participant's requirement to complete the tests (different procedures by different participants

192 required different sediment masses). All jars were closed with aluminum foil-lined lids and

193 shipped in cooled containers to the participants, along with the standard solution and coded

194 autosampler vials and glassware for the standardized experiments. Dry weight and organic

195 carbon content, as well as total concentrations of the target chemicals in the sediments were

196 determined by the coordinating laboratory as previously described.[9] The results are provided

197 in Table S2. This information was shared with the participants before initiating the

198 experiments.

199

200 **Determination of C_{free} based on the Participants' Own Procedures**

201 In this first experiment, all participants performed C_{free} determinations according to their

202 own procedure(s) and analyzed the resulting extracts themselves. Each measurement was

203 performed five-fold. A summary of the materials used and methods applied by all 11

204 participants is (anonymously) listed in Table S3. Procedures clearly differed in terms of type

205 of exposure (i.e., static vs. dynamic), exposure duration, verification of equilibrium

206 conditions (i.e., use of performance reference compounds (PRCs), multiple sampler

207 thicknesses, or multiple time points), sampler mass, sampler/sediment/water ratio, washing

208 and extraction of samplers, and solvents used.

209

210 **Determination of C_{free} based on Standardized Procedures**

211 After completing the above experiment, participants received a standardized protocol and

212 were asked to repeat the five-fold C_{free} determinations, strictly adhering to the prescribed

213 procedure. The protocol was method-, sediment-, and participant-specific, but all aspects and

8

214 steps (except the chemical analysis) were standardized, including sampler/sediment and

215 sediment/water ratio, sampler washing, glassware, composition of the added water, exposure

216 duration (6 weeks), method of shaking and shaking speed, and sampler cleaning and

217 extraction procedures after finishing the exposures. The sampler/sediment ratio was

218 dependent on the sediment properties and the polymer used, and the sampler washing and

219 extraction procedures were specific for different polymers. Furthermore, the sampler

220 extraction was tuned to the solvent used during chemical analysis by the particular

221 participant. A general description of the standardized protocol is presented on pages S20-

222 S21 of the Supporting Information. As outlined under 'Study Design', this experiment was

223 performed in duplicate. One set of extracts was analyzed by the participant, the other set was

224 shipped in a cooled container to the coordinating laboratory, where internal standards were

225 added and the extracts analyzed. The standardized protocol was also applied by the

226 coordinating laboratory to all 14 sampler types (as provided by the participants), in order to

227 quantify the inter-method variability.

228

229 **Supplementary Tests**

230 Supplementary tests focusing on additional sources of variation in C_{free} (polymer mass

231 determination, sediment heterogeneity, and sediment storage time) are described in the

232 Supporting Information.

233

234 **Chemical and Data analysis**

235 Target chemicals were analyzed by the participants as described in Table S3. GC-MS or GC-

236 ECD was used for PCB quantification, whereas PAHs were analyzed by either GC-MS or

237 HPLC-FLD. Concentrations in the sampler extracts were converted to concentrations in the

238 sampler material (C_s), using the sampler's mass (sheets) or polymer volume (SPME fibers).

239   C_{free} was then obtained by dividing C_s by a polymer- and chemical-specific K_{pw}. In the first

240   experiment (participants' own procedures), participants applied their own K_{pw}s (measured

241   themselves or taken from the literature) and some used PRCs in their calculations. In the

242   standardized experiment, a fixed set of K_{pw} values as measured by the coordinating

243   laboratory according to previously-published methods[10] was applied. Variability in each

244   experiment was quantified by averaging the five-fold C_{free} measurements of each participant

245   and subsequently calculating a variation factor (*VF*) for each target chemical. This factor

246   was calculated by assuming the experimentally-determined concentrations exhibited normal

247   distributions and then taking the ratio of the 95th percentile (*PCTL*) value of the averaged

248   C_{free} values per target chemical, to the 5th percentile value:

249

250
$$VF = \frac{95th\ PCTL}{5th\ PCTL}$$

251

252   Using this statistic, the range in C_{free} was quantified and expressed intuitively as a factor,

253   while excluding outliers. In order to compare experiments and sediments in a simple way,

254   the chemical-specific *VF* values were averaged per sediment for each experiment (VF_{av}).

255

256   Results and Discussion

257   State of the Science in Passive Sampling Sediment Pore Water

258   The results of the first experiment, in which all participants performed C_{free} determinations

259   according to their own procedures, are presented in panels A1-3 of Figure 1. In these three

260   figures (one for each sediment), the averaged C_{free} data for all target chemicals are plotted

261   against C_{free} values obtained by averaging all chemical-specific data produced by the

262   coordinating laboratory (referred to as Lab UU in Figure 1; all passive sampling formats;

263   standardized protocol. Note that using these values as independent variables does not imply

264 they are the target or actual values; they are solely used as reference). This way, the data are

265 presented in a straightforward and understandable manner, without any data manipulation,

266 yet clearly demonstrating the data variability. Additionally, in Figure S1, box plots are

267 presented per chemical. Nearly all data points fall within the 10:1 and 1:10 interval, but there

268 is a clear tendency towards under-predicting the averaged data of the coordinating

269 laboratory. Overall, the observed inter-laboratory variation is quite large; larger than the

270 variability reported for a previous small-scale inter-laboratory passive sampling comparison

271 (i.e., a factor of 2).[11] Note, however, that in ref *11* fewer samplers and target analytes (3 and

272 8, respectively) were tested, using a single sediment. Figure 1 may be also somewhat

273 misleading as the apparent concentration ranges in some cases seem to cover a factor of 100,

274 whereas they are actually composed of data for more than one chemical. The largest

275 variation in the present study was observed for PCB-77 in the BB and FD sediments, where

276 the concentration ranges did indeed span a factor of 100 and even 2400, respectively (see

277 Figure S1 and Table S4, in which ranges for all chemicals are presented). The cause for the

278 deviating behavior of this particular chemical is as follows. PCB-77 was a target chemical,

279 which was added to the SP sediment, but it was not present at detectable concentrations in

280 the field-contaminated BB and FD sediments (Table S2), as revealed by dedicated GC-MS

281 analyses at the coordinating laboratory. Nevertheless, several participants reported

282 considerable C_{free} values for the chemical in these sediments. The large concentration ranges

283 observed can thus be explained by the different detection (MS; ECD) and separation (GC

284 columns) approaches applied by different participants, which will have resulted in

285 inconsistencies in interfering/mis-identified peaks. Because the coordinating laboratory did

286 not report a value, the C_{free} of PCB-77 as reported by one of the participants was used as x-

287 axis value in Figure 1, in order to be able to visualize the variability of this chemical. Since

288 the data for PCB-77 in the BB and FD sediments obscure the average variability, they were

289   excluded from the data analysis when calculating VF_{av} values. These VF_{av} values are listed

290   in Table 1. Values for the first experiment are 9.7±4.1, 9.4±6.3, and 10.8±4.5 for the BB,

291   FD, and SP sediment, respectively. Apparently, when omitting the PCB-77 data, there are no

292   obvious differences in variability among the three sediments, even though they were

293   selected/composed based on differences in complexity for passive sampling. This may imply

294   that passive sampling produces results, which are independent of the type of sediment

295   studied, but it should be stressed that the overall variability is so large that any subtle

296   differences between results for the various sediments may be fully obscured.

297   Note that the variation observed in Figures 1 A1-3 includes variability as introduced by: (i)

298   different laboratories, applying different protocols carried out by different people (inter-

299   laboratory variability), (ii) the use of different K_{pw} values by different participants, (iii)

300   different ways of analyzing the chemicals, (iv) potential sediment heterogeneity and

301   contaminant instability; and, (v) the use of different passive sampling approaches (inter-

302   method variability). The contribution of each of these sources will be discussed in a semi-

303   quantitative manner in the subsequent sections.

304

305   **Impact of Standardizing K_{pw} values**

306   Since most of the measurements performed by the participants involved equilibrium passive

307   sampling, and inaccuracies in the K_{pw} of target analytes under equilibrium conditions are

308   considered "a major source of concern",[12] one would expect a clear contribution to reducing

309   the overall variability by standardizing the K_{pw}s used for calculating C_{free} values. After all,

310   the participants applied K_{pw} values measured in their own laboratory or values taken from

311   the literature. As such, there were considerable differences between the values that were

312   used. For PDMS, the largest difference between the lowest and the highest chemical-specific

313   K_{pw} values was a factor of 7, whereas for PE and POM this was even a factor of 13 and 20,

314 respectively. The impact of standardizing K_{pw}s was investigated by using K_{pw} values that had

315 been determined for each sampler/chemical by the coordinating laboratory (manuscript in

316 preparation). Remarkably, the impact of using K_{pw} values from a single source on the overall

317 variability was negligible, as shown in Figure S2. The VF_{av} values did not significantly

318 change after recalculating the C_{free} data as reported by the participants, using K_{pw} values

319 from the single source (see Table 1). The position of the data points, however, did change in

320 many cases, which makes sense, as K_{pw} determines the absolute value for C_{free}. In other

321 words, standardizing K_{pw}s does not reduce the variability of C_{free} measurements, but still is

322 of utmost importance, because of the final accuracy of C_{free} data. Using inaccurate K_{pw}s will

323 yield biased C_{free} data, which is an unwanted situation when applying passive sampling for

324 assessing risks of contaminated sediments. Therefore, it is recommended that high-quality,

325 accurate (standardized) K_{pw} values be used by the passive sampling community.[5,13]

326

Impact of Standardizing Experimental Protocols

328 Standardizing the experimental protocols, in addition to the K_{pw} values, had a clear impact

329 on the C_{free} inter-laboratory variability. Figures 1 B1-3, Figure S1, and Table 1 demonstrate

330 that the variability roughly halved, with the VF_{av} values being reduced to between 4 and 5

331 for all tested sediments. This obviously implies that the methodology of passive sampling

332 measurements influences the outcomes and that standardization of passive sampling

333 methods is definitely desirable. Because multiple issues and steps were standardized in the

334 protocols, it is not possible to attribute the variation reductions to a specific aspect of the

335 protocols; there are several likely candidates. The most important aspects that were

336 standardized (thus changed for certain participants) included the sampler/sediment and

337 sediment/water ratios, sampler washing procedure, glassware used, exposure duration,

338 method of shaking and shaking speed, and the sampler cleaning and extraction procedures

after finishing the exposures. Smedes et al.[14] showed that the sampler/sediment ratio may influence the equilibrium concentration in the sampler (and thereby the calculated C_{free}), as it was observed to be inversely related to this metric, due to depletion of the system. Also for the presently-investigated sediments a similar relationship was observed (see Figure S3). In the standardized protocol, the ratio was set such that chemical depletion from the three sediments was always below 2% for all chemicals and samplers.[14] However, when performing the measurements according to their own procedure(s), some participants applied (much) higher ratios, which will have resulted in higher depletion ratios (theoretically up to about 70%). Therefore, standardization of this step most probably will have contributed to the variation reduction. Likewise, Smedes et al.[14] demonstrated that the sediment/water ratio can affect the system's kinetics. Higher ratios yielded faster equilibration. Optimization of this ratio, together with a sufficient equilibration time and shaking regime, assured (near) equilibrium in all cases during the standardized experiment, as illustrated in Figure S4. In the first experiment in which the participants followed their own procedures, several participants (presumably) did not achieve equilibrium for all chemicals. PRCs were used to correct for this in several cases, following different calculation approaches, but such a correction may introduce uncertainties and inaccuracies.[15-16] This particularly applies to the more hydrophobic chemicals, for which the correction by some participants was based on extrapolation from released fractions of less hydrophobic PRCs only. It should be stressed though that correction for the degree of non-equilibrium based on PRCs does not necessarily introduce substantial error, as demonstrated by the experiments from one participant (Figure S5). Whereas the standardized protocol prescribed thorough mixing and no PRCs, the procedure of this particular participant involved static exposures and included PRC corrections. Results of both approaches agreed within a factor of about 2 for all chemicals and sediments.

364 Standardization of some of the other aspects of the protocols may also have contributed to

365 the variability reduction, but their contributions are probably less substantial. Sampler

366 extraction after the exposures to sediments may be an exception, as the specific solvent used

367 or the handling of samplers/extracts (e.g. cleanup or evaporation steps) by participants may

368 have introduced variability through, for instance, variable extraction recoveries or losses of

369 contaminants.

370

371 **Contribution of Analytical Chemistry to the Variability**

372 Even after standardizing K_{pw} values and experimental protocols, considerable variability in

373 the inter-laboratory C_{free} data remained (Figures 1 B1 to B3). This variability again roughly

374 halved when all passive sampling extracts were analyzed by one laboratory (see Figures 1

375 C1 to C3 and Figure S1). The VF_{av} decreased to about 2.5 for all three sediments (Table 1).

376 As such, chemical analyses had a substantial contribution to the overall variability. A similar

377 conclusion was also drawn for other inter-laboratory comparison studies on passive

378 sampling in surface waters,[17-18] but certainly is not restricted to passive sampling

379 measurements. Each experiment involving chemical analyses will suffer from errors

380 introduced through inaccuracies in the identification, integration, and calibration of

381 compounds. The case of PCB-77, as discussed above, already demonstrated that

382 identification is the first crucial step and, if not performed correctly, can result in huge inter-

383 laboratory variability. Peak area quantification generally may not be considered as the step

384 that contributes most to the overall variability introduced through chemical analysis.

385 However, in complex chromatograms with co-eluting peaks, baseline selection is subjective

386 and poor integrations may add from a few percent of error to perhaps a factor of two or

387 more. Any error will strongly depend on the sediment, the chemical, the analytical

388 separation power, the selectivity of identification, the integration approach (i.e.,

389   quantification based on peak area or height), and the efficacy of any clean-up procedure. The

390   major source of error introduced by analytical chemistry is likely related to calibration.

391   Apart from correct application of internal standards, final concentrations quantified in the

392   analyzed extracts are directly related to the accuracy of calibration standards. Even for PAHs

393   and PCBs, i.e., compounds that are often routinely analyzed, this accuracy may be

394   insufficient. The analysis of the standard solution in the present study demonstrated that the

395   variation in PCB concentrations was characterized by a VF of 2 to 3, while for PAHs it was

396   3 to 4.5 (see Figure 2). From Figure 2 and the difference between Figures 1B and 1C, it can

397   thus be concluded that a major part of the present inter-laboratory variability in C_{free} data

398   originates from a step that basically has nothing to do with passive sampling measurements,

399   but is part of every experiment involving the measure of chemicals, and is often overlooked

400   as a source of error in experimental results. Therefore, including a standard solution in inter-

401   laboratory comparison studies involving chemical quantification is a prerequisite.

402

403   **Other Sources of Variability**

404   Figure 1 C shows the variability in the results of experiments that were standardized and of

405   which the extracts were analysed by one laboratory. The observed variability will therefore

406   only be caused by (i) inter-method variability, which will be discussed below, (ii) variability

407   in the performance of the standardized procedure by different laboratories, and (iii) other

408   sources of variability. Two other sources of variability were investigated in the present

409   study: the accuracy of sampler mass and fiber coating volume (i.e., analytical weighing and

410   the use of nominal fiber coating thicknesses), and sediment heterogeneity (originating from

411   insufficient mixing and different storage times). Generally, sheet samplers are weighed on a

412   balance and the concentrations quantified in polymers are expressed on a sampler mass

413   basis. Hence, an inaccurate balance or weighing procedure may introduce error and

414 consequently increase data variability. The results of the weighing test, however

415 demonstrated that sampler weights generally were within 1% of the weights recorded by the

416 coordinating laboratory. Only one participant reported weights deviating up to 4.7%. These

417 differences are small and, consequently, weighing did not contribute significantly to the

418 experimental variability in the present study.

419 When deriving the coating volume of a SPME fiber, product specifications provided by the

420 manufacturer are rarely questioned, although it often remains unsure how these were

421 established. A comparison of coating volumes calculated based on nominal, manufacturer-

422 provided thicknesses versus measured ones (Table S5) demonstrated considerable

423 differences, which amounted up to 16%. As such, fiber coating volumes may be a potential

424 source of variability in C_{free}. However, two of the fibers showing the largest deviations (S30-

425 1 and PAc) were used by the coordinating laboratory only, which used actual volumes

426 throughout the different experiments. Therefore, in the present study, the use of nominal

427 coating volumes may only have been a potential source of variability for the S10-1 fiber,

428 albeit not in the experiments where the chemical analyses (and subsequent calculations)

429 were performed by the coordinating laboratory.

430 Results of the sediment heterogeneity experiment showed that even after mixing for several

431 weeks, sediment heterogeneity may also have contributed to the observed overall variability

432 in C_{free}. VF_{av} values of 1.1 to 1.4 for the field-contaminated BB and FD sediments and 1.2 to

433 even 2.4 for the spiked SP sediment were calculated (see Table S6). The VF values are

434 rather chemical-independent for the BB and FD sediments, but for the SP sediment, they

435 increase with chemical hydrophobicity (see Table S6). Apparently, mixing this spiked

436 sediment for up to 4.5 weeks in a concrete mixer was insufficient to allow full chemical

437 homogenization for the most hydrophobic compounds. Note that the results presented here

438 were obtained by analyzing several sediment sub samples (n=10) taken directly from the

17

439   concrete mixer. These samples do not necessarily perfectly represent the sediment samples

440   as received by the participants, considering the large sediment volume that was contained in

441   the mixers. After filling all the jars with sediment required by the participants, excess

442   sediment was placed in spare jars. The *VF$_{av}$* values thus do not *per se* exactly quantify the

443   actual variability caused by sediment heterogeneity in the experiments and cannot be

444   directly deduced from the values in Table 1. They do indicate, however, that sediment

445   heterogeneity potentially may have contributed to the variability observed in Figures 1A-C.

446   Apart from that, sediment heterogeneity within a single sediment batch as received by a

447   participant is expected to be much smaller, as will be shown below (intra-method

448   variability).

449   Measurements performed with sediments stored for 4.5 months in the refrigerator, as

450   compared to measurements initiated directly after sampling from the concrete mixers

451   demonstrated that C_{free} of the target PAHs and PCBs decreased with about 20 % in the FD

452   sediment and 10 % in the BB and SP sediments. This suggests that sediment storage also

453   cannot be excluded as a source of variability. However, the time between the first participant

454   starting the first experiment and the last participant starting this experiment, was only one

455   month. Therefore, it is not very likely that storage time contributed significantly to the

456   variability in Figure 1A. The first and last started standardized experiments were, however,

457   three months apart and storage time thus may have been an additional source of variability in

458   Figure 1B. It should be stressed though that the two measurements (i.e., before and after

459   storage) were performed with two different sediment batches (jars); as such, sediment

460   heterogeneity may also have caused (part of) the difference in C_{free}. Assuming the

461   concentration decrease is a real phenomenon, progressive sorption (re-distribution) may be

462   the underlying mechanism for the SP and FD sediment. However, for the field-contaminated

463   BB sediment this process is improbable. Degradation is unlikely in all cases (chemicals,

464   sediments).

465

466   **Intra-method and Inter-method Variability**

467   The last experiment included C_{free} measurements with all sampler formats by the

468   coordinating laboratory. From this experiment, both the intra- and inter-method variability

469   could be deduced. As observed before,[19] the intra-method variability appeared very low. For

470   sheet samplers (PE, POM, SR), relative standard deviations (RSDs) of the five-fold

471   measurements were generally < 5% and for the (homogeneous) BB sediment, RSDs were

472   often below 2 or even 1 %, indicating very high repeatability. Prerequisites for low RSDs are

473   that the measurements are performed by skilled personnel, trained to work with passive

474   samplers and to perform high-quality chemical analyses (including highly consistent

475   integrations). For SPME fibers, RSDs of the five-fold measurements by the coordinating

476   laboratory were somewhat higher, with the values increasing with decreasing coating

477   thickness: RSDs S10 > S30 > S100 > sheets (see Table S7). The cause of this order most

478   probably relates to (i) the fact that the uncertainty in the sampling phase volume increases

479   with decreasing coating thickness (because of increased uncertainties in the actual coating

480   thickness, inaccurate cutting of the fibers, or coating wear during equilibration) and (ii) the

481   thinner the coating, the higher the probability for artifacts to occur through 'fouling' (i.e.,

482   particles or NAPLs sticking to the coating), potentially causing over-estimation of the

483   polymer-sorbed concentration.[8]

484   Owing to the high method precision, it was possible to accurately quantify the inter-method

485   variability. The resulting VF_{av} values (see Table 1, last row) demonstrate that on average the

486   results of all 14 passive sampling formats (both sheets and SPME fibers of different

487   polymers, sources, and thicknesses) match within a factor of 1.7. Thus, differences in C_{free}

488   determined with a suite of passive samplers were very small (see also Figure 1 D and Figure

489   S1 D). The underlying VF values do slightly increase with target chemical hydrophobicity,

490   in particular for the PCBs (Table S8). This increase is probably caused by the fact that K_{pw}

491   values become more uncertain for very hydrophobic chemicals, due to increasing

492   experimental difficulties related to reduced solubilities and slow kinetics.[10] Lower C_{free}

493   values for the more hydrophobic chemicals cannot explain the observation, as the underlying

494   measured concentrations in the extracts were not related to chemical hydrophobicity.

495   The data variability is practically identical for the different sediments, here indicating that

496   passive sampling is a robust technique, with which C_{free} can be determined precisely,

497   irrespective of the sediment under study. A comparison of the results of the different

498   samplers shows that the highest C_{free} values generally were measured with the S100, S30-2,

499   and S10 SPME fibers, whereas the lowest values generally were determined with POM, PE-

500   6, and SR. However, because the differences are so small, in particular relative to the

501   average (See Figure S1 D), it can be concluded that there are no specific polymers behaving

502   substantially differently to all the others and that their usage should be avoided. Different

503   methods do have their specific 'pros' and 'cons' though (e.g., practicability of handling, ease

504   of K_{pw} determination, detection limits, etc.). A detailed discussion of these factors is beyond

505   the scope of the present paper.

506   Overall, it can be concluded from the present study that passive sampling is ready for

507   implementation in actual risk assessment and management practices of contaminated

508   sediments. The technique is robust, as it produces results that are independent of the

509   sediment studied and sampling polymer or format used. However, standard protocols should

510   be applied (most importantly ensuring non-depletion, taking steps to deduce equilibrium

511   concentrations in the polymers, and full sampler extraction) and the analytical chemistry

512   element be carefully quality-controlled (e.g., by means of (certified) external standards). The

preparation and use of a passive sampling reference sediment may also be considered as a quality check in future work. Based on the standardized procedure, practical guidance and a proposed standard protocol for passive sampling in sediments will be presented in a follow up paper.

Supporting Information

Description of supplementary tests; Tables listing sampling formats, sediment characteristics, participants' own procedures, concentration ranges, fiber volumes, *VF*s characterizing sediment heterogeneity and inter-method variability, and intra-method variability (RSDs); Figures showing box plots, the effects of K_{pw} standardization and

538 sampler/sediment ratio, uptake kinetics in different samplers, and a static/dynamic exposure

539 comparison. This material is available free of charge via the Internet at http://pubs.acs.org.

540

References

542
543
544 (1)   Di Toro, D. M.; Zarba, C. S.; Hansen, D. J.; Berry, W. J.; Swartz, R. C.; Cowan, C.
545 E.; Pavlou, S. P.; Allen, H. E.; Thomas, N. A.; Paquin, P. R. Technical basis for establishing
546 sediment quality criteria for nonionic organic chemicals using equilibrium partitioning.
547 *Environ. Toxicol. Chem.* **1991,** *10* (12), 1541-1583.
548 (2)   Parkerton, T. F.; Maruya, K. A. Passive sampling in contaminated sediment
549 assessment: building consensus to improve decision making. *Integr. Environ. Assess.*
550 *Manag.* **2014,** *10* (2), 163-166.
551 (3)   Mayer, P.; Parkerton, T. F.; Adams, R. G.; Cargill, J. G.; Gan, J.; Gouin, T.;
552 Gschwend, P. M.; Hawthorne, S. B.; Helm, P.; Witt, G.; You, J.; Escher, B. I. Passive
553 sampling methods for contaminated sediments: scientific rationale supporting use of freely
554 dissolved concentrations. *Integr. Environ. Assess. Manag.* **2014,** *10* (2), 197-209.
555 (4)   Lydy, M. J.; Landrum, P. F.; Oen, A. M.; Allinson, M.; Smedes, F.; Harwood, A. D.;
556 Li, H.; Maruya, K. A.; Liu, J. Passive sampling methods for contaminated sediments: state
557 of the science for organic contaminants. *Integr. Environ. Assess. Manag.* **2014,** *10* (2), 167-
558 178.
559 (5)   Ghosh, U.; Kane Driscoll, S.; Burgess, R. M.; Jonker, M. T.; Reible, D.; Gobas, F.;
560 Choi, Y.; Apitz, S. E.; Maruya, K. A.; Gala, W. R.; Mortimer, M.; Beegan, C. Passive
561 sampling methods for contaminated sediments: practical guidance for selection, calibration,
562 and implementation. *Integr. Environ. Assess. Manag.* **2014,** *10* (2), 210-223.
563 (6)   Burkhard, L. P.; Mount, D. R.; Burgess, R. M. *Developing sediment remediation*
564 *goals at superfund sites based on porewater for the protection of benthic life from direct*
565 *toxicity to nonionic organic chemicals.* U.S Environmental Protection Agency. Washington,
566 2017, 74 pp.
567 (7)   Kupryianchyk, D.; Rakowska, M. I.; Roessink, I.; Reichman, E. P.; Grotenhuis, J. T.
568 C.; Koelmans, A. A. In situ treatment with activated carbon reduces bioaccumulation in
569 aquatic food chains. *Environ. Sci. Technol.* **2013,** *47* (9), 4563-4571.
570 (8)   Van der Heijden, S. A.; Jonker, M. T. O. PAH bioavailability in field sediments:
571 Comparing different methods for predicting in situ bioaccumulation. *Environ. Sci. Technol.*
572 **2009,** *43* (10), 3757-3763.
573 (9)   Jonker, M. T. O.; Smedes, F. Preferential sorption of planar contaminants in
574 sediments from Lake Ketelmeer, The Netherlands. *Environ. Sci. Technol.* **2000,** *34* (9),
575 1620-1626.
576 (10)   Jonker, M. T. O.; Van Der Heijden, S. A.; Kotte, M.; Smedes, F. Quantifying the
577 effects of temperature and salinity on partitioning of hydrophobic organic chemicals to
578 silicone rubber passive samplers. *Environ. Sci. Technol.* **2015,** *49* (11), 6791-6799.
579 (11)   Gschwend, P. M.; MacFarlane, J. K.; Reible, D. D.; Lu, X.; Hawthorne, S. B.;
580 Nakles, D. V.; Thompson, T. Comparison of polymeric samplers for accurately assessing
581 PCBs in pore waters. *Environ. Tox. Chem.* **2011,** *30* (6), 1288-1296.
582 (12)   Booij, K.; Robinson, C. D.; Burgess, R. M.; Mayer, P.; Roberts, C. A.; Ahrens, L.;
583 Allan, I. J.; Brant, J.; Jones, L.; Kraus, U. R.; Larsen, M. M.; Lepom, P.; Petersen, J.;

584 Pröfrock, D.; Roose, P.; Schäfer, S.; Smedes, F.; Tixier, C.; Vorkamp, K.; Whitehouse, P.
585 Passive Sampling in Regulatory Chemical Monitoring of Nonpolar Organic Compounds in
586 the Aquatic Environment. *Environ. Sci. Technol.* **2016,** *50* (1), 3-17.
587 (13)   Booij, K.; Smedes, F.; Allan, I. J. *Guidelines for determining polymer-water and*
588 *polymer-polymer partition coefficients of organic compounds.* ICES. Techniques in Marine
589 Environmental Sciences. 2017, 32 pp.
590 (14)   Smedes, F.; Van Vliet, L. A.; Booij, K. Multi-ratio equilibrium passive sampling
591 method to estimate accessible and pore water concentrations of polycyclic aromatic
592 hydrocarbons and polychlorinated biphenyls in sediment. *Environ. Sci. Technol.* **2013,** *47*
593 (1), 510-517.
594 (15)   Apell, J. N.; Gschwend, P. M. Validating the use of performance reference
595 compounds in passive samplers to assess porewater concentrations in sediment beds. **2014,**
596 *48* (17), 10301-10307.
597 (16)   Fernandez, L. A.; Harvey, C. F.; Gschwend, P. M. Using performance reference
598 compounds in polyethylene passive samplers to deduce sediment porewater concentrations
599 for numerous target chemicals. **2009,** *43* (23), 8888-8894.
600 (17)   Vrana, B.; Smedes, F.; Prokeš, R.; Loos, R.; Mazzella, N.; Miege, C.; Budzinski, H.;
601 Vermeirssen, E.; Ocelka, T.; Gravell, A.; Kaserzon, S. An interlaboratory study on passive
602 sampling of emerging water pollutants. *TrAC, Trends Anal. Chem.* **2016,** *76*, 153-165.
603 (18)   Booij, K.; Smedes, F.; Crum, S. Laboratory performance study for passive sampling
604 of nonpolar chemicals in water. *Environ. Tox. Chem.* **2017,** *36* (5), 1156-1161.
605 (19)   Hawthorne, S. B.; Jonker, M. T. O.; Van Der Heijden, S. A.; Grabanski, C. B.;
606 Azzolina, N. A.; Miller, D. J. Measuring picogram per liter concentrations of freely
607 dissolved parent and alkyl PAHs (PAH-34), using passive sampling with polyoxymethylene.
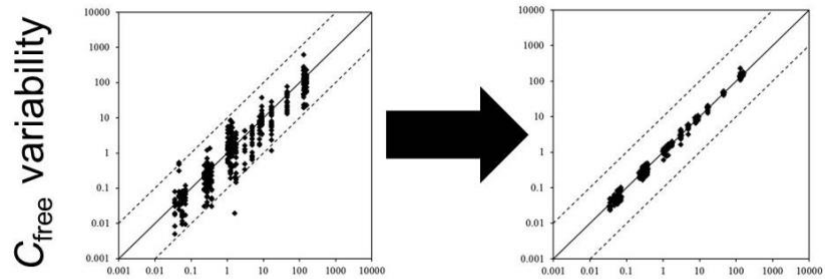608 *Environ. Sci. Technol.* **2011,** *83* (17), 6754-6761.
609
610

611   TOC art

612

613

614

615

25 chemicals
3 sediments
14 passive samplers
11 laboratories

25 chemicals
3 sediments
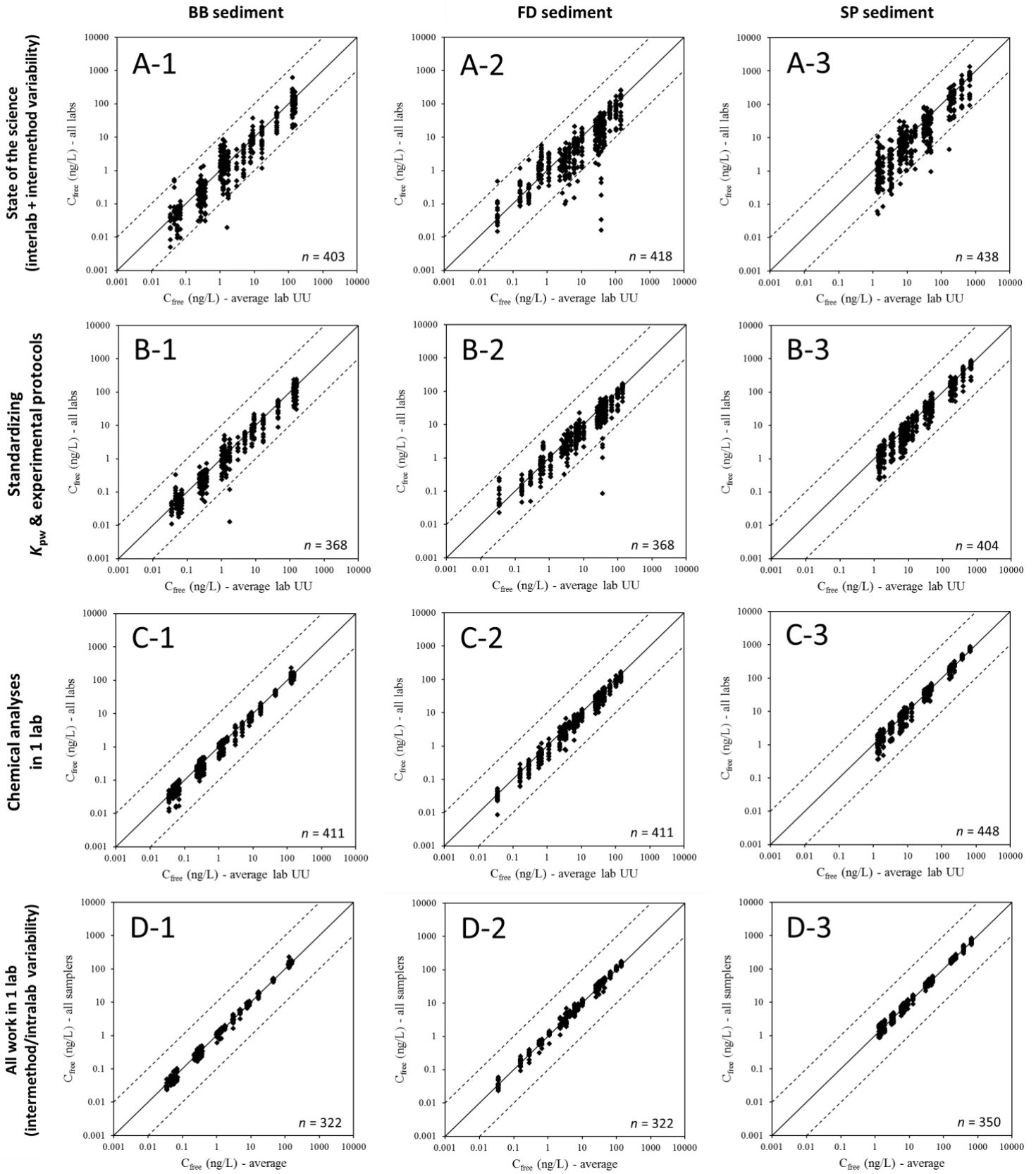14 passive samplers
1 laboratory

Table 1. Averaged Variation Factors (VF_{av}; ± standard deviations) per sediment and per experiment.[a]

	BB sediment[b]	FD sediment[b]	SP sediment
Measurements based on own protocols	9.7 ± 4.1	9.4 ± 6.3	10.8 ± 4.5
Standardizing K_{pw} values	8.9 ± 3.6	9.3 ± 4.6	10.8 ± 5.6
Standardizing protocols & K_{pw} values	4.4 ± 1.4	4.6 ± 2.2	4.5 ± 1.2
Standardizing & chemical analyses in one lab	2.4 ± 0.89	2.4 ± 0.72	2.6 ± 0.82
All work performed in one lab	1.6 ± 0.35	1.7 ± 0.42	1.7 ± 0.31

[a] The VF_{av} values are calculated by averaging the VF values of all chemicals for one sediment in a specific experiment. The number of chemicals included is 23 - 25, depending on the sediment and experiment.

[b] Data for PCB-77 are excluded (see text for explanation).

Figure 1.

628 **Figure 1.** Variability in freely dissolved concentrations (C_{free}) determined in three sediments
629 as measured with passive sampling methods (A) when the participants of the inter-laboratory
630 comparison followed their own protocols, (B) after standardization of K_{pw}s and experimental
631 protocols, (C) when, in addition to B, all chemical analyses were performed in one
632 laboratory, and (D) when both experiments and analyses with all samplers were performed
633 in one laboratory. Solid lines represent the 1:1 relationships; dashed lines indicate ± a factor
634 of ten. The *n* number in each plot indicates the number of data points.
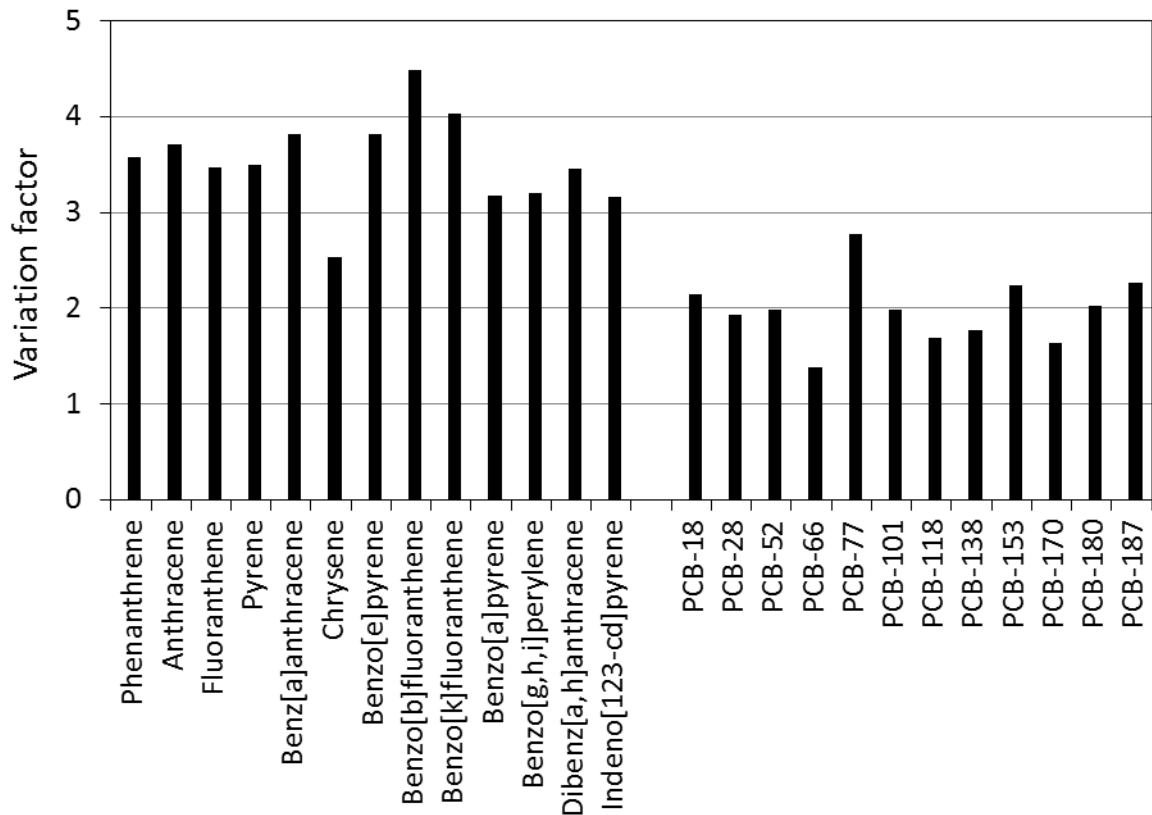635

636

637

638

639   Figure 2.

640

641

642   **Figure 2**. Variation factors (95th PCTL/5th PCTL) calculated based on the (range of)
643   concentrations of the target chemicals in the analytical standard, as reported by the
644   participants of the inter-laboratory comparison.
645