Chemosphere 288 (2022) 132441

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

Assessment of four rounds of interlaboratory tests within the UNEP-coordinated POPs projects

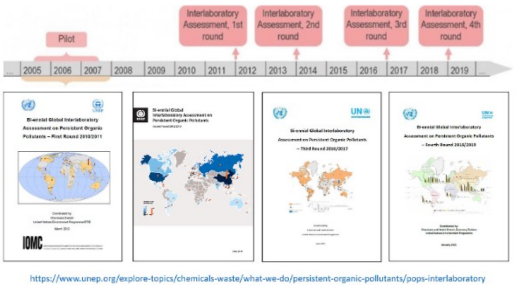
Heidelore Fiedler[a,*], Ike van der Veen[b], Jacob de Boer[b]

[a] Örebro University, School of Science and Technology, MTM Research Centre, SE-701 82, Örebro, Sweden
[b] Vrije Universiteit, Dept. Environment & Health, De Boelelaan 1085, NL-1081HV, Amsterdam, the Netherlands

HIGHLIGHTS

- Africa for dl-POPs, PBDE, and PFAS lacks capacity; Latin American countries for PFAS and PBDE.
- Overall, performance for dl-POPs and to lesser extend PBDE and PFAS satisfactory.
- More than 60% of the z-scores for OCP and PCB were not satisfactory.
- Human milk and fish posed the biggest challenge to the laboratories.
- Regular – and successful – participation in interlaboratory assessments remains essential.

GRAPHICAL ABSTRACT



ARTICLE INFO

Handling Editor: Derek Muir

Keywords:
Persistent organic pollutants
Stockholm Convention on POPs
Core matrices
Performance trends and indicators
Regular participation

ABSTRACT

Since 2005, the United Nations Environment Programme (UNEP) has supported developing countries with capacity building, including on-site training courses and provision on laboratory materials and consumables, and analysis of samples from Africa, Asia-Pacific and Latin America and the Caribbean (GRULAC) in designated expert laboratories. In order to check the performance of laboratories analyzing persistent organic pollutants (POPs) and giving trust into chemical analytical results, four rounds of interlaboratory assessments (ILs) were organized between 2010 and 2019. These were open to all POPs laboratories. In total, 41 575 z-scores, as indicators of performance, were generated in these four ILs; of these, 8 912 were from laboratories in countries supported by UNEP projects and 3 923 were from expert laboratories; these together constitute 31% of the total. 69% of all z-scores came from laboratories not participating in the UNEP projects, especially from China, who recognized the importance of such exercises for quality control at an early stage.

The results showed that POPs analytical capacity has increased over the years, but some gaps could not be closed especially not when LC-MS/MS equipment is essential such as for PFAS and congener-specific HBCD. Use of mass spectrometers provides better results than ECD instruments due to the broad spectrum of organochlorine pesticides to be analyzed under the Stockholm Convention. The main conclusion for all laboratories is that interlaboratory assessments provide important and objective snapshots of performance for the laboratories themselves and for external clients. The methods used in any interlaboratory testing should be the same as during routine analysis between such tests.

* Corresponding author.
E-mail address: heidelore.fiedler@oru.se (H. Fiedler).

https://doi.org/10.1016/j.chemosphere.2021.132441
Received 31 July 2021; Received in revised form 25 September 2021; Accepted 30 September 2021
Available online 5 October 2021

1. Introduction

The Stockholm Convention on Persistent Organic Pollutants (POPs), among others requires environmental monitoring data of high quality to assess the effectiveness of the measures taken under the Convention. In order to identify changes of POPs concentrations in the environment as a consequence of Stockholm Convention implementation, laboratories delivering data for the monitoring should be able at any time to provide accurate data. The differences in measured data should be distinctive and not caused by differences in quantification between two or more laboratories.

The laboratories assessed in this paper are from countries that are Party to the Stockholm Convention on POPs and therefore committed to implement Article 16 of the Convention, which established a global monitoring plan (GMP) and subject to the effectiveness evaluation. These countries have indicated in their national implementation plans (NIPs) the development of national monitoring capacity. Since many countries struggled with the development of monitoring capacity, the United Nations Environment Programme (UNEP) with funds from the Global Environment Facility (GEF) and others has coordinated four regional projects in Africa, Asia, Pacific Islands, and Latin America and the Caribbean to assist 42 eligible countries to strengthen their monitoring capacities. One of the components of the projects are the interlaboratory assessments of POPs to demonstrate proficiency in the analysis of POPs in core matrices of the Convention, which are ambient air, human milk or blood, and water (for perfluorinated compounds only). Four rounds of interlaboratory assessments have been conducted during the past ten years; they were open to any laboratory interested in participation. Although the UNEP/GEF projects had a strong capacity building component and so-called 'expert laboratories' assisted them throughout the project duration in analytical and other questions, no training or guidance beyond the usual information provided to PT participants was given. The Department of Environment & Health of the Vrije Universiteit Amsterdam, the Netherlands (VU E&H) and the Man-Technology-Environment (MTM) Research Centre, School of Science and Technology at Örebro University, Sweden, have organised all four 'Bi-ennial Global Interlaboratory Assessment on Persistent Organic Pollutants (POPs)', named 'IL1', 'IL2′, 'IL3′, and 'IL4′ for short. The set-up of the studies and the assessment of the results followed internationally agreed methods and have been described elsewhere (UNEP et al., 2012, UNEP et al., 2014, UNEP et al., 2017, UNEP et al., 2021).

This paper presents and assesses the performance of the POPs laboratories located in any of the 42 countries participating in the UNEP projects. Their results are weighted against the results from the 'expert laboratories', *i.e.,* experienced laboratories and with an active role in the UNEP projects or providers of data to the Stockholm Convention's data warehouse (GMP data warehouse; www.pops-gmp.org). The evaluation was done between laboratories, countries and regions, overall but also between the four rounds and for type of the test sample, e.g., test solution, abiotic or biotic matrices. These results should assist UNEP, the Stockholm Convention Parties and the financial mechanism in the evaluation of the effectiveness in the field of POPs monitoring but before all, provide an objective picture of the capacity and capabilities of the POPs laboratories for their own country and region.

2. Materials and methods

2.1. Test samples

Test materials consisted of:

(a) mixtures of analytical standards in an inert solution grouped into classes of POPs using similar analytical instrumentation, such as
   (i) Organochlorine pesticides (OCP) consisting of aldrin, dieldrin, endrin, DDTs, chlordane, chlordecone, endosulfan, heptachlor, hexachlorobenzene (HCB), hexachlorobutadiene (HCBD), hexachlorocyclohexanes (α-HCH, β-HCH, γ-HCH), pentachlorobenzene (PeCBz),
   (ii) Polychlorinated biphenyls (PCB): containing the six indicator PCB
   (iii) Dioxin-like POPs containing polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans, and dioxin-like PCB
   (iv) Polybrominated diphenyl ethers (PBDE) containing eight congeners
   (v) Hexabromocyclododecane (HBCD) with α-, β-, and γ isomers for determination
   (vi) Perfluoroalkane substances (PFAS) consisting of perfluoroalkane acids and perfluorooctanesulfonic acid (PFOS) precursors
   (vii) Toxaphene: containing the three Parlar congeners 26, 50, 62
   (viii) Hexabromobiphenyl PBB153

(b) Naturally contaminated abiotic and biotic matrices, such as
   (i) Sediment (Sed): for determination of all POPs
   (ii) Air extract: in toluene for determination of brominated and chlorinated POPs (extract fortified with native OCP and PBDE), in methanol for PFAS
   (iii) Fish: for determination of all POPs (in IL4, test sample B was fortified with native toxaphene for determination of toxaphene)
   (iv) Human milk (HM): for determination of all POPs
   (v) Water: for determination of PFAS
   (vi) Human plasma (HP) for determination of PFAS
   (vii) Transformer oil (TO): for determination of indicator PCB (IL2 only)
   (viii) Ash: for determination of dl-POPs (IL1 as a proxy for air extract)

Test samples were distributed by the coordinators of the ILs, either Örebro University or Vrije Universiteit together with information on the test matrices of the IL, instructions for filling the reporting form (as MsExcel®). Each laboratory was asked to use its own routine method and materials for analysis; there was no provision of analytical standards, clean-up/extraction or analytical columns or other materials to assist any laboratory in the IL. The test materials were sufficient to perform the required analyses but not enough to repeat the whole procedure many times. Typically, the results should be reported within eight to twelve weeks.

2.2. Assessment of the laboratories

This assessment was done on the test results as provided in the appendices 'z-score assessment' of the four reports published by UNEP in the appendices 4 (UNEP et al., 2012, UNEP et al., 2014, UNEP et al., 2017, UNEP et al., 2021). Therein are given z-scores – as their interpretation 'S' for satisfactory performance (2 z = ±25%), 'Q' for questionable performance (>2 z and <3 z = >25% and <37.5%), 'U' for unsatisfactory (>3z = ±37.5%) or extreme performance (>6 z = ±75%), and in the case of left-censored values (LCV), 'C' designates for being consistent and 'I' for being inconsistent with the limit of detection (LOD) – for each combination of POP analyte in the respective matrix.

The identity of the laboratories is not disclosed by the organizers but they can be assigned to a country and to a UN region. Country names are provided as ISO3 alpha codes for harmonization and simplicity; the regions are defined according to the UN.

For comparison and as a 'request', the expert laboratories providing backstopping and expertise in the projects and to analyse a large number of real samples for the global monitoring plan participated in the international laboratory assessments and are evaluated individually and as a group ('Expert' as designation for Region and Country).

The expert laboratories were specialized according to the following

Table 1
Overview of number of laboratories from each Region submitting results to the Rounds Percentages in parenthesis.

Region	IL1 (N = 36)	IL2 (N = 22)	IL3 (N = 53)	IL4 (N = 57)	Overall (N = 168)
Africa	9 (25.0%)	4 (18.2%)	10 (18.9%)	12 (21.1%)	35 (20.8%)
Asia	5 (13.9%)	7 (31.8%)	13 (24.5%)	16 (28.1%)	41 (24.4%)
GRULAC	18 (50.0%)	8 (36.4%)	22 (41.5%)	22 (38.6%)	70 (41.7%)
Expert	4 (11.1%)	3 (13.6%)	8 (15.1%)	7 (12.3%)	22 (13.1%)

core matrices and POPs (but may have participated in more than those listed below):

L037: Core matrix: air; POPs: all except PFAS.

L101: Core matrix: air; POPs: all except PFAS.

L104: Core matrix: air; POPs: all except dl-POPs and PFAS.

L105: Core matrix: air; POPs: all except dl-POPs.

L124: Core matrices: air, water, human milk; POPs: dl-POPs and PFAS.

L126: Core matrix: human milk; POPs: all except PFAS.

L195: Core matrix: air; POPs: all POPs.

L257: Core matrix: water; POPs: PFAS.

The database was assessed using a standard procedure to allow direct comparison between participants. The approach of the assessment is based on the standard ISO 13528 (ISO, 2015) and the International Union of Pure and Applied Chemistry International Harmonised Protocol for Proficiency Testing (Thompson et al., 2006) and using the Cofino model (Molenaar et al., 2018). The assigned value (AV), the between-lab coefficient of variation (CV) values and the laboratory assessment using z-scores are based on the principles employed in the Quality Assurance of Information for Marine Environmental Monitoring in Europe (QUASIMEME) proficiency testing. The performance assessment is described in the full reports of the studies (UNEP et al., 2012, UNEP et al., 2014, UNEP et al., 2017, UNEP et al., 2021) and in the previous publications (Abalos et al., 2013; de Boer et al., 2021; Fiedler et al., 2017, 2020, 2021).

2.3. Visualization

Statistical evaluations and visualization were made with R version R-4.0.3 (as of 2020–10-10) and R Studio Version 1.3,1056. For unification and simplification, alpha-3 codes are used to represent a country name (ISO).

3. Results

3.1. General overview and summary

Our database of laboratories registered for the four rounds of inter-laboratory assessments (IL1-IL4) contains 289 laboratories. There were always more laboratories registered than delivering results. The number of laboratories that delivered results and had achieved at least one z-score for one POPs and one test sample were 78 in IL1, 88 in IL2, 133 in IL3 and 117 in IL4. These numbers were larger than in other proficiency tests. Across all ILs, 41 575 z-scores were assigned. As described in the Introduction, this paper includes the performance of the laboratories from the 42 participating countries in the UNEP GMP2 projects (2016–2020) (UNEP, 2015a, b, c, d) and the expert laboratories. Included are results from laboratories in IL1 and IL2 for countries that did not participate in the UNEP-supported capacity building and data generation projects from 2008 to 2014, typically referred to as 'GMP1 projects' (such as MAR, TUN, THA, MNG, VNM, COL).

For this assessment, 88 laboratories with 80 from 27 countries participating in the latest UNEP-GEF GMP2 project (2016–2019) and eight expert laboratories (defined 'Expert' and as 28th country) had submitted 168 sets of analytical results. 15 countries from the UNEP-GMP projects did not have laboratories participating in any of the rounds (eight countries from the Pacific Islands project: KIR, MHL, NIU, PLW, SLB, TUV, VUT, WSM; two from Asia: KHM, LAO; four from Africa: COD, ETH, TGO, TZA, and BRB from GRULAC), mainly because they had no qualified laboratories for POPs analysis. Multiple or repeated participation is detailed in section 4.2.

In total, 168-times, a laboratory provided at least one result to have one z-score in one of the test matrices in any the four ILs. Of these, expert laboratories reported 22-times (Table 1). The highest number with 70 sets of results reporting was from the Group of Latin American and the Caribbean countries (GRULAC). The number of laboratories from each country per IL is shown in Table S 1. Most of the countries had only one laboratory per round participating; however, the number of laboratories peaked in distinct rounds such as Vietnam with eight laboratories in IL3

Table 2
Overview on number of z-scores per Round, Region, Type, Matrix and POP Group Percentages in parenthesis.

	IL1 (N = 1812)	IL2 (N = 1834)	IL3 (N = 4331)	IL4 (N = 4858)	Overall (N = 12 835)
			Region		
Africa	385 (21.2%)	162 (8.8%)	866 (20.0%)	624 (12.8%)	2037 (15.9%)
Asia	247 (13.6%)	483 (26.3%)	843 (19.5%)	1240 (25.5%)	2813 (21.9%)
GRULAC	821 (45.3%)	504 (27.5%)	1250 (28.9%)	1487 (30.6%)	4062 (31.6%)
Expert	359 (19.8%)	685 (37.4%)	1372 (31.7%)	1507 (31.0%)	3923 (30.6%)
			Type		
Test solution (TS)	723 (39.9%)	640 (34.9%)	1611 (37.2%)	2105 (43.3%)	5079 (39.6%)
Abiotic	557 (30.7%)	710 (38.7%)	1512 (34.9%)	1870 (38.5%)	4649 (36.2%)
Biotic	532 (29.4%)	484 (26.4%)	1208 (27.9%)	883 (18.2%)	3107 (24.2%)
			Matrix		
Test solution (TS)	723 (39.9%)	640 (34.9%)	1611 (37.2%)	2105 (43.3%)	5079 (39.6%)
Air extract		361 (19.7%)	791 (18.3%)	954 (19.6%)	2106 (16.4%)
Sediment (Sed)	369 (20.4%)	305 (16.6%)	710 (16.4%)	838 (17.2%)	2222 (17.3%)
Fish	327 (18.0%)	235 (12.8%)	726 (16.8%)	322 (6.6%)	1610 (12.5%)
Human milk (HM)	205 (11.3%)	230 (12.5%)	451 (10.4%)	500 (10.3%)	1386 (10.8%)
Water		1 (0.1%)	11 (0.3%)	78 (1.6%)	90 (0.7%)
Human plasma (HP)		19 (1.0%)	31 (0.7%)	61 (1.3%)	111 (0.9%)
Ash	188 (10.4%)				188 (1.5%)
Transformer oil (TO)		43 (2.3%)			43 (0.3%)
			Group		
OCP	592 (32.7%)	602 (32.8%)	1564 (36.1%)	1642 (33.8%)	4400 (34.3%)
PCB	331 (18.3%)	332 (18.1%)	800 (18.5%)	700 (14.4%)	2163 (16.9%)
dl-POPs	889 (49.1%)	699 (38.1%)	1486 (34.3%)	1487 (30.6%)	4561 (35.5%)
PBDE		134 (7.3%)	229 (5.3%)	306 (6.3%)	669 (5.2%)
PBB153			7 (0.2%)	19 (0.4%)	26 (0.2%)
HBCD			23 (0.5%)	36 (0.7%)	59 (0.5%)
Toxaphene			20 (0.5%)	30 (0.6%)	50 (0.4%)
PFAS		67 (3.7%)	202 (4.7%)	638 (13.1%)	907 (7.1%)

Table 3
Summary by z-score according to Region, Round, Type, Matrix, and Group (IL1-IL4).

	S (N = 6185)	Q (N = 1232)	U (N = 4388)	C (N = 232)	I (N = 798)	Overall (N = 12 835)
			Region			
Africa	402 (6.5%)	180 (14.6%)	1215 (27.7%)	14 (6.0%)	226 (28.3%)	2037 (15.9%)
Asia	1141 (18.4%)	333 (27.0%)	1091 (24.9%)	54 (23.3%)	194 (24.3%)	2813 (21.9%)
GRULAC	1901 (30.7%)	377 (30.6%)	1430 (32.6%)	80 (34.5%)	274 (34.3%)	4062 (31.6%)
Expert	2741 (44.3%)	342 (27.8%)	652 (14.9%)	84 (36.2%)	104 (13.0%)	3923 (30.6%)
			Rd			
IL1	944 (15.3%)	205 (16.6%)	663 (15.1%)			1812 (14.1%)
IL2	870 (14.1%)	213 (17.3%)	547 (12.5%)	48 (20.7%)	156 (19.5%)	1834 (14.3%)
IL3	2044 (33.0%)	344 (27.9%)	1615 (36.8%)	51 (22.0%)	277 (34.7%)	4331 (33.7%)
IL4	2327 (37.6%)	470 (38.1%)	1563 (35.6%)	133 (57.3%)	365 (45.7%)	4858 (37.8%)
			Type			
TS	2840 (45.9%)	599 (48.6%)	1528 (34.8%)	6 (2.6%)	106 (13.3%)	5079 (39.6%)
abiotic	2102 (34.0%)	412 (33.4%)	1671 (38.1%)	93 (40.1%)	371 (46.5%)	4649 (36.2%)
biotic	1243 (20.1%)	221 (17.9%)	1189 (27.1%)	133 (57.3%)	321 (40.2%)	3107 (24.2%)
			Matrix			
TS	2840 (45.9%)	599 (48.6%)	1528 (34.8%)	6 (2.6%)	106 (13.3%)	5079 (39.6%)
Air	994 (16.1%)	192 (15.6%)	697 (15.9%)	41 (17.7%)	182 (22.8%)	2106 (16.4%)
Sed	944 (15.3%)	201 (16.3%)	845 (19.3%)	50 (21.6%)	182 (22.8%)	2222 (17.3%)
Fish	649 (10.5%)	109 (8.8%)	634 (14.4%)	79 (34.1%)	139 (17.4%)	1610 (12.5%)
HM	525 (8.5%)	102 (8.3%)	533 (12.1%)	51 (22.0%)	175 (21.9%)	1386 (10.8%)
Water	41 (0.7%)	6 (0.5%)	34 (0.8%)	2 (0.9%)	7 (0.9%)	90 (0.7%)
HP	69 (1.1%)	10 (0.8%)	22 (0.5%)	3 (1.3%)	7 (0.9%)	111 (0.9%)
Ash	103 (1.7%)	10 (0.8%)	75 (1.7%)			188 (1.5%)
TO	20 (0.3%)	3 (0.2%)	20 (0.5%)			43 (0.3%)
			Group			
OCP	1311 (21.2%)	397 (32.2%)	2100 (47.9%)	124 (53.4%)	468 (58.6%)	4400 (34.3%)
PCB	842 (13.6%)	242 (19.6%)	913 (20.8%)	20 (8.6%)	146 (18.3%)	2163 (16.9%)
dl-POPs	2923 (47.3%)	451 (36.6%)	1026 (23.4%)	44 (19.0%)	117 (14.7%)	4561 (35.5%)
PBDE	381 (6.2%)	57 (4.6%)	177 (4.0%)	23 (9.9%)	31 (3.9%)	669 (5.2%)
PBB153	14 (0.2%)	1 (0.1%)	5 (0.1%)	3 (1.3%)	3 (0.4%)	26 (0.2%)
HBCD	54 (0.9%)	1 (0.1%)	1 (0.0%)	3 (1.3%)		59 (0.5%)
Toxaphene	39 (0.6%)	3 (0.2%)	8 (0.2%)			50 (0.4%)
PFAS	621 (10.0%)	80 (6.5%)	158 (3.6%)	15 (6.5%)	33 (4.1%)	907 (7.1%)

whereas the usual number was four. A similar stimulating effort through the UNEP projects is seen for Thailand in IL4 (eight laboratories participating). Detailed information is provided in Figure S 1 and Table S 2.

The general overview on the number of z-scores achieved for the 168 laboratories according to UN region (Region), type of test sample (Type), matrix (Matrix, as subgroups within abiotic and biotic matrices) and POPs analytes (Group) is shown in Table 2; percentages for each entry are given in parenthesis. Empty cells indicate that the parameter was not included in the IL; for example, 'Ash' was offered as a test sample only in IL1 as a proxy for the core matrix air. 'TO' was included in IL2 due to the need to have PCB tested in transformer oil. The 'new POPs' which are the POPs beyond the initial 12 POPs, were incorporated as the analytical techniques advanced; therefore, they were not offered in the IL1 and to a lesser extend in IL2 (PBDE and PFAS included from IL2).

As a region, GRULAC obtained most of the z-scores (32%), followed by Asia (22%) whereby it must be mentioned that laboratories from China (28 of the total of 289 laboratories) and Japan (6) were not included in this assessment since China was not a beneficiary in these UNEP GMP projects. The eight expert laboratories, here included as one Region, contributed with almost ¹⁄₃ to the total of z-scores.

For the groups of OCP and dl-POPs, the largest number of z-scores with about 35% of all for each of the POP analytical groups were assigned. Also, analysis of PCB had a high number and among the new POPs, only PFAS had a broader presence (907 z-scores, 7%). The total number of z-scores alone does not give an indication on the number of laboratories performing this kind of analysis rather is the consequence of

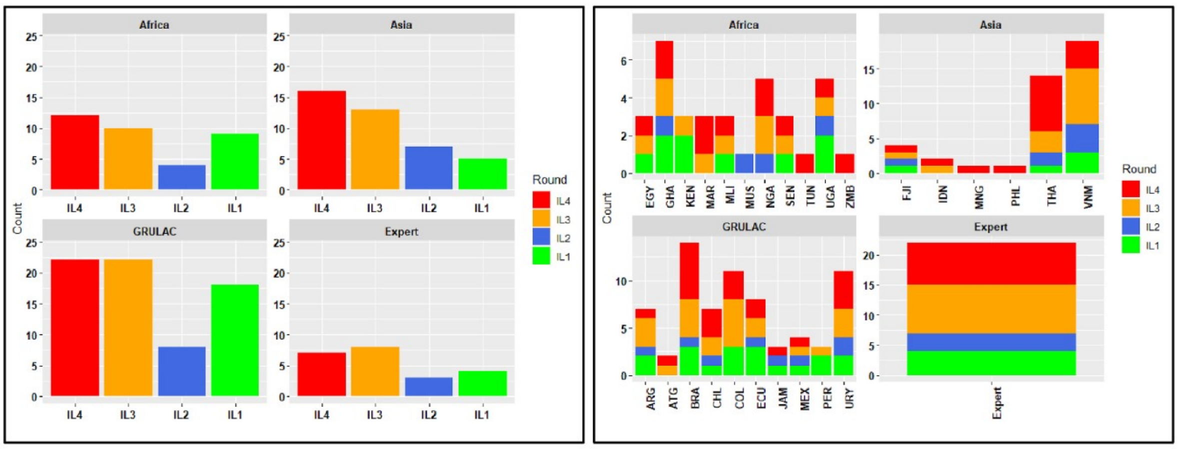


Fig. 1. Bar plots for laboratories per Region or country and Region according to Round (n = 168).

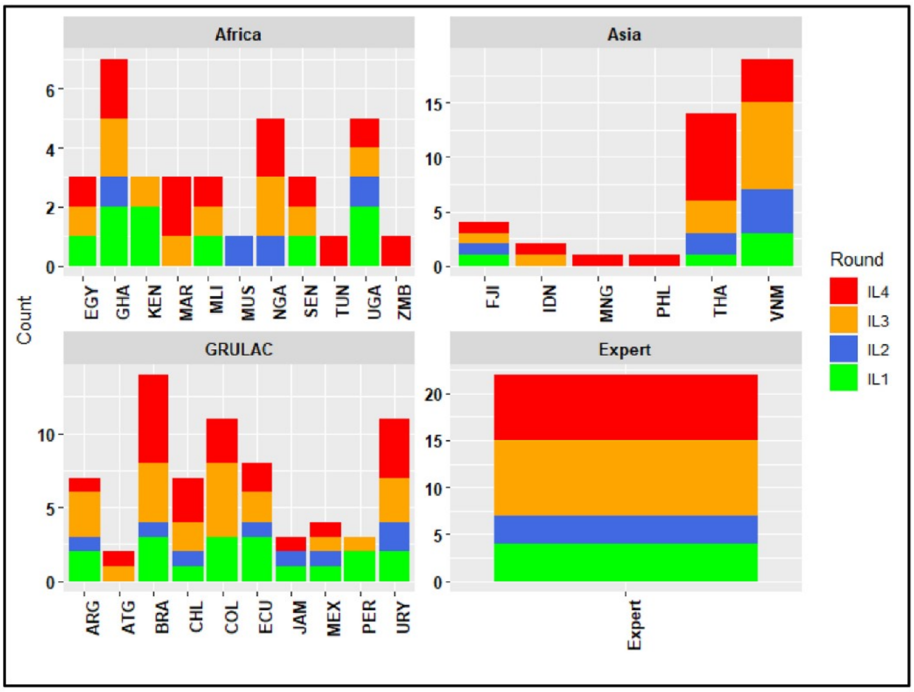


Fig. 2. Bar plots for number of laboratories per country within the Region colored for Round (n = 168).

the number of compounds to be reported within the POP group. For PFAS, in addition, there were two test samples included (human plasma and water).

Among the type of sample and the matrices, TS (40%) had the largest share followed by abiotic (36%) and biotic samples (24%). Among the core matrices of the GMP, 'Air' had most results (16.4%) although offered only in three of the four ILs; human milk ('HM') had 11% and 'Water' limited to PFAS had only 0.7% of the total.

The quality or interpretation of the z-scores is summarized in Table 3; rows and columns add up to 100%. Important for any assessment are the contents of the columns 2 (showing the 'S' performances) and 4 ('U' performances). The most successful laboratories were the Experts, which achieved 44% of all satisfactory results and only 15% of all unsatisfactory results. African laboratories had the smallest number of 'S' and GRULAC the highest number of 'U'. For the Type, 'TS' always had the highest percentage of 'S' results (46%) ad biotic the lowest (20%). The outcomes for test solutions and real samples are further detailed in section 3.3.

As to the distribution of the z-scores across all POPs, it is striking that the dl-POPs have the highest percentage of 'S' and the lowest percentage of 'U'. This demonstrates the overall good quality of the dioxin analyses. On the other hand, the quality of the OCP data is poor when between the POPs groups, 48% of the results were unsatisfactory and only 21% of all data were satisfactory. Details are discussed in section 3.3.

All the individual amounts of POP reported, their numeric z-scores as well as the z-score interpretation (as 'S', 'Q', 'U', 'C', 'I') are available in the appendices to the four reports. The reports are available as reports from UNEP and describe in detail the organization, test samples, and the results for all POPs for the four rounds of IL (UNEP et al., 2012, UNEP et al., 2014, UNEP et al., 2017, UNEP et al., 2021). The appendices for the IL3 are available online from (UNEP, 2017) and appendices for the IL4 are available online from (UNEP, 2019).

3.2. Participation and capacity in the regions

The number of laboratories from the regions in each round varied and is shown in Fig. 2, left, for the regions (summary numeric data in Table 1) and Fig. 2, right, for the number of laboratories within the

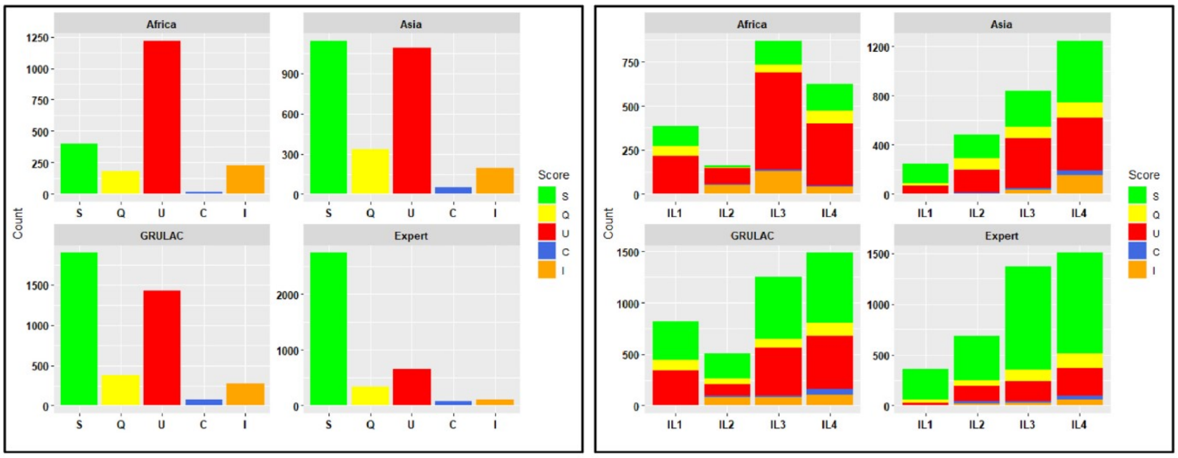


Fig. 3. Overview for quality of z-scores according to Region or Round.

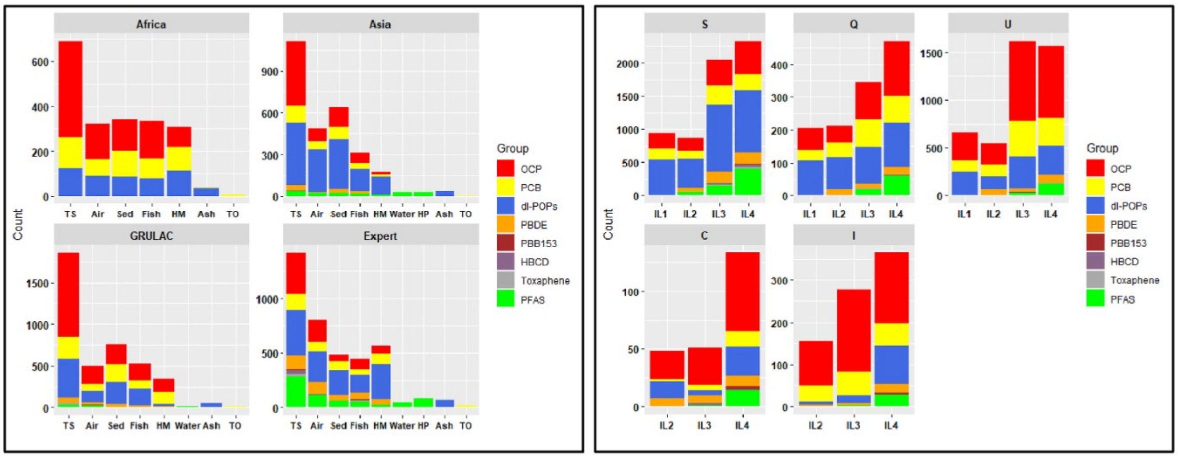


Fig. 4. Regional distribution of total z-scores according to matrix and POP group (left); z-scores grouped into POP Group and Round.

country according to Region (more information in Table S 1). The highest participation was from GRULAC in IL3 and IL4. The number of Asian laboratories increased steadily from IL1 to IL4. It can be concluded that the interlaboratory assessments had a stimulating effect in Asian countries so that finally, a UNEP-GEF project could be implemented from 2016 to 2019. Interestingly, in Africa, the number of laboratories (from participating countries) decreased from IL1 to IL2 but increased from IL3 to IL4. Also, for the expert laboratories, it took some time to join the interlaboratory assessments and not all of them have started in Round 1(see Fig. 1).

As to the countries, in Africa Ghana and Uganda laboratories participated in all four rounds, as did Fiji, Thailand, and Vietnam in Asia, and Argentina, Brazil, Chile, Ecuador, Mexico, and Uruguay in GRULAC. The participation of the individual laboratories and their assignment to Region is visualized in supplementary information, Figure S 1. The assessment of the quality on an individual laboratory basis is done in section 4.2.

3.3. *Performance*

In the four rounds of interlaboratory assessments, a total of 12 835 z-scores were generated. The following sections assess the quality of the analysis for the various aspects.

3.3.1. *Regional basis*

In terms of the quality of analysis on a regional basis, Fig. 3 shows that the ratio of 'S' to 'U' varies highly with Region; for numeric data, see Table 3 (and details in Table S 6). At right, the distribution of the z-scores in each Round is displayed. The 'Expert' laboratories justified their expertise as can be seen from the plots in Figure S 2 and the numeric values contained in Table 2): They had the highest number of 'S' (2741) and the lowest number of 'U' (652). Also, laboratories in GRULAC performed better than those in Asia, 'S' was greater than 'U'. The overall result from Africa is poor since the number of 'U' (1215) was about 3-fold above 'S' (402). On the positive site (and except for IL2), there appears to be a stable number of 'S' results (green color) in the remaining rounds. This can mainly be attributed to L053 (Figure S 2).

3.3.2. *POPs group*

Fig. 4 and Table S 3 show the total absence of PBDE and PFAS analytical capacity in Africa and poor representation of PFAS in GRULAC (TS and Water). Besides the expert laboratories that together covered all POPs and all matrices, toxaphene was analyzed only in GRULAC and HBCD only in Asia.

3.3.3. *Type and matrix*

All laboratories were encouraged to analyse the TS to demonstrate their capabilities to identify, separate and quantify the analyte in an inert solution without interference from any matrix effect and at comparatively high concentrations. Therefore, low levels do not matter and sensitivity of the instrument would not be a limiting factor nor exclusion criterion. As a consequence, 5079 z-scores could be attributed to TS (Tables 2 and 4).

Since not all laboratories analyzed all types of samples, Fig. 5 shows the choice of the samples and the performance for each region. A preference for a certain sample type cannot be seen and it is a positive outcome that the broader spectrum offered in the ILs was accepted. Also, most laboratories used the TS as a starting point for their quality assessment (left site). Individual numbers are provided in Table S 4 and in Fig. 5. The expert laboratories in general performed very well on all sample types whereas on a regional basis, Asia and GRULAC have more 'S' than 'U' for the TS and the abiotic samples and only for the biotic samples the ratio is turned; problems in Africa can already be seen with the TS and the abiotic samples. Thus, improvement of the POPs analysis in all aspects is necessary. It shall be noted that singular exemptions for individual laboratories apply. For details, see Table S 6.

The results for the quality of the national laboratories in each country for either TS, abiotic or biotic samples is shown in Fig. 6. It can be seen that in some countries there is good capacity for a multitude of matrix types with good performance (BRA and VNM also with high numbers; THA and URY) but there are also some countries where there is no (ATG, MUS, ZMB only for one Type; MNG and MAR not for biotic) or very poor performance (SEN, TUN, NGA, JAM, KEN, FJI). From the gaps (missing Type on x-axis) or failures ('U' or red color in Fig. 6), recommendations as to improving performance, building new capacity or

Table 4
Overall performance according to test matrix (core matrices are highlighted in grey color; TS as well), n = 12 835.

	TS (N = 5079)	Air (N = 2106)	Sed (N = 2222)	Fish (N = 1610)	HM (N = 1386)	Water (N = 90)	HP (N = 111)	Ash (N = 188)	TO (N = 43)
S	**2840 (55.9%)**	**994 (47.2%)**	**944 (42.5%)**	**649 (40.3%)**	**525 (37.9%)**	**41 (45.6%)**	**69 (62.2%)**	**103 (54.8%)**	**20 (46.5%)**
Q	599 (11.8%)	192 (9.1%)	201 (9.0%)	109 (6.8%)	102 (7.4%)	6 (6.7%)	10 (9.0%)	10 (5.3%)	3 (7.0%)
U	1528 (30.1%)	697 (33.1%)	845 (38.0%)	634 (39.4%)	533 (38.5%)	34 (37.8%)	22 (19.8%)	75 (39.9%)	20 (46.5%)
C	6 (0.1%)	41 (1.9%)	50 (2.3%)	79 (4.9%)	51 (3.7%)	2 (2.2%)	3 (2.7%)		
I	106 (2.1%)	182 (8.6%)	182 (8.2%)	139 (8.6%)	175 (12.6%)	7 (7.8%)	7 (6.3%)		

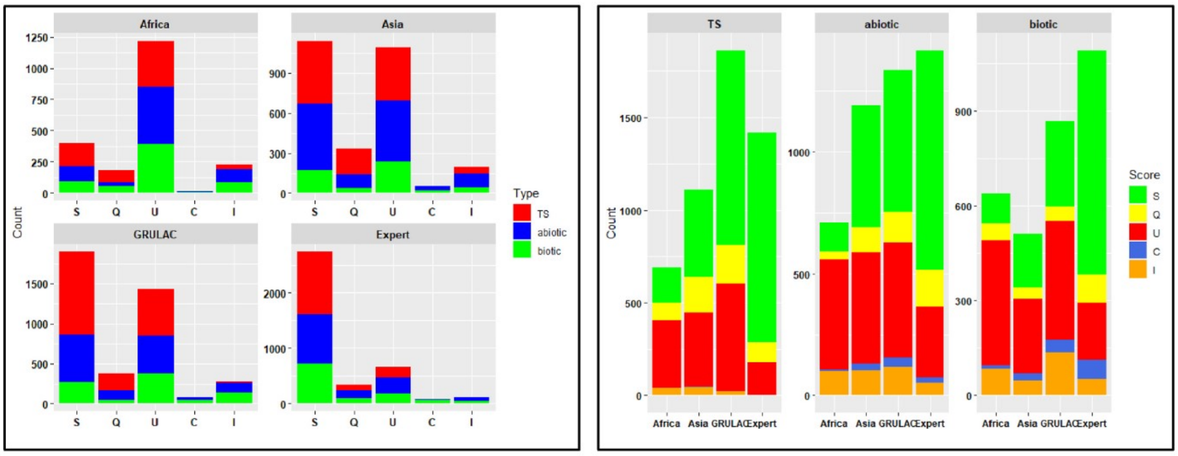


Fig. 5. Choice of matrix types within each Region and for each Round (left) and overview of z-scores according to sample type (Type) for each region (right).

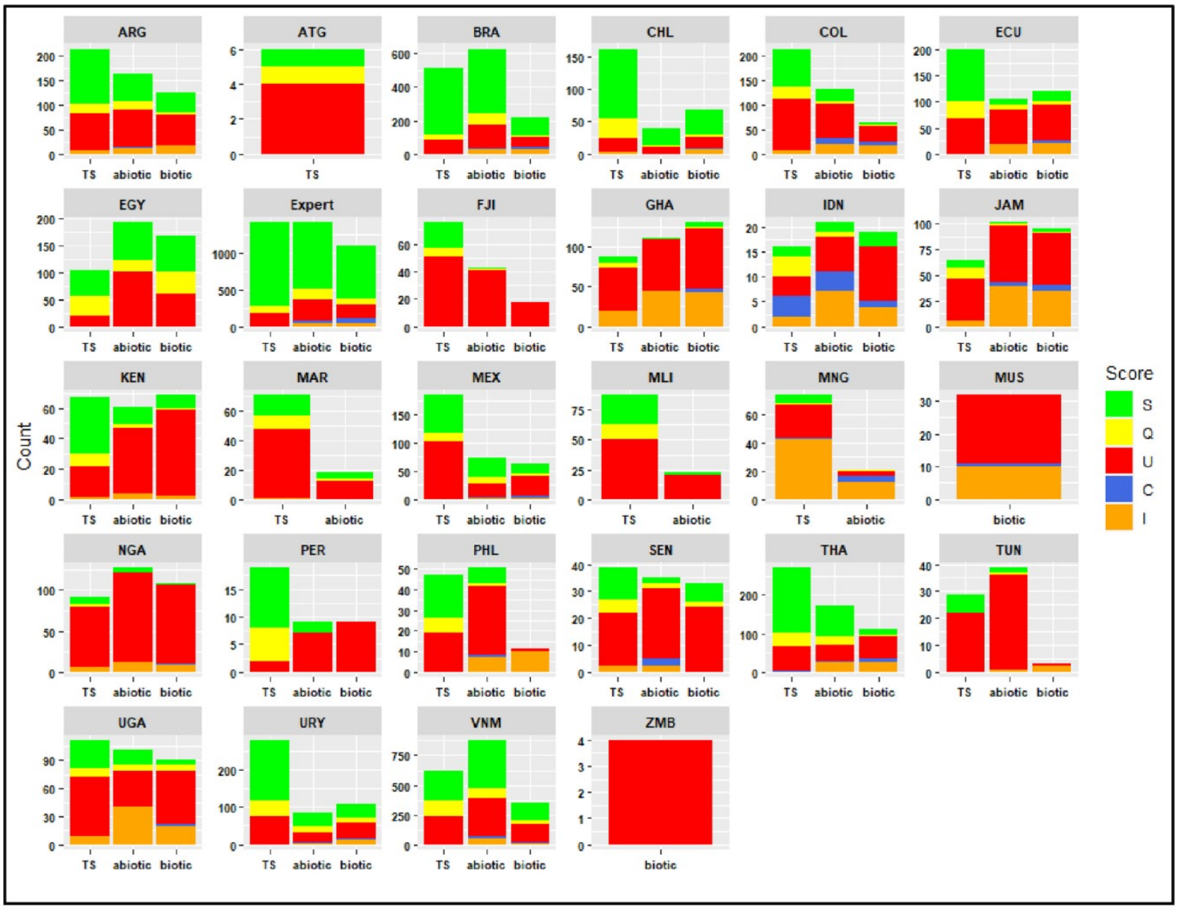


Fig. 6. Performance of country (n = 27 + 1) according to Type.

terminate POPs analysis should be drawn at national level. More specific information as to the matrix and POP with weaknesses and strengths is shown in Fig. 7 and the individual laboratories are displayed in Table S 4 and Figure S 4).

The following Table 4 summarizes and Fig. 7 shows the performance for the test matrices. Details for each region are available in Table S 5. Of special interest is the ratio between 'S' and 'U' for the 'TS' and the core matrices 'Air', 'HM' and 'Water' for PFAS. It can be seen that capacity for the analysis of the core matrices exists in all regions but not in Africa for water. For other matrices, there is no capacity for human plasma in Africa and GRULAC. These gaps coincide with PFAS analysis, which is not or rarely present in these two regions (see Fig. 4 and Table S 3).

For the three core matrices, 2 106 z-scores were attributed to Air, 1 386 to HM, and 90 to Water (Table 4). The high number for Sed (N = 2 222) indicates the high interest or need for laboratories to demonstrate expertise for this matrix; to a lower extend this is also true for Fish (N = 1 610), which had more z-scores than human milk HM (N = 1 386). For HM it is disappointing that 'U' results (533) are more than 'S' (525); for a core matrix, better results have been expected. From Fig. 7, it can be seen that most regions attempted to analyse all matrices and with

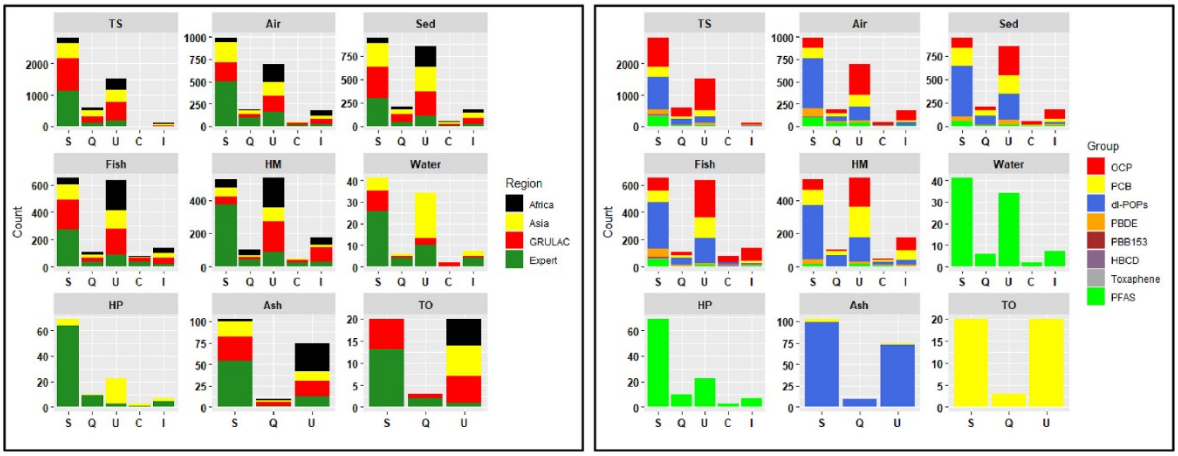


Fig. 7. Distribution and quality of z-scores for all matrices in the regions and for POP groups.

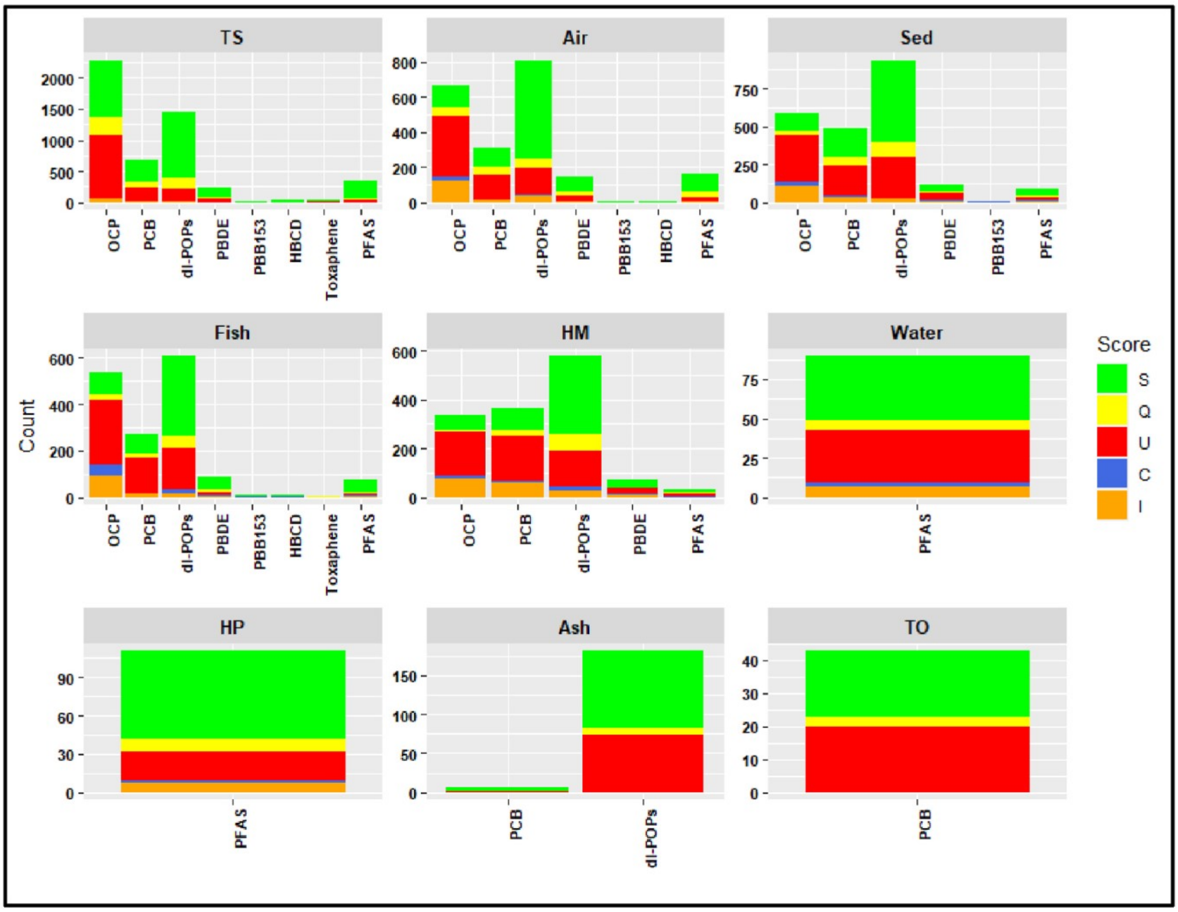


Fig. 8. Distribution and quality of z-scores for all matrices and POP groups.

respect to the core matrices there is only a lack of capacity for Water in Africa. With a view on performance, in all matrices in Africa it must stated that 'S'<'U'; for human milk also in GRULAC. For Air, GRULAC has about the same number of 'S' as 'U'. A positive ratio ('S'>'U') was found in GRULAC for air and Asia for all three core matrices (air, human milk and water) (Fig. 7, left). From Fig. 7, right, it is evident that the negative ratio on performance in air and especially in human milk is due to the high number of 'U' for OCP and PCB. For dl-POPs, there are always much more 'S' results than 'U' results.

4. Discussion

4.1. Specialization and performance

In this section, we investigate if specialization has an effect on performance by assessing the z-scores for the different matrices and the POPs groups. Fig. 8 displays the number and quality of z-scores according to the matrix for all the POPs groups analyzed. The green parts of the stacked columns refer to 'S' performance and the red parts to 'U' performance. For dl-POPs and all matrices, there was always more green

Table 5
Overview on z-score distribution according to number of participations; percentages in parenthesis.

Score	4x P (N = 4334)	3x P (N = 4270)	2x P (N = 2358)	1x P (N = 1873)	Overall (N = 12 835)
S	2288 (52.8%)	2256 (52.8%)	1012 (42.9%)	629 (33.6%)	6185 (48.2%)
Q	431 (9.9%)	441 (10.3%)	196 (8.3%)	164 (8.8%)	1232 (9.6%)
U	1274 (29.4%)	1290 (30.2%)	965 (40.9%)	859 (45.9%)	4388 (34.2%)
C	101 (2.3%)	49 (1.1%)	46 (2.0%)	36 (1.9%)	232 (1.8%)
I	240 (5.5%)	234 (5.5%)	139 (5.9%)	185 (9.9%)	798 (6.2%)

than red (quotient S/U for overall = 2.9); including the core matrices (Air = 3.7, HM = 2.3) and TS (5.1). For sediment and fish, the satisfactory results were 2.0- and 1.9-times greater than the unsatisfactory results. The opposite is the case for OCP, where always more ‘U’ are counted than ‘S’; best achievement was for TS (0.92) and worst results for HM (0.35) within the core matrices and fish (0.34) overall 0.62. For PCB, the relation is positive only for TS (1.6.) Always positive ratios were also obtained for PBDE, where the Sed was the matrix with the lowest ratio (1.1). For PFAS, also all ratios were positive with the exception of HM (0.93) and for PFAS. HBCD, toxaphene and PBB153 had so few z-scores overall and in the matrices so that it seemed that mostly competent laboratories analyzed these samples and achieve good results; they are not discussed further. These data are visualized in Fig. 8, right.

4.2. Experience and expertise

It is recommended that laboratories prove their expertise through regular analysis and regular – and successful – participation in inter-laboratory assessments. It is assumed that performance improves with repeated participation. In Figure S 5 and Table S 8, the laboratories are ranked by descending number of participation and colored by Region. Table 5 provides the summary of the z-scores achieved by number of participations; also visualized in Figure S 5. Across all laboratories and Rounds, 48.2% of the results were satisfactory whereas 34% were unsatisfactory.

Fig. 9 details the number of the participation by the laboratories and it can be seen that 11 laboratories participated 4-times, 14 laboratories 3-times, 19 twice, and 44 laboratories only once. The regional distribution is shown in Fig. 10. Both figures include the number and quality of the z-score. It can be seen that the number of z-scores is highest for laboratories with four or three participations. From optical inspection it can also be seen that most of these laboratories were successful and had more ‘S’ scores than ‘U’ scores. Some exceptions were found as well for certain laboratories; however, in such assessment, the total number of z-scores should be taken into account when assessing capacities or expertise and laboratories with relatively low numbers of z-scores, such as for L091 or L062 (less than 250 z-scores whereas L101, L126, and L105 had 924, 803 and 802 z-scores, resp.) be valued with some caveats (see Figure S1 and Table S 2).

The stimulating effect from the UNEP/GMP2 projects can be seen in Fig. 10, where Asian countries had participated for the first time. In the Asia region dominated the 1-time and 2-times participation, in the later Rounds dominate, showing increasing interest or needs.

The four ILs refer to different years and therefore, the composition of

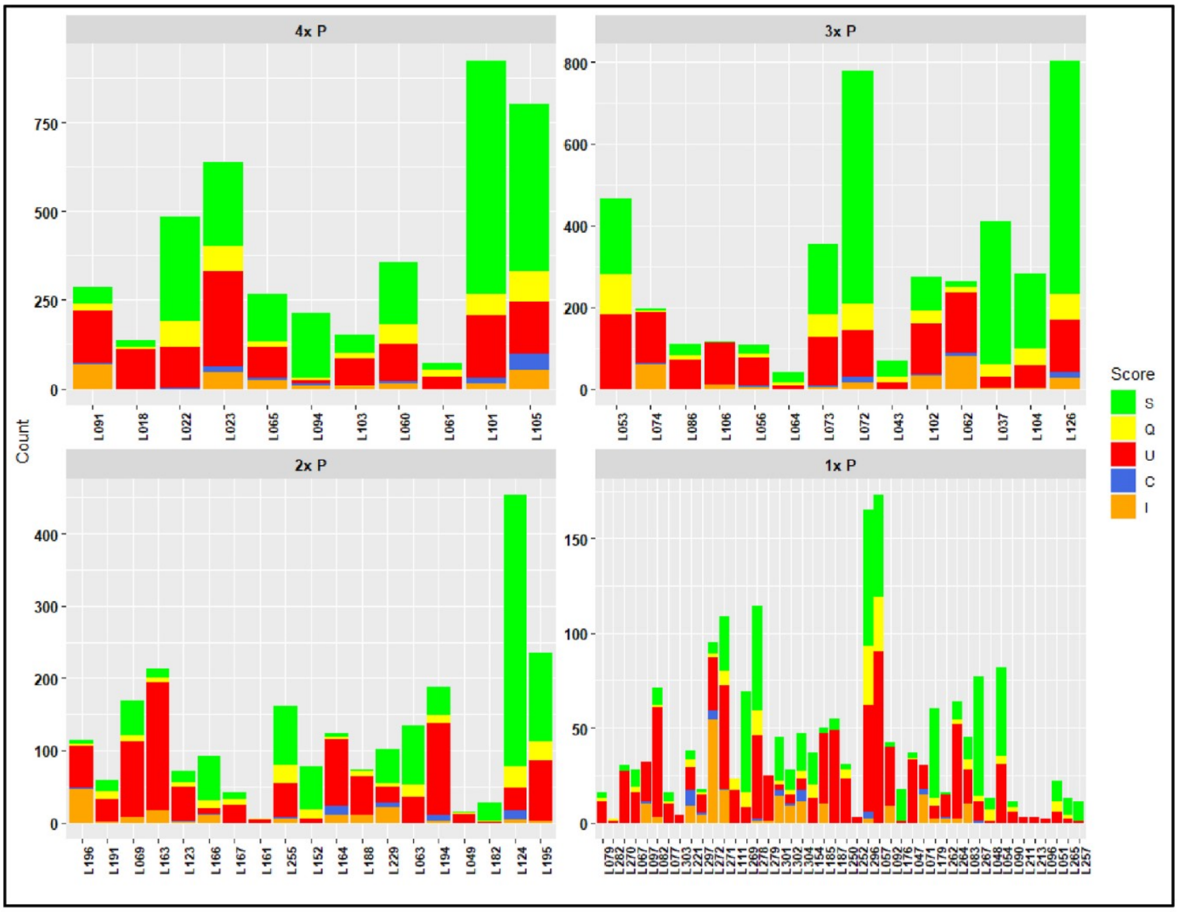


Fig. 9. Grouping of laboratories according to number of participation and z-score.

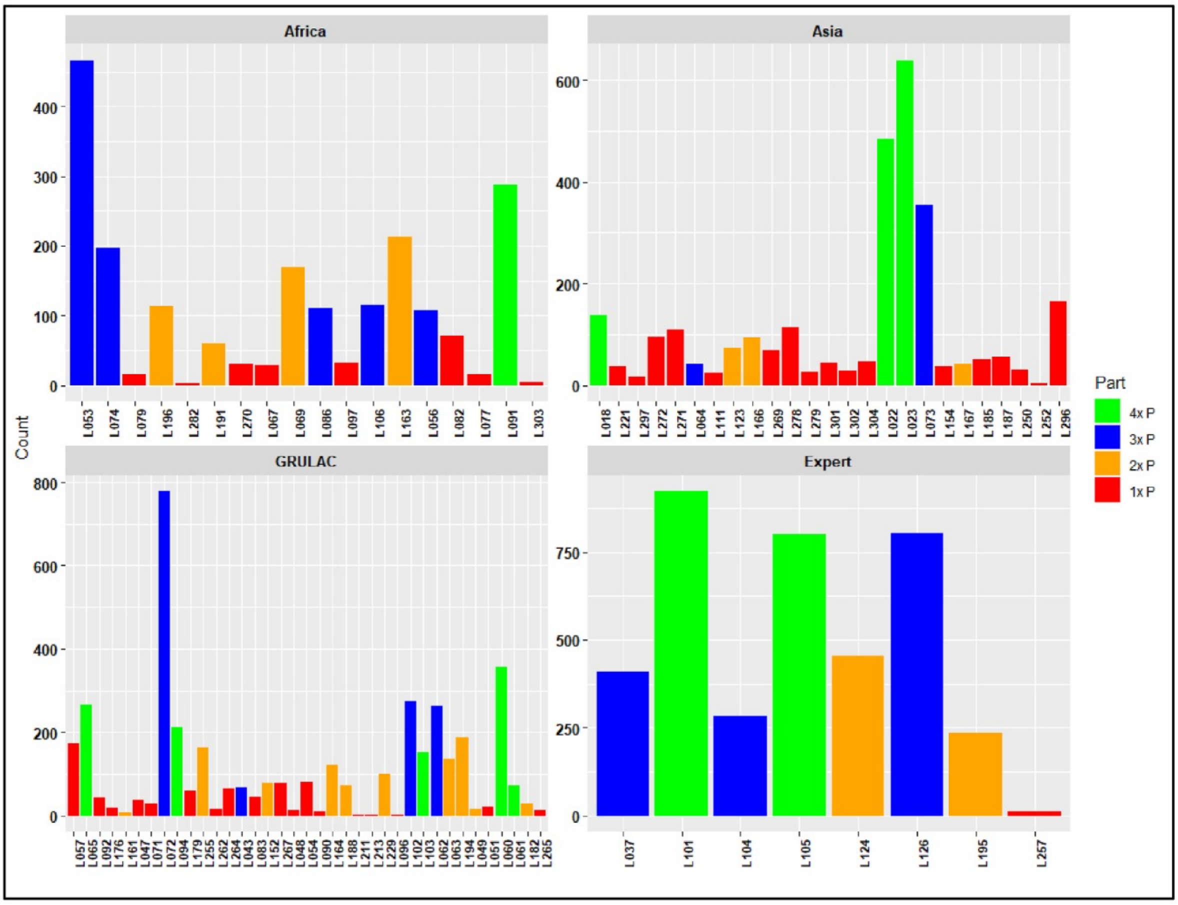


Fig. 10. Overview of z-scores per laboratory in each Region.

Table 6
Overall performance as z-scores in matrices ('S'/sum of z-scores, %) Cells have a color palette with strong green color indicating the highest percentage of 'S' results and strong red the lowest percentage.

	TS	Air	HM	Water	Overall		Sed	Fish
	%S	%S	%S	%S	**%S**		%S	%S
OCP	40	18	18		**30**		19	17
PCB	51	36	25		**39**		39	30
dl-POPs	73	70	56		**64**		58	57
PBDE	62	61	46		**57**		43	62
PBB153	57	60			**54**		40	56
HBCD	98	100			**92**			69
Toxaphene	80				**78**			60
PFAS	85	60	41	46	**68**		55	73

the laboratories was not the same in each IL. Details are shown in Figure S 5 and Table S 7.

Interestingly, the outcome was very similar, especially with respect to the percentage of 'S' results, which was around 47% for IL1, IL3 and IL4 and had the highest percentage in IL1 with 52.1%. The lowest percentage of 'U' with 29.8% was in IL2.

4.3. Overall quality

Table 6 displays the percentages of satisfactory z-scores within all z-scores achieved for a POP in a matrix. Overall, less than half of the z-scores were not satisfactory – or outside ±25% of the assigned value – for OCP and PCB. The numbers state that 70% of the OCP results that had been submitted by the laboratories in the four rounds did not reach the ±25% goal; for PCB only 39% of all results were satisfactory. In addition, for OCPs and PCB the relatively good performances on the test solutions (TS) drives the overall performance up to better values for the overall performance.

For dl-POPs (64%; with the largest share of the results) and PFAS (68%), ca $^2/_3$ of the results were satisfactory and for PBDE, more than half of the results were satisfactory (57%).

For the core matrix Air, provided as extract from PUF disks, the success rates for dl-POPs, PBDE, and PFAS were higher than for HM, where a positive ratio was obtained only for dl-POPs (56%). The absence of an interfering matrix may have played a role here.

5. Conclusion

Over approximately ten years and four rounds, the UNEP-coordinated interlaboratory assessments have gained international reputation as can be seen by the broad participation by laboratories from countries that were not participating in UNEP projects supporting the global monitoring plan of the Stockholm Convention on POPs. With 289 laboratories participating in the ILs and the large number of test matrices (eight test solutions of analytical standards and eight test matrices) and up to approximately 190 determinands per IL, these UNEP-coordinated proficiency tests were larger than other comparable studies. Whereas the performance assessments followed international practice, thanks to financial support from the POPs or chemicals and waste focal areas of the Global Environment Facility (GEF, Washington, DC, Unites States of America), the Quick Start Programme of the Strategic Approach to International Chemicals Management (SAICM, UNEP, Geneva, Switzerland), and the Thematic Programme for Environment and sustainable management of natural resources, including energy (ENRTP, European Union, Brussels, Belgium), all laboratories from developing countries or countries with economies in transition could participate free of charge.

The ILs were organized in 2-year rhythms allowing laboratories to check their performance regularly. Expert laboratories, as defined in section 2.2, were encouraged to participate. Overall, the results showed that the expert laboratories had very good performances in all ILs.

Newly listed POPs have been gradually included into the ILs (note inclusion of PFAS and PBDE from IL2, HBCD from IL3); however, short-chain chlorinated paraffins and polychlorinated naphthalenes were not yet included.

The z-scores, which are communicated to all laboratories in each round serve as performance indicators. They should be assessed by each laboratory for comparison and lessons learned drawn by each laboratory. Others include the following:

- Laboratories with regular or frequent participation and a broad spectrum of POPs and test matrices perform better than laboratories with single participation.
- For certain POPs, mass spectrometry seemed to be the only option to successfully analyse PFAS and isomer-specific HBCD (together with liquid chromatography) and dl-POPs, PBDE, toxaphene and ΣHBCD (together with gas chromatography).
- For dioxin analysis, the vast majority of the results as well as the high percentage of satisfactory results is obtained using capillary gas chromatographs coupled to sector-field instruments (high-resolution mass spectrometers) (see also for details Fiedler et al., 2021; UNEP et al., 2017, 2021); all other instrumentation gave poorer performance. The performance of ‘newer laboratories’ or ‘1- or 2-time participating laboratories’, which was less successful may be due to use of single MS or MS/MS detectors, which may not be able to handle complex mixtures of compounds and at low concentrations in biotic matrices.
- Whereas for the six indicator PCB, electron-capture detectors (ECD) may still be a valid option, mass spectrometers seemed necessary for the analysis of the more complex and larger number of OCP determinands (de Boer et al., 2021 in preparation).
- Some laboratories stated that they were capable to analyse more OCP when upgrading from ECD to MS; in addition, performance improved.

Besides the quantitative results presented before, the following descriptive and qualitative conclusions can be drawn:

- Through the UNEP projects, it was not possible to establish analytical capacities to have POPs laboratories confident to submit results for PFAS in Africa and GRULAC; in addition, in Africa, there was no capacity for the analysis of brominated flame retardants (PBDE and HBCD).
- Interlaboratory assessments should be undertaken regularly and for all combinations of POP (or chemical) with test matrices. They are for control of assessment and the analytical approach for generating and reporting results be the same as in routine analysis.
- Instrumentation and skilled personnel, materials and consumables including analytical standards, high-purity solvents, and high-quality gases should be always present in a laboratory.
- Laboratories should have and follow a business plan to obtain routines for POPs analysis to ensure sustainability of their work and maintenance of their instrumental infrastructure not only work towards an international check every two years.
- Self-control measures and quality controls including quality charts, laboratory and certified reference materials, should be applied throughout normal operation.
- Supply of analytical standards for identification and quantification of the POPs, new GC columns, extraction cartridges for a presumed better performance in the proficiency tests is no option.

Credit author statement

Heidelore Fiedler: Conceptualization; Data curation; Supervision dl-POPs and PFAS; Visualization, Writing – original draft. Ike van der Veen: Data curation; Formal analysis. Jacob de Boer: Project administration, Supervision OCPs, PCB and BFR; Validation; Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The contribution of the projects to support POPs monitoring in developing country regions financed by the Global Environment Facility (GEF) and implemented by the United Nations Environment Programme (UNEP; Geneva, Switzerland) is greatly acknowledged. The financial contribution from the European Union through the Environment and Sustainable Management of Natural Resources including Energy project ‘Second Worldwide UNEP Interlaboratory Study for Persistent Organic Pollutants’ for IL2 is greatly acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2021.132441.

References

Abalos, M., Abad, E., van Leeuwen, S.P.J., Lindstrom, G., Fiedler, H., de Boer, J., van Bavel, B., 2013. Results for PCDD/PCDF and dl-PCBs in the first round of UNEPs biennial global inter laboratory assessment on persistent organic pollutants. Trac. Trends Anal. Chem. 46, 98–109.

de Boer, J., van der Veen, I., Fiedler, H., 2021. Global Interlaboratory Assessment on PCBs, organochlorine pesticides and brominated flame retardants in various environmental matrices 2017/2019. Chemosphere preparation.

Fiedler, H., van der Veen, I., De Boer, J., 2017. Bi-ennial Global Interlaboratory Assessment on Persistent Organic Pollutants – Third Round 2016/2017. United Nations Environment Programme (UNEP), Geneva, Switzerland), p. 128.

Fiedler, H., van der Veen, I., de Boer, J., 2020. Global interlaboratory assessments of perfluoroalkyl substances under the Stockholm Convention on persistent organic pollutants. Trac. Trends Anal. Chem. 124.

Fiedler, H., van der Veen, I., de Boer, J., 2021. Interlaboratory assessments for dioxin-like POPs (2016/2017 and 2018/2019). Chemosphere, 132449. https://doi.org/10.1016/j.chemosphere.2021.132449.

ISO No Year. ISO 3166 - Country Codes. International Organization for Standardization, Vernier (GE), Switzerland. (Accessed 31 August 2021).

ISO, 2015. ISO 13528:2015 Statistical Methods for Use in Proficiency Testing by Interlaboratory Comparison.
Molenaar, J., Cofino, W.P., Torfs, P.J.J.F., 2018. Efficient and robust analysis of interlaboratory studies. Chemometr. Intell. Lab. Syst. 175, 65–73.
Thompson, M., Ellison, S., Wood, R., 2006. The international harmonized Protocol for the proficiency testing of analytical Chemistry laboratories: (IUPAC technical report). Pure Appl. Chem. 78, 145–196.
UNEP, 2015a. Continuing Regional Support for the POPs Global Monitoring Plan under the Stockholm Convention in the Africa Region, United Nations Environment Programme (UNEP). Global Environment Facility, Geneva, Switzerland (GEF),).
UNEP, 2015b. Continuing regional support for the POPs global monitoring plan under the Stockholm convention in the Latin American and caribbean region. In: United Nations Environment Programme (UNEP). Global Environment Facility, Geneva, Switzerland (GEF),).
UNEP, 2015c. Continuing Regional Support for the POPs Global Monitoring Plan under the Stockholm Convention in the Pacific Region, United Nations Environment Programme (UNEP). Global Environment Facility, Geneva, Switzerland (GEF),).
UNEP, 2015d. In: Regional Support for the POPs Global Monitoring Plan under the Stockholm Convention in the Asian Region, United Nations Environment Programme (UNEP. Global Environment Facility, Geneva, Switzerland (GEF),).
UNEP, 2017. IL3 Appendices (Örebro, Sweden).
UNEP, 2019. IL4 Appendices (Örebro, Sweden).
UNEP, Fiedler, H., Van der Veen, I., de Boer, J., 2017. Bi-ennial Global Interlaboratory Assessment on Persistent Organic Pollutants – Third Round 2016/2017. United Nations Environment Programme (UNEP), Geneva, Switzerland. https://wedocs.unep.org/handle/20.500.11822/21743.
UNEP, Fiedler, H., Van der Veen, I., de Boer, J., 2021. Bi-ennial Global Interlaboratory Assessment on Persistent Organic Pollutants – Fourth Round 2018/2019. United Nations Environment Programme (UNEP), Geneva, Switzerland. https://wedocs.unep.org/handle/20.500.11822/30923.
UNEP, Nilsson, H., van Bavel, B., Van der Veen, I., de Boer, J., Fiedler, H., 2014. Bi-ennial Global Interlaboratory Assessment on Persistent Organic Pollutants – Second Round 2012/2013. United Nations Environment Programme (UNEP), Geneva, Switzerland. https://wedocs.unep.org/20.500.11822/31377.
UNEP, van Bavel, B., van Leeuwen, S., de Boer, J., 2012. Bi-ennial Global Interlaboratory Assessment on Persistent Organic Pollutants – First Round 2010/2011. United Nations Environment Programme (UNEP), Geneva, Switzerland. https://wedocs.unep.org/handle/20.500.11822/31374.